It From Constrained Bit*

By Karl Coryat

Abstract. The Shannon-derived measure of surprisal, or the self-information of a message, is calculated relative to some contextual framework. Enriching context constrains a message's potential interpretation, typically enriching its information content in the process. This may have implications for a general informational theory in physics: The receipt of information by a system creates boundary conditions that constrain further new information, the receipt of which then imposes further boundary conditions, and so on. Such ever-tightening informational constraint, iterated over billions of years, may drive the evolution of complexity in an "it from bit" universe.

THE SCENE: A ship somewhere in the middle of the ocean, sometime in the middle of the 20th century. A student of the science of information theory intercepts a strange, repeating signal of long and short tones:

---———---   ---———---   ---———---   ---———--- ....

Although bizarrely ignorant of Morse code, the student is familiar with Claude Shannon's theory relating information and entropy [1]. He notices that the message uses an alphabet of three characters — short, long, space — and that it can be represented in full by a string of only ten characters: short short short long long long short short short space. As an exercise, he calculates the self-information value or "surprisal" of the message [2]:

$$\mathrm{I}(\omega) = -\log_2 p(\omega) \tag{1}$$

where the $\mathrm{I}(\omega)$ is the content of the message ω in bits, and $p(\omega)$ is the probability of the string of characters. Given that the student has no conception of the characters' meaning or their typical frequency in this strange language, he must assign each character's probability based only upon its appearance frequency in the ten-character message: short, 0.6; long, 0.3; space, 0.1. A quick calculation reveals that $p(\omega)$ equals $(.6^6)(.3^3)(.1^1)$, or .000126; therefore, the message carries 12.95 bits of information according to this interpretation scheme. The student declares that the message is unremarkable, and tosses it into the trash.

Fortunately, the ship's captain also receives the message. For the captain, who knows not only Morse code but also international distress signals, the message communicates that another vessel is in danger. Furthermore, the captain has independent knowledge that the only other ship within range is the S.S. John Wheeler, and finally, that the Wheeler has 2,048 souls onboard. From only 12.95 bits of generic information — properly interpreted within the context of Morse code, historically agreed-upon distress signals, the ships' positions, and the Wheeler's known passenger capacity — emerges a rich and terrifying reality: that thousands of human lives are in imminent peril.

---

* Entry for 2013 FQXi essay competition, "It from bit or bit from it?" Contact: tenrec@pacbell.net

Wheeler’s Its & Shannon’s Bits

When John Wheeler proposed that the universe is ultimately informational [3], that every thing, or “it,” emerges from fundamental informational units, or “bits,” he launched a discussion that continues to gain momentum today. But even in the information age, in many ways the enquiry into the possibly fundamental role of information has barely gotten off the ground [4]. Given our primitive understanding, it might seem as if a chunk of cosmic holographic surface needs to fall from the sky — *The Truman Show* meets *Contact* — in order to make any serious headway with “it from bit.” So, baby steps. Might our tale of the student, the captain, and the SOS signal shed light on the complexity of the universe? Examining the manner in which information seems to generate more information may get us “a bit” closer to realizing Wheeler’s vision.

Claude Shannon greatly expanded the understanding of information with his 1948 paper [1], which recognized that the informational capacity of a channel is optimized when the signal entropy is high, and that a channel carries no information when the entropy is zero because the configuration of all messages is known in advance. Shannon entropy has been extended to quantify the content of individual messages, in a measure known as self-information or surprisal [2]. The surprisal of a message (Equation 1) expresses its probability or expectation value as a quantity. For example, information about the flipping of a fair coin will provide the maximum Shannon entropy of 1 bit per flip, so a message of Heads-Heads-Tails-Heads has a surprisal of 4 bits. However, if the coin is not fair and favors tails, then the same message, which now has a lower probability, would have a *higher* surprisal due to the negative log relation. Intuitively, more “surprising” messages have higher surprisal values. A transmission of no information corresponds to a zero surprisal. No surprise there.

