Why is there Math in my Archaeology?

John Justeson and the Limits of Archaeological 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.

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 a book 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 – and, of course, archaeology.

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 of the 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., $lim_{p(x)\to 0}\ I(x) = \infty$).

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

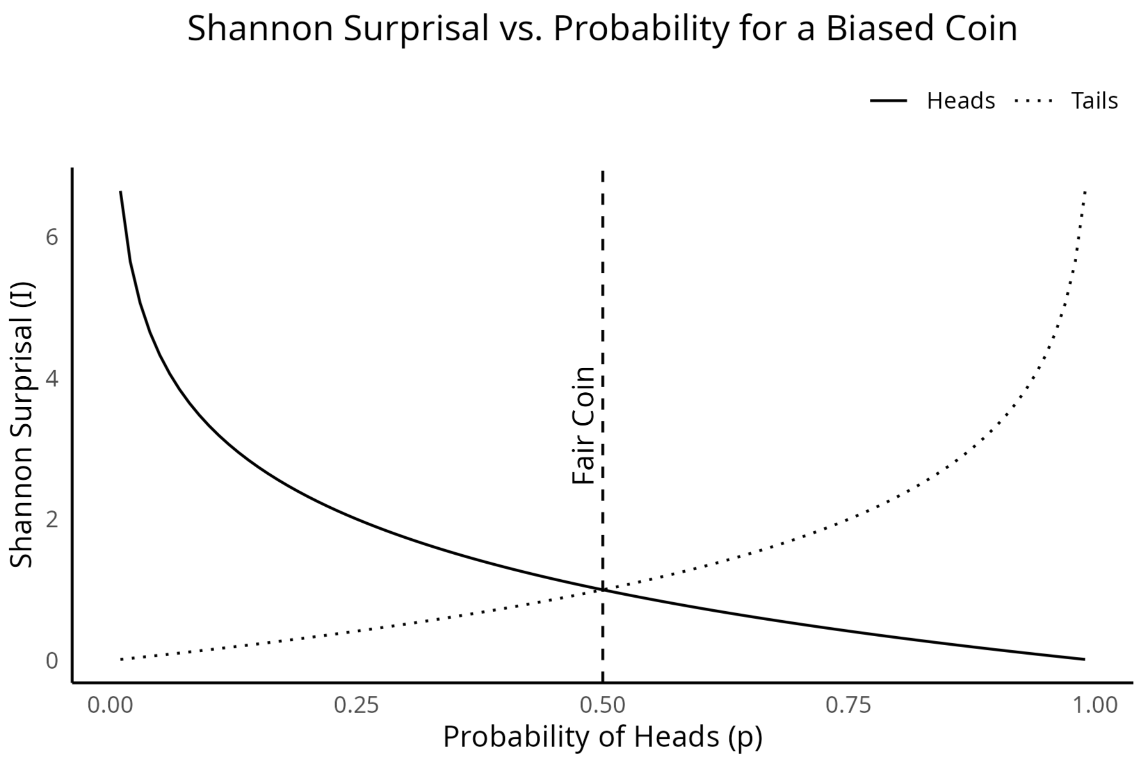

Figure 1: The surprisal $I(x)$ of a coin flip x (i.e., lands "heads" or "tails") as the probability $p(x)$ of landing "heads" ranges from 0 to 1 for a "biased" coin. A "fair" coin would land on heads or tails with equal chances or $p(x) = 0.5$.

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.

