UC Berkeley School of Information

Theorizing Information

Concepts of Information i218

Geoff Nunberg

March 31, 2010

The Theory-of-Information Food Chain

A rough but useful distinction: TOFI "Producers" and "Consumers" (Brian Smith)

TOI Producers:

> People concerned with "theories or inquiries that address information as a phenomenon in its own right, and who therefore bring forward specific theories about it. … people or theories or investigations that analyse what information is."
>
> > I.e., information theory, philosophers, theories of computation, documentalists & information studies,

The Theory-of-Information Food Chain

TOI Consumers (or "developers")

"People, theories, fields, etc.,… which employ the notion of information substantively but who more rely on information itself, or a concept of information, 'being available' for substantive use."

E.g., geneticists who theorize DNA as an information-carrier, psycholinguists who deploy information theory in studies of information processing, economists…

These theories may be "grounded" in a prior theory of information, linked to it, or rely on it, but usually wind up reinterpreting the notion for their own purposes.(GN)

Two Producers' Theories of Information

Two "producers" of theories of information

"Information theory" ("mathematical theory of communication).

Some applications of MTC

Philosophical accounts of (semantic) information

definitions

and antecedents

"[I]nformation is a very elastic term ... a more specific meaning ... the sender mentally selects ... At each selection there are eliminated all of the other symbols which might have been chosen ... more and more possible symbols sequences are eliminated ... the information becomes more precise ... Inasmuch as the precision of the information depends upon what .... might have been .. reasonable to hope to find in the number of these sequences the desired quantitative measure .... desirable ... to eliminate the psychological factors involved and to establish a measure of information in terms of purely physical quantities."

R[alph] V.L. Hartley, "Transmission of Information," *Bell System Technical Journal*, 1928

Narrowing the problem down

"The word communication will be used here in a very broad sense to include all of the procedures by which one mind may affect another. This, of course, involves not only written and oral speech, but also music, the pictorial arts, the theater, the ballet, and in fact all human behavior."
Weaver, *The Math. Theory of Communication*

Narrowing the problem down

"Relative to the broad subject of communication, there seem to be problems at three levels. Thus it seems

reasonable to ask, serially:

LEVEL A. How accurately can the symbols of communication be transmitted? (The technical problem.)

LEVEL B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)

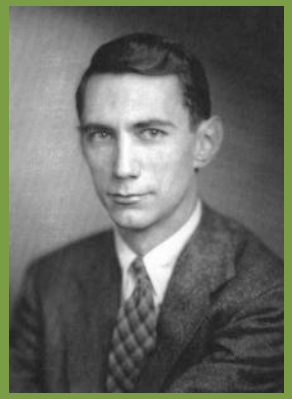
LEVEL C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.)

Cf syntax/semantics/pragmatics; form/meaning/use…

"The technical problem"

The technical problems are concerned with the accuracy of transference from sender to receiver of sets of symbols (written speech), or of a continuously varying signal (telephonic or radio transmission of voice or music), or of a continuously varying two-dimensional pattern (television), etc. Mathematically, the first involves transmission of a finite set of discrete symbols, the second the transmission of one continuous function of time, and the third the transmission of many continuous functions of time or of one continuous function of time and of two space coordinates.

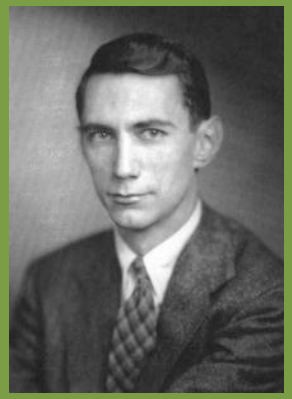
The Mathematical Theory of Communication



Rarely does it happen in mathematics that a new discipline achieves the character of a mature developed scientific theory in the first investigation devoted to it... So it was with information theory after the work of Shannon.

