

Information in *Seventy-Two Letters*

By Hayden Siebers

In Ted Chiang's short story *Seventy-Two Letters*, Robert studies nomenclature where he discovers names for forms called "euonyms". These "euonyms" can animate objects like his clay doll or, later, automatons which perform labor. The central problem of the story is "that the human species has the potential to exist for only a fixed number of generations" (Chiang 165). Robert struggles to find a name with such a high "degree of order" that it can contain many "tiny analogue[s]" or generations (Chiang 180). The discussions with his coworkers immediately invoke the real life counterpart to this order problem: the second law of thermodynamics. The solution was to create an autonym for humans which enables lexical duplication of itself. The real analog to this name is DNA! The aim of this paper is to compare euonyms and the reproduction problem to DNA by making the concept of order more rigorous using information theory.

Information Theory Introduction

The term information can be counterintuitive unless it is understood from its origins: understanding communication. Information is a measure of the freedom of choice of a sender in constructing a message for a receiver. Information can only refer to a system (Weaver 256). If we say a message has a certain amount of information, this really refers to the information of the system in context. This can cause confusion as it is not always obvious what the system in context is.

In order for a receiver to learn something from a message they must have uncertainty about what the message will be. If you already know what somebody will say to you, you will not gain any knowledge from them saying it. More information in a system refers to more freedom of choice for the sender, more uncertainty for the receiver, and more possible messages to be distinguished between. Part of the system's context is how much knowledge the receiver has about the system which corresponds exactly to the sender having less freedom of choice. Order in a system gives the receiver knowledge about the system. Low information indicates a highly ordered system.

For example, take a system which has two states: 0 or 1. If one state is sent, the receiver can distinguish between two messages such as yes or no. When the receiver cannot predict anything about the state that will be sent, 0 and 1 have equal probability, and the system has maximum information. If 0 has no chance of being sent, the system has no information because the sender has no freedom of choice: they must send 1. Claude Shannon developed a formula for evaluating the information of a system with multiple states and any probability distribution between them.

$$H = \sum_{i=1}^n p_i \log\left(\frac{1}{p_i}\right) \text{ (Shannon 393)}$$

where H is Shannon Entropy and p_i is the probability of the i th state occurring

Shannon called this quantity entropy because John von Neumann apocryphally told him “You should call it entropy for two reasons: first because that is what the formula is in statistical mechanics but second and more important, as nobody knows what entropy is, whenever you use the term you will always be at an advantage!” (The Adami Blog). To relate this entropy measurement of information back to what we already know about messages and communication we should notice a few important properties.

- (1) If one state is certain ($p_i = 1$), then entropy $H = 0$. (Shannon 394)
- (2) Maximum entropy H is achieved with probability distributed equally across each state. (Shannon 394)
- (3) The entropy of a composite system is smaller or equal to its additive entropy. It is equal iff the systems are independent. $H_{12} \leq H_1 + H_2$ (Shannon 395)

Thus entropy follows our intuition about information. The more uncertain the receiver is about the state space, the higher the entropy. In this paper, information and entropy will be synonymous. In the wider field, the term information is allowed to roam between multiple meanings while entropy is always fixed to this formula.

When the logarithm used is base two, the unit for information is a bit. In this case, entropy takes on the meaning of the average number of yes or no questions needed to narrow down the state space to one message.

Comparing the Information of a Euonym and DNA

In *Seventy-Two Letters*, the names are described as follows: “seventy-two tiny Hebrew letters, arranged in twelve rows of six, and so far as he could tell, the order of the letters was utterly random” (Chiang 148). The Hebrew alphabet has 22 symbols with no capitalization. To calculate the maximum entropy of this system the 72 letters can be thought of as 72 independent systems. Also, the probability of each of the 22 symbols in the alphabet should be equivalent.

$$H = 72 \sum_{i=1}^{22} \frac{1}{22} \log\left(\frac{1}{22}\right) \approx 321 \text{ bits}$$

When a euonym is put in the right environment—such as inside an automaton—it can bring life to an object. The system for euonyms can distinguish between approximately 2^{232} names; each of which could tell an automaton to do different actions. If it is the wrong euonym, the name won’t work.

DNA uses four different nucleotides with bases cytosine (C), thymine (T), adenine (A), or guanine (G). The human genome consists of about 3,200,000,000 choices of nucleotides from these options (Brown). The maximum entropy of the human genome can be found by assuming all of these nucleotides are independent and have equal probability.

$$H = 3.2 \times 10^9 \sum_{i=1}^4 \frac{1}{4} \log\left(\frac{1}{4}\right) \approx 6.4 \times 10^9 \text{ bits}$$

DNA has 6,400,000,000 bits of information. The reason for using the logarithm is that it reduces the measure down to a number which is more comprehensible. DNA has so much information that it is still an enormous number. 6,400,000,000 bits is equivalent to 800 MB of data. DNA has even been suggested to be a volume efficient storage of data instead of computers, but the practicalities are too difficult (Ionkov and Settlemeyer).

The euonyms in Chiang's story are said to be "utterly random", so it is likely that the actual entropy of that system is similar to the maximum entropy. In this way, DNA is quite different from euonyms. The nucleotides of DNA are not independent. They do not have equal probabilities. There is a lot of predictability in the system of the human genome so the actual entropy is far less than the maximum. DNA is not a random sequence of nucleotides; it has a large amount of order in it.

Transformation and Order

Two systems with the same entropy can be losslessly transformed into each other (Yockey 21). This is what we are doing when we assign yes or no to 0 or 1. We are simply transforming the 0 or 1 system into the yes or no system. DNA is transformed into RNA. The base nucleotides are not identical, but they are transformed one to one so there is no loss or gain of entropy. Transformations can take place between strange systems. DNA can be thought of as a transformation of the space of possible pairings of nucleotides to the space of possible constructions of matter starting with a zygote. Euonyms are transformed into metaphysical roles.