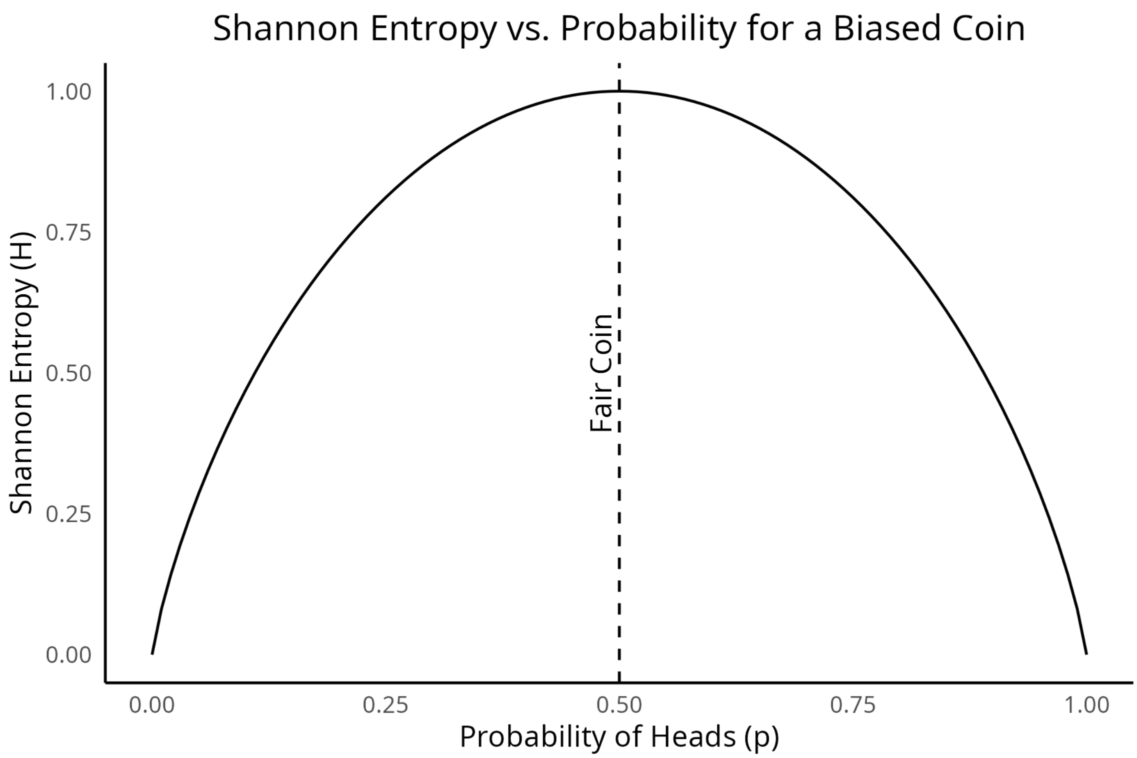

Figure 2: The overall system entropy $H(X)$ for biased coin flips as the probability $p(x)$ of landing "heads" ranges from 0 to 1. A "fair" coin with $p(x) = 0.5$ is the system with the most uncertainty, since either outcome ("heads" or "tails") is equally possible.

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.

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:

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

where $I(X|Y)$ 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.

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.

[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 $I(X|Y)$ is the information provided by X when we *already* know that Y has occurred.

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.

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, methodological innovations in quantifying archaeological analysis, in works such as Brainerd (1951) or 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).

[3] This method of calculating 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.

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 demonstrated how 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 such 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), employing 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 between methods and theories (Nolan 2020; Gheorghiade et al. 2023).

From the 1950s to today, quantitative analysis in archaeology has progressed somewhat independently from larger theoretical movements and critiques. Rather than the overarching epistemological ambitions of the early "New" or Processual archaeologists for a scientific objectivism, quantitative methods have instead become part of the standard toolkit of contemporary archaeological practice.

The Evolution of Information Theory in Archaeology

The integration of *formal* information theory into archaeological research has had a much slower evolution. In the late 1960s and early 1970s, inspired by Shannon's ideas, archaeologists began to experiment with these concepts to analyze the transmission of cultural traits, the integrity of artifacts, and the uncertainty inherent in archaeological data. More explicit and formal applications Shannon's model were much slower to appear. The mathematical and computational complexity of such models largely exceeded the capabilities at the time, and there remained substantial debate regarding their limitations in addressing the complexities of human behavior and cultural evolution.

Although not the first, Michael Schiffer's work (1972) is likely the best known of the early applications of a theory of information in archaeology. He tried to formalize the disruption of information flow caused by post-depositional processes, termed as "systemic and archaeological contexts." Schiffer emphasized how the information contained within artifacts could degrade over time due to various environmental and cultural factors, introducing "noise" into the archaeological record. This idea aligned with Shannon's theory of communication, where noise distorts messages as they pass through a channel. Schiffer's subsequent work (1983) on formation processes expanded on this, demonstrating how entropy, a measure of disorder, influences the amount of reliable information that survives in archaeological contexts.

Schiffer's book "Formation Processes of the Archaeological Record" (1987) is still among the most influential applications of information concepts to archaeology, even though Schiffer did not explicitly use Shannon's framework. Schiffer introduced the idea that archaeological sites are the result of two key processes – i.e, cultural formation processes involving human behaviors that create and modify archaeological deposits, and natural formation processes through non-human agents such as erosion or animal activity that affect the archaeological record.