A. I. Khintchin, 1956

The Mathematical Theory of Communication



Information theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press. In part, this has been due to connections with such fashionable fields as computing machines, cybernetics, and automation ; and in part, to the novelty of its subject matter. As a consequence, it has perhaps been balloned to an importance beyond its actual accomplishments. … Applications are being made to biology, psychology, linguistics, fundamental physics, economics, the theory of organization, and many others. In short, information theory is currently partaking of a somewhat heady draught of general popularity…What can be done to inject a note of moderation in this situation? Claude Shannon, 1993

Elements of the Theory

"The fundamental problem of communication is that of reproducing at one point a message selected at another point."

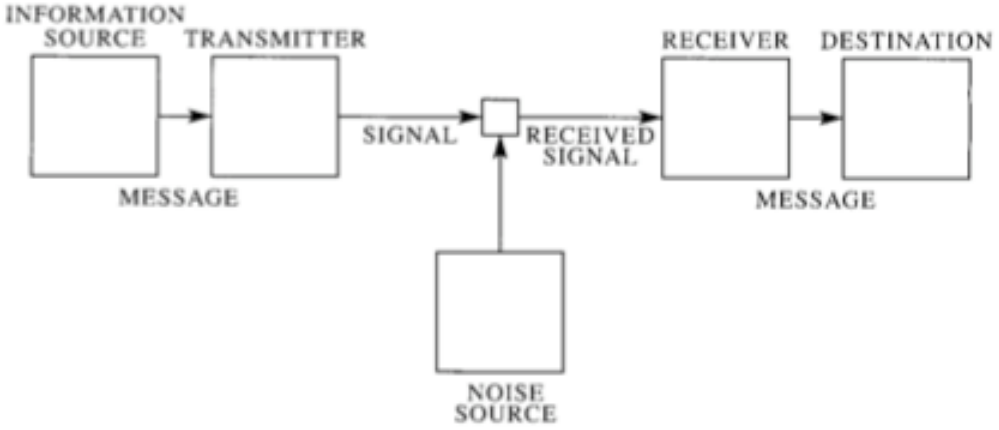
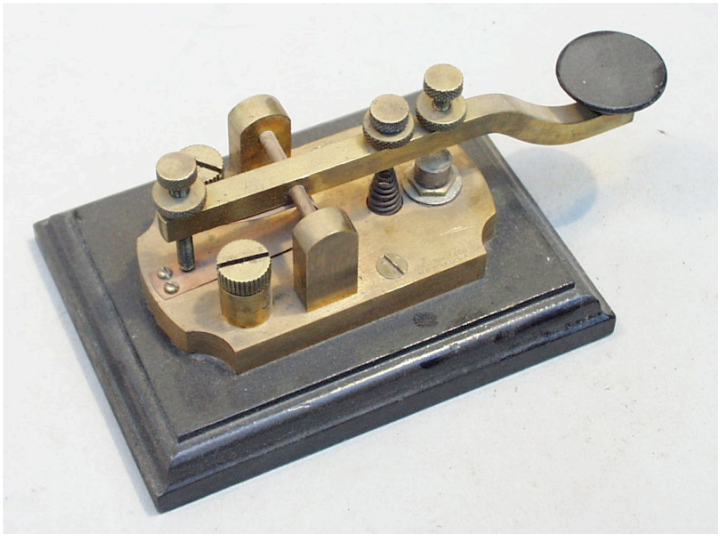


Fig. 1—Schematic diagram of a general communication system.

