Does information have an intrinsic value?

Pieter Adriaans
Universiteit van Amsterdam

www.pieter-adriaans.com
Info-metrics Institute
Workshop The Value of Information
April 26 2013
The value of information

• Intrinsic versus Instrumental Values
• Instrumental: could you tell me the way to the station?
• Intrinsic?
Possible Intrinsic Values

- Truth
- Beauty
- Love
- Honesty
- Friendship
- Trust
- Human life
- Freedom
- Individuality
- Equality

- Authenticity
- Sustainability
- Passion
- Creativity
- Biodiversity
- Ecological balance
- Health

- Information?
Does information have an intrinsic value?

- Based on the probability of the data (Shannon)
- Based on the descriptive complexity of the data (Kolmogorov)
- Based on the amount of computation involved (Classic complexity theory)
- A mix between 2 and 3 (Levin)
- Based on the amount of model information in the data (Sophistication, Facticity)
Tension between being and change

[What exists] is now, all at once, one and continuous...

Everything streams....
Plato’s theory of forms
the allegory of the cave
In-form-ation

• ‘Information’ as the process of being informed.
The Information of a student, a piece of marble...

• ‘Information’ as a state of an agent,
A man of information ....

• ‘Information’ as the disposition to inform,
Data sets that contain information...
Problem 1: Universalia

• What is the ontological status of these forms?

Plato
idealism
*(universalia ante res)*

Aristotle
Realism
*universalia in rebus*
Occam’s Razor: Nominalism

*entia non sunt multiplicanda praeter necessitatem*

entities must not be multiplied beyond necessity
Problem 2: Individuality

Ideal | Faults
---|---

Ideal | Faults

Same model
Different types of information

- Information as probability
- Information as structure
Information as probability: an experiment

Which word is this?

A N T I E
Information as probability: an experiment

Which word is this?

X A T I P
Information as probability: an experiment

Which word is easier to guess?

ANTINE

XATIP
Shannon information

• The entropy, $H$, of a discrete random variable $X$ is a measure of the amount of uncertainty associated with the value of $X$ (Shannon 1948, Shannon & Weaver 1949).

\[ I(x) = -\log P(x) \]

• The less we expect it the more value it has. Scarcity as a source of value.
Information is dependent on probability

• Letter frequencies in Dutch per 10,000 tokens

E = 1586 \quad I(E) = -\log(1586/10000) = 2.66 \text{ bits}

N = 858 \quad I(N) = -\log(858/10000) = 3.54 \text{ bits}

P = 123 \quad I(P) = -\log(123/10000) = 6.34 \text{ bits}

X = 3 \quad I(X) = -\log(3/10000) = 11.70 \text{ bits}

Low frequency = less probable = more information!!
Information as probability: an experiment

Here we miss 24.48 bits information

Here we miss 12.54 bits information
Information as structure: An experiment

Look at both configurations for 5 seconds
Which is easier to remember?
Kolmogorov complexity

• The information in a binary string $x$ is the length of the shortest program $p$ that produces $x$ on a reference universal Turing machine $U$ (Solomonoff 1960, 1964a, 1964b, 1997, Kolmogorov 1965, Chaitin 1966, 1987).

• The closer it is to randomness/maximal entropy the more value it has. Random strings are valuable in cryptography.
Information as structure: An experiment

\[ \log 20 + \log 32 \approx 10.3 \text{ bits} \]

\[ \log_2 20 + \log_2 \left( \frac{100}{20} \right) \approx 76.5 \text{ bits} \]
Information = entropy

Low entropy =
Much order =
Little information =
Easy to remember

High entropy =
Much disorder =
Lots of information =
Hard to remember
Problem: Max Entropy = zero Free energy
Measure(s) of meaningful/useful information

- Esthetic Measure (Birkhoff, Bense)
- Sophistication (Koppel)
- Logical Depth (Bennet)
- Statistical complexity (Crutchfield, Young)
- Effective complexity (Gell-Mann and Lloyd)
- Meaningful Information (Vitanyi)
- Self-dissimilarity (Wolpert and McReady)
- Computational Depth (Antunes et al.)
Some conventions

