PSA 2016 symposium contribution: “Data in Time: Epistemology of Historical Data”
Organized by Sabina Leonelli | 5 November 2016

LEGACY DATA, RADIOCARBON DATING AND ROBUSTNESS REASONING
Alison Wylie, University of Washington and Durham University

Introduction

Archaeologists put a premium on pressing “legacy data” into service, given the notoriously destructive nature of their practices of data capture. Legacy data consist of material and records that been assembled over decades, sometimes centuries, often by means and for purposes long since discredited or superseded. The primary strategies by which archaeologists put the data to work for new purposes are, I argue elsewhere (Wylie 2016), secondary retrieval, recontextualization, and experimental modelling. I focus here on a particularly telling example of secondary retrieval: the extraction of new data from old by means of radiocarbon dating.

As a physical dating technique underpinned by warrants of unimpeachable scientific pedigree, radiocarbon dating was expected, at the outset, to establish an empirical foundation for absolute chronologies that would render obsolete the local and relative chronologies on which archaeologists had long relied. And, indeed, it was applied to legacy data with dramatic effect, upending established archaeological chronologies and, with them, closely worked models of continent-wide culture processes like the spread of farming and the formation of early states and empires. In the event, however, many of the initial ^{14}C results were reassessed and amended as problems with contamination, the replicability of results, worries about a growing number of confounding factors, and questions about the cultural interpretation of ^{14}C dates came sharply into focus. Sixty years of standardizing applications and refining calibration curves has resolved many of these problems but two issues remain. One is that radiocarbon dates are a probabilistic estimate of a range of dates that an originating event could have generated often at a time scale and with margins of error that require considerable refinement to serve as the basis for cultural chronologies. And the second is that, even when problems of precision and scale are resolved, radiocarbon analysis dates a natural event – the point at which an organic sample ceased to exchange carbon with the atmosphere – so that its use in archaeological contexts requires a series of inferences about how the datable event relates to the cultural contexts and events of archaeological interest.

Now described as developing through not one but three revolutions – the initial introduction of radiocarbon dating (the Libby revolution) being the first, and the long process of calibration the second – this tortuous history reinforces a point made by internal critic/advocates in the early 1990s: no matter how much it is refined, radiocarbon dating typically cannot resolve, on its own, the chronological problems that archaeologists address. The challenge that animates the third radiocarbon revolution now under way is to ‘fully integrate archaeological information with ^{14}C dating in order to address archaeologically relevant … timescales and episodes’ (Manning 2015: 151). This is a genre of “robustness” reasoning that illustrates its epistemic risks as well as its appeal. To unpack what this means in an archaeological context I begin with an overview of the practices that have evolved through the three radiocarbon revolutions. What emerges is a history of dealing with the various kinds of error that can afflict robustness reasoning, on the basis of which I articulate a set of conditions that must be met if virtuous rather than vicious tangles of evidence are to be constructed.

Multiple radiocarbon revolutions

The formation of archaeology as a discipline in the 19th century turned on the successful development of chronologies that were originally founded on artefact typologies that capture patterns of association among artefacts found buried together (e.g. in burials and hoards, Trigger 1996: 124; Rowley-Conwy 2007: 32-47), and seriations that document the orderly succession of form and design within classes of artefacts (e.g. Deetz and Dethlefsen 1967). These were anchored temporally, where possible, by means of links to textual or epigraphic records, and with reference to the stratified deposition of this material (Renfrew 1973: 24). Tree-ring and varve analysis were used to build absolute chronologies of limited scope, but for the most part the dating of archaeological material was an internal affair until the advent of radiocarbon dating. When radiocarbon dating was first introduced there was enormous enthusiasm for the prospect that it would solve a range of chronological problems in archaeology, supplanting dependence on these local, uncertain and relative chronologies.[1] Willard Libby, the physical chemist who recognized that the rate of decay of radioactive carbon isotopes might be exploited for archaeological purposes, particularly emphasized the stability of the physical process of radioactive decay as the crucial warrant for its use as the anchor for archaeological dating:

> The rate of disintegration of radioactive bodies is extraordinarily immutable, being independent of the nature of the chemical compound in which the radioactive body resides and of the temperature, pressure, and other physical characteristics of its environment' (Libby 1952: 9, as cited by Francis 2002: 297).

Libby's insight was that, given this stable decay rate, if you know the ratio of radioactive (^{14}C) to stable carbon (^{12}C and ^{13}C) in the atmosphere in which a sample of organic material originated, you can use the difference between the proportion of carbon in the sample and this baseline ratio to estimate the time elapsed since the decay process began.