Varieties of Signals

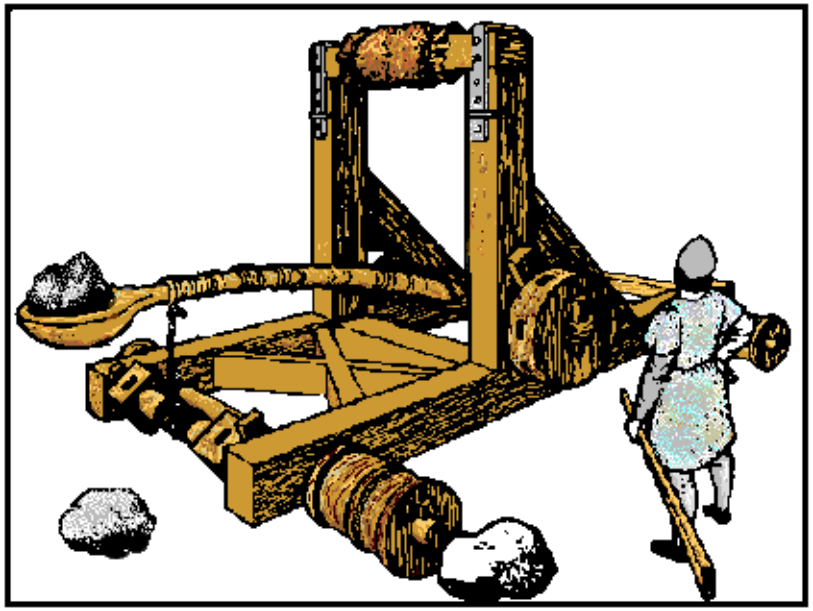


Telegraph

Varieties of Signals

Earlier Telegraph

Varieties of Signals



Very Early Telegraph

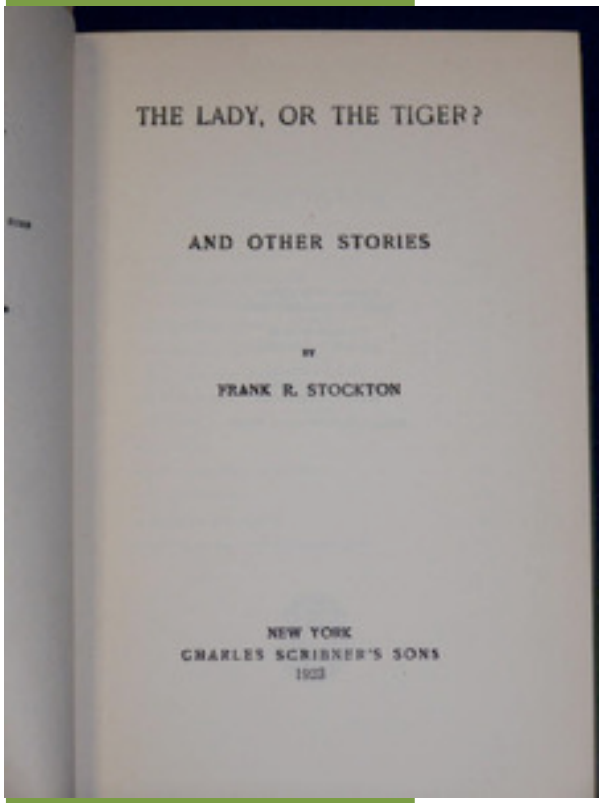
Reducing Uncertainty

Information = "Patterned matter-energy that affects the probability of choosing a particular alternative in a decision-making situation" Everett M. Rogers & Thomas W. Valente, after Shannon

Information is that which reduces uncertainty in a choice situation.

Weaver: "… in this new theory the word information relates not so much to what you *do* say as to what you *could* say. That is, information is a matter of your freedom of choice when you select a message."

A simple instance



In the very olden time there lived a semi-barbaric king, whose ideas, though somewhat polished and sharpened by the progressiveness of distant Latin neighbors, were still large, florid, and untrammeled, as became the half of him which was barbaric...
"The Lady or the Tiger," Frank Stockton, 1882

A simple instance

"Did the tiger come out of that door, or did the lady?"

"The Lady and the Tiger," Frank Stockton, 1882

Arm up → right door
Arm down → left door

A simple instance

Arm up → right door
Arm down → left door

...the amount of information is defined, in the simplest cases, to be measured by the logarithm of the number of available choices. It being convenient to use logarithms to the base 2... This unit of information is called a 'bit' ... a condensation of 'binary digit'.