As pointed out in any introduction to information theory, bit counts such as these say nothing about the *meaning*, or real useful content, of a message. Instead, they reflect clearly quantifiable statistical measures. Although Shannon’s theory has vast applications in data compression and other technologies, quantifying meaning seems out of reach. Or is it? In several places in his seminal “Information, Physics, Quantum” essay [3], Wheeler looks for clues into how meaning emerges from bits. So, some definition of meaning is in order here.

Information is often described as a reduction of uncertainty about the world [e.g., 5]. If your phone tells you that a bus is arriving in four minutes, your uncertainty about bus times is reduced. In this case your phone is giving you useful information: The information emanating from the screen has meaning to you, and it can be applied in decision-making processes and therefore physical action. If your phone were running a program that rearranged the display’s pixels according to some algorithm, the new pixel arrangement might correspond to exactly the same clock time; it may even have the same surprisal, analyzed with regard to the statistics of pixels displayed on your phone. But without access to a decoding algorithm, you cannot make sense of the information, so it is meaningless to you. It reduces your uncertainty about nothing, beyond the raw arrangement of pixels presently appearing on your screen. For you as the message receiver, it contains little or no semantic information, at least compared to a proper clock display. Even though we don’t yet know how to quantify semantic information or meaning, we *can* say that a jumbled display contains less of it, relative to a person’s needs as they walk to the bus stop.

Semantics is not limited to human beings communicating in language form, or even phones programmed to display the time. Consider the DNA code. The chemical structure of DNA does not necessarily “mean” anything in and of itself. Viewed on a computer printout, out of context, a short sequence of nucleotides tells us very little; it might as well be a random arrangement. A strand of DNA floating alone in the ocean is as functional as a starch molecule. But its structure does produce complex and functional protein structures and even entire organisms, *in the context* of biological mechanisms that read the code and assemble proteins according to pre-existing sets of rules as well as epigenetic influences [6]. The sequence is the same, whether read by human eyes or an RNA polymerase molecule. In the proper context, however, the information in the sequence gains semantic value: It has a meaning, relative to the message's receivers, in this case the biological mechanisms that transcribe and utilize the sequence, as well as those that benefit from the proteins manufactured. The DNA code is virtually nothing without this context; it transforms the molecule from a sugar-based double polymer that might as well be random, to one that transmits the message, "Build proteins like this!"

A distinction can be made between the statistical expectation value of a message (surprisal) and its semantic value, the message's ability to actually *inform* a particular receiver — to reduce uncertainty about some aspect(s) of the world, beyond the statistical expectation value. And, quite clearly, the enrichment of semantic value is related to the enrichment of context, whether that context is RNA polymerase, a pixel-decoding algorithm, knowledge of Morse code, etc. In cases where there is strong context — e.g., possessing knowledge of what is referenced by a time display on a phone, knowing the bus schedule, and so on — we would expect the semantic content (a quantification of meaning) to be greater than the statistical surprisal.

If we are to seriously pursue "it from bit" and the emergence of semantic meaning — arguably, what Wheeler really meant by the "it" — we will first need to confront the problem of quantifying context. Perhaps Shannon can help with that, too.

Quantifying Context

In recent years there has been much discussion regarding the emergence of higher forms of meaning conveyed by information in particular contexts. Unfortunately, context has not been rigorously defined, so it is presently of little use in theories of information. As Vlatko Vedral has put it, "We need some kind of a 'relative information' concept, information that is not only dependent on the probability, but also on its context" [7]. Context may be seen as having no place in a physical theory — that context (like meaning) is a function of language and intelligence and sociology, presently too complex to be quantified in physical interactions. However, notice the parallel between the enrichment of functional usefulness in a DNA sequence, *in context*, and the enrichment of the usefulness of an SOS Morse code message, *in context*. In relation to the student naïve about Morse code, "SOS_" comprises a bland assortment of dots, dashes, and a space, not unlike a DNA sequence examined out of context. For the captain, however, the message becomes greatly enriched in meaning, in the context of the knowledge he brings to the situation. In both cases, context is materially efficacious upon the physical world: When interpreted in context, the DNA code causes proteins to be synthesized, and when interpreted in context, an SOS message causes the launching of rescue efforts and the saving of lives. Contextual information is both real and identifiable, so we ought to be able to quantify it in some sense.

