The Pursuit of Meaning:

Information, Language and the Tangles of Complexity

David J. Galas

"Frequently the messages have meaning."

--*Claude Shannon* (1948)

Since the dawn of information theory a protracted discussion about what information really is and what it means to be complex has refused to die. What it means to be simple or random can be equally perplexing, and there is good reason for much of the discussion and confusion. These are hard questions, and definitions have been either vague or very restrictive and the major issues actually lie beyond the usual discourse on complexity and information theory.

The ideas about transmission of messages, entropy and information, randomness and the notion of computability in the Turing sense all entered into this discourse and debate on simplicity, complexity and information, and most of the key questions were raised early on and remain largely unchanged, with many unanswered. Modern computing, and the rise of machine learning and artificial intelligence have heightened the tenor of the debate, and some would say, have sharpened the consequences, but it remains clear now as then that the notions of meaning and Shannon's information are clearly entirely different concepts. James Gleick wrote a popular book in 2011 called "The Information" [1]. The very fact that a popular book has been written to some acclaim underlines the new penetration of some of these ideas and questions into the public consciousness, least into the public's mild curiosity about what information means and how the word and the substance is properly used in modern life. Gleick pulls together a networked fabric of stories and ideas of the modern conception of information, algorithmic complexity and related topics, in an engaging, rich and very informative book (no pun intended). The questions of meaning continue to lurk among the ideas, advances and subtleties, however, and do not get answered. Perhaps something new is needed.

Before trying to advance my own unorthodox approach to the problematic questions, it might be useful to put the information ideas, and the computing with information into some context. Before information theory and its ideas reached the general community of scientists, mathematicians and engineers[1], a controversy had raged in the halls of mathematics and philosophy about logic, the foundations of mathematics and whether machines could be devised to find and prove all knowable mathematical truth. The argument was that logic, if carefully formulated could be used to determine what is consistent and true and what is false. Russell and Whitehead struggled to lay the foundations of mathematics in logic and had run into some problems by 1900 when the German mathematician, David Hilbert, laid out his list of the 26 major problems of 20th century mathematics, and espoused the view that all could be clear if mathematicians simply went to work proving and disproving the truths and falsities among them. This was a notable event in the history of mathematics though not everyone agreed with Hilbert. A young man from Vienna, Kurt Gödel, however, permanently changed the world of mathematics shortly after this. He proved, not guessed, not opined, but proved, that there are mathematical truths that are unprovable in any way, no matter how advanced and well systematized our logic may be. To be clear, this means that there are mathematical truths, things that are real, valid and reliable – true - that cannot be proven. We may suspect they are true because there are so many examples, and no known counterexamples, but we can never know if there are counterexamples somewhere out there. This result, shocked, depressed and confused many, and the field seemed to drift in some ways, until many solid new results were advanced and were proved, new ideas formulated and mathematicians resigned themselves to Godel’s theorem and simple got on with it. Absolute truth was elusive, it seemed, but only in the sense that it was never complete. Proof of a mathematical statement was still truth, it was just not always available. Another young mathematician got on with new ideas

[1] Note that Shannon was not the first to advance the basic ideas here. Nyquist had already formulated the key concepts almost 20 years before [10, 11]. They were largely ignored.

about computing machines. This was Allan Turing who simply got on with working on what could be computed and what could be proven. A few years after this, in 1948, information theory was expounded by Shannon [2]. The central idea here was that any given message, a series of letters, transmitted over a channel of some kind, like a telegraph, could be quantified in terms of all of the possible messages of that same length using the same number of letters. The idea of quantitating the probability that a particular message could be received by chance was the machine behind quantifying information. It was, of course, completely independent of what was in the message in terms of English, Russian, German, or code words. The quantity was based on the "unlikelihood" of receiving exactly that message. Shannon showed how specific alphabets of letters, the lengths of messages, and the possibility of noise in the channel, possible mistakes in the transmission could affect the information in a consistent, quantitative theory. This enormously influential body of work was called information theory.