$\text{Log}_2\ 2 = 1$

Signal contains 1 bit of information

uncertainty

Reducing Uncertainty

Arm up → right door
Arm down → left door

...the amount of information is defined, in the simplest cases, to be measured by the logarithm of the number of available choices. It being convenient to use logarithms to the base 2... This unit of information is called a 'bit' ... a condensation of 'binary digit'.

$\text{Log}_2\ 2 = 1$

Signal contains 1 bit of information

Less certainty

Reducing Uncertainty

"Did the tiger come out of that door, or did the lady, or did the wart hog, or did the mortgage broker?"

$Log_2\ 4 = 2$ → 2 bits of information

L down R up→ door 1
L up R down→ door 2
L down R down → door 3
L up R up→ door 4

It's not about semantics

"The significant aspect is that the actual message is one selected from a set of possible messages."

Arm up → one signal
Arm down → another signal

Anti-semantics

First off, we have to be clear about the rather strange way in which, in this theory, the word "information" is used; for it has a special sense which... must not be confused at all with meaning. It is surprising but true that, from the present viewpoint, two messages, one heavily loaded with meaning and the other pure nonsense, can be equivalent as regards information.

Warren Weaver, 1949

Anti-semantics

"The fundamental problem of communication is that of reproducing at one point a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages." Shannon, 1948

I.e., "Communication" ends when it is determined *which* message was sent.

The effects of probability

Signals are not equiprobable…

Freq. of initial letters of English words

T	15.2%
A	11.4%
H	8.5 %
W	7.0%
F	3.5 %

Knowing that a message begins with T reduces the set of possible messages by 84.8%

Knowing that a message begins with F reduces the set of possible messages by 96.5%

→ F__ is more informative than T___

The effects of probability

What every "Wheel of Fortune" viewer knows:

_o_e__ a_e _o_e __e_i__a__e __a_
_o__o_a___

The effects of probability

What every "Wheel of Fortune" viewer knows:

_o_e__ a_e _o_e __e_i__a__e __a_ _o__o_a___

V_w_ls _r_ m_r_ pr_d_ct_bl_ th_n c_ns_n_nts

I.e., consonants are more informative than vowels

The effects of probability

Morse Code: fewer bits for most common letters:

A: .- B: -... C: -.-. D: -.. E: .

F: ..-. G: --. H: .... I: .. J: ---.

K: -.- L: .-.. M: -- N: -. O: ---

P: .--. Q: --.- R: ._. S: ... T: -

U: .._ V: ...- W: .-- X: -..- Y: -.-- Z: --..

1-BIT	2-BITS	3-BITS	4-BITS
E,T	I,A,N,M	S,U,R,W,D,O,G,K	H,V,F,L,P,J,B,X,C,Z,Q,Y

The effects of state-dependence (transitional probabilities)

Signals are not ergodic (zero-memory).

Overall frequency of *h* = .06

Overall frequency of s = .06

But probabilities are not the same after an initial t:

P(h) /#t__ > P(s) /#t__

i.e., a signal string beginning with *Ts*… is more informative than one beginning with *Th*…

The effects of probability

Signals are not equiprobable…

…if we are concerned with English speech, and if the last symbol chosen is "the," then the probability that the next word be an article, or a verb form other than a verbal, is very small. This probabilistic influence stretches over more than two words, in fact. After the three words "in the event" the probability for "that" as the next word is fairly high, and for "elephant" as the next word is very low. Weaver, *Math. Theory of Comm.*

Consequences of low entropy

Redundancy permits compression

*_abc_* _ _ _ _

Consequences of low entropy

Redundancy permits compression

dabchick

Consequences of low entropy

Redundancy permits compression

dabchick

These are the times that try men's souls.

