What is the relation between representation and information?

Manolo Martínez – Universitat de Barcelona – Logos – BIAP

Introduction

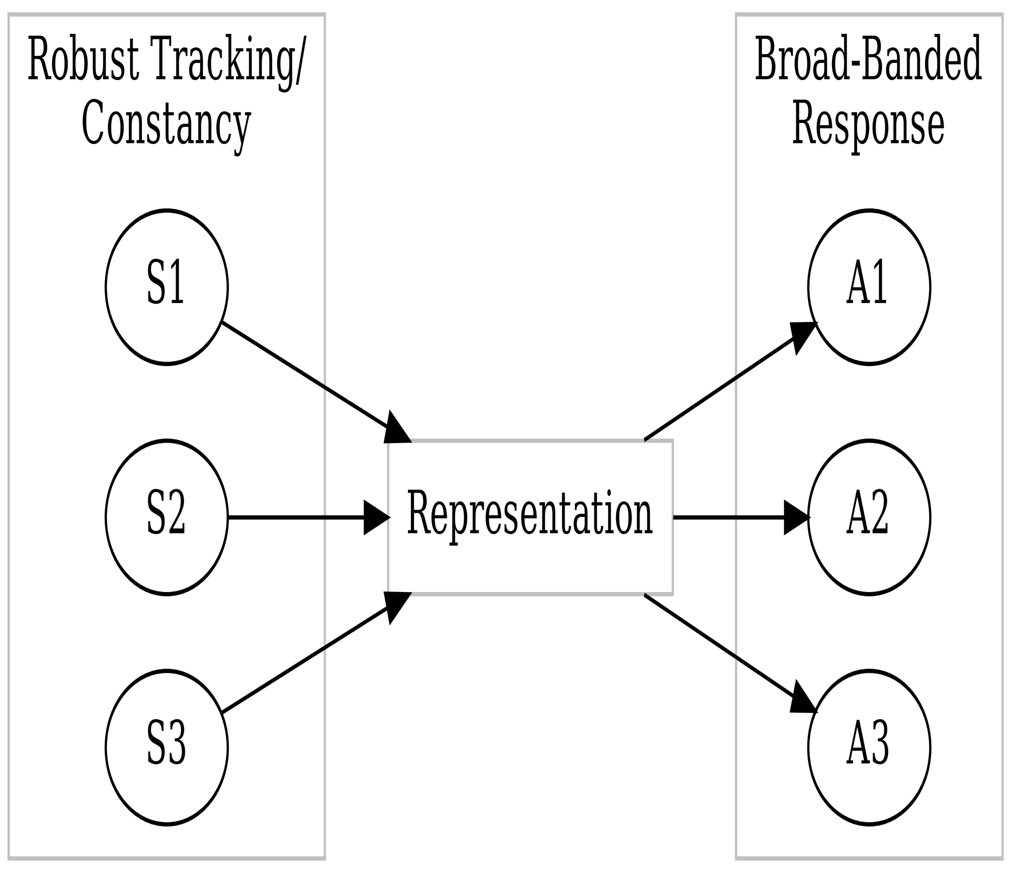
The basic informational relation in contemporary Dretskean accounts is that of *making a probabilistic difference* (Scarantino 2015). E.g., in Skyrms (2010, p. 36) the "information in M in favor of W" is defined as the *pointwise mutual information* between state and message.

$$\log_2 p(W_1 \vee M_i) - \log_2 p(W_1), \ldots, \log_2 p(W_n \vee M_i) - \log_2 p(W_n)$$

Unsurprisingly, from this focus on the probabilistic relation between individual states and messages it is routinely concluded that there is much more to representation than information. This conclusion is premature: informational content in the Dretskean tradition is not all there is to information theory.

This should not be taken to imply that information is all there is to representation---for one thing, I believe with teleosemanticists (Millikan 1984; Papineau 1987) that teleofunctions have a role to play in a complete theory of representation---but it does mean that no Dretske-style "semanticized information" needs to be recognized, over and above the quantities studied in information theory proper. It also means that some prominent proposals as to ways to bridge the information-representation gap are, in fact, unwittingly appealing to informational structure.