In the wake of these ideas a number of mathematicians provided extended analyses and new ideas. Andrei Kolmogorov offered multiple ways of thinking about and quantitating information, and Gregory Chaitin, and Ray Solomonoff independently came up with their own ideas about what is now called algorithmic complexity, one of the ways Kolmogorov presented to calculate information. This should be, and often is, now called Kolmogorov-Chaitin-Solomonoff complexity, or KCS complexity. A major catch, for all the elegance and incisive power of this idea is that it used the ideas of Turing to describe how the quantity should be computed, but was actually incomputable. We will briefly come back to this interesting point at the end of this essay.

We recently proposed[2] a mathematical pathway between the calculation of functional measures of information, like Shannon's and Kolmogorov's, and the theory of finite groups, a foundational structure of modern mathematics. Our proposal is actually a simple idea, but it seems to carry some significant potential for making some new connections on several levels. What this informal essay and reflections is intended to convey is just some of these connections. My intent is to expose the possibility of some really new ideas and new connections on these important topics. The connection between information and meaning, and the exposure of new ideas in this area is

[2] The paper referred to here is the unpublished manuscript, "The Group Theoretic Basis of Entropy: Information, Symmetry and Complexity" [1]

where we would like to focus attention. As Kolmogorov says, it is often very difficult to see where the boundary is between the simple and the complex, between the trivial and the impossible.

> *"At each moment there is only a fine layer between the trivial and the impossible. Mathematical discoveries are made in this layer"*
>
> *--Andrei Kolmogorov*

The realization that the original mathematical quantity, and I mean original in several significant ways, Shannon information-entropy, is actually the limiting case of a much more general concept about information and structure is an exciting and provocative notion. You must forgive me for seeing this as fundamental, and anything but trivial. The idea has the potential force to make us ask about what it really means, and whether it may extend beyond the ideas of Shannon and Kolmogorov. This possibility is certainly interesting, but on examination it could have an annoying flaw. It may inevitably draw us back to some time-worn arguments about the meaning of Shannon information, some we mentioned before. But I now argue that some of the value of the arguments and ideas about generalized integer entropy may lie in raising the time worn arguments again, albeit with the inherent confusion that accompanies much of the abstractions about information and its meaning. I think it is fully worth re-stirring the pot of the old controversy about the flaws of attribution of meaning to information, so we will.

The time-worn arguments I referred to above actually arise from the problematic attributions that were improperly injected into "information theory" in the very early days. There were serious misconceptions about what he had proposed that were properly rejected by Shannon, Wiener and others as completely irrelevant to Shannon's "information theory." These misconceptions were based on the idea that "information theory" had anything to do with the common ideas in the vernacular about information, its informal meaning, and the evocations of ideas that the word invoked, as it was thought of in the 1950's anyway. These flawed notions were opposed to the rather simple idea of Shannon's that information was just the measure of the transmission of bits distinguishing among all possible messages. The notion of what number of messages is possible versus what is actually transmitted is the fundamental idea here and much simpler than the evocation of meaning. This notion of "communications theory", Shannon's idea, versus "information theory," was a confusion, not really an argument of substance, but of mathematical reality versus wishful thinking. Many who engaged this debate in the 1950's wanted desperately to know that finally there was a theoretical basis that opened to door for them to learn about what information, knowledge or meaning was by applying Shannon's theory. This key insightful idea was, of course, a nonsensical misunderstanding of the ideas, and Shannon and Norbert Weiner argued strongly against it.

"The past is a foreign country. They do things differently there."

--L.P. Hartley

In this spirit we now ask whether the realization that Shannon information is a limiting case of something larger could possibly be a useful idea, not just in dealing with these rather shallow misunderstandings, but in extending the ideas of "information theory" beyond the ideas and formulation of Shannon's communications theory, and actually dealing with some of the misunderstood meaning. The applications of the ideas of "information theory" to statistical physics by Jaynes, had a dramatic impact well beyond the ideas of the statistics of communications. It would be powerful indeed if such an extension were possible, even if it were a small excursion into new and foreign ground, as it would have repercussions in many areas. I think it might be possible that we have found the pathway to such an extension, and perhaps a brief non-mathematical divagation here may help justify this suggestion.