Thes ar th time tha tr men' soul

The ar th tim th tr men' soul

Redundancy & pattern permits compression

Cf Dictionary compression: "small bits"

chromate
chromatic
chromatin
chromatogram
chromatograph
chromatography
chrome
chromic
chromium
chromosome (= 86 characters)

Redundancy & pattern permits compression

Cf Dictionary compression: "small bits"

chromate
chromatic
chromatin
chromatogram
chromatograph
chromatography
chrome
chromic
chromium
chromosome (86 characters)

→ chromate F7ic F8n F7ogram FBph FDy F5e F5ic F6um F5osome (27 characters)

Consequences of low entropy

Redundancy facilitates error detection and signal recovery in noisy channels

Redundancy facilitates processing

Shannon's Theorems of Signal Transmission

Propositions deal with

- Channels with & without noise
- Entropy of a source
- Channel capacity
- Discrete vs continuous information

Applications of Information Theory

An example: "Optimizing information density through syntactic reduction" (Roger Levy & Florian Jaeger)

Assumption (per Shannon): speakers structure utterances to minimize information density (amount of information per utterance unit). I.e. speakers try to spread out the surprisal.

"speakers structure their utterances in ways that buy them time to prepare difficult words and phrases."

Cf effects on contraction, speech rate, etc.

Phonetics: speakers lengthen syllables of less familiar words. *antimetabole* vs *antidepressant*…

Applications of Information Theory

An example: Optimizing information density through syntactic reduction (Roger Levy & Florian Jaeger)

Syntactic reduction:

(1) How big is [NP the family$_i$ [RC (*that*) you cook for i]]?

Assume information density is higher when relativizer is omitted. (because then the 1st word of the rel. clause does double work.)

Then "full forms (overt relativizers) should be used more often when the information density of the RC *would be high if the relativizer were omitted.*"

1. I believe (that) that drug makes you sleepy.
2. I believe (that) this drug makes you sleepy.

Applications of Information Theory

An example: Optimizing information density through syntactic reduction (Roger Levy & Florian Jaeger)

Syntactic reduction:

(1) How big is [NP the familyi [RC (*that*) you cook for i]]?

Assume information density is higher when relativizer is omitted. Then "full forms (overt relativizers) should be used more often when the information density of the RC *would be high if the relativizer were omitted.*"

1. I believe (that) that drug makes you sleepy.
2. I believe (that) this drug makes you sleepy.

Table 1 Rate of optional that (% OPT) before pronoun that and this in CCs

	WSJ		**BC**		**SWBD**		**Total**	
Subject	% OPT	Total	% OPT	Total	% OPT	Total	% OPT	Total
Pronoun this	26%	39	58%	19	18%	50	28%	108
Pronoun that	2%	41	33%	6	9%	675	9%	722
Fisher's Exact	**p<0.01**		**n.s.**		**p<0.05**		**p<0.001**	

Applications of Information Theory

Assume information density is higher when relativizer is omitted. Then "full forms (overt relativizers) should be used more often when the information density of the RC *would be high if the relativizer were omitted.*"

Contrast the effects of pronoun case…

I believe (that) she took the test.

I believe (that) the girl took the test.

… and NP complexity:

I believe (that) the student who failed the test has dropped the course.

I believe (that) the student has dropped the course.

Applications of Information Theory

Assume information density is higher when relativizer is omitted. Then "full forms (overt relativizers) should be used more often when the information density of the RC *would be high if the relativizer were omitted.*"

Contrast the effects of pronoun case…

√I believe (that) she took the test.

I believe (that) the girl took the test.

… and NP complexity:

I believe (that) the student who failed the test has dropped the course.

√I believe (that) the student has dropped the course.

Limits of MTC as Account of "Information"

Talk of "information" normally presupposes meaningfulness.

Counter-intuitive results of Math. Theory of Communication:

> Signals with high entropy (i.e., random number sequences, audio static) convey more information than signals with low entropy (well-formed English texts, musical passages)

Limits of MTC as Account of "Information"

Talk of "information" normally presupposes meaningfulness.