Bridging the Gap



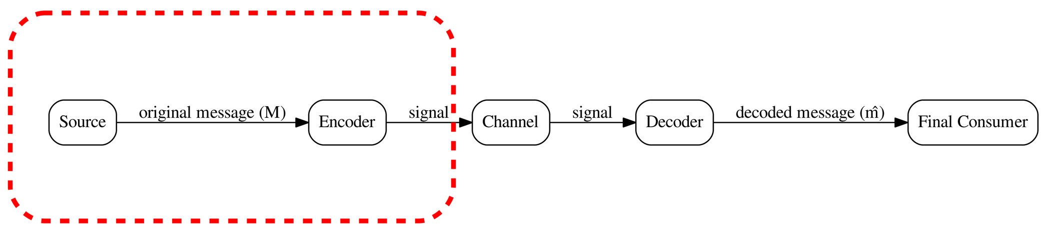
Many-to-one-to-many architectures favored by Burge (perceptual constancy) and Sterelny

Reference Magnetism: Focus on the *entities* that should figure in the content of simple representations---they should be appropriately *natural*, or *real*.

- Dan Ryder (2004, 2006) has argued that neurons become attuned to *sources of correlation*.
- Richard Boyd's *homeostatic property clusters* (also HPC henceforth, Boyd 1989): clusters of properties which tend to be instantiated together
- Lewis (1983): "among the countless things and classes there are ... [o]nly an elite minority are carved at the joints, so that their boundaries are established by objetive sameness and difference in nature. Only these elite things and classes are eligible to serve as referents"

Information Theory is a Source-Channel Theory

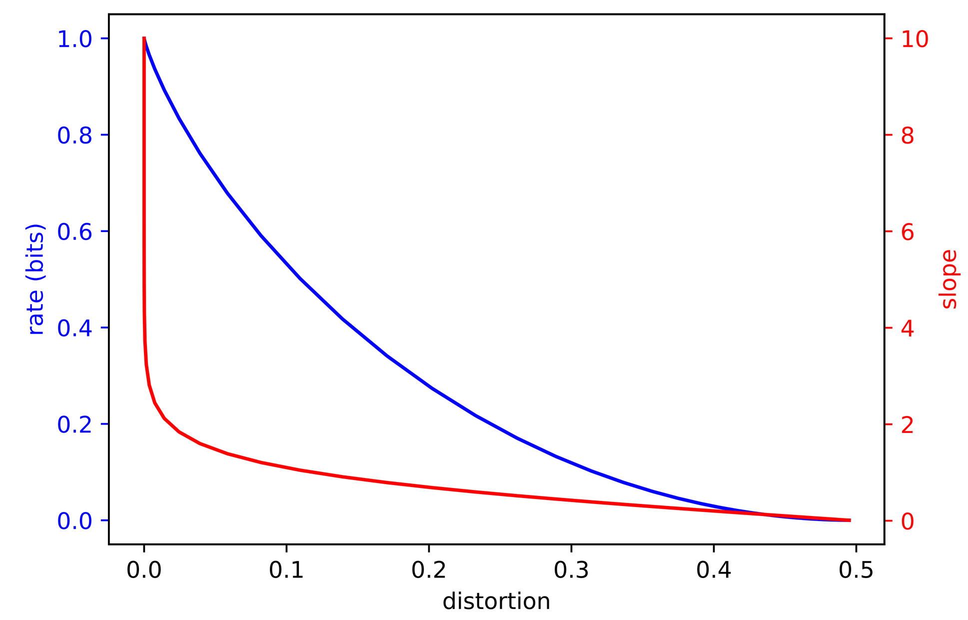
The most celebrated results in information theory have to do with specifying how faithful the transmission of information from a source can be, when it happens over a (typically noisy, typically narrow) channel. These results have played absolutely no role in informational accounts of representation.



An information-processing pipeline, and the fraction considered in Dretskean analyses

Rate-Distortion Theory

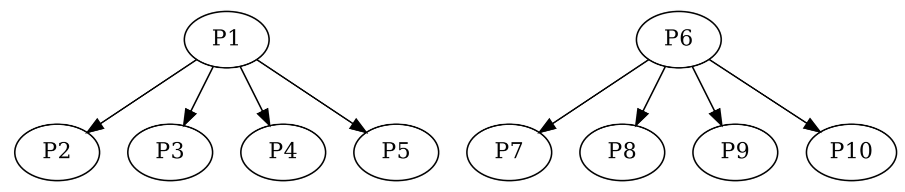
In typical cases of representation, channel rate is consistently smaller than ideal. The way in which information theory deals with lossy transmission is by defining a *distortion measure* (Cover & Thomas 2006, p. 304) that gives a score to a pair composed of a certain original message and the decoded version thereof. The *rate-distortion function* calculates the minimum rate necessary to achieve a certain distortion.