The student on the ship calculates that the surprisal of the SOS message is 12.95 bits. But suppose he learned a party trick and can decode the Morse code characters, even if he doesn't know the letters' significance. Naïve also in matters of linguistics, he has no notion of the frequencies of letters in English, and assumes that all letters and the space are equally probable. (For an information theorist, he is exceptionally uninformed.) In that case, he calculates the surprisal of the four-character message "SOS_" to be $-\log_2(27^{-4})$, or 19.02 bits. Here, having information about Morse code letter equivalents puts the message in a form that leads to a higher calculated surprisal. Surprisal is context-dependent; it is not an absolute quantity.

“SOS_” in Morse code

Context	Surprisal	Informational boost (in context)
None	12.95 bits	–
Letter equivalents only	19.02 bits	6.07 bits
Letter equivalents + frequencies	15.00 bits	–4.02 bits (2.05 bits total)

“QXZ_” in Morse code

Context	Surprisal	Informational boost (in context)
None	16.84 bits	–
Letter equivalents only	19.02 bits	2.18 bits
Letter equivalents + frequencies	32.39 bits	13.37 bits (15.55 bits total)

Figure 1: The surprisal value of Morse-code signals depends not only on the characters in the signal (SOS vs. QXZ), but also upon the context used in the calculation. In most cases, stipulating additional context boosts the calculated surprisal, most significantly in the QXZ case when we consider letter frequencies in English.

Now imagine that the distress signal were QXZ instead of SOS.[1] Naturally, if the student assumes that all letters are equally probable, he will calculate the same surprisal for QXZ as for SOS: 19.02 bits. However, if he knows Morse code *as well as* letter frequencies in English,[2] then the calculated surprisal of “QXZ_” rises to a whopping 32.39 bits (see Fig. 1). This higher surprisal value expresses the fact that the letters Q, X, and Z are rare in English. Therefore, the presence of additional context (letter-frequency statistics) increases the message’s potential to inform: Previously meaningless messages now mean two different things, where SOS = blandness, and QXZ = rarity. The informational boost supplied by additional context can now be quantified: For a receiver of QXZ who accounts for letter frequencies, uncertainty regarding the message’s expectation value is reduced 13.37 bits *further* than for a receiver who interprets with all letters equally likely. Given that this represents a more “meaningful” message, we might refer to the surprisal boost as *semantic entropy*:[3] It measures the increase of a message’s potential to inform, when analyzed in one context relative to another.

As Bernd-Olaf Küppers puts it, “To understand a piece of information of a certain complexity, one always requires background information that is at least of the same complexity” [6]. After all, if we measure the appearance of a particle with a particle detector, any *meaningful* description of the experiment results must include contextual information that frames the event: When did the experiment begin? When did it end? Where was the photomultiplier tube? What particle(s) are we registering? Do we know that a particular sample was nearby? And so on. Without specifying such contextual boundary conditions, merely knowing that a detector clicked is a bare, primitive, meaningless bit. Adding boundary conditions reduces the scope of the event’s possible interpretations. (The experiment started on Wednesday? Then we know the particle didn’t appear on Tuesday.) With each additional constraint, the experiment result’s potential to reduce our uncertainty about the world goes up — presumably with an accompanying increase in its semantic value, were we to quantify that.

---

[1] For simplicity, assume that the distress signal QXZ (like SOS) is transmitted without spaces between the letters, but that the letters can nonetheless be distinguished for translation.

[2] These calculations use statistics from http://www.data-compression.com/english.html.