The radiocarbon revolution that Libby set in motion has, indeed, been 'sensational'; as Manning puts it in a recent retrospective appraisal, it has 'entirely restructured the practice and understanding of prehistoric archaeology around the world' (Manning 2015: 128). However, realizing the promise of this first radiocarbon revolution was no means straightforward. By the late 1950s questions were being raised about the reliability of ^{14}C results and within a decade it was clear that the radiocarbon dating could not be treated as a 'silver bullet' in a more consequential epistemic sense. It took some time for radiocarbon laboratories to refine methods for measuring ^{14}C in archaeological samples that control for the effects of elecromagnetic impurities, ambient radiation, radon contamination and fractionation (in reactions that do not go to completion), and to standardize count-time and conventions for calculating and reporting error. By the early 1980s protocols ensuring inter- and intra-lab reliability had been instituted,[2] and archaeologists had established procedures for minimizing contamination by younger or older organic material when recovering and handling samples. But in the process, as radiocarbon dating became widespread, a number of anomalies were identified that could not be attributed to contamination or processing error, making clear just how complex the physical processes are that underpin the method. It was this realization that catalyzed the second revolution: the long process of calibration that began in the mid-1960s (Manning 2015: 129).

It was discovered early on that Libby's original estimate of the half-life of ^{14}C was out by 162 years; improved estimates available by the late 1960s set it at 5730 ± 40 years rather than 5568 ± 30 years (Renfrew 1973: 288), but for pragmatic reasons it persisted as the standard long after the correction was made.[3] The most significant insight where archaeological applications are concerned was, however, the growing realization that the proportion of ^{14}C to ^{12}C and ^{13}C in the atmosphere is not uniform over time or space, or in its uptake by different types of organic matter. The industrial and bomb effects are particularly strong; 160 years of burning fossil fuels has dumped steadily growing amounts of 'dead' carbon into the atmosphere, depressing the proportion of radioactive to stable carbon isotopes, while above-ground nuclear tests in the Cold War era dramatically increased the proportion of ambient radioactive ^{14}C (Gillespie 1986: 20). Even when these effects are controlled for, samples from different types of organic material have different concentrations of ^{14}C depending on whether they are terrestrial or marine (e.g. whether they absorb carbon in the form of bicarbonate rather than carbon dioxide, or occur in carbon sinks created by ocean currents), what kind of

[1] The history of radiocarbon dating is an immensely complex story of enthusiasm and ambivalence, institutional manoeuvring and competition for access and authority. Marlowe (1999) gives a detailed account of its initial years; Francis (2002) outlines the impact of the first radiocarbon revolution on interdisciplinary research on Quaternary extinctions; and various aspects of this history are reported in a number of articles that assessed ^{14}C dating as it was being developed. Many of these appeared in the journal *Radiocarbon*, or in proceedings of the International Radiocarbon Conferences (e.g. Long 1992, Taylor, Long and Kra eds. 1992, Stuiver and van der Picht 1998) as well as in archaeological journals (Browman 1981, Bronk Ramsey 2008, Chippindale 2002 and Shott 1992). Manning (2015) provides a contemporary overview of how these debates have unfolded.

[2] Even so, in a review of *Radiocarbon After Four Decades* (Taylor *et al.* 1992), Browman (1994) observed that, while 'error magnitude is no longer linked clearly to lab type', differences in the standards employed by different laboratories was still an issue (p. 378). In response to these issues, Shott (1992: 219) emphasized the need for ongoing scrutiny of how different laboratories handle length of count-time, conventions for estimating counting errors, fractionation effects (a function of technique and count-time), and how they normalize results, despite the fact that, by the early 1990s, archaeologists had been advised not to worry about inter-lab variation.

[3] See, for example, Gillespie's discussion: the 'new value is sometimes used for geophysical research but should not be used for age reports. To convert from the old to the new half-life, multiply by 1.03. There is very little point in making this correction in isolation' (1986: 27). This is one example Francis cites of the conventions on which radiocarbon dating relies (2002: 300). See also Renfrew (1973: 288).

photosynthetic pathway they use to fix carbon (this differs between arid, succulent or temperate zone plants), and whether their metabolic processes discriminate against heavy isotopes (e.g. in bone collagen). So from the outset there was a very real question about what standard to use as the atmospheric baseline for determining how long the ^{14}C in a particular sample had been decaying since it stopped exchanging carbon with its environment. This was initially Cretaceious cabonate (Pee De Belemnite, PDB) and subsequently oxalic acid corrected to the average count rate for terrestrial wood dating to 1950. The discussion of this point in the Oxford *Radiocarbon User's Handbook* of 1986 is particularly interesting:

> The choice of this value [the average value for terrestrial wood used to normalize the measured carbon-14/carbon-12 ratio] *is arbitrary*, and other values could have been used with perhaps more theoretical justification. This normalization procedure, however, has been agreed internationally by the radiocarbon community, and the user is encouraged to check whether the laboratory does in fact use it. (Gillespie 1986: 18, emphasis added)

In addition, ^{14}C production is affected by sunspot activity and dipole movement,[4] and these effects are sometimes amplified by associated changes in temperature that have an impact on the atmospheric mixing and circulation of ^{14}C as well as its rate of absorption into carbon reservoirs. The result was recognition by the early 1980s that there are global differences in the concentration of ^{14}C between the northern and southern hemispheres, given proportionately more ocean surface in the southern hemisphere (this allows for more rapid transport of ^{14}C into ocean reservoirs), and also local variation that results from geological events such as volcanoes and geysers or in cases where climatic factors affect the rate of ^{14}C exchange between atmosphere and ocean (Browman 1981: 249-67; Gillespie 1986: 26-7). A series of reports that appeared in *Science* in 2001 documented 'a regional, time-varying ^{14}C offset [that] can occur within a hemisphere' (Kromer, Manning, Kuniholm, Newton, Spurk and Levin 2001; Manning, Kromer, Kuniholm and Newton 2001; Reimer 2001), in this case in securely dated tree-ring samples from Anatolia and southern Germany that grew at the same time (fifteenth to seventeenth centuries AD). The authors hypothesize that this is a consequence of a solar minimum which raised ^{14}C levels, depressing radiocarbon relative to calendric ages, and an associated cooling effect that had seasonally different impact on trees characterized by different growth periods (Kromer *et al*. 2001: 2529-30; Manning *et al*. 2001: 2533). Identifying, measuring and building these effects into estimates of radiocarbon dates is an on-going process.

