Why is there Math in my Archaeology?

John Justeson and Calculating the Limits of Interpretation

James Scott Cardinal Jennifer Ann Loughmiller-Cardinal

Fifty years ago, what was arguably the most important paper ever written for modern work in quantitative archaeology was published in American Antiquity. Unfortunately for its author, and generations of archaeologists, few took notice of it at the time. With few citations, more than half of which have occurred in just the last few years, its elegance and mathematical precision went largely unappreciated—even by the growing cohorts of computational and quantitative archaeologists whose work would have greatly benefited from it. In this paper, we demonstrate that John Justeson's 1973 article "Limitations of Archaeological Inference" was not only accurate and precise in its implications, but also very much still at the forefront of archaeological thought… even if the field at large doesn't yet realize it.

A Gentle Introduction to Information Theory

What is now known as "Information Theory" began with a paper written by Claude Shannon, titled "A Mathematical Theory of Communication" (1948), resulting from his work in cryptography at Bell Labs. At the heart of Shannon's theory was the idea that *information* is fundamentally tied to the reduction of *uncertainty*. Shannon approached information not in terms of meaning, but as a measure of the *reduction of uncertainty* within a system of communication.

Information, Entropy, and Surprisal

Shannon proposed a particular relationship between information and uncertainty in terms of statistical probabilities. He derived a quantitative measure of that uncertainty derived from the concept of *entropy* used to describe disorder in the thermodynamics of physical systems. Shannon, however, repurposed entropy to refer to the average uncertainty contained in a system given by the equation:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\ \log_2\ p(x_i)$$

What this equation is describing is the total entropy H of some system X that contains n discrete attributes or elements $(x_1, x_2, \ldots x_n)$. The entropy is equal to the negative sum, over all n features, of each element's probability of occurrence $p(x_i)$ times the $\log_2$[1] of that probability.

The higher the entropy of a system, indicated by a larger value for H, the more uncertainty or randomness there is to the elements of X. Somewhat counterintuitively, the more uncertain or random a system the more information it conveys. Remember that Shannon defines information as the reduction of uncertainty. The greater the uncertainty (i.e., high entropy), the more potential information the system is capable of producing because there is greater uncertainty to reduce.

To see how, we need to understand what Shannon defined as *surprisal.* Surprisal, also known as self-information, is a measure of how surprising or unexpected a specific event is based on its probability. In essence, surprisal measures the information content of a specific outcome – i.e., rare events carry more information than common ones because they are less expected. Low probability events, those that occur infrequently, are highly surprising. Conversely, high probability events are not.

Consider it this way – if an event is nearly certain to occur, you would *already* be expecting it to happen when it does. Its occurrence tells you nothing that you did not already know. It is only when something happens that we did *not* expect (i.e., we are surprised) that it is providing *new* information. Therefore, surprisal (denoted as $I(x)$) is the potential *information* contained in a single event based on its probability $p(x)$:

$$I(x) = -\log_2 p(x)$$

Surprisal is zero for events that are certain (i.e., the probability $p(x) = 1$), and grows larger as the probability of the event decreases (Figure {#figure:surprisal_example}). Exceedingly rare events, by contrast, would be very surprising to witness and approaching "infinitely" surprising as the probability of the event goes to zero (i.e., $lim_{p(x)\to 0}\ I(x) = \infty$).

Entropy represents the *average* surprisal over all possible outcomes from a probability distribution. It quantifies the overall uncertainty or unpredictability of a system or source of information. The higher the entropy, the more information the system is capable of producing, since there is greater uncertainty about which outcome will occur.

Entropy is highest when all outcomes are equally likely, and decreases as we gain more information to anticipate whether or not that event is likely to occur (Figure {#figure:entropy_example}). Information is therefore the reduction of that uncertainty or entropy when a new event is observed. We have learned more about the underlying probabilities for future events.

For the first time, scientists had a way to *quantify* information. Shannon had defined information in a way that made it possible to measure and *analyze* it mathematically, based solely on its statistical structure and independently of its content or meaning.

[1] $\log_2$ refers to the base-2 logarithm.