The notion of information loss in these processes echoed Shannon's concepts of entropy and noise. Schiffer's emphasis on understanding how archaeological data are transformed before and after deposition mirrors the concerns of information theory regarding how signals are distorted through transmission. By conceptualizing the archaeological record as a series of transformations from its original state, Schiffer advanced a model that paralleled Shannon's information processing, where each formation process acts as a filter, introducing "noise" and altering the original "message."

John Justeson (1973), however, explicitly applied Shannon's concepts directly to archaeological inference by demonstrating and quantifying the theoretical limits of reconstructing past human behavior through fragmentary data. He focused on how entropy could quantify uncertainty and signal degradation, although he cautioned that oversimplification sometimes results when human complexity is reduced to mathematical models. Justeson's objective was to try and formalize the analysis of the inherent *limitations* of such interpretations. He derived a complete mathematical formulation to assess whether any given assemblage of archaeological features contained *sufficient* signal to rigorously identify underlying patterns. It directly addressed the tension between abstract quantitative frameworks and the nuances of particular cultural trajectories, a critique that has persisted in the field, but sought out a methodological compromise that could actually *quantify* that inherent uncertainty.

Justeson's work helped begin a dialogue within archaeology about the inherent limitations of inference from incomplete datasets (e.g., Sullivan 1978; Plog 1978; Hayden 1984), contributing to the development of more cautious and methodologically rigorous approaches to interpreting the archaeological record. His use of Shannon's ideas encouraged some archaeologists to critically evaluate the reliability of their data, and the extent to which they could justifiably infer past behaviors or cultural practices. Unfortunately, the sophisticated mathematical and computational understanding required for the article's *quantitative* implications and applications seems to have relegated it to relative obscurity. [4]

It would not be until the 1980s that other scholars substantially applied formal information theory to model cultural interactions. Dickens and Fraser (1984; notably citing Justeson 1973) used Shannon's idea of channel capacity to study the flow of cultural information in the Middle Woodland Period, seeking to quantify how much cultural interaction could be detected within the archaeological record. Similarly, Renfrew (1983) explored the idea of culture as a communication system, where information flows between individuals and groups. He applied Shannon's concept of information transmission to study how cultural signals travel and degrade over time, though he acknowledged the complexity of non-linear dynamics in human societies, which challenge the assumptions of equilibrium-based models.

More recently, the use of information theory in archaeology had broadened, particularly in studies of cultural transmission. Crema, Kandler, and Shennan (2016) advanced Shannon's ideas by applying equilibrium and non-equilibrium models to study cultural transmission from frequency data. They used these models to reveal how cultural traits spread and stabilize within populations, providing quantitative insights into processes that are often difficult to observe directly in the archaeological record. Similarly, Carrignon, Alexander Bentley, and O'Brien (2023) used information theory to estimate transmission rates, applying Shannon's communication model to measure the uncertainty associated with the diffusion of cultural traits.

Gheorghiade et al. (2023) expanded Shannon's concept of entropy into a framework they called "Entropology" that posits entropy measures to better understand archaeological data. They critique the traditional applications of information theory for focusing too narrowly on entropy without accounting for the broader complexity and uncertainty of archaeological contexts. This critique echoes the central debate surrounding the use of information theory in archaeology – i.e., while it offers valuable tools for formalizing the study of cultural transmission and data integrity, it does not capture the intricate and chaotic nature of human historical exigencies.

Another major critique of these applications, such as that by Raab and Goodyear (1984), concerns the oversimplification of human behaviors when abstract models like those derived from Shannon's theories are applied. They argue that middle-range theory, which often uses these models, fails to capture the full complexity of human action. Zubrow (1972) similarly critiqued the difficulty of accounting for environmental and social variables when applying information-theoretic frameworks. Despite this, some scholars such as Nolan (2020) have proposed to refine these models. Nolan assessed entropy, noise, and channel capacity to evaluate the significance (in the technical and regulatory cultural resources sense of the term)

[4] It is worth noting that John had published this article while still a graduate student at Stanford University, before completing his Masters. The article has seen a recent and substantial resurgence of attention, garnering more citations within the last ten years than it had in the previous four decades.

of archaeological data, particularly focusing on how much information about past societies could be accurately recovered from the fragmented and noisy record.