In short, the inferential warrants necessary to make effective use of the ^{14}C decay rate as a basis for archaeological dating include a much wider and diverse range of domain-specific 'material postulates' (Norton 2003: 648) than initially realized by advocates of the first radiocarbon revolution.[5] The second radiocarbon revolution has been process of calibrating the atmospheric ^{14}C dates for specific time periods and regions against for samples of known age. Initially the basis for calibration was tree-ring data, but it has also included artefact sequences, stratigraphic data and historical records, just the kinds of chronological evidence that radiocarbon dating was expected to displace,[6] and at this point several ^{14}C calibration systems are available online (e.g. CALIB 7.1, Stuiver, Reimer and Reimer 2016; OxCal 4.2, Bronk Ramsey 2015). As these were refined, 'wiggle effects' were identified such that, for some periods of archaeological interest, samples with different true ages correspond to the same radiocarbon ages, or the spread in their true ages is exaggerated, compressed or even reversed. This reinforced the now conventional wisdom – the catalyst for the third radiocarbon revolution – that, in any archaeological application of the method, radiocarbon results must be interpreted in light of other contextual and chronological evidence.

The Childers Site

The implications of this growing recognition of the complexity of radiocarbon dating were articulated in the early 1990s by Shott (1992) in terms that anticipate the third revolution. He considers a puzzling

[4] These factors change the extent to which cosmic rays are deflected before they can reach the upper atmosphere and produce thermal neutrons.

[5] I am using the terminology of 'warrants' in the sense proposed by Toulmin (1958), as domain-specific norms of inference and substantive assumptions that mediate, in this case, the arguments by which conclusions about the evidential significance of empirical data are established. See the appended 'Toulmin scheme' illustrations, from *Evidential Reasoning in Archaeology* (Chapman and Wylie 2016).

[6] The work of refining and integrating regional calibration curves is on-going and is regularly reported at International Radiocarbon Conferences and in special issues of the journal *Radiocarbon*; for example, a recent issue of *Radiocarbon* is devoted to the new IntCal13 calibration dataset (Reimer 2013).

suite of ^{14}C dates for the Childers Site, a Late Woodland site in the Ohio Valley; the radiocarbon dates available at the time suggested that the site was occupied for 600 years (1050-1650 BP) with a discontinuous later occupation of 200 years (750-950 BP), while the archaeological evidence suggested that Childers was the result of a single relatively short occupation of ten to fifty years sometime in the period AD 400 - 800 (Shott 1992: 204, 207), results that, 'at face value ... resist simple interpretation' (p. 208). Rather than take the physics-backed radiocarbon dates as given, Shott scrutinized each line of evidence, assessing its security in its own terms and its plausibility in relation to the others (Figure 4.1).[7] He started with the radiocarbon samples, rejecting a third of them on grounds of poor provenance or risk of contamination, and then developed several strategies for evaluating the competing hypotheses about the Childers occupation suggested by the remaining ^{14}C results in light of the archaeological evidence. Given cultural practices of reuse, curation, trade and other forms of circulation, the ^{14}C datable organic samples may have been burned or cut long before they were deposited in the contexts from which they are recovered. To establish a connection between the natural event and the cultural target requires close attention to the context and association of the dated samples (Shott 1992: 203); it is a matter of using what Shott refers to as 'independent evidence concerning a site's antiquity' to determine how radicoarbon results should be interpreted in relation to the cultural events of archaeological interest (p. 203).[8]

The archaeological evidence for a single, relatively short occupation includes characteristics of the site itself – the relative homogeneity of assemblages in all major classes of cultural material and the low rate of feature overlap compared with other Late Woodland sites and assemblages – as well as background knowledge about the rates of decay typical for the types of wood found at the site, patterns of resource depletion associated with the foraging and horticultural activities documented for Childers, and ethnohistoric evidence that suggests a typical span of site occupation in the region. Although this establishes no precise length of occupation for Childers, they do reinforce Shott's initial conclusion that Childers was not occupied for anything like as long or in the discrete periods suggested by the radiocarbon results.