[3] I have avoided using the term “entropy” in place of surprisal, as many do, but perhaps it is appropriate here.

We can now propose three principles regarding context:

1. A distinction can be made between *new* information (e.g., a particle detector finding something) and *context* — historical, spatial, temporal, or other legacy information that bears upon the interpretation of new information (e.g., the location of the particle detector).

2. Context tends to constrain the functional significance, potential interpretation, or meaning of new information. For example, in the context of biological (or technological) transcription mechanisms, a DNA sequence can self-replicate or build specific proteins. It cannot do these things without the presence of this context.

3. The constraint of new information can be quantified by the increase of its information content in the presence of contextual constraints, compared to the content that would be calculated without those constraints. For more on this idea, see Endnote 1.

Contextual Constraint & "It From Bit"

Legacy information, in the form of contextual constraint, can quantifiably increase new information's potential to inform. When old bits meet new bits they seem to multiply in some sense, an idea that could lead to insights into the generation of complexity. But first comes a critical question: Can the above principles be generalized to include not just thinking beings and the molecules of living systems, but indeed everything in the universe?

If John Wheeler is right about "it from bit," then it seems nature does not have much of a choice. Contextual constraint is ubiquitous in the physical world. Consider a photon from deep space reaching some system. Upon impact, the photon (representing information new to the system) has a wavelength that is a function of context: not only the physical nature of the photon emission, but also the advancement or recession velocity of the system impacted, relative to the source velocity. The photon's wavelength, an objective property, is constrained by these equally objective boundary conditions, which are in place at/prior to emission and impact; the photon may be emitted at one frequency, and then absorbed at a different frequency, each constrained by the objective context of the associated events, relative to one another.

Now it may be argued that assuming context or contextual constraint in such a physical situation is a useless metaphysical exercise. The interaction occurs whether or not we consider context; if there is no interpretation or analysis going on upon the photon's impact, then the notion of contextual constraint may be seen as inappropriate. However, consider this: If the receiving system were sufficiently complex to seek information about the emission event (e.g., a photon detector connected to a computer program), then such information could *only* be arrived at by incorporating the context surrounding the event. It is in this sense that I mean the interaction is a function of context: For any entity that would interpret the significance of a photon's arrival — i.e., how the arrival reduces uncertainty regarding the bigger picture of the systems involved — then context must be taken into account. Without that context, the arrival is a meaningless single bit. Meaning invariably derives from context; no context, no meaning.

Consider also quantum mechanics, where the contextual specifics of an experiment (apparatus orientation, the observable being probed, and so on) specifies the measurement basis. If we set up a double-slit experiment to measure the wavelength of an ensemble of particles, then the positions of the slits, their separation, and the position of the screen all form a contextual basis that physically constrains the pattern on the screen. But, this context also constrains the *meaning* of our measurements. Eliminate any of the contextual boundary conditions, for example by knowing nothing about the slit-separation distance, and our calculations are hopeless. The wavelength remains uncertain without the full contextual description, so the pattern of spots is essentially meaningless to us.

This brings up an important question. In at least some scenarios, as in Endnote 1, enriching context causes an equal enrichment of informational content. It seems generally intuitive that a more constrained interpretation of a message would correspond to the receipt of more information (or perhaps, more useful information). However, in some cases. informational content quantifiably *decreases* under enriched context. In our Morse code example, taking letter frequencies under consideration when analyzing the message SOS causes the calculated surprisal to go *down* by 4.02 bits (see Fig. 1). On one hand this makes sense, because S, O, and the space are common in English text, so their appearance should be relatively unsurprising. But it also says something deeper about expectation, regularity, and order in general — a point that has deep implications for a universe that is highly complex while also being highly ordered.

