O. LOMBARDI

DRETSKE, SHANNON'S THEORY AND THE INTERPRETATION OF INFORMATION

1. INTRODUCTION

In his well known book, *Knowledge and the Flow of Information*, Fred Dretske (1981) attempts to build a bridge between philosophy and cognitive sciences by introducing the concept of information in the theory of knowledge. He distinguishes between sensory processes (seeing) and cognitive processes (recognizing) in terms of the different ways in which the received information is coded. In the final part of his book, he analyzes the capacity of physical systems to hold beliefs and to develop concepts on an informational basis.

For these purposes, Dretske begins by examining the notion of information as characterized in Shannon's theory. But although this is his starting point, Shannon's theory has, he argues, two main limitations: first, it is unable to handle the information contained in *individual* messages and, second, it is a quantitative theory dealing only with *amounts* of information and ignoring its *content*. For these reasons, Dretske proposes some formal modifications of the standard theory to make room for individual amounts of information. On the basis of the proposed changes, he elaborates a semantic theory which attempts to capture what he considers to be the nuclear sense of the term 'information', that is, information as something capable of yielding knowledge.

In this article it is shown, first, that Dretske's modifications suffer from some formal defects. It is then indicated precisely how these defects can be remedied in order to preserve Dretske's general proposal. In fact, it is shown that, if the changes are introduced in a formally correct way, Shannon's theory can express much more than what Dretske himself assumes. Finally, it is argued that the semantic character of Dretske's theory relies neither on the definition of informational content nor on the intentionality of the natural laws underlying the transmission of information. What confers a semantic dimension to Dretske's theory is a particular

Synthese (2005) 144: 23–39

interpretation of the very nature of information, which differs widely from the interpretation usually adopted in physical sciences.

2. BASIC CONCEPTS OF SHANNON'S THEORY

The theory of information was originally proposed to solve certain specific technological problems: in the early 1940s, it was thought that the increase of the transmission rate of information over a communication channel would increase the probability of error. Claude Shanon's paper 'The Mathematical Theory of Communication' (1948) surprised the community of communication engineers by proving that this was not true: as long as the communication rate is below the channel capacity, which is easily computed from the characteristics of the channel, information can be transmitted with no errors. This work has immediately led to many applications in technological fields such as radio, television and telephony. At present Shannon's theory has become a basic element in the training of communication engineers.

According to Shannon's theory, communication requires a source *S*, a receiver *R* and a channel *CH*:

If *S* has a range of possible states $s_1, \ldots, s_n$ whose probabilities of occurrence are $p(s_1), \ldots, p(s_n)$, the *amount of information generated at the source by the occurrence of* s_i is defined as:[1]

(1) $$I(s_i) = log\, 1/p(s_i),$$

where '*log*' is the logarithm to the base 2, and the resulting unit is called '*bit*' – a contraction of *binary unit*.[2]

But Shannon's theory is not concerned with the occurrence of specific events; rather it is concerned with the communication process as a whole. Hence, the average amount of information generated at the source is defined as the average of the $I(s_i)$ weighted by the corresponding probability:

(2) $$I(S) = \Sigma_i\, p(s_i) I(s_i) = \Sigma_i\, p(s_i) log\, 1/p(s_i).$$

The maximum value of $I(S)$ is $log\ n$, which obtains when all the $p(s_i)$ have the same value, $p(s_i) = 1/n$.

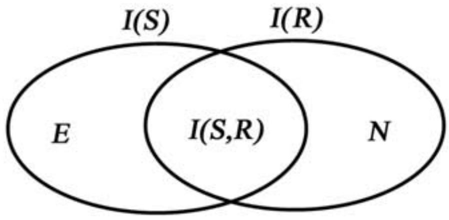
Analogously, if R has a range of possible states $r_1, \ldots, r_m$ whose probabilities of occurrence are $p(r_1), \ldots, p(r_m)$, the *amount of information received at the receiver by the occurrence of* r_j is:

(3) $I(r_i) = log\, 1/p(r_i)$,

and the *average amount of information received at the receiver* is defined as:

(4) $I(R) = \Sigma_j p(r_i) I(r_i) = \Sigma_j p(r_i) log 1/p(r_i)$.

The relationship between $I(S)$ and $I(R)$ can be represented by the following diagram:



where:

- $I(S, R)$: *transinformation.* Average amount of information generated at S and received at R.
- E: *equivocation*. Average amount of information generated at S but not received at R.
- N: *noise*. Average amount of information received at R but not generated at S.

As the diagram shows, $I(S, R)$ can be computed as:

(5) $I(S, R) = I(S) - E = I(R) - N$

E and N are measures of the amount of dependence between the source S and the receiver R:

- If S and R are completely independent, the values of E and N are maximum ($E = I(S)$ and $N = I(R)$), and the value of $I(S, R)$ is minimum ($I(S, R) = 0$).
- If the dependence between S and R is maximum, the values of E and N are minimum ($E = N = 0$), and the value of $I(S, R)$ is maximum ($I(S, R) = I(S) = I(R)$).