The crucial element of Shott's argument that anticipates the strategies of reasoning associated with the third radicocarbon revolution is his reanalysis of the radiocarbon dates. This includes includes pair-wise tests for contemporaneity, the calculation of a mean occupation date and measures of dispersion from the most credible radiocarbon results (calibrated to AD 585), and a strategy of modelling the dispersion of dates that, given standard sources of error, could be generated by samples that originated in a ten- to fifty-year occupation whose hypothetical true date is the average suggested by the pooled radiocarbon dates. Shott's modelling exercise shows that the wiggle effects built into calibration curves for the period in question, and the implications of normalizing ^{14}C dates for the kinds of material that make up the Childers samples, could well produce radiocarbon dates that range over 200 years for samples with the same cutting or burning dates. He concludes on this basis that the dispersal of the most reliable radiocarbon dates is consistent with the archaeological hypothesis for Childers; the samples could all have originated in a single short ten- to fifty- year occupation, as the archaeological evidence suggests, but most likely toward the end of the Late Woodland in the 200-year date range suggested by the ^{14}C results.[9] The upshot is that the although the radiocarbon results 'override our prior beliefs about the site's age' (p. 219), interpreted in light of the archaeological, ethnohistoric and ecological data they 'warrant archaeological conclusions [about the length of occupation] that an uncritical reading of all radicarbon results would not support' (p. 225). Given that 200 years is as close a determination of the occupation dates for Childers as radiocarbon dating can be expected to yield – at the time, for this period and for the types of sample analysed – Shott urged archaeologists to redouble their efforts to refine and extend existing local and relative chronologies, in this case, chronological sequences based on the seriation of ceramics and

[7] Shott's analysis of the integrity of these ^{14}C samples is a classic example of source criticism of the I characterize elsewhere as central to practices of secondary retrieval (Wylie 2016: 6-14; Chapman and Wylie 2015: 94-100). Considered in terms of Toulmin's analysis of 'arguments for use', Shott is articulating qualifications of the scope and strength of the evidential claim in question. He is also engaged in responding to potential rebuttals (in Toulmin's terms), systematically testing alternative hypotheses which is, on Reiss's pragmatic theory of confirmation, a matter of building indirect evidence for one among the suite of hypotheses Shott considers by a process of elimination (Reiss 2015: 347).

[8] Shott cites, in this connection, what Schiffer had described as a 'strong case' approach (Schiffer 1986; Shott 1992: 203).

[9] Shott refers here to estimates of typical error that, at the time, suggested that 'results even in the AD time interval can be reliably resolved only to approximately a 200-year range' (p. 226).

other classes of tools and artefacts. Reflecting on the ‘vagaries’ of calibration, he sees this use of multiple lines of evidence as ‘a method that not only controls the time dimension’, reducing reliance on radiocarbon dating, but also ‘tracks subtle cultural variation’ (p. 226).

Shott’s assessment of radiocarbon dating in the early 1990s was cautiously optimistic; he acknowledges that its importance for archaeology ‘is almost impossible to exaggerate’ but observes that it had ‘failed to meet the high expectations we developed for it’ (p. 202). The contributions of the second revolution to that point are reflected in his use of the growing body of background knowledge that underpins the calibration of radiocarbon dates to appraise sample integrity and margins of error. But he is clear that further refinement in the calibration of radiocarbon sequences will not, on its own, solve the problem of reconciling dissonant chronologies and establishing the cultural significance of ^{14}C dates. Establishing culturally relevant as well as secure and precise chronologies requires ‘archaeological observation and judgment’ (1992: 203).

Robustness reasoning and the third radiocarbon revolution

Taking up these themes two decades later, Manning notes a fundamental shift in approach that marks the advent of the third radiocarbon revolution. Rather than seek incontrovertible, physics-backed empirical foundations that can displace reliance on archaeological chronologies and their ‘web’ of background assumptions, the challenge is to ‘fully integrate archaeological information with ^{14}C dating in order to address archaeologically relevant timescales and episodes’ (Manning 2015: 151). Its distinctive contribution is the development of systematic analytical techniques for using multiple lines of evidence to assess margins of error in physical dating, and delimit, within the range of physically possible dates, a subset of archaeologically plausible dates. These include strategies of secondary retrieval and source criticism of the kind that Shott used to assess the provenance and integrity of samples from which ^{14}C dates are drawn, as well as the analysis of stratigraphic data, design sequence seriations, typological convergence and spatial distributions which, together, generate a range of chronological models for the target event or context of archaeological interest. These are then subjected to sensitivity anlaysis:

> One component of a model is changed, and the model is rerun. The outputs from the original model and its variant are then compared. When these are very similar, then the model can be regarded as insensitive to the component of the model that has been varied. When the outputs differ markedly, the model is sensitive to that component. Sensitivity analyses are useful not only in determining how far the outputs of a model are stable, but also help us to identify which components of a model are most critical to the resultant chronology. (Bayliss and Whittle 2015: 234)

Bayliss and Whittle, prominent advocates of this approach, emphasize the capacity of diverse lines of evidence to both constrain and reinforce one another, and articulate the rationale for what they describe as triangulation strategies in terms of an informal Bayesian model of confirmation. They argue that any assessment of the bearing of (new) evidence on a hypothesis must take into account how well supported the hypothesis is on other grounds (its prior probability), as well as the degree to which the evidence in question is discriminating: whether it would hold regardless of the truth or falsity of the test hypothesis (an appraisal of the prior and posterior likelihood of the evidence cited).[10] Construed in these ‘pragmatic Bayesian’ terms, the strategies characteristic of the third radiocarbon revolution are a classic example of the use of methods of ‘multiple determination’ (Wimsatt 1981: 123-4; Soler 2012: 3) that Wimsatt has influentially described as various forms of ‘robustness’ reasoning,[11]