Consider transmitting a piece of English writing, like Keats' poem "Ode on a Grecian Urn," over a noise-free communications channel. The ones and zeros representing the letters of the words of this poem can easily be produced, transmitted electronically and extracted on the other end. If we try to quantitate the amount of "information" transmitted, and adhere to Shannon's rigorous mathematical constraints, we can indeed arrive at a number, and this number has a precise meaning. What this number tells us is how many bits we need to distinguish this poem from all other possible transmissions of 1's and 0's of a similar length, a very large set of possibilities, but a clear process. We could alter this calculation slightly and lump the bits into clumps based on words. Then we would ask what are all possible words, in order to quantitate the Shannon information from this vantage point. If we specify, for example, that the unabridged Oxford English Dictionary contains all possible words, this could work, but perhaps not easily.

Nonetheless, this is in the same spirit and the same bit-based distinctions are central. If we are allowed to throw in a few words of slang, or French or Spanish words that most people can recognize it makes things better, perhaps a bit fuzzy, but it is not conceptually different. What should we do, however, with a clump that is not recognized by our criteria as a "word"? This problem is technically hard, but this is still not fundamentally difficult, since it amounts to decisions about a look-up dictionary of words, however large. If we ask, on the other hand, that we only consider proper English sentences in calculating the information content the problem suddenly gets much worse. Here I need an agreed upon English grammar that includes rules for all possible sentences using all possible words and all possible punctuation, in addition to a dictionary. This is difficult, and is much harder, though not at all impossible.

Now we have to plunge into the woods even more deeply since what we would really like to include is only any possible "meanings" of an English sentence. So a grammatically correct sentence that is ambiguous or nonsensical has no content of this kind. This is harder yet, by a much larger margin. But now I want to go the full distance and include all possible poems, and even further I want to include only all possible poems that have a literary quality above a minimal threshold of meaning and artistic value, conveying thoughts and invoking feelings based on the shared experiences of a certain class of educated English-speaking readers who are inured to poetry. Can I use a Shannon-like criterion to do this? I think we have now ventured far beyond Shannon's measures with these criteria, even with any known modifications and extensions. We have crossed from communications theory into the most complex issues of "meaning," and while these issues are poorly formulated here at the moment, it's clear that Shannon and Wiener would probably rage over the idea that there's anything that "information theory" can provide to this problem of meaning. I would agree, however, that it's a pointless exercise to try to fit the Shannon mathematics to these problems. It is clear that we are facing here a vastly more fundamental problem that just enumerating all examples of the messages we may need to deal with. Perhaps the difficulty can be illustrated and epitomized by quoting from Keats' poem:

What leaf-fring'd legend haunts about thy shape
Of deities or mortals, or of both,
In Tempe or the dales of Arcady?
What men or gods are these? What maidens loth?

What mad pursuit? What struggle to escape?
What pipes and timbrels? What wild ecstasy?
Heard melodies are sweet, but those unheard
Are sweeter; therefore, ye soft pipes, play on;
Not to the sensual ear, but, more endear'd,
Pipe to the spirit ditties of no tone.

What mad pursuit, indeed. We need something new, well beyond Shannon to begin grappling with the function space of poems like Keats', a worthy challenge. The general integer entropy ideas here are, I think, a step in the right direction, but surely do not address the full slate of conceptual issues raised in our notes and discussion. It actually doesn't deal directly with any of them, but it suggests, I would argue, a direction in which a solution might lie, and that is worth a lot.

An interesting viewpoint that we might take away from the ideas of entropy and the notion of Shannon information as a concept that emerges only at the limit of a more general range of ideas and measures, is in this picture. The essential pairing of a group, defining a specific class of symmetries and structures, and an information measure, a function of the distribution of objects of a set, is the central idea. Shannon-like entropy has its pairing with the Symmetric group (S_N) because this group, the group of all possible permutations shuffles objects in an ordered set in all possible ways, and thus generates no information in itself (I am now using the term information in a very loose and more general way, not confining the term to what can be measured by Shannon.) Since S_N consists of all possible permutations, I need nothing to specify "all" possible ways of jumbling things up. If I limit the possible jumbling by disallowing some permutations or classes of permutations, thus defining a subgroup, I need to specify which permutation operators, elements of the group, are disallowed, and this is essential information in defining the information measure that is paired with the group, and its structure. Structure can emerge as one disallows certain kinds of jumbling. So to summarize this discussion, we could say that the entropy results from the collision of the information in the data itself, ordered sets, with the information in specifying the paired group. Another metaphor is that the two kinds of "information" come together in a complementary process to give a more significant, deeper measure of "information." The metaphor attributed to Michelangelo captures this meaning well.