In his paper on the entropy of English-language text [8], Shannon wrote: “The new method of estimating entropy exploits the fact that anyone speaking a language possesses, implicitly, an enormous knowledge of the statistics of the language.” This knowledge improves the predictability of letters in a string of English text, lowering the entropy to only about 1 bit per letter by his calculations. (Random strings of equally probable letters plus the space would have the maximum entropy of $\log_2 27$, or 4.75 bits per letter.) Understanding a highly complex organizational system such as the English language, with all of its rules and conventions (e.g., Q is always followed by U), decreases the entropy of strings of letters.

This is intriguing when we consider that much of the information coming to us from the world is relatively expected, due to the regularities and physical laws of our universe. Weeks or months into a double-slit experiment to measure wavelength, the new spots on the screen that we find each morning do not provide us with any new information. In the context of the larger ensemble, the surprisal value of any individual spot vanishes to zero — as much as the surprise of the Sun rising each morning vanishes to zero. Events that are predicted, due to the constraint imposed by known physical laws, provide us with no information. Hence, we can see a parallel between the constraint of *information* due to context, and the more conventional *physical* constraint imposed on objects by the universe’s classical and quantum mechanical laws.

It makes sense that if a new result of a well-known experiment is unexpected, in that the result diverges from prediction, then the informational value of that result will be high. Think of the discovery of apparent cosmic acceleration in the late 1990s: Red-shift data from distant supernovae were discovered to be inconsistent with the models. This discovery became extremely meaningful — it led to the 2011 Nobel Prize in Physics, which would not have occurred had the data fit the models. The known laws of physics tell us what results to expect from repeated observations. If these laws appear to be violated, then such new information becomes extremely meaningful to us. The greatly unexpected is greatly meaningful.

The Evolution of Complexity & the Arrow of Time

Let us imagine that John Wheeler was right about “it from bit,” and try to see how contextual constraint might play a role in such a world. The physical interactions of any system with other systems appear to occur in a specific sequence, obeying the principles of causality and locality. We can think of these interactions as iterated measurements of the environment, occurring in a sequence. If the system is indeed a fundamentally *informational* system in some way, then new information reaching the system (with nonzero surprisal) would tend to enrich its cumulative informational content, as it does in Endnote 1. Presumably such enrichment would create additional historical context that was not part of the system at the time of the measurement. This context would create boundary conditions that bear upon *subsequent* interactions along the sequence; this is all in accordance with ordinary causality. With each interaction, the system acquires history, which creates boundary conditions that constrain further interactions — something like the way hands of blackjack become increasingly constrained as the deck gets used up, or how our knowledge of the value of π becomes constrained with each decimal place that we calculate. If this is an "it from bit" world, we now have a way to account for the evolution of complexity, and indeed of meaning, broadly defined. As any system has interactions with the environment and gains history in the form of contextual constraint, the "meaning" or "semantic value" carried by new information (as measured relative to the system) becomes more defined.

Much of the information from the world comes to us in the form of patterns, as subtly ordered as a spiral galaxy or as regular as a crystal. Other information, meanwhile, is as surprising as a gamma-ray burst or as indeterminate as neutron decay. Over billions of years, this broad diversity of surprisals, this iterated interplay of both ordinary and extraordinary information, could create an evolving universe of increasing complexity, as well as deep order and symmetry that "unreasonably" resembles mathematical models — just as we conscious intelligent beings observe [9]. And, since the accumulation of contextual constraint is local and in accordance with causality, and therefore must occur in the same cause–effect direction everywhere, time gets an arrow: For any observer, time points in the direction of increasing information content, what I earlier called semantic entropy.

Conclusion

Are we any closer to answering the question, "it from bit, or bit from it?" Matter/energy is quantized according to Planck's constant, and information is also quantized into bits. Meanwhile, context is real in both mechanical and informational systems. The physical wavelength of a photon at absorption, or the physical pattern of spots on a screen, is inextricably dependent upon the particular configuration of the systems involved; we call this dependency *physical law*. Take away boundary conditions from an experiment involving such laws, and our uncertainty of any variable goes up. Similarly in purely informational scenarios, context produces more defined reduction of uncertainty; we need to incorporate some contextual information if we expect to get more than 12.95 bits out of an SOS message. Such resemblance, between physical-law constraint and more "meaning-oriented" contextual constraint, could be purely coincidental, or an artifact of anthropocentric projection. Or possibly, it could point to something else, for example that our universe is a "simplest-case scenario" and may contain far less information than is conventionally believed [10].

