centro de ciencias de la complejidad

Information processing and the problem of meaning

Questions (16) and Assertions (8)

Chris Stephens,
C_3 Centro de Ciencias de la Complejidad y
Instituto de Ciencias Nucleares, UNAM

Workshop: Statistical Physics, Information Processing and Biology
Santa Fe Institute 16-18/11/2012

To be, or not to be--that is the question:
Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune
Or to take arms against a sea of troubles
And by opposing end them. To die, to sleep--
No more--and by a sleep to say we end
The heartache, and the thousand natural shocks
That flesh is heir to. 'Tis a consummation
Devoutly to be wished. To die, to sleep--
To sleep--perchance to dream: ay, there's the rub,
For in that sleep of death what dreams may come
When we have shuffled off this mortal coil,
Must give us pause.

Human nucleotide sequence

```
AAAAGAAAAGGTTAGAAAGATGAGAGATGATAAAGGGTCCATTTGAGGTTAGGTAAT
ATGGTTTGGTATCCCTGTAGTTAAAAGTTTTTGTCTTATTTTAGAATACTGTGACTA
TTTCTTTAGTATTAATTTTTCCTTCTGTTTTCCTCATCTAGGGAACCCCAAGAGCAT
CCAATAGAAGCTGTGCAATTATGTAAAATTTTCAACTGTCTTCCTCAAAATAAAGAA
GTATGGTAATCTTTACCTGTATACAGTGCAGAGCCTTCTCAGAAGCACAGAATATTT
TTATATTTCCTTTATGTGAATTTTTAAGCTGCAAATCTGATGGCCTTAATTTCCTTT
TTGACACTGAAAGTTTTGTAAAAGAAATCATGTCCATACACTTTGTTGCAAGATGTG
AATTATTGACACTGAACTTAATAACTGTGTACTGTTCGGAAGGGGTTCCTCAAATTT
TTTGACTTTTTTTGTATGTGTGTTTTTTCTTTTTTTTTAAGTTCTTATGAGGAGGGA
GGGTAAATAAACCACTGTGCGTCTTGGTGTAATTTGAAGATTGCCCCATCTAGACTA
GCAATCTCTTCATTATTCTCTGCTATATATAAAACGGTGCTGTGAGGGAGGGGAAAA
GCATTTTTCAATATATTGAACTTTTGTACTGAATTTTTTTGTAATAAGCAATCAAGG
TTATAATTTTTTTTAAAATAGAAATTTTGTAAGAAGGCAATATTAACCTAATCACCA
TGTAAGCACTCTGGATGATGGATTCCACAAAACTTGGTTTTATGGTTACTTCTTCTC
TTAGATTCTTAATTCATGAGGAGGGTGGGGGAGGGAGGTGGAGGGAGGGAAGGGTTT
CTCTATTAAAATGCATTCGTTGTGTTTTTTAAGATAGTGTAACTTGCTAAATTTCTT
ATGTGACATTAACAAATAAAAAAGCTCTTTTAATATTAGATAA
```

49

INGREDIENTS

- 1/3 cup sugar plus additional for sprinkling
- 5 oz bittersweet chocolate (not unsweetened), chopped
- 3 large egg yolks at room temperature
- 6 large egg whites
- Accompaniment: lightly sweetened whipped cream
- Special equipment: a 5 1/2- to 6-cup glass or ceramic soufflé dish

PREPARATION

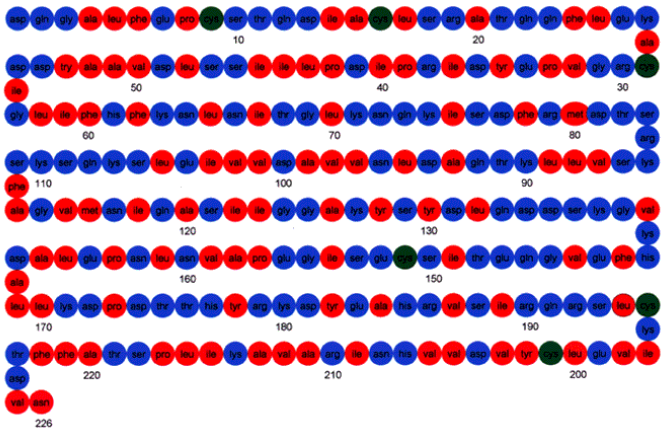
Preheat oven to 375°F. Generously butter soufflé dish and sprinkle with sugar, knocking out excess.