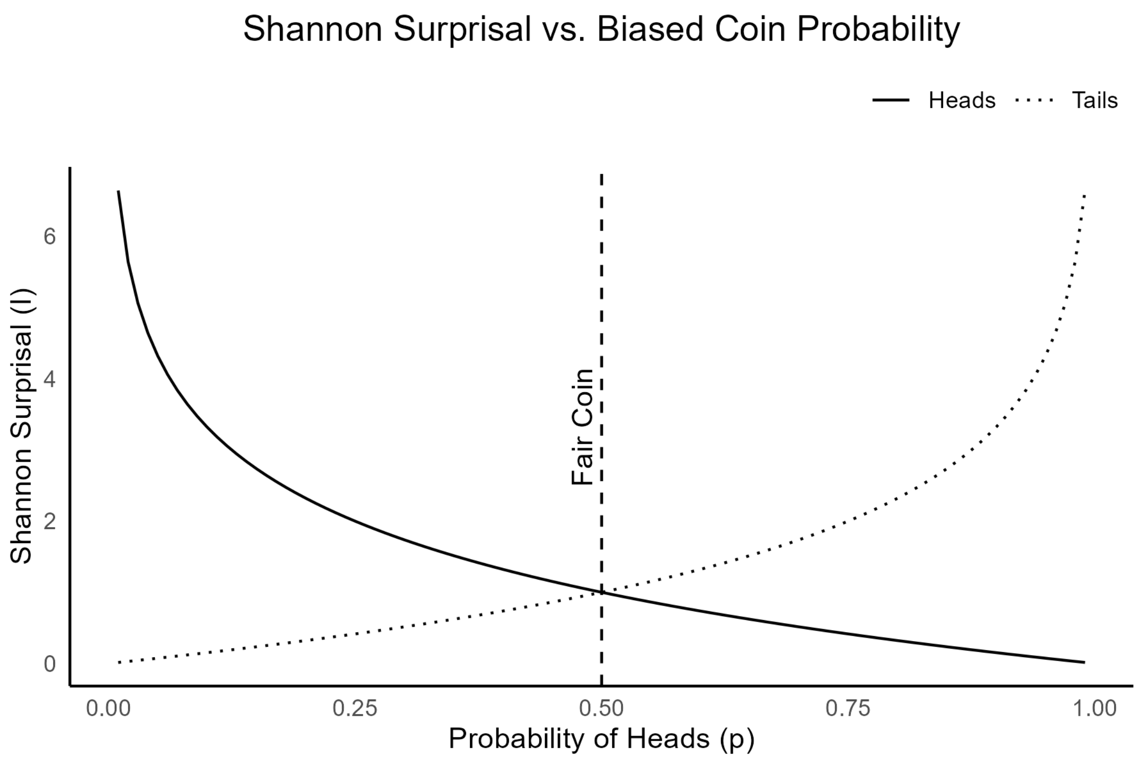

Figure 1: The surprisal $I(x)$ of a coin flip x (i.e., lands "heads" or "tails") as the probability $p(x)$ of landing "heads" ranges from 0 to 1 for a "biased" coin. A "fair" coin would land on heads or tails with equal chances or $p(x) = 0.5$.