[10] In an analysis of the role of Bayesian statistical reasoning in radiocarbon calibration, Steel (2001) describes archaeologists as ‘eclectic and pragmatic’ in their use of statistical tools (p. 154). For practical reasons of computational tractability as well as substantive reasons to do with the wiggles in calibration curves, packages such as CALIB make use of Bayesian statistics alongside classical statistics (p. 162): ‘Bayesian computational algorithms … more easily accommodate the complexities raised by the irregular form of the calibration curve’ (2001: 160). A discussion of the use of Bayesian models and methods in a special issue of *Radiocarbon* on the internationally agreed 2013 calibration curves supports this claim (see Niu, Heaton, Blackwell and Buck 2013). Steel’s analysis is addressed to philosophical critics of Bayesian confirmation theory, like Mayo (1996), who argue that, in practice, scientists do not make explicit use of Bayesian methods but, rather, rely on classical statistics (Steel 2001: 153-4).
[11] Wimsatt describes these as serving a wide range of purposes: to establish ‘the existence and character of a common phenomenon, object, or result’ (pp. 123), the reliability of the instruments and systems of measurement used to detect and to probe these phenomena, and the models built of them (Wimsatt 2012: 93-4)

specifically in application to the kind of problem that Hacking explores in connection with microscopy (Hacking 1981, 1983: 186-203).[12] The principle at work here is that we believe what we see through optical, acoustic and scanning electron microscopes not just because each of these instruments depends on well understood physical principles — evidential claims about the entities observed are backed by inferential warrants that render them individually secure[13] — but because, when used in conjunction with one another, it is implausible that images produced by such different means would converge as a consequence of confounding influences that generate compensating error in each of these very different lines of evidence. By extension, when multiple lines of evidence fail to converge they have a capacity to expose error that might not be detected in an assessment of the security of the backing and inferential credibility of each taken on its own. In an archaeological context I have described this as a matter of building cables rather than chains of evidential reasoning (1989); evidential reasoning is at its strongest when archaeologists can exploit the epistemic independence of distinct lines of evidence that rely on causally independent processes of trace generation and on conceptually independent detection techniques and inference-warranting bodies of background knowledge.[14]

This rationale for robustness reasoning has been described as a type of 'no-miracles' argument – 'it would be miraculous if multiple independent experiments showed *x* (where *x* is an entity, or a process, or a constant, or a relation) and *x* was not real' (Stegenga 2009: 653) – and it has recently been challenged by a number of critics who are sceptical of Wimsatt's more ambitious claims about the epistemic virtues of convergence (e.g. Stegenga 2009: 654-6; Hudson 2014; see also Soler 2014). I am not concerned here with the question of how ubiquitous robustness reasoning strategies are and whether or not they are ultimately reducible to or superceded by other more decisive modes of evidential reasoning, although I note that the third radiocarbon revolution took shape precisely because this physics-backed dating method proved to be incapable of displacing reliance on the scaffolding of pre-existing methods of chronological reckoning, and that the appeal to multiple lines of evidence is advocated as a necessary supplement to calibration. That said, this philosophical debate has directed the attention of both critics and advocates to two sets of reasons for caution about robustness reasoning that are relevant here. One is a general caution that the rhetorical force of convergence arguments can be misleading; convergence may add little epistemic weight to that provided by distinct lines of evidence considered on their own.[15] Another is an appreciation that, in practice, the processes of calibration and mutual adjustment required to integrate diverse lines of evidence carry a very real risk of artificially producing convergence. Taken together, these objections suggest a number of conditions that must be met if the risks of spurious convergence are to be avoided, all of which figure prominently in archaeological debate about the robustness of evidential reasoning from radiocarbon dates.[16]

---

[12] These parallels are developed in more detail in my analysis of epistemic security and independence in evidential reasoning in archaeology (2000, 2011).

[13] For philosophical discussion of 'security' as distinct from robustness in the context of physical and biomedical sciences, see Stegenga (2012: 212-13) and Staley (2004).

[14] I characterize this as horizontal independence between lines of evidence, as distinct from vertical independence between a test hypothesis and the evidence invoked in its support (Wylie 2011: 381, 387).

[15] Soler refers to these as worries about 'illusions of robustness' (2014: 210). The more specific objections she addresses, as developed by Hudson (2014), are that the evidence that is used to calibrate a measurement technique, or built into the scaffolding that enables a targeted test of contending hypotheses, should be understood to be superseded by the results of the measurement or test result that it makes possible. See Soler's discussion of these objections as developed by Hudson (2014) in *Seeing Things: The Philosophy of Reliable Observation* (Soler 2014: 204-5). For an especially trenchant critique of spurious convergence in an archaeological context, see Ullmann-Margalit's analysis of a pernicious interdependence between textual and material evidence in interpretations of the Dead Sea Scrolls (2006: 40-55).

