# *What Information is given by that sentence?*

Carnap and Bar-Hillel on Semantic Information

We intend to explicate the presystematic concept of information, insofar as it is applied to sentences or propositions and inasmuch as it is abstracted from the pragmatic conditions of its use. We shall then define, on the basis of this systematic concept of semantic information, various explicata for the presystematic concept (or concepts) of amount of semantic information and shall investigate their adequacy and applicability.

# *Sentences, not symbols*

The first thing to note is that Carnap and Bar-Hillel (CBH) try to analyze the *content* and the *amount* of information as it is carried by *sentences* (linguistic entities) or *propositions* (nonlinguistic entities, expressed by sentences), rather than the average amount of information produced by a source via choices between different symbols.



# *Reductionism*

Nevertheless, CBH hint already at a reduction to apply their analysis to the information carried by physical types or tokens:

Instead of talking about the information carried by a sound wave, one could instead talk about the information carried by the sentence:

[1] 'The sound wave ... has been transmitted.'

# *Idealization*

From the beginning, CBH emphasize that their theory should be understood as making certain simplifying assumptions. It's a theory of information that idealizes away from all pragmatic aspects and cognitive limitations:



# *Idealization*

"The semantic information carried by a sentence with respect to a certain class of sentences may well be regarded as the "ideal" pragmatic information which the sentence would carry for an "ideal" receiver whose only empirical knowledge is formulated in exactly this class of sentences. By an "ideal" receiver we understand, for the purpose of this illustration a receiver with a perfect memory who "knows" all of logic and mathematics together with any class of empirical sentences, all of their logical consequences."

# *Some Definitions* - $L^n_\pi$

n different individual constants

π different one-place predicates

connectives: ¬, v, &, →, ≡

standard definition of wff's

Centre for the Study of Logic, Language, and Information

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF



# *Some Definitions* - $L^n_\pi$

Every sentence is either $L^n_\pi$-true, $L^n_\pi$-false, or factual. Logical relations are:

i $L^n_\pi$-implies j $=_{Df}$ i → j is $L^n_\pi$-true

i is $L^n_\pi$-equivalent to j $=_{Df}$ i ≡ j is $L^n_\pi$-true

i is $L^n_\pi$-disjunct with j $=_{Df}$ i v j is $L^n_\pi$-true

i is $L^n_\pi$-exclusive of j $=_{Df}$ i & j is $L^n_\pi$-false

Centre for the Study of Logic, Language, and Information

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

*Manuel Bremer, Daniel Cohnitz*
*Information Flow and Situation Semantics*
*ESSLLI 2002*

# *Some Definitions - Q-predicators, Q-properties*

A *Q-predicator* is a conjunction in which every primitive predicate occurs either unnegated or negated and no other predicate occurs at all. A property designated by such Q-predicator is a *Q-property*.



# *Some Definitions - State descriptions*

A *state-description* is a conjunction of n Q-sentences, one for each individual.

*Centre for the Study of Logic, Language, and Information*

HEINRICH HEINE UNIVERSITAT DUSSELDORF

*Manuel Bremer, Daniel Cohnitz*
*Information Flow and Situation Semantics*
*ESSLLI 2002*

## *Some Definitions - range of sentences*

For any sentence j of the system, the class of those state-descriptions in which j holds is called the *range* of j, R(j).



## *Content of a sentence*

Now, what a sentence i says, is that the universe is not in one of those states which are described by the Z (class of state descriptions) in $V_z – R(i)$, where $V_z$ is the class of all Z. In other words: i $L^n_\pi$-implies the negation of every Z in $V_z – R(i)$.

*Centre for the Study of Logic, Language, and Information*

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

*Manuel Bremer, Daniel Cohnitz*
*Information Flow and Situation Semantics*
*ESSLLI 2002*

# Content of a sentence

These negations are called the content-elements E of i and their class the content of i, Cont(i).



# Content of a sentence

An analytic statement has minimum content, and a self-contradictory statement maximum content. In CBH words:

"A self-contradictory statement tells too much, it excludes too much, and is incompatible with any state of the universe, whereas an analytic statement excludes nothing whatsoever and is compatible with everything."

# *Content of a sentence*

According to the scholastic dictum, *omnis determinatio est negatio*, the content of a sentence is taken to be the class of those possible states of the universe (state-descriptions) which are excluded by this sentence.

In other words, the class of those states whose being the case is incompatible with the truth of the statement/sentence.



# *Content as the explicatum of information*

Now the information conveyed by a statement i is explicated as the class of state-description excluded by that statement, Cont(i).

## *proper m-functions*

CBH go on to define measure functions over ranges, one of which, $m_p$ is supposed to be the logical probability on no evidence. The logical probability of a sentence i is 1 iff i is L-true and 0 iff i is L-false. The content measure of i, cont(i), is by definition the logical probability of ¬i, $m_p(\neg i)$.



## *proper m-functions*

The choice of this very function can easily be motivated: there is one clear adequacy criterion for a proper m-function, the greater the logical probability of a statement, the smaller its content measure.

Now, the mathematically simplest relationship that fulfills this requirement is obviously the complement to 1. Let $m_p(i)$ be the logical probability of i. Then 1 – $m_p(i)$ can be taken as the plausible measure for the content of i!