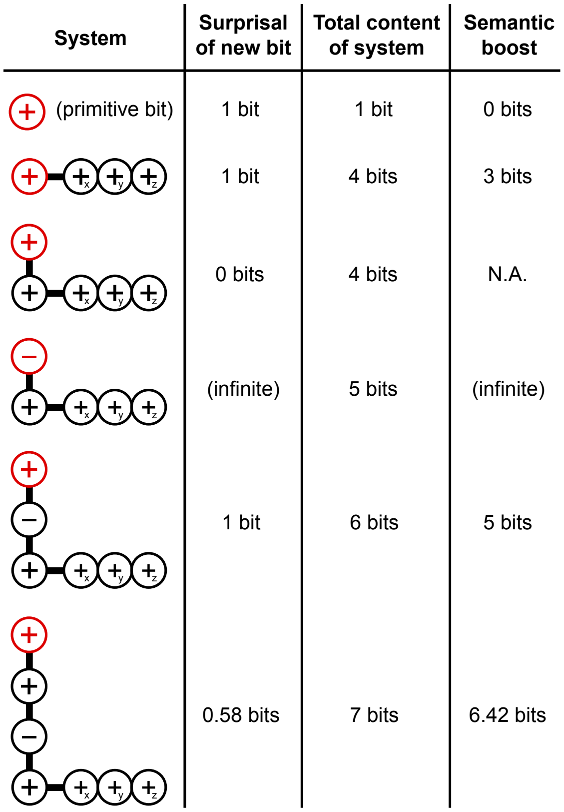
Understanding the physics of information and the ways in which information interacts with other information, on levels much higher than statistical analysis, can seem to be an opaque problem. Wheeler wrote that "every *it* — every particle, every field of force, even the spacetime continuum itself — derives its function, its meaning, its very existence entirely" from bits and the "apparatus-elicited answers to yes-or-no questions, binary choices, *bits*" [3]. But how can we build theoretical frameworks and models to test such a bold hypothesis as "it from bit"? Perhaps quantifying the effects of contextual constraint can open a window.

Endnote
Building a Context, Bit By Bit

To better understand how context constrains new information, let us examine the semantic content of one-bit messages, under near-"absolute zero" contextual conditions.

Consider one bit of information coming to us with no context whatsoever — a click of some particle detector, somewhere, sometime; an event in some completely undefined measurement process. Faced with this one primitive bit, how is our uncertainty reduced? Not by much — only regarding whether the click occurs[4] or not (it does). Now consider the click in a minimal spatial context: We stipulate that the measurement is localized somewhere in the *z*-positive partition created by the *xy* plane. Given this stipulation, the new bit now reduces our uncertainty by two bits, as two yes/no questions are answered (does the click occur? yes; does it occur in the *z*-positive partition? yes). With a further two bits, for a total of four, we can specify that the click occurs in one octant of our three-dimensional space, e.g., (+ + +). It is still the same bit — but by attaching three bits' worth of spatial context, the end result is four bits of potential uncertainty reduction, i.e., "the bit occurs in the (+ + +) octant." Of course, this is nothing like a real-world experiment. In order for the (+ + +) spatial context to mean something real, we need to attach *additional* context that defines the basis of our axes and fixes our frame of reference. If we know that the *xy* plane is defined by the Earth's equator, for example, we become equipped to localize the event in relation to a well-defined feature of our planet.