[16] The set of conditions outlined here expands on those I identified in Wylie (2000 and 2011), and is informed by Soler (2012: 15-22; 2014: 210-12) and by contributors to Soler, Trizio, Nickles and Wimsatt (2012), especially Stegenga (2012). See, in particular, their discussion of the need to ensure that each line of evidence is credible in its own right (Soler 2012: 8; Stegenga 2012: 212-13, 219), and their treatment of the conditions necessary to establish that 'the plurality [of distinct lines of evidence] must be real and not just an illusion' (Soler 2012: 27). In her discussion of conditions for independence between lines of evidence, Soler distinguishes between content and historical/genetic independence (pp. 27-8), two dimensions of assessment that are captured here by conditions 3 and 4. Considerations of independence between the 'epistemic spheres' (as Soler refers to them, 2012: 28) in which distinct lines of evidence and their warrants are developed are a particular focus of attention in my earlier discussions of conditions for horizontal independence and are presupposed here. See also Stegenga's analysis of independence between modes of evidence (2012: 217-19). He and Soler both emphasize the importance of recognizing that assessments of security and of independence, and judgments about how to

1) ***Security***: each line of evidence – including its anchoring facts or observations and the warrants for their interpretation as evidence – must must be credible in its own right.
2) ***Causal anchoring and causal independence***: for a suite of evidential claims it must also be demonstrated that the material traces anchoring them are causally produced (in the first instance) by the same target of inquiry but that their subsequent transmission is causally independent.
3) ***Conceptual independence***: the warrants backing the interpretation of each line of evidence must also be independent. In particular, they must not depend on common assumptions (implicit or explicit) that could produce artificial convergence in the interpretation of distinct anchoring facts as evidence.
4) ***Grounds for calibration***: the calibration of warrants backing each line of evidence must be justified on independent grounds, not because they ensure the convergence of these lines of evidence.
5) ***Addressing divergence***: when lines of evidence fail to converge, each must be assessed for sources of error in their warrants and the backing for these warrants; no one line of evidence can be assumed secure and exempted from critical scrutiny.

The trajectory of the multiple radiocarbon revolutions I have described can be read as a sustained process of addressing concerns arising from these conditions. The initial enthusiastic reception of radiocarbon dating reflected confidence that, given its backing by nuclear physics, it met the first condition with a vengeance: it seemed uniquely secure. So long as it could be assumed that radiocarbon dates are determined exclusively by the invariant decay rate of ^{14}C, it was plausible that the method could deliver a universal, non-contingent evidential foundation for dating archaeological material. There was no need to rely on multiple lines of evidence except when questions of relevance arose about the bearing of radiocarbon dates on the cultural events of archaeological interest, and there seemed no question but that the second two conditions of causal and conceptual independence were met. The physical processes that give rise to a distinctive, time-sensitive ratio of radioactive to stable carbon in a sample of organic material are, in an obvious sense, causally independent of the cultural and material processes that produced and preserved the sample in an archaeological context. Moreover, the background knowledge from nuclear physics on which radiocarbon dating depends could be assumed to play no role in the construction of archaeological chronologies anchored in historical records, stratigraphic data, or stylistic seriation. Combined with the assumption of unimpeachable security, these considerations underwrote the expectation that radiocarbon dating could (and should) supplant reliance on local, contingent, conceptually entangled chronologies based on archaeological and historical evidence.

The second radiocarbon revolution was a response to concerns about whether, in fact, radiocarbon dating met the first condition when it was recognized that a great many factors other than Libby's decay rate affect the measured proportion of stable to radioactive carbon in archaeological samples. Establishing the security of this singular line of evidence put a premium on strategies of secondary retrieval and source criticism that involve scrutinizing the anchoring facts, and on the painstaking process of building the warrants that underpin the inference of evidential claims from these facts. Conceived as a process of calibration this was, in the first instance, a matter of identifying alternative lines of chronological evidence that are sufficiently secure in their (limited) domains of application that they could be used to cross-check radiocarbon dates. This, in turn, required that the second and third conditions be met: these distinct lines of evidence must be shown to originate in the same target event, but to follow causal pathways that are not vulnerable to the same types of distortion as affect the ratio of stable to radioactive carbon in tested samples, and they must depend on conceptually distinct ranges of background knowledge. Dendrochronology seemed to meet these conditions: the annual growth of tree-rings might be distorted by climatic fluctuations but not by the factors that affect the decay rate of radioactive carbon, so comparing the count of growth rings with the ^{14}C date for a well preserved sample of wood should provide just the kind of causally and conceptually independent control required, at least for some stretches of radiocarbon-based chronologies. This picture becomes complicated, however, when it is recognized, for example, that sunspots not only have an impact on the baseline ratios of stable to radioactive carbon in the atmosphere but also can affect climate which, in turn affects the growing period and growth rate of trees. In the process of assembling the

---

weigh different lines of evidence, depend on context-specific considerations and come in degrees; 'pragmatic, context-sensitive judgments are pervasively involved in scientific research, especially when degree and relevance appraisals are in play' (Soler 2014: 212).

scaffolding of warrants necessary to calibrate radiocarbon curves it became necessary to interrogate and substantiate the assumptions of independence that had underwritten the optimism of the initial revolution. The credibility of the calibration curves developed over decades of painstaking international, cross-field collaborative work ultimately depends on meeting the fourth condition: bringing into play background knowledge about the nature and effects of potential confounds and their interaction that justify fine-tuning radiocarbon results but do not already figure in the generation of these results.[17]