- $U = \text{Universe}$
- $M = \text{Model}$
- $D = \text{Data Set}$
- For the moment $M$ and $D$ are finite
- $|M| = m$
- $|D| = d$
Selecting the best model according to Bayes

\[ M_{map} = \arg \max_{M \in U} \frac{P(M)P(D|M)}{P(D)} \]

\[ = \arg \max_{M \in U} P(M)P(D|M) \]

\[ = \arg \max_{M \in U} \log P(M) + \log P(D|M) \]

\[ = \arg \min_{M \in U} -\log P(M) -\log P(D|M) \]

Model Code Data to Model Code

Optimal Shannon Code
The minimum description length (MDL) principle: J. Rissanen

- The best theory to explain a set of data is the one which minimizes the sum of:
  - the length, in bits, of the description of the theory and
  - the length, in bits, of the data when encoded with the help of the theory
Turing two-part code optimal compression

\[ K(D) \quad \text{Kolmogorov Complexity} \quad \delta(D) \quad \text{Randomness Deficiency} \]

\[ \phi(D) \quad \text{Facticity} \quad \rho(D) \quad \text{Residual Entropy} \]
Definition of facticity: the amount of self-descriptive information in a dataset

\[ U(x) = \min \{ |i| + |p| : \min \{ |i| + |p| : U(ip) = x \} \} \]

- The facticity \( \phi(x) \) of a string \( x \) is the amount of self descriptive information in \( x \).
- Facticity is definite but not robust!
A possible synthesis

\[ \phi^U(x) = \arg \min_{|i|} \{ i : \min_{i,p} \{ |\tilde{i}| + |p| : U(\tilde{i}p) = x \} \} \]

The facticity of a data set is the amount of factual information that can be extracted from it.
Philosophy of information: tentative foundational research program

• **Given**: a universe that can store information and in which computation can take place.
  – Space, Time
  – Stable data configurations with low energy transformations
  – Medium entropy

• **Interpret**: traditional notion of universals in terms of two part code optimization

• Empirical Particularism: The particular is the universal (Between idealism and nominalism)
What does it mean to be an individual? The classical view
What does it mean to be an individual?
Computational View:

Data Peter U

Data Pieter A

Model Code  Data to Model Code

Model Code  Data to Model Code
Idealization in art as two part code optimization

- **Realism**: representation is isomorphic to the data
- **Idealization**: ideal schemas optimally compress the description of a set of examples with errors
- **Schematization**: optimal compression under bounded complexity
Characterization

- **Characterization**: optimal bounded compression of an individual example conditional to the optimal general theory
Brain: Modularity, Compression

130 million receptors in the eye

1.5 million channels in the optic nerve
More formal Definitions

• The model code:

$$\varphi^*(x) = \arg \min_{i} \left\{ \min_{i,p} \{|i| + |p| : U(ip) = x \} \right\}$$

• Conditional facticity:

$$\varphi(x \mid y) = K(\varphi^*(x) \mid \varphi^*(y))$$

• The factual value of information:

$$\nu(x \mid y) = \max\{ \varphi(xy \mid y), \varphi(y \mid xy) \}$$
Some other applications

• The notion of an ideal teacher
• game playing (digital bluff poker)
• Honing’s problem of the ‘surprise’ value of musical information
• dialectic evolution of art
• the problem of interesting balanced esthetic composition
• optimal product mix selection
• Schmidhuber’s notion of low complexity art
• and Ramachandran’s neuro esthetics
Genetic algorithms optimizing facticity in images (Den Heijer 2013)

• Experiments
• Shannon entropy, brightness of pixels, 32 bins
• Shannon entropy, brightness of pixels, 256 bins
• Kolmogorov, with different compressors jpeg, jpeg2000 and png
Data tree A

Data to Model

Model Code

Data to Model Code

Data Tree B

Model Code

Data to Model Code
Conclusions

• intrinsic value of information: facticity = the amount of model information
• Philosophy of information as a foundational discipline
• Rethinking history of philosophy
• Consequences for:
  – Political organization of society
  – Ethics and human rights
  – Scientific Methodology
  – Esthetics