Introductory questions about symbolic values

S. Halayka *

July 25, 2011

Abstract

Many distinct concepts including data, memes, and information are introduced. This manuscript aims to highlight the important role that unconscious physical reality plays in the creation and transmission of symbolic values.

1 Discussion

What is a datum? A datum is a single symbolic value. A few example types of symbolic values are numbers, characters/letters, words, sentences, paragraphs, gestures/movements, songs, images, sculptures, galaxies, etc. Symbolic values are also known as ideas or concepts (e.g., the concept of zero, the concept of the Milky Way).

What is a meme? Traditionally, a meme is defined as a unit of "cultural" transmission, where a single symbolic value is transmitted from one entity to another via some form of physical interaction. Memes emerge from the symbolic values, insomuch that a symbolic value still exists even if it is never once transmitted.

What is information / entropy? Information is a measure of the distinctness of a group of symbolic values. Information emerges from the symbolic values, insomuch that a symbolic value still exists even if its information content is zero.

Where m is the number of *distinct* symbolic values in a group of n symbolic values, and p_i is the relative frequency of the ith distinct symbolic value

$$\sum_{i=1}^{m} p_i = 1, \tag{1}$$

the average information content per symbolic value (Shannon entropy) is

$$S = -\sum_{i=1}^{m} p_i \ln(p_i). \tag{2}$$

At the lower limit where all symbolic values are equal (where $m = 1$), the entropy is

$$S = \ln(1) = 0. \tag{3}$$

Oppositely, at the upper limit where all symbolic values are distinct (where $m = n$), the entropy is

$$S = \ln(n). \tag{4}$$

Consider the following group of $n = 5$ not quite fully distinct symbolic values as a general case:

- the English word "hello",
- the French word "bonjour",

*Saskatoon, SK Canada – email: shalayka@gmail.com

- the English word “goodbye”,
- the English word “hello”,
- the cat word “meow”.

Only $m = 4$ of the $n = 5$ symbolic values are distinct, given that the English word “hello” is repeated once. The relative frequency of the English word “hello” is $2/n = 2/5$ (i.e., the word occurred $2/5$th of the time), and the relative frequencies for the remaining 3 distinct symbolic values are all $1/n = 1/5$. Note that the relative frequencies $2/5+1/5+1/5+1/5$ sum to equal 1, as is required by equation (1). The entropy is

$$\begin{aligned} S &= -2/5\ln(2/5)+ \\ &\quad -1/5\ln(1/5)+ \\ &\quad -1/5\ln(1/5)+ \\ &\quad -1/5\ln(1/5), \\ &\approx 1.33. \end{aligned} \tag{5}$$

This general case serves to illustrate that the entropy is not necessarily integer-valued. Unlike the symbolic values and memes, the entropy is not necessarily quantized into discrete bundles.

Another interesting case is that of the identical entropy $S = \ln(3)$ of the two different groups {‘A’, ‘B’, ‘C’} and {‘A’, ‘B’, ‘C’, ‘A’, ‘B’, ‘C’}. Although the groups are of different size in terms of n, both groups contain the same $m = 3$ distinct symbolic values and relative frequencies $p_{\text{all}} = 1/3$. This case serves to illustrate that more symbolic values (i.e. 6 *vs* 3) does not necessarily mean more entropy.

Because the natural logarithm function was used in the preceding calculations, the entropy has been given in terms of natural digits (base-e). For illustration purposes only, the entropy can also be given in terms of binary digits (base-2)

$$\begin{aligned} S_b &= S/\ln(2), \\ &\approx 1.92. \end{aligned} \tag{6}$$

Just like the “natural” entropy, the binary entropy is not necessarily integer-valued. The binary entropy is integer-valued only in some very special cases, such as where $m = n$ and n is a power of 2. For instance, the binary entropy of the group of $m = n = 2^3 = 8$ distinct symbolic values {0, 1, 2, 3, 4, 5, 6, 7} is

$$S_b = \ln(8)/\ln(2) = 3. \tag{7}$$

This very special case serves to illustrate that a 3-bit integer can represent 8 distinct symbolic values, which is a familiar concept from computer science.

What is the dictionary-based compression of a group of symbolic values? A dictionary-based compression algorithm uses a method called database normalization to reversibly cull all repetitious symbolic values from a group. Thus, the post-cull group is effectively compressed, smaller. Decompression reverses the cull process.

The compressibility (repetitiveness) of a group of n symbolic values is

$$C = 1 - e^S/n. \tag{8}$$

For instance, the common .zip computer file format will very much compress a large group of low-entropy symbolic values (i.e. $C \approx 1$ where $n \gg 1$ and $e^S \ll n$).