Melt chocolate in a metal bowl set over a saucepan of barely simmering water, stirring occasionally until smooth. Remove bowl from heat and stir in yolks (mixture will stiffen).

Beat whites with a pinch of salt in a large bowl with an electric mixer at medium speed until they just hold soft peaks. Add 1/3 cup sugar, a little at a time, continuing to beat at medium speed, then beat at high speed until whites just hold stiff peaks. Stir about 1 cup whites into chocolate mixture to lighten, then add mixture to remaining whites, folding gently but thoroughly.

Spoon into soufflé dish and run the end of your thumb around inside edge of soufflé dish (this will help soufflé rise evenly). Bake in middle of oven until puffed and crusted on top but still jiggly in center, 24 to 26 minutes. Serve immediately.

Amino Acid Sequence of hJHBP



What do these sequences have in common? What “information” is in these sequences?

They can all be inputs or outputs or programs

What do they have in common?

1) They are sequences of symbols that can represent physical structures and that can combine to form a hierarchy of higher order structures - Building Blocks - at different scales

2) These Building Blocks obey different rules at different levels

3) Very, very few Building Blocks are "valid", very few combinations appear

A. Are the individual symbols (chemical compounds) in the alphabet?
   - Very few symbols (chemical compounds) out of all possibilities are used (combinatorics).

B. Are the combinations of symbols (Building Blocks) (words; exons/introns/genes…) in the lexicon?
   - Very few combinations of symbols are in the lexicon

C. Are the combinations of words "grammatically" correct?
   - Very few combinations of words are such

D. Are the combinations of words/sentences etc. "semantically" meaningful?
   - Very few combinations are semantically meaningful

Q1: What are the physical constraints/laws/rules/… that govern which sequences (combinations of Building Blocks) are allowed or preferred?
Q2: How do these physical constraints/laws/rules/… affect the information content?
Q3: How does one assign a "fitness" to a sequence?
Q4: What is the right basis and how do we transform between bases?
A1: They both are characterised by highly modular and highly rugged landscapes.

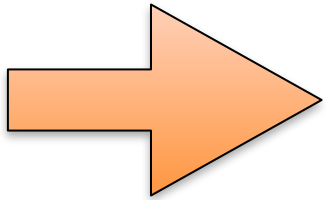
What information is in these sequences?

Intrinsic measures of information…

These are single numbers!

Shannon entropy (needs an ensemble)
Kolmogorov complexity
Any amount of correlation functions on symbols or Building Blocks…

$$< \alpha_i > \quad i \in a,b,...,z$$
$$< \alpha_i \alpha_j > \quad i,j \in a,b,...,z$$
$$< \alpha_i \alpha_j \alpha_k > \quad i,j,k \in a,b,...,z$$



Information independent of context and universal

What can the statistics of Building Blocks tell us?

.

A2: No amount of statistical analysis of a text nor any intrinsic information measure will lead us to meaning.

Q5: Is there any meaning to meaning in biology?

Meaning

Still a very debated topic in linguistics. No generally accepted theory. Perhaps partly because human language can be so abstract.

[**mee**-ning]

noun

1. what is **intended** to be, or actually is, expressed or indicated; signification; import:
2. the end, **purpose**, or significance of something:

'In most cases, the meaning of a word is its use',
Wittgenstein, Philosophical Investigations

Q6: Is meaning a scientific concept?

Q7: Can we measure meaning?

Q8: What measuring apparatus measures meaning?

Q9: Can we compute meaning?