"The sculpture is already complete within the marble block, before I start my work.
It is already there, I just have to chisel away the superfluous material."

-Michelangelo Buonarroti

He said that to produce a great sculpture one simply starts with a fine, large block of white Carerra marble, and removes all parts that are not part of a magnificent sculpture. The block is S_N. If one removes all the marble (all of S_N) one has nothing, but if one disallows only the right, forbidden strokes of the chisel one removes just the superfluous permutations and is left with a magnificent piece of art.

In the discourse on integer entropy we have used the idea of the cycles induced by permutations as the key to identifying the subgroup, but keep in mind that Caley's theorem reveals the amazing truth that all finite groups, that is, all possible finite symmetries, can be represented by subgroups of the permutations. This idea is amazing – all possible symmetries, all, are hiding in the block of marble we call the symmetric group. The notion of all possible symmetries is a stunning notion. A visual analogy illustrates. If one has looked upon, for example, the patterns in Islamic and other art traditions, all discrete rotations, the patterns in cellular automata, the drawings of M.C. Escher, and many other artists, one can begin to see the enormity of this idea. A door into a garden with enormous potential, the basic tools, the definition and exploration of functional representations of order, disorder and meaning, is opened by this notion. I think this general idea is worth repeating. Caley's theorem says that the block of marble is S_N, in which reside, in potential, all possible sculptures, all possible finite groups, and the general integer entropy functions may then take their measure in some sense that is not yet concrete and clear, to define what must chiseled away. I readily acknowledge that this metaphor and its implications needs to be much more carefully worked out, but I think the door is evidently in this vicinity, and is just a little open. The functions we are discussing have only N bits (including the identity) in that it tells us which sets of permutations to include in the subgroup based on how many disjoint, non-empty cycles that set induces. The finite groups are an enormous set that represent all possible symmetries of a finite set of objects. We consider these ideas as specific tools for the implementation of new ideas, a new approach to characterizing information.

How could the mathematical idea just summarized address the first steps of meaning we used to limit the Shannon construct? Take the idea of considering only words in a message. How could we use the information in a subgroup to deal with this? If we start with a set of words in my allowed messages then I want a subgroup that allows all permutations of letters (the basic alphabet) that shuffles words, or creates new words, but on the other hand excludes all permutations that create non-words. The specification information of the subgroup must be what provides this distinction. The synthesis of this information with the mathematical structure of the measure that is specifically paired with the subgroup produces a meaningful measure. We don't really know much about these measures, how the constraints in the subgroups interact with the measure functions, and what the properties of these measures are.

The escalation of this general idea up the ladder of levels of meaning, to sentences and beyond is clearer in principle, but very unclear in practice. Given a sentence I want a subgroup that can create other sentences, but I want to delete all permutation operators that create non-sentences, and so on. Now with this idea on the table many questions arise. How does one actually do this in practice, in the sense of defining subgroups with the appropriate properties? How far can this be taken? Is there something like an "alphabet of meaning"?

In exploring the mathematical literature about pattern-avoiding permutations I found many interesting connections indicated between permutations, specifically subgroups of S_N, and various mathematical objects and structures. I recently came across a link with formal languages, which would seem to be a fruitful path to follow in a quest for meaning. This paper [5] suggests that pattern avoiding permutations generate groups that can be mapped to languages, and therefore, indirectly, that the information measure we defined as integer entropy can be used to describe the information content of context-dependent formal languages, definitely a step in the direction we are advocating. The measures are defined by the subgroup of permitted permutations, and the pattern-avoiding permutations [4] define such a group. The connection between the pattern-avoiding permutations and context-sensitive formal languages made by M. Elder [5] connects general integer entropy and these languages (see the abstract.) While we can't see the detailed connections with our ideas at this stage, I feel that Elder's ideas should prove to be a significant step towards moving information theory towards the inclusion of meaning in its scope of measures. What this seems to suggest is that a finite group, or groups, and the measures of a given passage in a language [6,7] together would measure somehow the meaning in that passage.

