Extracting sunbeams from cucumbers: Reproducibility as a necessity for the future of archaeological science

## Introduction

In their paper celebrating the 40th anniversary of this journal Torrence et al (2015) noted that reproducibility was an issue important to the reputation and sustainability of the discipline, and necessary for archaeological science to behave like a science. As part of the celebration of the 50th anniversary, and of Torrence’s leadership of the journal, my contribution revisits these topics of archaeology’s status as a science, this journal’s place in the landscape of archaeological science, and how the journal has responded to a growing recognition of the importance of reproducibility. I first present bibliometric evidence of the position of archaeology as a whole, and this journal in particular, in the hierarchy of the sciences. Next I report on the journal’s progress in supporting reproducible research, and my work doing a new kind of peer review for JAS, one that evaluates the computational reproducibilty of the research submitted for publication.

The question of archaeology’s status as a science usually comes up in the context of what the discipline should or should not be. One of the first landmarks in tackling this question is the debate published by *Antiquity* between classical archaeologist Jacquetta Hawkes and palaeoanthropologist Glynn Isaac. Hawkes (1968), advocating a humanistic archaeology, was concerned that scientific approaches to archaeology were causing researchers to be “swamped by a vast accumulation of insignificant, disparate facts, like a terrible tide of mud, quite beyond the capacity of any man to contain and mould into historical form”. More optimistic about the integration of science and archaeology, Isaac (1971) counters that “New levels of precision in presenting data and in interpreting them can surely lead to briefer and more interesting technical reports as well as providing the basis for more lively literary portrayals of what happened in prehistory. Expanding on Isaac’s perspective, Binford (1962) argued that archaeology should operate as a science after the model proposed by philosopher Carl Hempel, which prescribed hypothesis-driven approaches, leading to generalizable laws of human behavior. Counter-arguments came from numerous directions, notably Hodder (1985) who rejected the quest for generalisations and instead argued that archaeology should be subjective and reflective, focussed on symbolic and relational meanings of material culture and the historical particularity of past human cultures. These debates, and the many more similar ones summarised by Martinón-Torres and Killick (2013), have become a genre in archaeological writing that can be characterized as mostly based on personal observations, microscopic dissections of a handful of cherry-picked case studies of good or bad practice, and discussion of various philosophers and sociologists.

What has been missing from these debates is a macroscopic observation of what the majority of archaeologists are actually doing, and an empirical comparison to a broad spectrum of relatively harder and softer disciplines. At the ‘hard’ end of the spectrum (e.g. physics and chemistry), scholars more typically share a large set of established set of theories, facts, and methods, facilitating fairly rapid agreement on the validity and significance of new results. At the ‘soft’ end of the spectrum (e.g. economics and psychology), the set of theories, facts, and methods on which there is widespread consensus is smaller, and agreement is slower and less frequently reached about the significance of new findings and the continuing relevance of previous work. Hardness and softness is a controversial distinction, in part because it is often used to imply a rank order of disciplines that encodes legitimacy, productivity, perceived value to society, and worthiness of funding (Editors 2012; Cole 1983). However, independent of these value judgments, empirical analysis of scholarly articles does indicate a spectrum of variation in practice linked to differing degrees of consensus in a discipline, for example in approaches to data visualisation (Cleveland 1984; Smith et al. 2000). Similarly, quantitative analysis of the frequency of positive results (ie. full or partial support for a research hypothesis) in publications is significantly correlated with hardness, consistent with a model where researchers in harder fields more readily accept any result their research produces, while those in softer fields have more freedom to choose which theories and hypotheses to test and how to interpret results (Fanelli 2010). The hard-soft spectrum is also evident in surveys of how researchers view their own work relative to those in other fields (Biglan 1973).

## Methods and materials

To objectively quantify the relative hardness or softness of archaeology, as an evaluation of its status as a science, and the place of this journal in context of other archaeology journals, I take a bibliometric approach. This approach is based on Fanelli and Glänzel (2013), who examined the hardness and softness of 12 disciplines using scholarly publication parameters. Fanelli and Glänzel (2013) found a spectrum of statistically significant variation in bibliometric variables from the physical to the social sciences, with papers tending at the softer end of the spectrum tending to have fewer co-authors, use less substantive titles, have longer texts, cite older literature, and have a higher diversity of sources. In Fanelli and Glänzel (2013)’s analysis harder sciences include Space Science, Physics, Chemistry, less hard sciences include social sciences include Psychiatry, Psychology, Economics, Business and General Social Sciences, and the Humanities define the soft end of the spectrum. Following Fanelli and Glänzel (2013), I quantify the number of authors, length of article, relative title length, age of references, and diversity of references for a large sample of peer-reviewed journal articles.