Counter-intuitive results of Math. Theory of Communication:

- Signals with high entropy (i.e., random number sequences, audio static) convey more information than signals with low entropy (well-formed English texts, musical passages)
  - Floridi: "According to MTC, the classic monkey randomly pressing typewriter keys is indeed producing a lot of information."

Limits of MTC as Account of "Information"

Talk of "information" normally presupposes meaningfulness.

Counter-intuitive results of Math. Theory of Communication:

The less you know, the more informative a signal is:

"In Chicago, Illinois, President Obama…"

Theories of Semantic Information

The word 'information' has been given different meanings by various writers in the general field of information theory…. It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field.
Shannon, 1993

Efforts to reframe "syntactic information" à la Shannon to deal with phenomena of meaning, but preserve some features (e.g., of information as quantifiable, & involving reduction of uncertainty).

"Semantic information" →information is always *about* something.

Theories of Semantic Information

Yehoshua Bar Hillel, 1955: "When [Shannon's predecessor Hartley] speaks of the 'measure of information,' or the 'amount of information,' or 'information content' in a signal sequence, he has nothing else in mind than a certain function of the relative frequency of this sequence among the set of all possible signal sequences of the same length." ***BUT***

"The event of transmission of a certain statement and the event expressed by this statement are, in general, entirely different events, and the logical probabilities assigned to these events, relative to certain evidences, will be as different as will be the frequencies with which events of these kinds will occur relative to certain reference classes... the concept of semantic information has intrinsically nothing to do with communication."

E.g. "Butler went to the NCAA f_____."

Naturalism and Semantic Information

Bar Hillel and Carnap offered a theory of semantic information that is nonnaturalistic; presupposed minds and language: "The theory we are going to develop will presuppose a certain language system and the basic concepts of this theory will be applied to sentences of that system."

Reframing classical issues in terms of information

Logical inference as "elucidation" (van Benthem)

E.g., the disjunctive syllogism ("the dog")

$$\frac{P \vee Q \qquad -P}{Q}$$

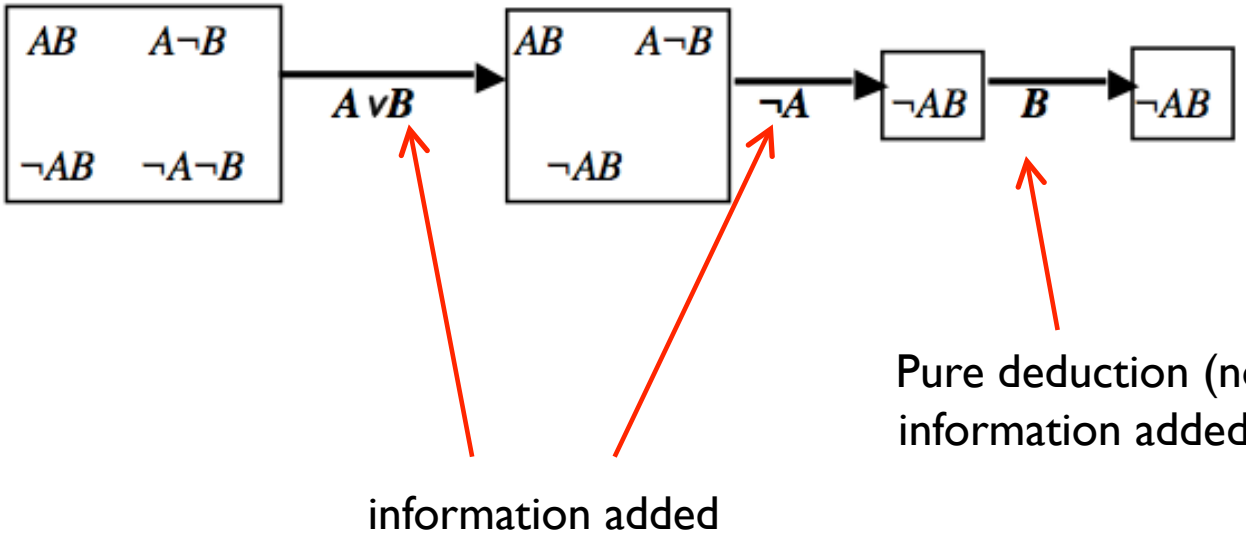
Premises do not drop out of the sky! In a café, your friend has ordered Applejack, while you took a Brandy. A new waiter comes back from the kitchen with two glasses. What we see around us every day is that the new waiter asks who has the Applejack (say), puts it down, and then puts the other glass without asking.
van Benthem