The use of Shannon's information theory in archaeology has evolved from early models of data degradation and cultural transmission to more sophisticated frameworks that incorporate the entropy and uncertainty in teh archaeological record. Scholars like Schiffer (1972), Justeson (1973), and Renfrew (1983) laid the foundation, while modern researchers like Nolan (2020), Crema, Kandler, and Shennan (2016), and Gheorghiade et al. (2023) have expanded these concepts to address the challenges posed by incomplete and noisy archaeological records. However, the ongoing debate highlights the tension between the precision offered by information theory and the complex realities of human history, questioning the extent to which these mathematical models can truly capture the richness of the past.

Ironically, the ensuing debates largely failed to recall that one of them already provided a roadmap for determining *exactly* that extent quite early on in the venture.

The Limitations of Archaeological Inference

In Justeson's 1973 article, he introduces Shannon's theory of communication as a means to formalize the analysis of inherent inferential limitations in archaeological interpretation. In the introduction, he firmly situates the paper within what was, at the time, growing theoretical tensions between conflicting goals within archaeology. Some were advocating for a "new" archaeology focused on "predictive behavioral science" while the "traditional" archaeology's aim was the reconstruction of "social and cultural histories" (Justeson 1973, 131). Justeson viewed the distinction as merely opposing "poles on a continuum of research commitments" and instead posed a slightly different question – is there a way to determine whether or not we were actually capable of doing *either*, given a particular archaeological source?

The article is presented in two parts. The first ("A Theoretical Framework") introduces the relevant aspects of Shannon's theory of communication and posits the analogous relationships between those processes and the nature archaeological data. He makes the argument that it is not sufficient just to describe the archaeological record as an information channel, but that is is specifically a channel of a particular *type* that allow its interpretation. The second part ("Application of Information-Theoretic Measures") illustrates how specific measures for the fidelity or integrity of the archaeological record can be derived from Shannon's model. He presents a set of formal tools through which archaeologists could calculate these specific measures from observations of artifact attributes to determine whether sufficient information existed within an assemblage to be *interpretable*.

Rather than present a predominately conceptual framework, as previous works had done (e.g., Doran 1970; Schiffer 1972; Clarke 1973), the objectives of the article were more ambitious. Justeson aimed to use Shannon's formal models to demonstrate the specific properties and capabilities of the archaeological record to transmit information. Working backwards from the observation that the archaeological record functioned as a communication channel, he showed that such a channel must also possess the formal properties of a certain type of communication network. Furthermore, he showed that the encoding of information carried by that channel would need to take a particular form in order for it to successful convey interpretable information about the past. In essence, "…the archaeologist is in the position of the code-breaker tapping a channel with whose code he is not fully familiar by means of another channel" (Justeson 1973, 134).

"Part I – A Theoretical Framework"

The paper builds from the hypothesis that archaeological interpretation is fundamentally limited by the quantity of information that can be extracted from the archaeological record. Like the other early archaeological invocations of information theory, Justeson described the archaeological record as a degraded and incomplete set of signals from past behavior that are transmitted through the "channel" of the archaeological record. Each artifact or feature would represent a small, noisy fraction of the original cultural system.

Justeson's main departure from the others was in that he applied Shannon's concept of entropy *directly* to the assessment of the degree of uncertainty that might be incorporated into archaeological interpretations simply by the nature of that channel. Remember, Shannon linked reduction of uncertainty to information. Justeson focused on highlighting how *noise* in the archaeological record – due to processes such as taphonomy or excavation biases – interacts with the inherent *entropy* entailed by the archaeological features or processes that *encode* past behaviors. Those interactions innately affect the *capacity* of the archaeological channel to reliably transmit information.

Therefore, the inherent limit of archaeological inference would be the limits (i.e., the "upper bound" in mathematical terms) of the channel's capacity given a certain amount of noise. Past those limits, *de*coding the source signal (i.e. the *behavior*) would become highly susceptible to more ambiguous, unreliable, or even spurious interpretations. To find – and *calculate* – that limit, Justeson needed to specify the nature of the transmission channel and its properties and identify (and prove) the existence of a coherent system of encoding.