As the security of radiocarbon dating improved, the need to address the relevance component of the second condition – establishing how the natural events dated by means of radiocarbon analysis relate to the cultural events that archaeologists investigate – came sharply into focus. In the context of the third radiocarbon revolution, multiple lines of evidence are used not just to calibrate radiocarbon dates but as an essential resource, alongside ^{14}C dates, for building and refining archaeological chronologies. Pragmatic Bayesians are explicit about this: what they advocate is a practice of robustness reasoning in which no one line of evidence is presumed to stand as a uniquely secure empirical foundation for answering questions about the 'time dimension' of the cultural past (Shott 1992 226). Considered in this light, the insight central to the third radiocarbon revolution is that robustness is by no means miraculous; it is the product of hard work on multiple, irreducibly local but widely networked fronts.

There is much more to say about the range of scaffolding conditions necessary to do this. In *Evidential Reasoning in Archaeology* (2016), Chapman and I argue that these include not only epistemic warrants – domain-specific material postulates of various kinds – but also the social and institutional conditions that make it possible to cultivate the cross-border 'interactional' and 'meta-' expertise (Collins and Evans 2002, 2007) necessary to sustain 'trading zones' (Galison 2007, Collins, Evans and Gorman 2007) that foster virtuous, non-circular, robustness reasoning. I leave the details of this account to future papers.

References

Bayliss, A., Bronk Ramsey, C., van der Plicht, J., & Whittle, A. (2007). Brandshaw and Bayes: Towards a Timetable for the Neolithic. *Cambridge Archaeological Journal, 17*(1), 1-28.

Bayliss, A., & Whittle, A. (2015). Uncertain on Principle: Combining Lines of Archaeological Evidence to Create Chronologies. In R. Chapman & A. Wylie (Eds.), *Material Evidence: Learning from Archaeological Practice* (pp. 213-242). London: Routledge.

Bronk Ramsey, C. (2008). Radiocarbon Dating: Revolutions in Understanding. *Archaeometry, 50*(2), 249-275.

Bronk Ramsey, C. (2015). OxCal version 4.2. Retrieved from http://c14.arch.ox.ac.uk/embed.php?File=oxcal.html

Browman, D. L. (1981). Isotopic Discrimination in Correction Factors in Radiocarbon Dating. *Advances in Archaeological Method and Theory, 4*, 241-295.

Browman, D. L. (1994). Review: *Radiocarbon After Four Decades,* edited by R. E. Taylor, A. Long, and R. S. Kra; and *Proceedings of the 14th International Radiocarbon Conference*, edited by Austen Long. *American Antiquity, 59*(2), 377-378.

Chapman, R., & Wylie, A. (2016). *Evidential Reasoning in Archaeology*: Bloomsbury.

Chippindale, C. (2002). Capta and Data. *American Antiquity, 65*(4), 605-612.

Collins, H., & Evans, R. (2002). The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science, 32*(2), 235-296.

Collins, H., & Evans, R. (2007). *Rethinking Expertise*. Chicago: University of Chicago Press.

Collins, H., Evans, R., & Gorman, M. E. (2007). Trading Zones and Interactional Expertise. *Studies in History and Philosophy of Science, 38*, 657-666.

Deetz, J. F., & Dethlefsen, E. S. (1967). Death's Head, Cherub, Urn and Willow. *Natural History, 76*, 29-37.

Francis, K. (2002). *"Death Enveloped All Nature in a Shroud": The Extinction of Pleistocene Mammals and the Persistence of Scientific Generalists.* (PhD), University of Minnesota, Minnesota.

Galison, P. (2010). Trading With the Enemy. In M. E. Gorman (Ed.), *Trading Zones and Interactional Expertise* (pp. 25-52). Cambridge MA: MIT Press.

Gillespie, R. (1986). *Radicarbon User's Handbook*. Oxford: Oxford University Committee for Archaeology.

Hacking, I. (1981). Do We See Through a Microscope? *Pacific Philosophical Quarterly, 18*, 305-322.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hudson, R. (2014). *Seeing Things: The Philosophy of Reliable Observation*. New York: Oxford University Press.

[17] These include, for example, knowledge of the production, sequestration and uptake of atmospheric carbon drawn from fields as diverse as atmospheric and marine science, paleoecology, biochemistry and botany.