Think of this in terms of incremental information

Reframing classical issues in terms of information

Logical inference as "elucidation" (van Benthem)

Information flow reduces options: "discarding of possibilities"



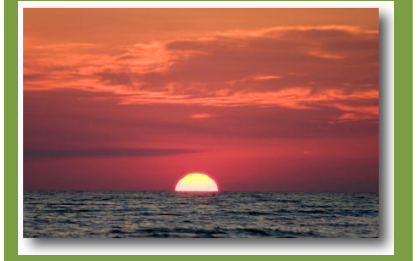
Dynamic semantics

Discourse Representation Theory (Hans Kamp), File Change semantics (Irene Heim):

> Sentences do not express truth-conditions; rather, recipes for updating a discourse representation.

> A farmer owned a donkey. He beat it.
>
> If a farmer owns a donkey, he beats it.
>
> Every farmer who owns a donkey beats it.
>
> Mary arrived home yesterday. She entered the house.

Naturalism and Semantic Information

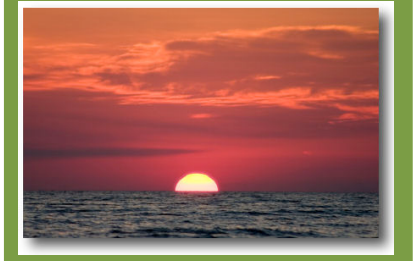


Shannon describes MCT as a solution to an "engineering problem" involving messages and signals, whether meaningful or not.

> But "signals" etc. presuppose agency/intervention (even if signals are produced, e.g. by a surveillance camera).

No way to talk about the mind-independent notion of the "information" contained in a red sunset, tree rings, or a footprint in the snow.

Naturalism and Semantic Information



No way to talk about the mind-independent "information" contained in a red sunset, tree rings, or a footprint in the snow. (Floridi's "environmental information")

"Naturalistic" theories of semantic info try to assimilate "nonnatural" (intentional) information to "natural" (nonintentional) information.

> Cf Grice's "natural meaning" ("Smoke means fire") and "nonnatural meaning" (e.g., "*Smoke* means particulate suspension emitted by burning substance")
>
> Cf reformulation of "signals" as "data"
>
> "This [omniscient] observer notes the errors in the recovered message and transmits data to the receiving point over a "correction channel" to enable the receiver to correct the errors." Shannon, p. 21

Theories of Semantic Information

Dretske: theory is intended to capture "what we normally or ordinarily mean by talking of some event, signal, or structure as carrying (or embodying) information about another state of affairs"

But "The theory is not a candidate for *Webster's Dictionary*"

Intuitive Features of Semantic Information

Events or states generate information.

Less probable events generate more information.

> E.g., "It's going to rain today" is more informative in Berkeley in July than January.

Dretske: Quantity of semantic information is reckoned relative to states of affairs that are ruled out.

> E.g., "Obama won" is more informative relative to primaries than to presidential election.
>
> Obama beat Clinton, Biden, Richardson, Edwards, Dodd, and Kucinich… "Obama won" contains 3 bits
>
> Obama beat McCain: "Obama won" contains one bit.
>
> Similarly Floridi: "There will be 3 guests tonight" is more informative than "There will be some guests tonight."

