

# The gametic entropy of a population

1

E. Castedo Ellerman

2

2022-01-27

3

## Abstract

4

This discussion document defines the gametic entropy of a population. It is a precise interpretation of a phenomenon occurring at the intersection of population genetics and information theory. This interpretation challenges minor statements in two journal articles. Gametic entropy is the application of Shannon entropy to the genetic information carried by gametes in the propagation of a population. This document serves as a reference to facilitate discussion. The practical utility of gametic entropy is not covered in this document. Understanding of Shannon entropy and genetics is required to understand the entire document.

5

6

7

8

9

10

11

12

13

## Background

14

An early application of Shannon entropy [1] in population genetics is the highly cited article "The Apportionment of Human Diversity" by Lewontin in 1972 [2]. Shannon entropy is a measure of information, expressed in units of *bits* (base 2 logarithm of probability). The idea of genetic information being stored and quantified in bits appears in the context of evolutionary genetics as early as 1961 [3].

15

16

17

18

19

20

Eight years after Lewontin's article, BDH Latter wrote:

21

"The Shannon information measure used by Lewontin (1972) ... is extremely difficult to interpret genetically." [4]

22

This claim is about possible interpretations, in the context of genetics, of Shannon entropy. In contrast to the context of genetics, Shannon entropy was established in the context of communication systems.

23

24

25

A mirror claim can be made about possible interpretations of genetic information in the context of communication systems. Indeed, CT Bergstrom & M Rosvall have claimed that genetic information lacks an obvious interpretation in the context of communication theory:

26

27

28

29

"Geneticists are not so fortunate. For them, the analogy to communication theory is less obvious. Efforts to make this analogy explicit seem forced at best, ..." [5]

Below we specify an interpretation and refer to it as the *gametic entropy of a population*. With an understanding of information theory, we propose this interpretation is both

- an easy genetic interpretation of Shannon entropy, and
- a natural example of communication of genetic information.

This interpretation challenges the statements in the two mentioned articles. It should be emphasized that the statements are peripheral and not core claims of their respective articles (which are excellent).

## The interpretation

The *gametic entropy of a population* is:

The amount of information transmitted by a gamete in the propagation of a population.

In the specific application of the 1972 Lewontin article [2], the interpretation is:

The amount of information transmitted by a gamete, **at a random locus**, in the propagation of a population.

Before getting into the biology, we review the core concepts upon which Shannon entropy is based. We follow by observing where these concepts appear in the propagation of a population. Only sexually reproducing populations are discussed initially. Clonal populations will be discussed later as a degenerate special case.

## Messages, senders, and receivers

Shannon entropy is the key measurement of information theory [1]. Entropy measures the degree of uncertainty on what messages a sender communicates to a receiver. Entropy measures a real tangible minimum capacity required of any channel or storage used to communicate those messages. This minimum requirement applies to all channel and storage mechanisms, no matter the physical mechanism.

Roughly speaking, entropy is the best possible score in a game of *20 questions*. That is the minimum number of yes/no questions that can be asked to determine a yet to be known message. One bit is the quantity of information gained (uncertainty reduced) by one yes/no question about two equally likely possibilities.

In population genetics, a key process of interest is the propagation of a population. 60  
In this process, what are the messages, senders and receivers? 61

## Gametic messages 62

The propagation of a population requires new members to be born. For sexually 63  
reproducing species, a new birth requires genetic information to be passed 64  
down from a mother and a father in the population. Each half of this genetic 65  
information is stored and transmitted by a gamete. It is at the unit of a gamete 66  
that we clearly see a complete message, sent by each of the two parents to new 67  
offspring, the receiver. 68

For a next possible birth in a population, there is a probabilistic distribution 69  
of possible genetic messages carried by each gamete. The Shannon entropy of 70  
the distribution of possible messages is the gametic entropy of the population. 71  
One can roughly think of gametic entropy as the number of yes/no questions 72  
offspring need to ask per parent to find out what alleles they are to inherit: "Hey 73  
parent, do I get an Rh+ or an Rh- allele from you?" 74

When only looking at the entropy of autosomal genetic information, the distinc- 75  
tion between maternal and paternal does not matter, but when considering 76  
the information in sex chromosomes, the gametic entropy can be specifically 77  
maternal or paternal. The unqualified gametic entropy is a per gamete entropy 78  
and thus the average of both maternal and paternal gametic entropies, since 79  
every birth requires one of each kind of gametic message. 80

## Toy illustration 81

We consider a toy example of a sexually reproducing unicorn population with 82  
biallelic chromosomes. Imagine a population of unicorns with 10 autosomal 83  
chromosomes of which the entire lengths of each chromosome consists of only 84  
one of two equally frequent haplotypes. In contrast, the X and Y chromosomes 85  
of these unicorns are completely fixed with only one X haplotype and only one 86  
Y haplotype in the population. The gametic entropy across the 10 autosomal 87  
chromosomes is exactly 10 bits. The maternal gametic entropy for the sex 88  
chromosome is zero because there is no uncertainty about which sex chromosome 89  
(or its haplotype) is transmitted. The total maternal gametic entropy is thus 90  
10 bits. In contrast, the paternal gametic entropy for the sex chromosome 91  
is 1 bit since there is an equal chance of an X or a Y chromosome being 92  
communicated (and no additional uncertainty regarding each sex chromosome's 93  
respective haplotype). Thus the total paternal gametic entropy is 11 bits. The 94  
overall gametic entropy of this unicorn population is 10.5 bits, the average of 95  
the parent-specific entropies. 96

## Decoupling medium of transmission

97

One of the key insights from information (communication) theory is that fundamental properties of information, such as entropy, exist regardless of the forms of storage or transmission.