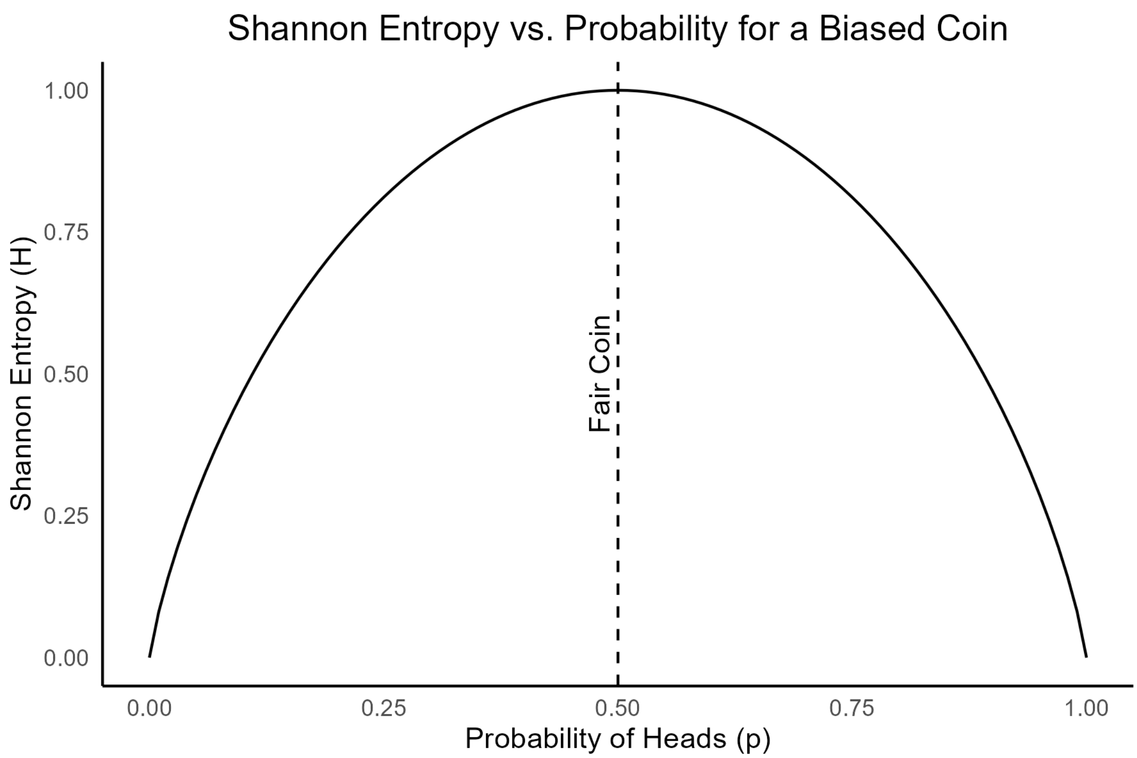

Figure 2: The overall system entropy $H(X)$ for biased coin flips as the probability $p(x)$ of landing "heads" ranges from 0 to 1. A "fair" coin with $p(x) = 0.5$ is the system with the most uncertainty, since either outcome ("heads" or "tails") is equally possible.

Information theory has evolved over the last few decades into a highly diverse discipline in its own right, with broad applications. Shannon, however, developed the theory towards one particular application – communication. Specifically, he was looking for a way to understand how information could be efficiently and reliably transmitted across communication systems, especially in the presence of noise or interference.

Channel, Signal, and Noise

Under Shannon's model of communication, the relationships between information, channel, signal, noise, and channel capacity define the core aspects for transmitting data effectively. Information represents the content or message that needs to be conveyed, which can vary in complexity. In this context, entropy is a measure of the inherent complexity of the information a message might contain – i.e., higher entropy indicates greater variability in *potential* messages.

A communication *channel* is the medium or system through which information is transmitted. Channels connect sender to receiver, and are characterized by their capacity to handle information. This *channel capacity* (C) represents that maximum rate at which information that may reliably be transmitted across that a channel. It places an upper bound on how much information can be sent through such a channel of communication with an arbitrarily low rate of errors, given by:

$$C = \max_{p(x)} I(X \mid Y)$$

where $I(X \mid Y)$ is the *mutual information*[2] between the transmitted variable X and the received variable Y. It measures the amount of information *shared* between X and Y or, put another way, how much of what was transmitted by the sender is retained and correctly *understood* by the receiver. The capacity C for the channel, then, is where maximum amount of information can be correctly transmitted with the fewest number of errors or misunderstandings by the receiver.

Information is transmitted across a communication channel as *signals*, which are encoded representations of the information. A signal is is defined as the physical embodiment of information that is transmitted across a communication channel from a sender to a receiver. Shannon treats signals as mathematical entities that *encode* data using a sequence of symbols, typically binary (0s and 1s), representing the discrete or continuous probability states of information. This encoding enables information to be manipulated, stored, and transmitted efficiently, with the ultimate goal of achieving maximum fidelity in the presence of noise or interference. Shannon's model abstracts signals into probabilistic terms, allowing for quantification of the information.

[2] Remember, $I(x) = -\log_2 p(x)$ is the *surprisal* value of event x that represents the information conveyed by that event. The *mutual information* between two events can be though of as the information conveyed when *both* events occur simultaneously.

Channels, however, are not perfect. They can introduce disturbances known as *noise*, which interferes with the signal and can alter the received message, creating a challenge in accurate data transmission. The more noise present, the harder it is to reliably convey information. Noise is essentially random disturbances or fluctuations in the transmission of information along a channel that interfere with the signal. Noise can distort or obscure messages, increasing the probability of errors in decoding them.

Since channel capacity is the maximum rate at which information can be transmitted over a channel without errors, excess noise degrades capacity by introducing errors. So, channel capacity depends on both the *bandwidth* of the channel (i.e., the allowable range of possible signal frequencies) and the *signal-to-noise ratio* (often simply called "SNR"). Shannon's theory shows that for a channel to transmit information efficiently, the signal must be strong enough to overcome noise, but *not* so strong that it leads to unnecessary redundancy in the message encodings.

This balance maximizes the channel's capacity, allowing the most efficient transfer of information while minimizing error. This gives us another way[3] to find a channel's capacity, given by:

$$C = B \log_2 \left(1 + \frac{S}{N}\right)$$

where B is the bandwidth of the channel, S is the power of the signal, and N is the noise. The signal-to-noise ratio (SNR) $\frac{S}{N}$ represents, a measure of how strong the signal is relative to the noise. As noise (N) gets larger relative to signal (S), the SNR starts dropping closer towards zero. Since $\log_2(1+0) = 0$ that means that, no matter how large its ideal bandwidth, the channel's capacity C goes towards zero as well.