"Basic Concepts of Information Theory and Their Archaeological Correlates" Whereas Shannon described information in terms of the reduction of uncertainty, Justeson notes that information can also be thought of in terms of *contrasts*. That is, information can be seen as a way we are able to distinguish the qualities or attributes of one type of thing from those of another. Information, then, is how we determine categories by reducing the uncertainty of correctly assigning a thing or event to a given category.

That information is, in the archaeological case, *encoded* through the deposition of an assemblage that reflects the material expressions of past behaviors as an input message or source *signal*. The message or signal is transmitted across a channel, which has particular characteristics and limitations. These properties allow for the introduction of *noise* that may affect the input signal and alter or obscure the original message. Innate channel properties, along with any noise or distortion of the input signal, results in an output signal that is received and decoded on the other end as the output message (see {#figure:info-channel-schema}).

Channel capacity, then, is a measure of how much information can be transmitted without errors (i.e., *input* = *output*).

Figure 3: Schematic representation of information transmission ("Fig. 1" in Justeson 1973, 133).

Justeson points out, however, that archaeologists are actually dealing with *two* distinct communication systems. Essentially, the archaeological channel input is itself the *output* of the cultural system that *produced* the assemblage. The formation of archaeological deposits is preceded by a cultural encoding resulting in the production of the original assemblage, which is itself the material consequent of human behaviors. The archaeological record (i.e., *channel*) is therefore the *secondary* channel through which we, as archaeologists, are attempting to "break" the code of the original *behavioral* encoding with which we are unfamiliar.

"Channel Classification, Channel Properties, and Codes" Describing the archaeological record as a "channel" is a useful abstraction, of course, but far more important is to identify the nature and properties of that channel. Shannon's work was describing a specific *type* of communication channel, one in which both the natural properties of the channel and the general system of encoding for signals were known. Telephone or radio transmissions are useful heuristics for understanding the basics of Shannon's theory, but neither are properly analogous to the archaeological case. For an archaeological channel, these are either unknowns or the *objective* of analysis.

To actually *model* the archaeological record as a communications channel, and to formalize how the potential capacity of such a channel might derived, required determining its specific type and properties. Futhermore, it required deriving a specification for the encoding of archaeological data that could generalize to the diversity of archaeological data types. Such a code would need to entail a probabilistic basis, as needed by Shannon's definitions for information, as well as reflect the particular nature of archaeological data and how it is recovered. Justeson discusses two general properties of communications channels that broadly distinguish channel types.

The first is whether or not the channel has what is referred to, in statistical terms, as *memory*. In a channel without memory, the elements of any signal or message are unaffected by any other or prior signals. Each transmitted signal is *independent* from any others. In probabilistic terms, this implies that each element or symbol in a message's transmission has a fixed probability distribution unaffected by earlier transmissions. This would make them statistically equivalent to a sequence of independent, identically distributed events. Conversely, in a channel with memory each transmission would potentially be altered by the conditional probability based on prior messages.

The second property is related to the allowable encoding of signals. These are distinguished by whether the set of symbols or elements used to form (i.e., encode) the signal are *finite* or *infinite*. This distinction is somewhat moot in the archaeological case since, as Justeson points out, "…the number of meaningful attributes in any material system is finite, and since both the input and output signals are material systems, then we can say that both the input and output alphabets are finite" (Justeson 1973, 135). Since there are a finite number of possible elements used to encode archaeological signals, the channel is described as *discrete* (as opposed to *continuous*). This primarily has the effect, for the purposes of analysis, of limiting the possible types of probability distributions involved. For archaeological assemblages, these would most often be attribute- or type-frequency probabilities. It also places certain constraints on the *memory* of the channel as well.

The memory of a discrete archaeological channel therefore requires a bit more consideration. In a discrete *memoryless* channel, the presence or absence of any given element or attribute would by definition be *independent* of the presence or absence of any other. That would mean, though, that "…their signals are undecipherable from our vantage point since, being independently transmitted, they show no differentiation or special clustering on any level; independence implies only random associations" (Justeson 1973, 135). Justeson also observes, however, that the memory of an archaeological channel would also have to be finite. Otherwise, the accumulation of noise within the channel over time would eventually render any subsequent signals uninterpretable.