98  
99  
100

At current levels of technology, DNA is the only medium of transmission of interest for the genetic information carried by gametes. Nonetheless, a thought experiment of hypothetical but almost plausible transmission mediums helps elucidate the independence of the "pure" information transmitted from the medium of transmission.

101  
102  
103  
104  
105

### A thought experiment

106

Reflecting on DNA sequencing technology, maternal spindle transfer in "three-parent" IVF [6] [7] and artificial synthesis of an entire (bacterial) genome [8], it is not hard to imagine a possibility of completely dematerialized transmission of the genetic information carried by gametes. This transmission can be over any of the channels described by information theory, including telegraph schemes.

107  
108  
109  
110  
111

We now can propose an answer to the following question posed by CT Bergstrom & M Rosvall:

112  
113

"Is there a clean mapping from informational processes in biology onto the telegraph schema?" [9]

114

The practical utility of such telegraphy, if any, is hard to imagine today. But as an entertaining thought experiment in science fiction, we can image the utility of interplanetary transmission of the genetic information in donor gametes for IVF. Rather than physically transporting gametes between planets, which could take months to years, the pure data can be transmitted in minutes or hours. All of the facts about transmission rates in telegraphy apply in a hypothetical interplanetary IVF system. For a given population of equally likely possible donors, the maternal (paternal) gametic entropy, is the best possible rate, measures in bits, for transmitting the genetic information of a egg (sperm) donor.

115  
116  
117  
118  
119  
120  
121  
122  
123  
124

## Degenerate case of clonal populations

125

In the case of asexually reproducing clonal populations, there is no distinction between the genetic information in a gamete vs a parent. The "gametic message" is simply the entire genome of a parent, which is sent in its entirety to its single-parent offspring. The gametic entropy of a clonal population is the same as the entropy of the distribution of distinct genomes in the population.

126  
127  
128  
129  
130

## Concluding questions 131

This document challenges claims in two previous journal articles regarding two  
respective questions: 132  
133

1. Is the gametic entropy of a population an extremely difficult genetic 134  
interpretation of Shannon entropy? 135
2. Is the gametic entropy of a population a forced analogy in communca- 136  
tion theory for geneticists? 137

## Acknowledgements 138

ECE thanks Steven Orzack and John Novembre for fruitful discussions about 139  
the Lewontin 1972 paper. 140

## References 141

1. Shannon CE, Weaver W (1998) The mathematical theory of communica- 142  
tion. Univ. of Illinois Press, Urbana 143
2. Lewontin RC (1972) The Apportionment of Human Diversity. In: 144  
Dobzhansky T, Hecht MK, Steere WC (eds) Evolutionary Biology.  
Springer US, New York, NY, pp 381–398 145
3. Kimura M (1961) Natural selection as the process of accumulating genetic 146  
information in adaptive evolution. *Genetical Research* 2:127–140. <https://doi.org/10.1017/S0016672300000616> 147
4. Latter BDH (1980) Genetic Differences Within and Between Populations 148  
of the Major Human Subgroups. *The American Naturalist* 116:220–237.  
<https://doi.org/10.1086/283624> 149
5. Bergstrom CT, Rosvall M (2011) The transmission sense of information. 150  
*Biology & Philosophy* 26:159–176. <https://doi.org/10.1007/s10539-009-9180-z> 151
6. Amato P, Tachibana M, Sparman M, Mitalipov S (2014) Three-parent 152  
in vitro fertilization: Gene replacement for the prevention of inherited  
mitochondrial diseases. *Fertility and Sterility* 101:31–35. <https://doi.org/10.1016/j.fertnstert.2013.11.030> 153
7. Zhang J, Liu H, Luo S, et al (2017) Live birth derived from oocyte spindle 154  
transfer to prevent mitochondrial disease. *Reproductive BioMedicine*  
*Online* 34:361–368. <https://doi.org/10.1016/j.rbmo.2017.01.013> 155
8. Gibson DG, Glass JI, Lartigue C, et al (2010) Creation of a Bacterial 156  
Cell Controlled by a Chemically Synthesized Genome. *Science* 329:52–56.  
<https://doi.org/10.1126/science.1190719> 157

9. Bergstrom CT, Rosvall M (2011) Response to commentaries on “The Transmission Sense of Information”. *Biology & Philosophy* 26:195–200. 158  
<https://doi.org/10.1007/s10539-011-9257-3> 159