Kromer, B., Manning, S. W., Kuniholm, P. I., Newton, M. W., Spurk, M., & Levin, I. (2001). Regional 14CO2 Offsets in the Troposphere: Magnitude, Mechanisms, and Consequences. *Science, 294*(21 December), 2529-2532.
Libby, W. (1952). *Radiocarbon Dating*. Chicago: University of Chicago Press.
Long, A. (Ed.) (1992). *Proceedings of the 14th International Radiocarbon Conference*. Tucson: Department of Geosciences, University of Arizona.
Manning, S. W. (2015). Radiocarbon Dating and Archaeology: History, Progress and Present Status. In R. Chapman & A. Wylie (Eds.), *Material Evidence* (pp. 113-127). London: Routledge.
Manning, S. W., Kromer, B., Kuniholm, P. I., & Newton, M. W. (2001). Anatolian Tree Rings and a New Chronology for the East Mediterranean Bronze-Iron Ages. *Science, 294* (21 December), 2532-2535.
Marlowe, G. (1999). Year One: Radiocarbon Dating and American Archaeology, 19747-1948. *American Antiquity, 64*(1), 9-32.
Mayo, D. (1991). Novel Evidence and Severe Tests. *Philosophy of Science, 58*(4), 523-552.
Norton, J. D. (2003). A Material Theory of Induction. *Philosophy of Science, 70*(4), 647-670.
Parker Pearson, M., Cleal, R., Marshall, P., Needham, S., Pollard, J., Richards, C., . . . Richards, M. (2007). The Age of Stonehenge. *Antiquity, 81*, 617-639.
Reimer, P. J. (2013). IntCal13: Letter from the Guest Editor. *Radiocarbon, 55*(4), v.
Reiss, J. (2015). A Pragmatist Theory of Evidence. *Philosophy of Science, 82*(3), 341-362.
Renfrew, C. (1973). *Before Civilization: The Radio Carbon Revolution and Prehistoric Europe*. New York: Penguin Books.
Rowley-Conwy, P. (2007). *From Genesis to Prehistory: The Archaeological Three Age System and its Contested Reception in Denmark, Britain, and Ireland*. Oxford: Oxford University Press.
Schiffer, M. B. (1986). Radiocarbon Dates and the 'Old Wood' Problem: The Case of the Hohokam Chronology. *Journal of Archaeological Science, 13*, 13-30.
Shott, M. J. (1992). Radiocarbon Dating as a Probabilistic Technique: The Childers Site and Late Woodland Occupation in the Ohio Valley. *American Antiquity, 57*(2), 202-230.
Soler, L. (2012). The Solidity of Scientific Achievements: Structure of the Problem, Difficulties, Philosophical Impilcations. In L. Soler, E. Trizio, T. Nichles, & W. C. Wimsatt (Eds.), *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science* (pp. 1-60). New York: Springer.
Soler, L. (2014). Against Robustness? Strategies to Support the Reliability of Scientific Results *International Studies in the Philosophy of Science, 28*(2), 203-215.
Staley, K. (2004). Robust Evidence and Secure Evidence Claims. *Philosophy of Science, 71*(4), 467-488.
Steel, D. (2001). Bayesian Statistics in Radiocarbon Calibration. *Philosophy of Science, 68*(Proceedings), S153-S164.
Stegenga, J. (2009). Robustness, Discordance, and Relevance. *Philosophy of Science, 76*(5), 650-661.
Stegenga, J. (2012). *Rerum Concordia Discors:* Robustness and Discordant Multimodal Evidence. In L. Soler, E. Trizio, T. Nichles, & W. C. Wimsatt (Eds.), *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science* (pp. 207-226). New York: Springer.
Stuiver, M., Reimer, P. J., & Reimer, R. (2016). CALIB Radiocarbon Calibration version 7.1. Retrieved from http://calib.qub.ac.uk/calib/
Stuiver, M., & van der Plicht, J. (1998). Editorial Comment: INTCAL98: Calibration Issue. *Radiocarbon, 40*(3), xii-xiv.
Taylor, R. E., Long, A., & Kra, R. S. (Eds.). (1992). *Radiocarbon After Four Decades: An Interdisciplinary Perspective*. New York: Springer-Verlag.
Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.
Trigger, B. G. (1989). *A History of Archaeological Thought*. Cambridge: Cambridge University Press.
Ullmann-Margalit, E. (2006). *Out of the Cave: A Philosophical Inquiry into the Dead Sea Scrolls Research* Cambridge, MA: Harvard University Press.
Wimsatt, W. C. (1981). Robustness, Reliability, and Overdetermination. In M. B. Brewer & B. E. Collins (Eds.), *Scientific Inquiry and the Social Sciences* (pp. 124-163). San Francisco: Josey-Bass.
Wimsatt, W. C. (2012). Robustness: Material, and Inferential, in the Natural and Human Sciences In L. Soler, E. Trizio, T. Nichles, & W. C. Wimsatt (Eds.), *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science* (pp. 89-104). New York: Springer.
Wylie, A. (1989). Archaeological Cables and Tacking: Beyond Objectivism and Relativism. *Philosophy of the Social Sciences, 19*(1), 1-18.
Wylie, A. (2000). Rethinking Unity as a Working Hypothesis for Philosophy of Science: How Archaeologists Exploit the Disunity of Science. *Perspectives on Science, 7*(3), 293-317.
Wylie, A. (2011). Critical Distance: Stabilising Evidential Claims in Archaeology. In P. Dawid, W. Twining, & M. Vasiliaki (Eds.), *Evidence, Inference and Enquiry* (Vol. Proceedings of the British Academy 171, pp. 371-394). London: Oxford University Press.
Wylie, A. (2016). How Archaeological Evidence Bites Back: Strategies for Putting Old Data to Work in New Ways. *Science, Technology & Human Values*. doi:10.1177/0162243916671200