The influence of telecommunication and cryptography on Shannon's theories are obvious, but the underlying concepts quickly found new applications and implications in other fields of study. By linking information to uncertainty and statistical probabilities, Shannon's abstracted and highly generalized model of information and communication could be adapted to studying all manner of systems. It would not be too long after the theories described in Shannon's technical paper were expanded and republished in book-length form a year later as "The Mathematical Theory of Communication" (Shannon and Weaver 1949) that they would begin to appear in disciplines ranging form physics to physiology.

A Brief History of Quantitative Analysis and Information Theory in Archaeology

The integration of quantitative methods into archaeology during the 20th century profoundly transformed the methods by which archaeologists tried to understand the past. By the 1950s,

[3] This way to calculate channel capacity is more common (and often much more practical) in telecommunication applications, such as those Shannon was studying, in which such things as "channel bandwidth" and "signal power" can be directly measured or otherwise experimentally ascertained.

methodological innovations in works by Brainerd (1951) and Heizer and Cook (1956) highlighted the value of statistical frameworks in chronology-building and site analysis, establishing a precedent for rigorous quantitative inquiry in archaeology. This push intensified during the 1960s with the advent of the "New Archaeology" championed by figures such as Lewis Binford.

New Archaeology, later termed Processual Archaeology, advocated for an explicitly scientific approach grounded in hypothesis testing, formal quantification, and systems theory (Kendall 1969; Binford 1981; Raab and Goodyear 1984). David Clarke's texts *Models in Archaeology* (1972) and *Analytical Archaeology* (1978) formalized these aspirations by advocating for complex statistical models and systems theory to provide explanatory frameworks in the study of past human behavior. By situating quantitative methods at the heart of analysis, Processual Archaeology sought to go beyond mere description to causal understanding, particularly through middle-range theories that sought correlations between artifacts and behavioral processes (Binford 1981; Schiffer 1983).

By the 1970s, quantitative archaeology had begun to consider some of the conceptual elements of Shannon's information theory, such as its introduction in Doran's (1970) early applications of systems theory and simulation. Justeson (1973) notes the work of Fred Gorman (1970) as possibly the first *formal* mathematical application of Shannon's theory to the quantitative analysis of archaeological collections. Justeson's own offering (1973) explored the potential for information-theoretic concepts as a tool for addressing inferential challenges in archaeology. He posited that Shannon's entropy measures could be applied to quantifying the integrity of archaeological *signals* by considering the formation of the archaeological record itself in terms of *channel* and *capacity.* The rapid developments in computer applications further fueled this expansion, facilitating the adoption of statistical methods across archaeological contexts (Kintigh and Ammerman 1982; Kintigh 1984; Richards 1998; Djindjian 2015).

The slow adoption and application of methods increasingly influenced by Shannon's information theory (albeit not often explicitly so) allowed archaeologists to assess patterns in artifact distribution and site organization with new mathematical precision, later inspiring applications in cultural transmission and inter-group interaction studies (Dickens and Fraser 1984). Through the 1980s, however, critiques of quantitative and rigidly "science-like" approaches to cultural phenomena began to emerge, primarily from post-processual theorists who argued for subjective interpretations and a focus on agency and meaning rather than structural functionalism (Klejn 1977).

Despite critiques, quantitative methods, including information-theoretic approaches, continued to evolve and become an indispensible feature of archaeological methods. In recent years, advances in entropy and information measures emerging from developments in machine learning and data science have been increasingly applied to artifact analysis, as seen in works such as Paige and Perreault (2022) or Río, López-Hernández, and Chaparro Velázquez (2024), who employed entropy to examine variability in stone tool production strategies. These newer studies align with a trend towards recognizing the flexibility of quantitative tools to address a broader array of archaeological questions, marking a shift away from the purely deterministic frameworks of early Processualism towards a more nuanced integration within contemporary theoretical paradigms (Nolan 2020; Gheorghiade et al. 2023).

From the 1960s to today, the trajectory of quantitative analysis in archaeology has seen waves of theoretical support and critique. Early Processualism lauded quantitative rigor for its perceived objectivity, while post-processual scholars raised concerns about the reductionist tendencies of such models. Over time, however, quantitative methods – including applications of Shannon's information theory – have found renewed relevance as tools capable of balancing empirical rigor with interpretative flexibility, a balance increasingly valued in today's archaeological discourse.