These parameters are useful because of how they signify consensus in a research community. A larger number of authors on a paper reflects collaboration of people working together on a common goal. Collaborators have specialized roles, each of whom has the ability to study a part of the problem with high accuracy and detail, with harder fields having larger groups of collaborators (Zuckerman and Merton 1972). Reflecting this collaboration group size, harder disciplines tend to have higher average numbers of authors on papers. Article length has an inverse correlation with field hardness. In low-consensus, or softer, fields, papers must be longer to present justification, nuance and contextualization of results. The number of substantive and informative words in an article’s title tends to be positively correlated with article length in harder disciplines (Moshe Yitzhaki 2002; M. Yitzhaki 1997), reflecting a focus on empiricism and efficiency that is characteristic of high-consensus disciplines. While Moshe Yitzhaki (2002) removed stop-words (e.g. prepositions, articles, conjunctions, etc.) to calculate article length, in order to generate results for comparison with Fanelli and Glänzel (2013) I follow his method of dividing the total word count of the article title by the total number of pages of the article to compute relative title length.

The age of works cited has long been used as a measure of a field’s hardness (Moed, Van Leeuwen, and Reedijk 1998; Börner 2010), based on the assumption that harder fields assimilate new results more rapidly that softer fields (Price 1970). I calculated a recency of references index for each article (also known as the Price index), which is the proportion of all cited works that were published in the five years preceding the paper. The diversity of references

While Fanelli and Glänzel (2013) analysed papers published in a single year (2012), I found only 303 papers for that same year, and 70% of papers in the sample published after that date. To make efficient use of the available data and ensure robust representation from different areas of archaeology, including those with lower frequencies of journal article publication, I analysed 9697 papers published during 1975-2025. This sample was collected from Clarivate’s Web of Science database by first selecting the Web of Science category ‘Archaeology’ and the Document type ‘article’ (n = 28,871). To focus on journals of broad relevance to most archaeologists, and that are representative of substantial communities of practice, I then filtered the results to keep only articles published in the top-ranking 25 journals according to their h-indices as reported by Clarivate’s Journal Citation Indicator. Finally I excluded journals with less than 100 articles in the database, resulting in 20 journals.

The entire R code (R Core Team 2024) used for all the analysis and visualizations contained in this paper is included in the Supplementary Online Materials at https://doi.org/xxx/xxx/xxx to enable re-use of materials and improve reproducibility and transparency (Marwick 2017). All the figures, tables, and statistical test results presented here can be independently reproduced with the code and data in this compendium (Marwick, Boettiger, and Mullen 2018). The R code is released under the MIT license, the data as CC-0, and figures as CC-BY, to enable maximum re-use.

## Results

### How does archaeology compare to other fields?

|  |
| --- |
| Figure 1: Distributions of article characteristics hypothesised to reflect the level of consensus. The boxplot shows the distribution of values of archaeology articles. The thick black line in the middle of the boxplot is the median value, the box represents the inter-quartile range (the range between the 25th and 75th percentiles, where 50% of the data are located), and individual points represent outliers. The thin grey horizontal lines in each boxplot indicate the median values computed by Fanelli and Glanzel (2013), where p = physics, s = social sciences, h = humanities. |

[Figure 1](#fig-compare-other-fields) shows the distribution of bibliometric variables for archaeology in context of data from other fields presented by Fanelli and Glänzel (2013). The most striking indicator of archaeology as a hard science is the number of authors, where it is between the social sciences and physics. Archaeology is a close fit with the social sciences in relative title length. It is between the the social sciences and humanities in recency of references and diversity of references. The clearest indicator of archaeology as a soft science is article length where it is similar to the humanities. Overall archaeology does not sit squarely at either end of the hard-soft spectrum. It is generally not a harder science than the social sciences, with the exception of collaborator group sizes.

