STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

SP 35

D. van Dantzig

Some informal information on "information".

Synthese (An International Journal for the Logical and the Psychological Study of the Foundationss of Science) 9 (1954), nos.3-5, p. 137-144.

Reprinted from

SYNTHESE

An International Journal for the Logical and the Psychological Study of the Foundations of Science

Volume IX, Issue 3, Nos. 3—5

PUBLISHER

F. G. KROONDER BUSSUM (NETHERLANDS)

D. VAN DANTZIG
Amsterdam

SOME INFORMAL INFORMATION ON "INFORMATION" [1]

In "An enquiry into meaning and truth" [2] Bertrand Russell tells a story about a doctor who comes home, late at night and tired.. His wife, somewhat talkative after having had already a good rest, asks: "And did Mrs. X have her baby?". "Yes", the doctor says, "Is it a boy or a girl?" "Yes", the doctor says.

"The last answer", Russell says, "though logically impeccable would be infuriating". Our first question, also discussed shortly by Russell in his stimulating book, is: "*Why* is this answer infuriating?".

The answer to the latter question is quite simple. After the doctor's first "Yes" the lady knows [3] already that the babe is "a boy or a girl", but she wants to know something more, and this further information the doctor withholds by affirming only what she knew already. Otherwise stated: she knew that among the two statements

"It is a boy"
"It is a girl"

(where 'It' is the new-born) one is true. The doctor confirms this, which is superfluous [3], instead of telling *which one* of the two statements is true. So the lady becomes angry because, though she gets an answer, she does not get the *information* she wants.

The socalled "theory of information", which since a few years is being developed, admits even a quantitative measure of an "amount of information". As we shall see below, the lady wants exactly [4] *one unit* of information.

1. This paper contains the main content of the introductory talk held before the 9th International Summer Conference on August 10, 1953. It was written post factum and somewhat elaborated and extended. The underlying ideas were scetched for the first time in a session of the Epistemological Section of the International Society for Significs, held on April 12, 1952. Since then a paper by Carnap and Bar-Hillel appeared, which in some respects overlaps ours.

2. London, 1940.

3. We leave the possibility of twins, as well as physical abnormities like a child having no sex, or both sexes, out of consideration.

4. Leaving out of consideration the slight excess of male over female births.

The theory of information has two sources. The first of these is a theory by which the renowned statistician and biometrician Ronald A. Fisher showed in 1925, how to make the best use of the information contained in a group of observations for the purpose of estimating an unknown quantity. The second of these is the "Communication Theory" in which Claude Shannon and Norbert Wiener, basing themselves on older researches, showed how to send as much information as possible per unit of time through a given "channel" (e.g. telegraph- or telephone-wire, radio- or television-transmitter, etc.). We shall not go here into the two quantitative definitions of "amount of information", due to Fisher and Shannon, which are not in complete accordance with each other, but which were recently (1951) unified in a paper by S. Kulback and R. A. Leibler.

For the question arises, why the *significists* are interested in this subject, as they are not immediately concerned either in mathematical statistics or in communication engineering.

There are two reasons for the significist's interest in the subject, even when leaving aside his perhaps sometimes somewhat awkward hobby of nibbling at significations of terms used in other fields.

For one thing the engineer's communication theory has rapidly developed into a theory of the transmission of sense data through the nervous system in animals and men. And, as the significist before all is interested in the phenomenon of mutual human understanding, it is of the utmost importance for him to keep abreast of the results of this research.

Even more directly in his line, however, is a second application of the new science, which until now has hardly been developed, and the possibility and desirability of which was first outlined by Warren Weaver in his appendix to Shannon's paper. It deals with the concept of "semantic information", and is intimately connected with some of the concepts, introduced into significs by G. Mannoury, in particular the concept of "indicative part" of an act of discourse. The main reason why the International Society for Significs has chosen the subject of Information Theory as a main theme for this conference, is therefore contained in its wish to find out how far and in which way the concepts developed and results obtained by workers in the different branches of information theory can be adapted to the needs of significs.

It is not possible as yet to outline such an adaptation of information theory to significs. I must therefore restrict myself to a few remarks and a rough scetch of what *may* become possible after further research.