References Cited

Binford, Lewis R. 1981. "Behavioral Archaeology and the "Pompeii Premise"." *Journal of Anthropological Research* 37 (3): 195–208. https://doi.org/10.1086/jar.37.3.3629723.

Brainerd, George W. 1951. "The Place of Chronological Ordering in Archaeological Analysis." *American Antiquity* 16 (4): 301–13. https://doi.org/10.2307/276979.

Clarke, David L., ed. 1972. *Models in Archaeology*. London: Methuen & Co.

———. 1978. *Analytical Archaeology*. 2. ed. New York: Columbia University Press.

Dickens, Roy S., and Martin D. Fraser. 1984. "An Information-Theoretic Approach to the Analysis of Cultural Interaction in the Middle Woodland Period." *Southeastern Archaeology* 3 (2): 144–52.

Djindjian, François. 2015. "Computers and Mathematics in Archaeology, Anatomy of an Ineluctable Success!" In *CAA2014: 21st Century Archaeology*, edited by F. Giligny, F. Djindjian, L. Costa, P. Moscati, and S. Robert, 1–6. Archaeopress Publishing Ltd. https://doi.org/10.2307/jj.15135883.4.

Doran, James. 1970. "Systems Theory, Computer Simulations and Archaeology." *World Archaeology* 1 (3): 289–98. https://doi.org/10.1080/00438243.1970.9979448.

Gheorghiade, Paula, Vaiva Vasiliauskaite, Aleksandr Diachenko, Henry Price, Tim Evans, and Ray Rivers. 2023. "Entropology: An Information-Theoretic Approach to Understanding Archaeological Data." *Journal of Archaeological Method and Theory* 30 (4): 1109–41. https://doi.org/10.1007/s10816-023-09627-4.

Gorman, Frederick. 1970. "Symbol as Strategy: Toward a Synthetic View of Material Culture." Field Museum of Natural History.

Heizer, Robert F., and Sherburne F. Cook. 1956. "Some Aspects of the Quantitative Approach in Archaeology." *Southwestern Journal of Anthropology* 12 (3): 229–48. https://doi.org/10.1086/soutjanth.12.3.3629082.

Justeson, John S. 1973. "Limitations of Archaeological Inference: An Information-Theoretic Approach with Applications in Methodology." *American Antiquity* 38 (2): 131–49. https://doi.org/10.2307/279360.

Kendall, David G. 1969. "Some Problems and Methods in Statistical Archaeology." *World Archaeology* 1 (1): 68–76. https://doi.org/10.1080/00438243.1969.9979428.

Kintigh, Keith W. 1984. "Measuring Archaeological Diversity by Comparison with Simulated Assemblages." *American Antiquity* 49 (1): 44–54. https://doi.org/10.2307/280511.

Kintigh, Keith W., and Albert J. Ammerman. 1982. "Heuristic Approaches to Spatial Analysis in Archaeology." *American Antiquity* 47 (1): 31–63. https://doi.org/10.2307/280052.

Klejn, Leo S. 1977. "A Panorama of Theoretical Archaeology." *Current Anthropology* 18 (1): 1–42. https://doi.org/10.1086/201846.

Nolan, Kevin C. 2020. "Bringing Archaeology into the Information Age: Entropy, Noise, Channel Capacity, and Information Potential in Archaeological Significance Assessments." *Quality & Quantity* 54 (4): 1171–96. https://doi.org/10.1007/s11135-020-00980-0.

Paige, Jonathan, and Charles Perreault. 2022. "How Surprising Are Lithic Reduction Strategies? The Information Entropy of the Modes a-i Framework." *Lithic Technology* 48 (3): 237–52. https://doi.org/10.1080/01977261.2022.2113699.

Raab, L. Mark, and Albert C. Goodyear. 1984. "Middle-Range Theory in Archaeology: A Critical Review of Origins and Applications." *American Antiquity* 49 (2): 255–68. https://doi.org/10.2307/280018.

Richards, Julian D. 1998. "Recent Trends in Computer Applications in Archaeology." *Journal of Archaeological Research* 6 (4): 331–82. https://doi.org/10.1007/bf02446083.

Río, Fernando del, Rodrigo López-Hernández, and Claudia Chaparro Velázquez. 2024. "On Information, Entropy, and Early Stone Tools." *Molecular Physics*, February. https://doi.org/10.1080/00268976.2024.2310644.

Schiffer, Michael B. 1983. "Toward the Identification of Formation Processes." *American Antiquity* 48 (4): 675–706. https://doi.org/10.2307/279771.

Shannon, Claude E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (4): 623–56. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x.

Shannon, Claude E., and Warren Weaver. 1949. *The Mathematical Theory of Communication.* Urbana, Illinois: University of Illinois Press.