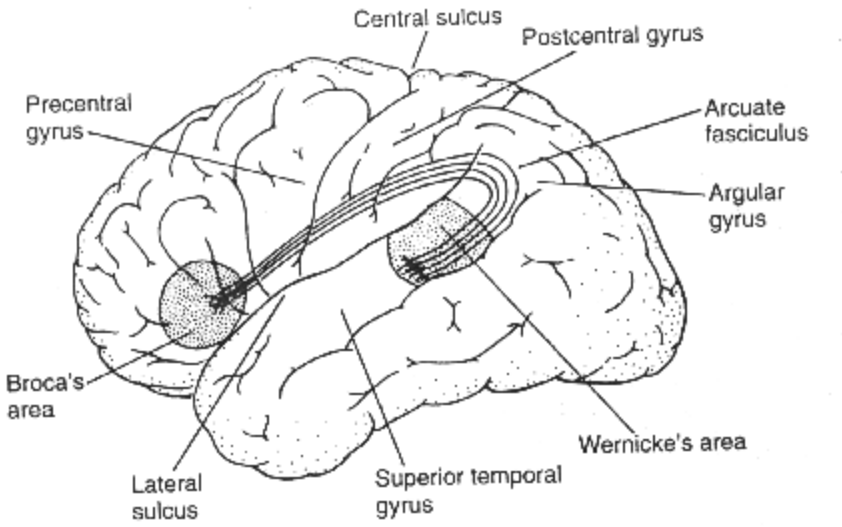
A3: Computers can't do semantics

A4: Semantics is the only thing that imposes long-range interactions in sequences - at least in language. Its what makes language complex.

The "Semantometer"

some call it a brain,
its also a "Complexometer"

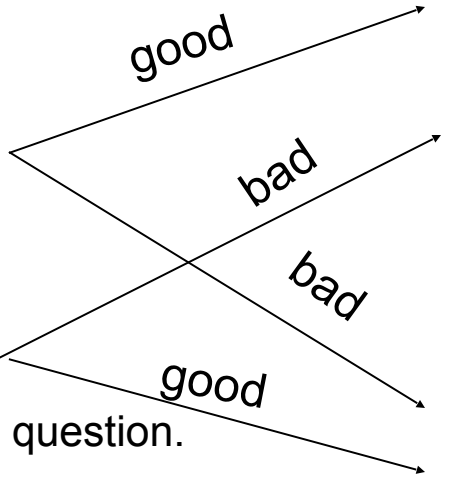
measuring meaning in biology



This apparatus is surely capable of measuring meaning. Or maybe not…?

How good is your apparatus?

It needs to be calibrated, just like any other measuring apparatus



- To be or not to be that is the question.
- Para ser o no ser que es la pregunta.
- Om te zijn of te zijn niet dat de vraag is.

あるためまたはないため質問である

- Because of a certain or because it is not, it is question.
- Because or it is not for the sake of, that having asked and being convinced.
- Being not to be for the sake of, or that that, you ask, are convinced.
- It is that without having for the sake of, or, you ask, are convinced.

Q10: How does the information of a sequence change as a Semantometer measures it?

Q11: What is the free energy cost of the neural computation that measures meaning?

A5: Meaning is partially captured by the mutual information between the sequence and the state of the Semantometer.

The Context Dependence of Meaning and its Value

A6: The (semantic) information of any sequence (Building Block) depends on its context (due to epistatic interaction both with other sequences (Building Blocks) and other inputs)

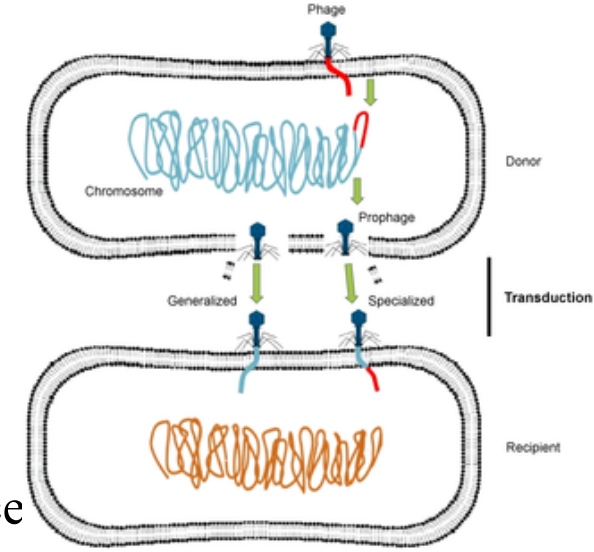
Give me your wallet or I'll shoot you.
Context: In a movie versus in real life changes the value. "Intrinsic" semantic content is the same. True meaning must account for. e.g., sensory input - pragmatics.
Shoots —> He shoots pigeons, El shoots palomas, She shoots pool, The shoots are green, The panda eats shoots and leaves, The panda eats, shoots and leaves

Meaning is associated with purpose, intentionality —> function —> decisions —> behaviour

A7: Different meanings lead to different decisions which can have different "fitness".
Q12: How do we give it a value or fitness?