In the following considerations the semantic concept of "information" is considered as belonging to what I might call the "Logic of partial knowledge", which is a part of semantics. The term 'semantics' is taken over from linguistics, where it denotes, in particular since Michel Bréal's "Essai de Sémantique" (1897) the study of words with regard to their signification. This is done in contradistinction to syntaxis and grammar, which study the rules according to which sentences are built up out of words and words out of letters. So it is a syntactical statement to say that "Mrs. X was delivered from a baby" is a linguistically correct sentence (or at least becomes so if the letter X is replaced by the name of the lady), whereas "Delivered X baby was Mrs. from a" is not. Neither syntaxis nor semantics in the linguistic sense deal with the question whether a sentence under consideration is *true* or not.

In symbolic logic (logical) syntaxis contains a set of rules according to which logical formulae (or "sentences") may be formed out of their elements, whereas semantics contains a.o. rules according to which formulae or sentences may be accepted as being "true". Without going into the (formal) concept of "truth" we remind here only that a disjunction of two sentences, e.g. "It is a boy or it is a girl" or shortly "B or G" is true if 'It is a boy' (shortly 'B') is true, and also if 'It is a girl' (shortly 'G') is true, *and in no other case*.

Returning now to the doctor's wife, we must remark that after her husband's first "Yes" she *knew* 'B or G' to be true, without knowing either 'B' or 'G' to be true. According to her "partial knowledge" therefore 'B or G' is true without either 'B' or 'G' being true — as yet. Of course, also neither of these two statements is known to be false, for if she knew e.g. 'B' to be false, then she would know 'G' to be true. Such statements to which neither the predicate 'true' nor the predicate 'false' has (as yet) been attributed might be called "amphoterous". [5]

Now, whereas in syntaxis no "truth values" (i.e. the predicates "true" of "false") are attributed at all, and in logical semantics they usually are attributed to a disjunction only if each of its constituents has a truth value, the semantic theory of information can be considered as dealing with a sequence of intermediate stages, in each of which some of the statements have obtained truth values, whereas other ones have not (have remained amphoterous), viz. in

5. The term has been used in a somewhat similar sense by E. W. Beth, "Wijsbegeerte der Wiskunde", 1948, p. 115.

such a way that in each subsequent stage the truth values accepted before remain valid, whereas some hitherto amphoterous statements obtain truth values. The "information" given then consists of the attribution of truth values to hitherto amphoterous statements. [6]

The quantitative measure of the information, sometimes called "the amount of information", is defined by requiring 1° that the information measure is additive (i.e. if subsequent informations are given, then the measure of the total information is the sum of the measures of its constituent partial informations) and 2° that a *unit* of information is given if the information determines whether any given statement is true or not, provided both possibilities have probability $\frac{1}{2}$. Such a unit is called a "binary unit", abbreviated as "bit".

It can be proved then that the determination of that one out of n mutually exclusive possibilities that is true, provided one of them is known to be true already and they have equal probabilities, requires an amount of information consisting of log n bits, the logarithms being taken with basis 2.

We illustrate this with an example. A chess-board consists of 64 $= 2^6$ fields. The logarithm of this number with basis 2 is 6. Hence 6 bits are required to determine anyone of the fiels, assuming these to have equal probabilities. We show this by an example, noting beforehand that every bit of information halves the number of available fields. The 6 bits may be successively:

1. The row-number is even (hence 2, 4, 6 or 8)
2. In the initial position of a game of chess the field is occupied by a chessman (hence the rownumber is 2 or 8)
3. This chessman is black (hence rownumber $= 8$)
4. The field is white (hence the chessman is the Queen's Knight, the Queen, the King's Bishop or the King's Rook)
5. It is next to the (black) king (hence Q or KB).
6. It is at the king's left hand (hence KB).

Hence the datum KB 8, or in the simpler continental notation, f8, and also the complete determination of any other field on the chessboard contains 6 bits of information.

The significist, of course, is interested in such simple examples for the sake of illustration only. His real interest lies with the infinitely more complex cases of information given in the communication between human beings in ordinary life. It is mainly this complexity

6. The exact definition, which we will not give here, is based on considering information as the passage of a Boolean algebra to another one homomorphic with it and may (but must not) be made quantitative by means of probability measures.

which makes the actual computation of the measure of some information given quite illusive.

Nevertheless, the concept of information measure is very useful for the significist, and closely related to Mannoury's concept of "indicative element". Both concepts cannot be said to be equivalent, if only because Mannoury's concept, which dates from a time long before the modern more exact theory was created, is less precise.