Instead, he suggests, the more appropriate channel type would be a discrete *decreasing-memory* channel, in which prior events have less influence over time. Channels with decreasing memory are often analyzed through the lens of *ergodic processes* and *Markov models*, which help in quantifying the rate at which the influence of the past diminishes. These models would be beneficial in applications where the channel environment evolves slowly over time, where the channel state may vary due to factors like mobility or environmental changes (Cover and Thomas 2001). Unless, however, the rate of decreasing memory is constant – i.e., they are *stationary*, which Justeson notes is not an appropriate assumption for human behavior – such models are highly complex and computationally difficult to analyze.

That leaves the specification of what an archaeological *code* might look like, and it is here where Justeson *really* starts to take a deep dive into the mathematical details.[5] The definition for a code that Justeson presents is derived from Wolfowitz (1961), which requires some explanation. Given that a channel is a discrete channel with finite memory, the code has to conform to certain minimum requirements for it to be predictably encoded and decoded as a signal.

The requirements for the code are deceptively simple:

[5] This is also the point, we expect, that many people (including this chapter's authors) might originally have had some difficulty in following the logic of the article. It requires a certain level of comfort with mathematical and set notation, some basic understanding of set theory, background in statistical and graphical analysis, and a general familiarity with rhetorical style of how mathematical models and proofs are presented. John himself had been a dual-major in anthropology and probability theory as an undergraduate at U.C. Berkeley, so was already well-versed in the "language" so to speak.

1. it must allow *uniquely* distinguishable (i.e., "disjoint") sequences of symbols, attributes, or events to be sent and received; and
2. the probability that any given message received over the channel is the same as the unique sequence that was sent must be ≥ 0.

The first requirement (i.e., *uniqueness*) means that any given sequence of symbols, whether sent or received, can only encode and be decoded as one (and only one) message. The same sequence of symbols cannot have multiple possible decodings. In mathematical terms, then, a signal consists of N combinations (i.e., "ordered pairs") of unique input sequences u_i (sent) and unique output sequences A_i (received), where $i = 1, ..., N$. This is written as

$$\Big\{(u_1, A_1), \ldots, (u_N, A_N)\Big\}$$

The second requirement brings in the probabilistic aspect of information theory. There must be a way to describe a *probability* for whether a sequence sent (u_i) will be the one received and that it matches the correct *de*coding (A_i). This is given by

$$P\left\{v(u_i) \in A_i\right\} \geqslant 1 - \lambda,\ i = 1, \ldots, N$$

…where $v(u_i)$ is what is *actually* received when u_i is sent, and $1 - \lambda \geq 0$ (i.e., to be a probability, $\lambda > 0$ and ≤ 1).

If those requirements are both met, then it is considered a legitimate code with parameters n (the length of sequences), N (the number of sequences or messages), and λ (the measure of probabilities). In the archaeological case, we don't have the source or input sequence u_i, but we do have the output A_i. If we can find a way to determine those three parameters, and they meet the two theoretical requirements, then we are able to identify that there *is* an underlying code and channel to the archaeological assemblage that can be decoded and interpreted.

The point that is easy to miss in this part of the paper is that Justeson's goal here is to provide a way to determine analytically whether or not an archaeological assemblage *meets those minimum requirements* to be considered a code *at all*. If not, "… there is really no basis for speaking of the existence of a channel" (Justeson 1973, 136). This, above all else in the paper, represents Justeson's true theoretical and methodological challenge to the ambitions of the "New Archaeology" as a project. He is saying that unless we can *prove* that there is an *analytically* feasible code entailed by archeological data, there can *be* no supportable interpretation of it.

> If the empirically measured parameters are not consistent with the relationship between them that is required by the theory for a given material or behavioral system, then the data by which that system is to be interpreted cannot have a consistent susceptibility to decoding; that is, there will be no basis for deriving a coherent archaeological interpretation of the data that will accurately reflect

the prehistoric situation. Thus, the question of the existence of a code is one of primary importance for our considerations. (Justeson 1973, 136)

Knowing what type of channel is being dealt with is simply a preliminary to the study of channel properties. It is important to know now what properties of channels are important to consider archaeologically. The major thing to look at here is the code. The code is defined mathematically as a system of N ordered pairs consisting each of an input sequence u_i of n alphabetic symbols and a set A_i of output sequences, where the N sets of output sequences do not have any members in common, and where the probability that the sequence received when u_i is transmitted will be among the members of the set A_i will always be greater than or equal to $1 - \lambda$ where λ is greater than 0 and less than or equal to 1. In symbols, it is "a system

$$\left\{(u_1, A_1), \dots, (u_N, A_N)\right\}$$

where the u_i are n-sequences, the A_i are disjoint sets of n-sequences and

$$P\left\{v(u_i) \in A_i\right\} \geqslant 1 - \lambda,\ i = 1, \dots, N$$

… we shall call it a code (n, N, λ)" (Wolfowitz 1961, 51–52). The expression $v(u_i)$ represents the signal received when u_i is sent, while the term *n-sequence* is an input or output signal of length n.