Let's return now to think a little more about Kolmogorov's ideas about algorithmic complexity, and provide a final note to our set of speculations. This thought is particularly speculative. The difference between the information concept of Kolmogorov-Chaitin-Solomonoff , which we now call Kolmogorov complexity, or KCS complexity, and Shannon's idea is profound in the sense that KCS is not based on the set of all possible messages within some constraints, but on the measure of the minimum size of a program that could output a given message. This idea is profoundly different, and leads to very useful ideas about computing, but does not provide any way of actually computing the complexity, unlike the Shannon information. The idea of a computer program to generate an output, and arrangement of symbols seems to have a parallel in the group elements that create arrangements out of previous arrangements, and the complexities of a formal language that can express meaning in subtle and context-sensitive ways. However, we are stopped here. More thought, new ideas perhaps, and certainly more work lies before us. There is, nonetheless, a sense in which our theoretical construct seems to settle nicely between Shannon and Kolmogorov. To understand this well will require an extensive investigation and discourse and much more thinking and digging, which we look forward to engaging. The idea here is not yet a program but simply a suggestion. Between the beautiful ideas of Shannon and Kolmogorov there lies a rich and fertile ground for exploration and meaning.

Can the time-worn debate we first referred to, meaning versus the statistics of messages, be injected with new life? I'm sure it can be, perhaps not by us, but certainly. Can a new direction here provide the ground for moving towards a theory of meaning grounded in information theory? Perhaps not yet, but the questions raised suggest there are some intuitive analogies, and possible directions indicated. After all, the structure of languages are devised to convey meaning, and there must be a way to measure this essential quantity, even if a complete alphabet of meaning remains inaccessible. Analogies are the wings that keep ideas aloft. They are the first direction signs of insight, but also sometimes the garlands of fantasy. We may already be in the layer that Kolmogorov cites between the trivial and the impossible, and that is almost enough to give meaning and interest to the quest.

References:

1. James Gleick, "The Information: A history, A theory, A flood", Pantheon Books, New York (2011)
2. Claude Shannon, "A mathematical theory of Communication," *Bell System Technical Journal,* (1948)
3. Doxiadis, Apostolos, and Papademetriou, Christos, "Logicomix: an epic search for truth," Bloomsbury (2017)
4. David J. Galas, "The Group theoretic basis of entropy: Information, symmetry and complexity", unpublished manuscript (2017)
5. Herbert S. Wilf, "The patterns of permutations," *Discrete Math*. 257 (2002) 575–583.
6. Murray Elder, "Pattern avoiding permutations are context-sensitive," arXiv0412019 (https://arxiv.org/pdf/math/0412019.pdf) *Abstract: We prove that a variant of the insertion encoding of Albert, Linton and Ruskuc for any class of pattern avoiding permutations is context-sensitive. It follows that every finitely based class of permutations bijects to a context-sensitive language.*
7. N. Chomsky, M. P. Schutzenberger, "The algebraic theory of context-free languages", in: *Computer programming and formal systems*, North-Holland, Amsterdam, (1963), pp. 118–161.
8. J. E. Hopcroft, J. D. Ullman, *Introduction to automata theory, languages, and computation*, Addison-Wesley Publishing Co., Reading, Mass., 1979, Addison-Wesley Series in Computer Science.
9. R. I. Grigorchuk, A. Mach`ı, "An example of an indexed language of intermediate growth", *Theoret. Comput. Sci*. 215 (1-2) (1999) 325–327.
10. Nyquist, Harry, "Certain factors affecting telegraph speed". *Bell System Technical Journal*, 3, 324–346, (1924); and Nyquist, Harry, "Certain topics in telegraph transmission theory", Trans. AIEE, vol. 47, 617-624, Apr. (1928) Reprinted as a classic paper in: *Proc. IEEE, Vol. 90, No. 2, Feb 2002*
11. Hartley, Ralph V. I. "Transmission of Information," Bell System Technical Journal, 7, 535-563 (1928)