### How has the hardness of archaeology varied over time?

|  |
| --- |
| Figure 2: Distribution of article characteristics for archaeology articles over time. Data points represent individual articles. The colour of the points indicates if the overall trend is toward softer (orange) or harder (green) Generalized Additive Models were computed to fit the lines summarising the relationships between the variables and the time series. |

[Figure 2](#fig-change-over-time) shows how the bibliometric indicators of field hardness have changed of time for archaeology articles. By two measures, the number of authors and relative title length, archaeology has become increasingly harder over time. On the other hand, three metrics indicate that archaeology has become softer (diversity of references, article length and recently of references). Although all the relationships are statistically significant, generally these temporal trends are very weak with low slope values, indicating very slow change over time. Similarly the r-squared values are very low, demonstrating that much of the variability in these metrics is independent of time. The most striking change over time is in the increase in the number of authors, which has the highest r-squared value of these metrics. One interesting detail evident in [Figure 2](#fig-change-over-time) is the increase in the range of diversity of references after about 2010. This may be due to some broader changes in academic publishing around this time, such as moves to digital-first continuous publishing, new journals appearing (e.g. Archaeological and Anthropological Sciences in 2009 and Journal of Island & Coastal Archaeology in 2010), and non-archaeology journals becoming more relevant to archaeologists. For example, *PLOS ONE* received its first impact factor in 2010 and in 2011 Nature’s *Scientific Reports* began publishing (Malashichev 2017).

### How do archaeology journals vary in hardness?

|  |
| --- |
| Figure 3: Variation in bibliometric indicators of hardness for 20 archaeological journals. The journals are ordered for each indicator so that within each plot, the harder journals are at the top of the plot and the softer journals are at the base. The lower right panel shows a bar plot that is the single consensus ranking computed from all five variables, using the Borda Count ranking algorithm. |

[Figure 3](#fig-variation-by-journal) shows the distribution of our bibliometric variables of hardness for each of the 20 journals in the sample. Overall agreement between these bibliometric variables in ranking these journals on a hard-soft spectrum is moderate to strong, with a Kendall’s coefficient of concordance (Wt) value of 0.64 (in a 0-1 range, where 1 is perfect agreement) and a p-value of 2.67 x 10-06. Panel F of [Figure 3](#fig-variation-by-journal) shows an overall consensus ranking of all journals in the sample. In this consensus ranking the *Journal of Archaeological Science* among the top five archaeology journals for hardness. It is placed at the harder end of the hard-soft spectrum especially by the number of pages and relative title length, and to lesser degrees by the number of authors and recency of references. However, according to the diversity of references, the *Journal of Archaeological Science* is at the middle of the spectrum.

The *Journal of Cultural Heritage* is the only journal that consistently ranks as hard across all variables, occurring in the top five journals for all five metrics. This journal primarily publishes materials science and computational analyses related to conservation and preservation of historic objects in museums and other collections. Authors of papers in recent issues have affiliations with museums, cultural heritage programs, and chemistry, engineering, and physics departments at European and Chinese universities. Notably, papers in this journal typically do not engage in questions or debates about past human behaviour or culture. The absence of these questions in research published in this journal makes it an outlier here, since these questions are central to a common definition of archaeology as ‘cultural anthropology of the past’, a phrase first found in Leroi-Gourhan (1946) and repeated in widely-used contemporary undergraduate textbooks such as Renfrew, Bahn, and DeMarrais (2024). Most archaeologists would likely be surprised at the decision by Clarivate to include the *Journal of Cultural Heritage* in their category of archaeology journals, leading to this result in [Figure 3](#fig-variation-by-journal) where the hardest archaeology journal publishes papers that are not very archaeological at all because they do not engage with anthropological topics.

The *Journal of Archaeological Research* is notable because it consistently ranks as soft, ranking as the softest journal for four of our five bibliometric variables. This is a predicable result for a review journal, which is a distinct type of journal dedicated to summarizing, analyzing, and synthesizing existing research in a particular field. The stated aim of the *Journal of Archaeological Research* is to ‘bring together the most recent international research summaries on a broad range of topics and geographical areas’ (Feinman and Parkinson 2024). A typical article is a long single-authored synthesis of archaeology in a region or on a topic. As the only review journal in this sample, this is a stark contrast to the other journals here that present original research findings, and like the *Journal of Cultural Heritage*, may be considered an outlier in this sample.

Biglan, Anthony. 1973. “The Characteristics of Subject Matter in Different Academic Areas.” *Journal of Applied Psychology* 57 (3): 195–203. <https://doi.org/10.1037/h0034701>.

Binford, Lewis R. 1962. “Archaeology as Anthropology.” *American Antiquity* 28 (2): 217–25.

Börner, Katy. 2010. *Atlas of Science*. The MIT Press. <https://mitpress.mit.edu/9780262014458/atlas-of-science/>.

Cleveland, William S. 1984. “Graphs in Scientific Publications.” *The American Statistician* 38 (4): 261–69.