The parameters n, N, and λ therefore specify the code for the channel. Without such a code we cannot really speak of information being transmitted or received, hence there is really no basis for speaking of the existence of a channel. We can find out if there is a code for he archeological channel by finding if values we compute for N, n, and λ are consistent with the requirements of a code for a discrete finite-memory channel. In particular, the value of N is related to that of n by the formula $N = 2^{n(C-\epsilon)}$, where C is the *channel capacity* – a measure of the ability of the channel to transmit information – and ϵ is a positive constant. C may be determined, often only with great labor, from the relation

$$C = \max_{\pi} \left\{ \sum_j \left[\sum_i \pi_i w(j|i) \log_2 \sum_i \pi_i w(j|i) - \sum_i \pi_i w(j|i) \log_2 \ \pi_i w(j|i) \right] \right\}$$

where $\pi = (\pi_1, \dots, \pi_k)$ is any probability distribution, $w(j|i)$ is the probability of receiving j if i is sent – j can be null – and k is the number of elements in the input alphabet. alphabet. A simple method for calculating an upper bound for the capacity may be found in Helgert (1967); it would require too much discussion for presentation here.

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.

Carrignon, Simon, R. Alexander Bentley, and Michael J. O'Brien. 2023. "Estimating Two Key Dimensions of Cultural Transmission from Archaeological Data." *Journal of Anthropological Archaeology* 72 (December): 101545. https://doi.org/10.1016/j.jaa.2023.101545.

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

———. 1973. "Archaeology: The Loss of Innocence." *Antiquity* 47 (185): 6–18. https://doi.org/10.1017/s0003598x0003461x.

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

Cover, Thomas M., and Joy A. Thomas. 2001. *Elements of Information Theory*. Hoboken, NJ: Wiley-Interscience.

Crema, Enrico R., Anne Kandler, and Stephen Shennan. 2016. "Revealing Patterns of Cultural Transmission from Frequency Data: Equilibrium and Non-Equilibrium Assumptions." *Scientific Reports* 6 (1). https://doi.org/10.1038/srep39122.

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.

Hayden, Brian. 1984. "Are Emic Types Relevant to Archaeology?" *Ethnohistory* 31 (2): 79. https://doi.org/10.2307/482057.

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.

Helgert, H. 1967. "On a Bound for Channel Capacity (Corresp.)." *IEEE Transactions on Information Theory* 13 (1): 124–26. https://doi.org/10.1109/tit.1967.1053939.

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.

Plog, Stephen. 1978. "Social Interaction and Stylistic Similarity: A Reanalysis." In *Advances in Archaeological Method and Theory*, 143–82. Elsevier. https://doi.org/10.1016/b978-0-12-003101-6.50011-1.

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.

Renfrew, A. Colin. 1983. "Divided We Stand: Aspects of Archaeology and Information." *American Antiquity* 48 (1): 3–16. https://doi.org/10.2307/279815.

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. 1972. "Archaeological Context and Systemic Context." *American Antiquity* 37 (2): 156–65. https://doi.org/10.2307/278203.

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

———. 1987. *Formation Processes of the Archaeological Record*. University of New Mexico Press.

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.

Sullivan, Alan P. 1978. "Inference and Evidence in Archaeology: A Discussion of the Conceptual Problems." In *Advances in Archaeological Method and Theory*, edited by Michael B. Schiffer, 1:183–222. Elsevier. https://doi.org/10.1016/b978-0-12-003101-6.50012-3.

Wolfowitz, Jacob. 1961. *Coding Theorems of Information Theory*. Englewood Cliffs, N.J.,:

Prentice-Hall.

Zubrow, Ezra B. W. 1972. "Environment, Subsistence, and Society: The Changing Archaeological Perspective." *Annual Review of Anthropology* 1 (1): 179–206. https://doi.org/10.1146/annurev.an.01.100172.001143.