*Centre for the Study of Logic, Language, and Information*

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

*Manuel Bremer, Daniel Cohnitz*
*Information Flow and Situation Semantics*
*ESSLLI 2002*

# *Amount of Information - An Alternative*

If we, for simplicity, assume the in-value of each incoming basic sentence as being 1 (rather than a decrease of this value, as cont predicts), we arrive at the following alternative formula for the amount of information, inf, for any sentence i:

$$\text{inf}(i) = -\text{Log } m_p(i)$$

Which is analogous to MCT (Mathematical Communication Theory).



# *Amount of Information*

If we want on these grounds to define the expectation of information in a given situation in which we have a number of mutually exclusive alternatives with the logical probabilities $p_i$, we arrive at the familiar entropy expression:

$$-\sum_{i=1}^{n} p_i \log p_i$$

# *Ambiguous Intuitions*

Nevertheles, inf is not the only possible definition of information. The basic idea behind inf was the Popperian intuition that the more alternatives a statement excludes, the more informative it is. Although inf presents one way of measuring this, there is, as we have seen already, the more direct way to define inf (remember 'cont') via the relative number of alternatives it excludes:

$$\text{cont}(i) = 1 - m_p(i)$$



# *Ambiguous Intuitions*

Intuitively, the difference is supposed to be this: Whereas cont might be viewed as a measure of the substantial information a statement carries, inf measures its surprise value, the prior unexpectedness of its truth.

# *cont and inf*

Let's get a feeling of this difference:

cont(i&j) = cont(i) + cont(j)
iff (i v j) is logically true

inf(i&j) = inf(i) + inf(j)
iff i and j are independent with respect to their logical probability.



# *cont and inf*

inf(i) = cont(i) = 0
iff i is L-true.

More interesting are their differences, though:

cont(i/j) = cont(i→j)
inf(i/j) = –log mp(i/j)

This is one reason for the preference given in MCT to the correlate of inf.

## *cont or inf*

It seems that one of our intuitions is that the amount of information of some statement i relative to some statement j should be the same function of the probability of i given j as the absolute amount of information of i of the absolute probability of i. To fulfill this requirement means to have a log type of function.



## *Well, it's explication*

To figure out that there are more than one explicata for our prescientific concept of the amount of information a message carries, is not too surprising or problematic. Explications are meant to uncover exactly such prescientific confusions.

# *Well, it's explication*

Another one of our prescientific intuitions is the following: Asked what we regard as the appropriate relation between the absolute amount of information of a given statement i and its amount of information relative to any j, we are normally very positive that no increase in the evidence should increase the amount of information, though it might not necessarily decrease it.



# *A plus for cont*

Now it can be shown thbat

$$\text{cont}(i/j) \leq \text{cont}(i)$$

whereas the corresponding statement for inf does not hold.

# *Difference in practical situations?*

Bar-Hillel tells the follwing story to illuminte the practical difference inf and cont might cause:

"There was a bridge party in A's villa, with B, C, D, and E participating; A was the host and only kibitzed. When the last rubber was finished and the guests were looking for A to take leave of him, they found him murdered in the garden. Every one of the four players had been the dummy at one time or another and had left the room for refreshments. Each one had, on the available evidence, an equal opportunity for murdering A.



# *Difference in practical situations?*

A reward was promised to those who could forward information leading to the identification of the murderer. A day later, X came and produced evidence sufficient to prove that B could not have been the murderer. The next day Y showed, to the district attorney's satisfaction, that C was innocent. The following day, Z did the same for D. Whereupon E was duly convicted and electrocuted.

*Centre for the Study of Logic, Language, and Information*

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

*Manuel Bremer, Daniel Cohnitz*
*Information Flow and Situation Semantics*
*ESSLLI 2002*

# *Difference in practical situations?*

The problem now for the district attorney was hoow the reward should be distributed; he had to adopt some numerical distribution. He could evaluate the information given by the three informants according to the absolute inf or absolute cont values of their statements, or according to their measures relative to the information he received, or according to any explicit function whatsoever.



# *Difference in practical situations?*

I asked six friends at MIT how they would have handled the situation. I received six different answers. One, a newcomer to MIT, with little previous contact with Information Theory, would have distributed equally between X, Y, and Z.

## *Difference in practical situations?*

Another claimed that the information supplied by Y was worth more than that supplied by X, since X's testimony excluded one suspect out of four, whereas Y's testimony eliminated one out of three, and similarly for Z. He was in favor of distributing the reward according to the proportion 1/4 : 1/3 : 1/2.



## *Difference in practical situations?*

A third agreed with the second's evaluation but argued for a distribution of the reward according to a logarithmic scale.

A fourth wanted to give all of it to Z, since he alone achieved the identification of the murderer.

A fifth, an Iranian, was sure that if the story had happened in his country some years ago, the attorney would have kept the reward for himsellf, which is probably exaggerated; and I have forgotten what the sixth had to say.

*Centre for the Study of Logic, Language, and Information*

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

*Manuel Bremer, Daniel Cohnitz*
*Information Flow and Situation Semantics*
*ESSLLI 2002*