Cole, Stephen. 1983. “The Hierarchy of the Sciences?” *American Journal of Sociology* 89 (1): 111–39. <https://doi.org/10.1086/227835>.

Editors. 2012. “A Different Agenda.” *Nature* 487 (7407): 271–71. <https://doi.org/10.1038/487271a>.

Fanelli, Daniele. 2010. ““Positive” Results Increase Down the Hierarchy of the Sciences.” *PLOS ONE* 5 (4): e10068. <https://doi.org/10.1371/journal.pone.0010068>.

Fanelli, Daniele, and Wolfgang Glänzel. 2013. “Bibliometric Evidence for a Hierarchy of the Sciences.” *PLOS ONE* 8 (6): e66938. <https://doi.org/10.1371/journal.pone.0066938>.

Feinman, Gary, and William Parkinson. 2024. “Aims and Scope: Journal of Archaeological Research.” *SpringerLink*. <https://link.springer.com/journal/10814>.

Hawkes, Jacquetta. 1968. “The Proper Study of Mankind.” *Antiquity* 42 (168): 255–62. <https://doi.org/10.1017/S0003598X00034451>.

Hodder, Ian. 1985. “Postprocessual Archaeology.” *Advances in Archaeological Method and Theory*, 1–26.

Isaac, Glynn LL. 1971. “Whither Archaeology?” *Antiquity* 45 (178): 123–29. <https://doi.org/10.1017/S0003598X00069283>.

Leroi-Gourhan, André. 1946. *Archéologie Du Pacifique-Nord: Matériaux Pour l’étude Des Relations Entre Les Peuples Riverains d’asie Et d’amérique*. Vol. 47. Travaux Et Mémoires de l’institut d’ethnologie. Paris: Institut d’ethnologie.

Malashichev, Yegor. 2017. “From Open Access to Open Science.” *Biological Communications*, no. 1: 3–5.

Martinón-Torres, Marcos, and David Killick. 2013. “Archaeological Theories and Archaeological Sciences.” <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199567942.001.0001/oxfordhb-9780199567942-e-004>.

Marwick, Ben. 2017. “Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation.” *Journal of Archaeological Method and Theory* 24 (2): 424450. <https://doi.org/10.1007/s10816-015-9272-9>.

Marwick, Ben, Carl Boettiger, and Lincoln Mullen. 2018. “Packaging Data Analytical Work Reproducibly Using r (and Friends).” *The American Statistician* 72 (1): 80–88. <https://doi.org/10.1080/00031305.2017.1375986>.

Moed, Henk F, Th N Van Leeuwen, and Jan Reedijk. 1998. “A New Classification System to Describe the Ageing of Scientific Journals and Their Impact Factors.” *Journal of Documentation* 54 (4): 387–419.

Price, DJS. 1970. “Citation Measures of Hard Science, Technology and Nonscience.” *Communication Among Scientists and Engineers*, 3–22.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Renfrew, Colin, Paul Bahn, and Elizabeth DeMarrais. 2024. *Archaeology: Theories, Methods, and Practice*. 9th ed. New York: Thames & Hudson.

Smith, Laurence D., Lisa A. Best, D. Alan Stubbs, John Johnston, and Andrea Bastiani Archibald. 2000. “Scientific Graphs and the Hierarchy of the Sciences:: A Latourian Survey of Inscription Practices.” *Social Studies of Science* 30 (1): 73–94. <https://doi.org/10.1177/030631200030001003>.

Torrence, Robin, Marcos Martinón-Torres, and Th. Rehren. 2015. “Forty Years and Still Growing: *Journal of Archaeological Science* Looks to the Future.” *Journal of Archaeological Science*, Scoping the future of archaeological science: Papers in honour of richard klein, 56 (April): 1–8. <https://doi.org/10.1016/j.jas.2015.03.001>.

Yitzhaki, M. 1997. “Variation in Informativity of Titles of Research Papers in Selected Humanities Journals: A Comparative Study.” *Scientometrics* 38 (2): 219–29. <https://doi.org/10.1007/BF02457410>.

Yitzhaki, Moshe. 2002. “Relation of the Title Length of a Journal Article to the Length of the Article.” *Scientometrics* 54 (3): 435–47. <https://doi.org/10.1023/A:1016038617639>.

Zuckerman, Harriet, and Robert K Merton. 1972. “Age, Aging, and Age Structure in Science.” *Higher Education* 4 (2): 1–4.