Value given by the Semantometer versus objective value
Smoking kills
Do you give up smoking or not?

Meaning in Biology (remember the brain and language is biology)



Q13: What does a gene, or any sequence, mean?

What is the information content of a gene in a glass of water versus in a genome. Same sequence, same intrinsic information/entropy.

What is the information content of a viral gene in a virus/bacterium/human? e.g., that can confer antibiotic resistance

Q14: What is a Semantometer at the cellular level?

The rest of the cell versus the nucleus?

Q15: Are there multiple Semantometers at different scales?

Q16: What is the free energy cost of measuring meaning or generating new meanings at the cellular level?

A8: New sequences and new meanings are generated principally through generalised recombination.

Conclusions

Biological information and information processing, and, indeed, complexity itself, can only be properly understood in the context of the meaning of that information

Meaning is contextual - environment (niche) dependent - and has a value - “fitness” (but multi-objective)

Its a scientific concept because its measurable Semantometers exist. Brains are a great example we should all think about.

We can speak and understand English/Spanish/German… but we don’t speak genome or cell, at least at the level of science

Both biological and human language are highly **recombinative** systems based on recombining lexically, syntactically and semantically meaningful Building Blocks within a very rugged landscape on sequence space and displaying a high degree of **modularity** and **redundancy**

We have been quite successful in **deducing** lower order structure from higher order but singularly unsuccessful at **inferring** (from first principles) the existence of higher order structure form lower order

Evolution as computation - too general - evolution is a particular type of computation - algorithmically it is search - and that search, both in physics and biology, is very biased - dominated by recombination of Building Blocks - and is massively parallel - specialisation and multi-tasking

Emergence of algorithmic language in genetic systems
OA Palacios, CR Stephens, H Waelbroeck
Biosystems 47 (3), 129-147

Shows how a model genetic system with a non-trivial, regulatory-like, GP map can lead to representations of genetic info that have several language like properties

Zero is not a four letter word: studies in the evolution of language
C Stephens, M Nicolau, C Ryan
European Conference on Genetic Programming, 371-380

Fitness Landscape Epistasis And Recombination
MB DEL RÍO, CR Stephens, DA Rosenblueth
Advances in Complex Systems 18 (07n08), 1550026

Shows how recombinative dynamics ("sex") is only advantageous on modular or negatively epistatic landscapes

Geometry the renormalization group and gravity
D O'Connor, CR Stephens
arXiv preprint hep-th/9304095

Fisher information matrix on the space of "all" theories and Renormalization Group transformations

What Isn't Complexity?
CR Stephens
arXiv preprint arXiv:1502.03199

Physics is "being" biology is "doing" and other such niceties

Who's Smart and Who's Lucky? Inferring Trading Strategy, Learning and Adaptation in Financial Markets through Data Mining
CR Stephens, JL Gordillo, EM Miranda
Natural Computing in Computational Finance, 95-114

The problems of statistical inference of information in an experimental market - efficient markets and all that - rational versus irrational

Is modularity the reason why recombination is so ubiquitous?
MB del Río, CR Stephens, DA Rosenblueth
arXiv preprint arXiv:1209.4142

Shows that hill climbing on fitness landscapes does not give an intuitive idea of how genetic dynamics proceeds. However, effective fitness landscapes do showing how certain genotypes can be favoured in the absence of a selective advantage. Symmetry breaking of the GP map

Landscapes and effective fitness
PF Stadler, CR Stephens
Comments® on Theoretical Biology 8 (4-5), 389-431

Coarse-grained dynamics for generalized recombination
CR Stephens, R Poli
IEEE transactions on evolutionary computation 11 (4), 541-557

How to write a population genetics with arbitrary shuffling of genes (homologous and non-homologous crossover, deletion, creation, translocation, transposition, inversion,…) in terms of Building Blocks

Market efficiency and learning in an artificial stock market: A perspective from Neo-Austrian economics
HA Benink, JL Gordillo, JP Pardo, CR Stephens
Journal of Empirical Finance 17 (4), 668-688

Shows how algorithms with Building Block are more efficient search algorithms

Building blocks and search
A Lozano, V Mireles, D Monsivais, CR Stephens, SA Alcalá, F Cervantes
Mexican International Conference on Artificial Intelligence, 704-715