What does the MTC and Theories of Sem. Info have in Common?

a. Identical—saying the same thing, just in very different language;

b. Compatible—focusing on different aspects of the same phenomenon

c. Incompatible—saying contradictory things about the same phenomenon;

d. Unrelated—talking about two independent phenomena, just using overlapping terminology;

Brian Smith

Theories of Semantic Information

Naturalistic semantic theories of info answer to various philosophical concerns.

E.g., "Gettier problem" theories of knowledge as justified true belief.

JTB:

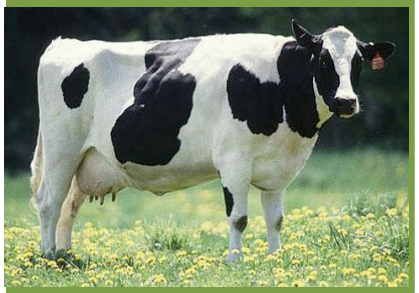
P is true

A believes that P is true,

A is Justified in believing that P is true

Problems with JTB

Farmer Brown's cow



Farmer Brown's Cow
(Martin Cohen)

Implications of Theories of Semantic Information

Naturalistic semantic theories of info answer to various philosophical concerns.

E.g., "Gettier problem" theories of knowledge as justified true belief.

> Fred Dretske: knowledge is "information-produced belief" i.e.,
>
> > K knows that s is *F* $=_{df}$ *K*'s belief that s is F is caused (or causally sustained) by the information that s is F. 6
>
> i.e. rule out accidentally true beliefs…
>
> > K believes that s is F.
> > If s were not F, K would not believe that s was F.
> >
> > But things get complicated…

Kripke's red barn

Theories of Semantic Information: Situation Theory

Perry, Barwise, Israel (Situation Theory): Information inheres in **situations**, relative to **constraints** (law-like regularities).

"A given tree has one hundred rings. That indicates that it is at least one hundred years old."

Information in situation is relative to constraint that holds between rings and seasons.

An X-ray indicates that Jackie has a broken leg.

The situation of an X-ray looking such-and-such indicates that its subject has a broken leg.

Incrememental information: Jackie was the dog X-rayed.

→ J has a broken leg

Note: information inheres in the X-ray. If X-ray is stored, information is stored. On these views, information is in the world.

The Metaphysics of Information

On these theories, information is in the world, independent of mind.

"Expanding metal indicates a rising temperature…. It meant that before intelligent organisms, capable of exploiting the fact by building thermometers, inhabited the earth." Dretske, "Misrepresentation"

Theories of Semantic Information: One Version

Note that these views presume that information must be true (i.e., to license knowledge).

Floridi: Information = "true meaningful data."

But in ordinary lg, "information" need not be true.

Results **1** - **50** of about **2,910,000** for "**false information**". (**0.25** seconds)

Results **1** - **50** of about **2,940,000** for "**incorrect information**". (**0.26** seconds)

Information and the Origin of the Mental

"In the beginning there was information. The word came later. The transition was achieved by the development of organisms with the capacity for selectively exploiting this information in order to survive and perpetuate their kind."

Dretske, "Knowledge and the Flow of Information"

I.e., Information is a resource that organisms use, like carbohydrate molecules. Sensory & cognitive mechanisms evolve so as to make use of this resource.

Information and the Origin of the Mental

How does information give rise to meaning?

> "Epistemology is concerned with knowledge: how do we manage to get things right? There is a deeper question: how do we manage to get things wrong? How is it possible for physical systems to misrepresent the state of their surroundings?... Unless we have some clue to how this is possible, we do not have a clue how naturally evolving biolgical systems could have acquired the capacity for belief."
>
> Dretske, "Misrepresentation," 1993

Information and the Origin of the Mental

How does information give rise to meaning?

"Natural signs are incapable of misrepresenting anything."

Dretske, "Misrepresentation," 1993