Also other ideas introduced by Mannoury find their counterpart in the modern theory. Mannoury has drawn attention to the fact that the concept of "meaning" of a word is not unique, but that (at least) a distinction between the meaning it has for the speaker and that for the hearer is necessary. On the other hand communication engineers are greatly occupied with the fact that a signal transmitted through a channel is always more or less disturbed by what he calls 'noise'. The difference between the signal emitted and the signal received, i.e. the noise, corresponds with Mannoury's difference between "speaker's meaning" and "hearer's meaning". Here also the analogy is not complete, in particular as the communication engineer is especially interested in *random* noise, whereas the significist would like to pay greater attention to the *systematic* deviations of the information received from the information emitted, due to the individual characteristics of the sending and the receiving apparatus (the speaker and the hearer). More in particular these deviations are due to the facts that the information they obtained *previously* and therefore their interpretation of the signal considered is different, and that they select and interpret the signal with respect to different *purposes*.

This leads immediately to a discussion of the concept of *relevant* information.

A large part of the information given in actual life is irrelevant to the purposes of the receiver. To mention a characteristic case: information given by a newspaper that the prime ministers of two countries A and B met may be of interest to a reader. Very often the newspapers add information about the place where they met, the duration of their talk, the dress and type of hat they wore, the make and colour of the cars they came in, about whether they smiled or not after the meeting, etc., most of which is almost completely irrelevant to the two questions which really interest the politically-minded reader: did they come to an agreement? If so, which one?

Information theory, even in its present stage, is able to deal with similar phenomea, though of a far simpler nature. In order to illustrate this, we give an example of a situation which can be dealt with completely in the semantic theory of information. Let the receiver R of the information know already that a point P is situated somewhere in the part of a plane C within a given curve cf. fig. 1), and let him be interested in the question whether it lies in the part A or the part B [7].

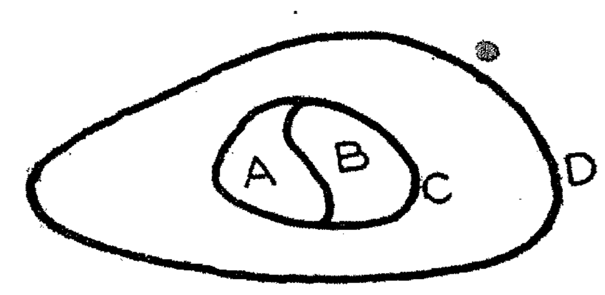

Fig. 1.

If the information given states that P is situated somewhere within the curve D, this is completely irrelevant, because it is known already to the receiver, as it lies within C and therefore a fortiori within D.

Sometimes, however, irrelevant information is given by stating that some point Q is situated somewhere within a curve E (fig. 2).

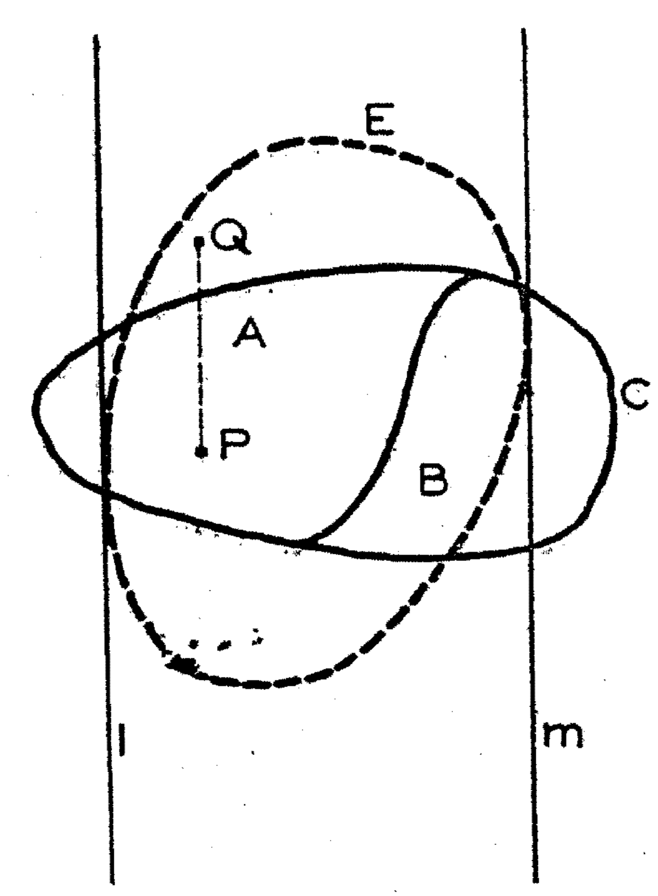

Fig. 2.