The rate-distortion function for a Bernoulli(0,5) source and Hamming distortion

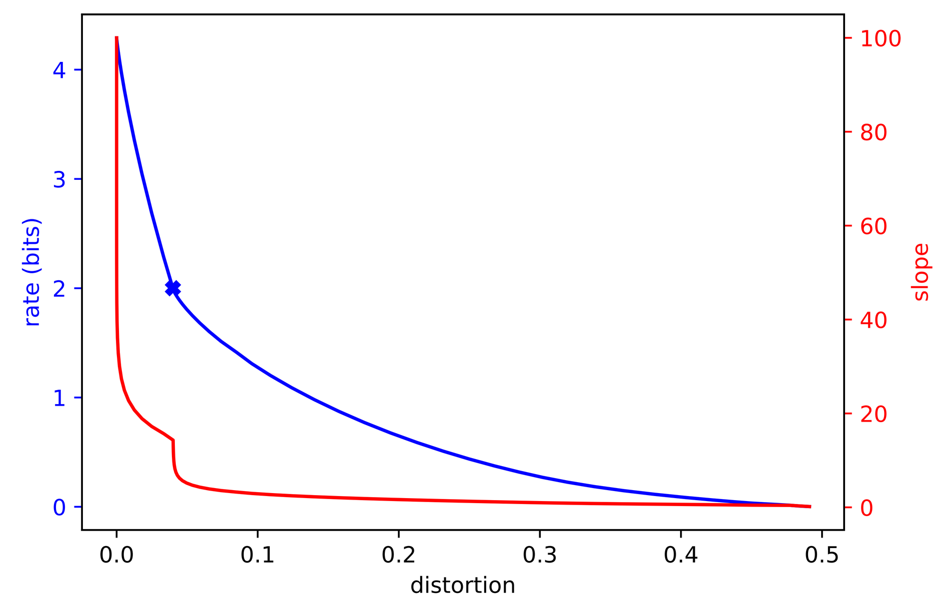
The situation this setup is modeling is one in which a single cue is present or absent, and a signal tries to keep track of whether it does. This is precisely the kind of situation where many theorists (certainly Sterelny and Burge, for the reasons reviewed above) would see the postulation of representations as entirely idle.

Vervet monkey alarm call (https://www.bbc.co.uk/programmes/p016dgw1)



A two-predator souce

The predators the vervet monkey targets are modeled as homeostatic property clusters, in their turn represented by two Bayesian networks, each with a parent node and four children.



The rate-distortion function for the two-predator source

There is a clear "sweet spot"---a sudden drop in the usefulness of extra rate, see the red curve---when the system hits a rate of 2 bit/use. I.e, there is, in a certain principled sense, an optimal level of lossy compression. I claim that this is no coincidence. Our representation-attributing practices gravitate towards this kind of situations.

Optimal Encoding Strategy:
First divide the incoming signal in two halves, one corresponding to properties through ; the other corresponding to properties through .
If there is a majority of 1s in the first half of the original signal set the first bit of the signal to 1. Otherwise set it to 0. Ditto for the second half of the original signal and the second bit of the signal.
Optimal Decoding Strategy:
If the first bit in the incoming signal is 1, set the first half of the decoded signal to 11111. Otherwise, set it to 00000. Ditto for the second bit and the second half of the decoded signal.

Optimal encoding-decoding pair

The Rate-Distortion Approach

A signal, *M*, in a certain information-processing pipeline, *P*, is a representation if the following two conditions are met:

Existence:
There are sweet spots in the rate-distortion curve associated with *P*.

Optimality:
M is produced as part of an encoder-decoder strategy that occupies the vicinity of one of these sweet spots.

mail@manolomartinez.net
manolomartinez.net
@manolophilo