Now we ask: In context of the discovery of the first click — which in turn is considered in the given context of the (+ + +) octant — *is there a second click in the octant?* Given the history of the first click, an otherwise naïve observer has to assume that a click is expected in the octant, so it has an apparent probability of 1. Since it happened before, there is literally no reason (in this minimal context) to expect that it won't happen again. An actual second click, then, should it occur, provides no new information; this second bit has zero surprisal, being identical to the first result. If however there is *no* other click, then this is unexpected in context, and we have new information: We now know that clicks may happen in the (+ + +) octant, and they also may not happen. *In full context,* the lack of a second click communicates five bits: three to define the octant in which we are looking for clicks, and two to define a click and a subsequent non-click.



System	Surprisal of new bit	Total content of system	Semantic boost
(primitive bit)	1 bit	1 bit	0 bits
	1 bit	4 bits	3 bits
	0 bits	4 bits	N.A.
	(infinite)	5 bits	(infinite)
	1 bit	6 bits	5 bits
	0.58 bits	7 bits	6.42 bits

Figure 2: The surprisal of arriving bits (red) depends on whether they are expected (lower surprisal) or unexpected (higher surprisal). The semantic boost of each new bit quantifies the "meaning" of its arrival, in historical and spatial context. The bottom system comprises 7 bits of information: 3 bits to specify the (+ + +) octant, and 4 to describe the experiment results thus far (3 positives out of 4 measurements). Thus, the semantic boost provided by the context is (7–0.58) bits or 6.42 bits.

In the case of a click and a non-click, we may then ask a third yes-or-no question, *is there another click in the octant?* Now we expect a click at 0.5 probability, so another click has a surprisal once again of 1 bit. In full historical context, however, the new bit tells us: In the octant (+ + +), two of our measurement intervals have had clicks, and one has not. Continuing to yet a fourth measurement, a click would tell us we have had a positive result in three out of four measurements, a story that requires 7 bits to tell (see Fig. 2).

As our naïve observer continues to collect information, he begins to build an understanding of the experiment playing out: clicks and non-clicks, with an increasingly calculable expected frequency of occurrence. If he performed 10,000 measurements and found them trending toward an average rate of 0.75 clicks per measurement, he might propose a "law of 0.75 in octant (+ + +)." In context of that law, any further ensemble of measurements, conforming to the law when coarse-grained, would reduce in semantic content to the law itself, which in this case requires 7 bits to express. In a statistical-mechanical universe that obeys physical laws, semantic content is more subtle than merely the linear sum of all bits involved.

---

[4] Present tense because no temporal context is specified.

References

[1] C. E. Shannon, A mathematical theory of information, Bell System Technical Journal **27**: 379–423 & 623–656 (1948).

[2] M. Tribus, *Thermodynamics and Thermostatics: An Introduction to Energy, Information and States of Matter, With Engineering Applications*. D. Van Nostrand Company (1961).

[3] J. A. Wheeler, Information, physics, quantum: the search for links, in *Complexity, Entropy & the Physics of Information*, W. H. Zurek, ed., Westview (1990).

[4] P. Davies, Universe from bit, in *Information and the Nature of Reality*, P. Davies & N.H. Gregersen, eds., Cambridge University Press (2010).

[5] Information entry, Stanford Encyclopedia of Philosophy online, accessed 20 May 2013.

[6] B. O. Küppers, Information and communication in living matter, in *Information and the Nature of Reality*, P. Davies & N.H. Gregersen, eds., Cambridge University Press (2010).

[7] V. Vedral interview by Paul Davies on www.amazon.com (on page for *Decoding Reality: The Universe as Quantum Information*).

[8] C. E. Shannon, Prediction and entropy of printed English, Bell System Technical Journal **30**: 50–64 (1951).

[9] E. P. Wigner, The unreasonable effectiveness of mathematics in the natural sciences, Communications on Pure and Applied Mathematics **13**: 1–14 (1960).

[10] K. Coryat, Toward an informational mechanics, entry in FQXi “Questioning the foundations” essay contest (2012).