Noise is information added to transformation from outside of either system in the transformation (Weaver 273). If noise is ignored, entropy can only be lost in a transformation. Due to this fact, finding a transformation from one system to another provides an upper bound on the entropy stored in the second system. This is useful if the system resists the traditional method of calculating entropy.

Take for example the number π . Although π is an irrational and transcendental number, it is not purely random. Entropy can be measured by the "size in bits of the shortest computer algorithm describing the sequence or bit string" (Yockey 167). This is a transformation from the system of 'n' bits to the system of real number digits which can be created with 'n' bits.

Conclusion

Finally we can view the reproduction problem and its solution with information insight! In order to create a human within a human within a human stacked for many generations you need a system with enough information to distinguish between that situation and every other situation that could happen in the given context. You would need a euonym or a DNA sequence which could not only distinguish the human body out of all possibilities, but also another human inside of that body. One might expect that an infinite sequence of humans would require a system with infinite entropy; however, there is an order in this sequence just like in π . One can create an autonym which contains the information needed to build the human body combined

with the extra information needed to duplicate the same autonym for the next generation. The autonym or DNA acts in the same way as the computer algorithm: setting a lower bound for the information of the human generational system by transforming into it from the name or DNA sequence system.

Robert's solution is an elegant one. Recursion is a simple yet powerful tool which is familiar to any programmer or mathematician. The strength of information theory is that it is general enough to apply to both the fictional world and the real world. It shows us both the power of Robert's solution and the ingenious solution of self-replication in our real world.

Five Fun Bonus Facts

1. Complete Purpose and Complete Randomness

A perfectly efficient message and a completely random message have the same amount of entropy and are indistinguishable (Yockey 168). A euonym is an example.

2. Thermodynamics and Conservation of Information

The formula for entropy in thermodynamics is equivalent to information entropy up to a constant. The states in this case would be the many microstates of particles (Çengel). Energy is conserved, but useful energy is constantly decreasing as entropy increases. This is similar for information where information is conserved, but usable information only decreases. Unfortunately usable information is hard to define or at least I don't understand how to.

3. The Central Dogma

Francis Crick established the idea of the Central Dogma stating that information is only transferred across or down from DNA to RNA to proteins (Yockey 20). This turns out not to be true, but it is rare for proteins to transfer information to genomes (Koonin). This might be predicted as DNA and RNA have equal entropy, but information is lost in the transformation from codons into amino-acids.

4. Redundancy in English and Crossword Puzzles

Relative entropy is actual entropy divided by max entropy. Redundancy is one minus relative entropy. Written English has a redundancy of about 50%. Yo mgt b abl t rnsct tis sntce w hf of th ltrs gn. This means there is an algorithm in your brain which is capable of transforming English from half information. This gives traditional English a lower bound of 50% redundancy. If English had max entropy, filling out random letters in a grid would create a valid crossword puzzle. 50% redundancy is the perfect amount to create a two dimensional crossword puzzle and 33% would be perfect for three dimensions. (Shannon 399)

5. Evolution and Information

DNA stores information about what body to make. In a sense, this information is really about the environment as it tells us what body is best adapted for the environment. Adami reports that "Information is the central currency for organismal fitness, and appears to be that which increases when organisms adapt to their niche. Information about the niche is stored in genes". (Adami 68)

Class Book

Chiang, Ted. *Stories of Your Life and Others*. Vintage, 2010. EBSCOhost, search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=592472&site=ehost-live.

Works Cited

Adami C. The use of information theory in evolutionary biology. *Ann N Y Acad Sci*. 2012 May;1256:49-65. doi: 10.1111/j.1749-6632.2011.06422.x. Epub 2012 Feb 9. PMID: 22320231.

Adami, Chris. "Whose Entropy Is It Anyway? (Part 1: Boltzmann, Shannon, and Gibbs)." *The Adami Blog: Chris's Thoughts On...*, Michigan State University, 26 June 2014, adami.natsci.msu.edu/blog/2014/6/25/whose-entropy-is-it-anyway-part-1-boltzmann-shannon-and-gibbs-#:~:text=von%20Neumann%20supposedly%20replied%3A,will%20always%20have%20the%20advantage.%E2%80%9D.

Brown TA. *Genomes*. 2nd edition. Oxford: Wiley-Liss; 2002. Chapter 1, The Human Genome. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK21134/>.

Çengel, Y.A. On Entropy, Information, and Conservation of Information. *Entropy* 2021, 23, 779. <https://doi.org/10.3390/e23060779>

Ionkov, Latchesar, and Bradley Settlemyer. "DNA: The Ultimate Data-Storage Solution." *Scientific American*, 21 May 2021, www.scientificamerican.com/article/dna-the-ultimate-data-storage-solution/.

Koonin EV. Does the central dogma still stand? *Biol Direct*. 2012 Aug 23;7:27. doi: 10.1186/1745-6150-7-27. PMID: 22913395; PMCID: PMC3472225.

Shannon, C. E. "A mathematical theory of communication," in *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423, July 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

Weaver, Warren. "RECENT CONTRIBUTIONS TO THE MATHEMATICAL THEORY OF COMMUNICATION." *ETC: A Review of General Semantics*, vol. 10, no. 4, 1953, pp. 261–81. JSTOR, <http://www.jstor.org/stable/42581364>. Accessed 10 Dec. 2023.

Yockey, Hubert P. *Information Theory, Evolution, and the Origin of Life*. Cambridge University Press, 2005. EBSCOhost, search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=132359&site=ehost-live.