Where do symbolic values originate from? Symbolic values (concepts) originate from unconscious physical reality and the unconscious mind. For example, the concept of the Tycho supernova event originated from unconscious physical reality, and the concept of the English word “hello” originated from the unconscious mind. The conscious mind is located at the boundary between unconscious physical reality and the unconscious mind.

What separates one symbolic value from another? The classification of a symbolic value is based entirely on the symbolic value's fine details, in particular which fine details are to be considered and which fine details are to be ignored. In other words, the classification of a symbolic value involves a balance between complexity and abstraction. For instance, is the *spoken* English word "hello" the same symbolic value as the *written* English word "hello"? No, the two symbolic values are not the same, if you consider the fact that they are meant to be transmitted using quite different forms of physical interaction. Oppositely, do two transmissions of the spoken English word "hello" represent the same symbolic value, even if the two transmissions are made using different voices? Yes, the two transmissions represent the same symbolic value, if you ignore the fact that the two transmissions consist of slightly different sound waves.

What is intelligence? One half of intelligence is the ability of the unconscious mind to create a symbolic value (concept) that closely mimics the complex nature of the corresponding concept created by unconscious physical reality. Consider the following chain of events:

- the Tycho supernova event SN 1572 occurs (a star explodes),
- 8,500 years or so later, a human finally sees and subsequently memorizes the image formed by the light pulse transmitted from SN 1572, and thus the concept of SN 1572 is successfully transmitted from unconscious physical reality to the unconscious mind,
- the concept of the English name "the Tycho supernova event SN 1572" is created by the unconscious mind,
- the concept of SN 1572 as modeled by the theories of general relativity and Standard Model particle physics is created by the unconscious mind,
- the concept of SN 1572 as modeled by the theory of supersymmetric strings is created by the unconscious mind.

At first, the concept of SN 1572 is drastically abstracted by the unconscious mind as the image of SN 1572 is memorized. Afterward however, the subsequent concepts of SN 1572 created by the unconscious mind begin to grow in complexity (and intelligence) over time, more closely mimicking the complex nature of unconscious physical reality.

The other half of intelligence is the ability of the unconscious mind to create a complex concept that is not primarily related to unconscious physical reality. For instance, a distinct expression of emotion bearing much hidden context (complexity) seems to be a product of intelligence.

Altogether, intelligence is the ability to perceive and quantify the fine details of unconscious physical reality and the unconscious mind.

What is self-awareness? Self-awareness is the ability to realize that there exists a gap of unconscious physical reality between one's own unconscious mind and the individual unconscious minds of all other entities. Secondary to self-awareness is the intelligence to realize that the transmission of (possibly complex) symbolic values can be effective at bridging this gap of unconscious physical reality.

For instance, self-awareness is exhibited by a cat that intentionally attempts to draw attention to itself by transmitting the symbolic value "meow", in anticipation of some kind of positive response from a second entity. Self-awareness is also exhibited by a cat that becomes jealous when the cat's favourite human transmits a lot of symbolic values to some third entity.

What is artificial intelligence? Artificial intelligence is the ability of some part of unconscious physical reality (i.e. a machine) to be able to automatically and consistently classify and transmit complex symbolic values. For instance, an artificial neural network is a relatively simple machine that can be taught to automatically and consistently classify complex symbolic values of (i.e. images of dogs *vs.* images of trees). At a much deeper level, a machine capable of passing the Turing test would be able to not only automatically and consistently classify complex symbolic values, but also be able to automatically and consistently transmit these complex symbolic values in order to convincingly draw attention to itself like a real intelligent human would.

What are some other examples of non-human transmitters of symbolic values?

- Apes, birds, and many other animals are transmitters of symbolic values.

- The Cheezburger Network[1] web servers. The web servers are literally a chunk of the collective unconscious, insomuch that the unconscious servers exist to store and transmit symbolic values (i.e. LOLcat images).
- Anything that reflects light and sound is a transmitter of symbolic values. For instance, an echo of the spoken English word "hello" is a transmission of a symbolic value.

References

[1] Shannon C. The Mathematical Theory of Communication. ISBN: 9780252725463

[2] Jung CG. The Archetypes and The Collective Unconscious. ISBN: 9780691018331

[3] Dawkins R. The Selfish Gene. ISBN: 9780199291151

[4] Loudon K. Mastering Algorithms with C. ISBN: 9781565924536

[5] Bourg DM, Seemann G. AI for Game Developers. ISBN: 978-0596005559

[6] Shieber S, editor. The Turing Test, Verbal Behavior as the Hallmark of Intelligence. ISBN: 9780262692939

[7] Motl L. Large black holes have a nearly continuous spectrum. `http://motls.blogspot.com/2011/07/large-black-holes-have-nearly.html`

[1] http://icanhascheezburger.com