If it were known that Q were the *same* point as P, the information would be highly relevant, because it would exclude a small part of A, and a large part of B as a possible place of P. Often, however, nothing is known about identity, or even a weaker relationship between the points P and Q, and then the information about Q is completely irrelevant to the question about P. The information about Q may be *partially* relevant to that question, if it is known e.g. that P and Q are situated on *the same* vertical line. In that case the information about Q implies that the parts of A and B outside the vertical strip bounded by the lines l and m is excluded.

[7]. We assume that it cannot lie on one of the curves drawn.

Irrelevant information is the main characteristic of a well-composed detective story. The normal situation is as follows. A murder has been committed, and it is asked to find out the murderer. Let us assume — as it sometimes, but not always, is the case — that the data are such that only one out of a given group of persons can be the murderer, and that they initially have equal probabilities of being so [8]. The story then consists of an enormous amount of completely irrelevant information, under which the relevant information is carefully hidden.

This leads us to the concept of *misleading* information, which also is one of the main characteristics of a good detective story. Let us again assume that the information required is an answer to the question whether a point P is situated in a part A or B of the plane (fig. 3) and let the information given at one stage state

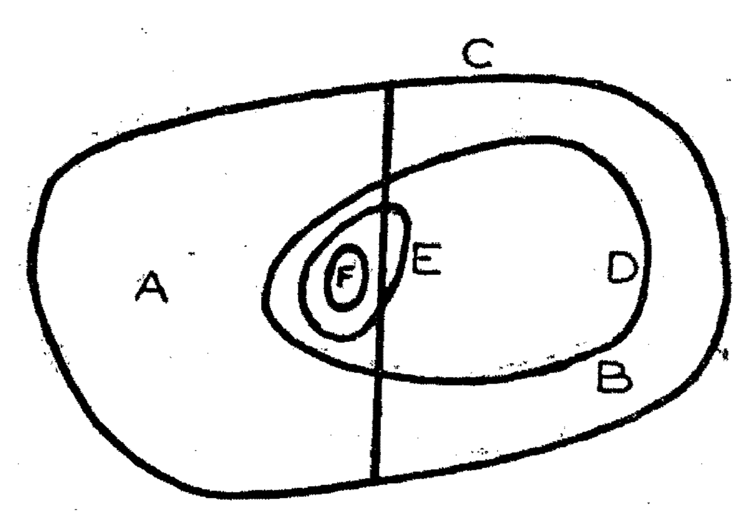

Fig. 3.

that it is situated within the curve D. Assuming that the probability that P is situated within some domain is proportional to its area, this information is highly relevant, as it makes the probability that P is in A very small and that it is in B very large. If, nevertheless, the point P is situated somewhere in the remaining part of A, this information is highly misleading. Mathematically, one might consider this as negative information relative to the question posed, although the information taken as such (namely that the domain in which P is situated is narrowed down from C to D) certainly is positive. Remarkable is the fact that the relevant information being positive or negative depends on the true position of P. If the subsequent information is given that P is situated within E, this again is highly relevant. The probability of P being within B which had become very large has suddenly become very small. It is apparently conflicting with the previous information. Whether the relevant information is positive or negative again depends on whether the true position of P is in A or in B.

8. Cf. Agatha Christy, Cards on the table, where their number equals four, so that the amount of information required is ${}^2\log 4 = 2$.

All these examples show that information implying only a change of probability, though certainly relevant, may be of very little use for getting knowledge about the true situation, unless special precautions are taken to ensure that only or mainly the probability of the true situation increases. This is done, to a certain extent, in mathematical statistics.

If finally the information is given that P is situated within the curve F, then it follows that P is in A and the required information is obtained.

Finally we must consider the possibility — also very frequent in detective stories — that some of the information given is conflicting, or even false. In order to deal with this situation, we must take account of the fact that the statement "this information is false" contains information *about* information. In the semantic theory of information this can be dealt with by passage to a meta-system of the one studied hitherto. We shall not go into this at present.

Also is this neither the place nor the moment to describe the precise symbolic logical form of the considerations given above, which, anyhow, does not lead to great difficulties.

Concluding I might state that the semantic theory of information can be built up without too great difficulties and provides the significist with a useful "mathematical model" for his researches, which admits a complete discussion in simple cases and a valuable insight in the complex ones in which he really is interested.