The values of E and N are functions not only of the source and the receiver, but also of the communication channel. The introduction of the communication channel leads directly to the possibility of errors resulting from the process of transmission. The channel CH is defined by the matrix $[p(r_j/s_i)]$, where $p(r_j/s_i)$ is the conditional probability of the occurrence of r_j given that s_i occurred, and the elements in any row must sum to 1. Thus, the definitions of E and N are:

$$(6)\qquad \begin{aligned} E &= \Sigma_j\, p(r_j)\Sigma_i\, p(s_i/r_j)\, log\, 1/p(s_i/r_j) \\ &= \Sigma_i\, \Sigma_j\, p(r_j, s_i)\, log\, 1/p(s_i/r_j) \end{aligned}$$

$$(7)\qquad \begin{aligned} N &= \Sigma_i\, p(s_i)\Sigma_j\, p(r_j/s_i)\, log\, 1/p(r_j/s_i) \\ &= \Sigma_i\, \Sigma_j\, p(s_i, r_j)\, log\, 1/p(r_j/s_i) \end{aligned}$$

where $p(s_i, r_j) = p(r_j, s_i)$ is the joint probability of s_i and r_j:

$$(8)\qquad p(s_i, r_j) = p(s_i)p(r_j/s_i) = p(r_j, s_i) = p(r_j)p(s_i/r_j).$$

The *channel capacity* is given by:

$$(9)\qquad C = max\, I(S, R)$$

where the maximum is taken over all the possible distributions $p(s_i)$ at the source.[3]

The strong relationship between the characteristics of the channel and the values of E and N allows us to define two special types of communication channels:

- *Equivocation-free channel (E = 0)*: A channel defined by a matrix with one and only one non-zero element in each column.
- *Noise-free channel (N = 0)*: A channel defined by a matrix with one and only one non-zero element in each row.

3. INFORMATION OF INDIVIDUAL MESSAGES

Traditional epistemology defines knowledge as justified true belief, considering justification and truth as independent conditions for knowledge. In his *Knowledge and the Flow of Information*, Fred Dretske proposes to replace this traditional account by an information-theoretic analysis.[4] In particular, he introduces a semantic concept of information and applies it to traditional questions in the theory of knowledge. By identifying knowledge with information-caused belief, he distinguishes between sensory processes and cognitive processes – between seeing and recognizing – in terms of the different ways in which information is coded, and analyzes the capacity of physical systems to hold beliefs and to develop concepts. But this is not the part of Dretske's work with which we are concerned. Here we are interested in his interpretation of the concept of information. Even though Dretske adopts Shannon's theory as a starting point, his approach introduces two new elements: first, a change in the basic formulas of the theory and, second, the introduction of the notion of informational content. Let us begin with the first point.

According to Dretske, one of the reasons why Shannon's theory is unable to deal with semantic issues is that semantic notions apply to *individual* items, while the theory of information deals with *average* amounts of information. Since Dretske is concerned with seeking an information-based theory of knowledge, he is interested in the informational content of particular messages and not in average amounts of information:

> if information theory is to tell us anything about the informational content of signals, it must forsake its concern with averages and tell us something about the information contained in particular messages and signals. For it is only particular messages and signals that have a content. (Dretske 1981, 48)

In order to focus on the information contained in individual messages, Dretske changes the usual interpretation of the relevant quantities of the theory: instead of considering the average amount of information $I(S)$ as the basic quantity (Equation (2)), he confines his attention to the amount of information generated at the source by the occurrence of s_a (Equation (1)):

$$(10) \qquad I(s_a) = log\, 1/p(s_a)$$

and instead of adopting the transinformation $I(S, R)$ as the relevant quantity, he defines a new 'individual' transinformation $I(s_a, r_a)$, amount of information carried by a particular signal r_a about s_a, by analogy with Equation (5) (Dretske 1981, 52):[5]

$$(11) \qquad I(s_a, r_a) = I(s_a) - E(r_a)$$

where:

$$(12) \qquad E(r_a) = \Sigma_i\, p(s_i/r_a)\, log\, 1/p(s_i/r_a).$$

According to Dretske (1981, 24), $E(r_a)$ is the contribution of r_a to the equivocation E because, given the definition of E (Equation (6)), it follows that:

$$(13) \qquad E = \Sigma_j\, p(r_j)\Sigma_i\, p(s_i/r_j)\, log\, 1/p(s_i/r_j) = \Sigma_j\, p(r_j)E(r_j).$$

Foreseeing that he might be accused of misrepresenting or misunderstanding the theory of information, Dretske emphasizes that "*the above formulas are now being assigned a significance, given an interpretation, that they do not have in standard applications of communication theory. They are now being used to define the amount of information associated with particular events and signals*" (Dretske 1981, 52). He immediately adds that, even though such an interpretation is foreign to standard applications of the theory, it is "*perfectly consistent*" with the orthodox uses of these formulas.

Dretske's aim of modifying the standard theory of information to make room for the information contained in individual messages is worth pursuing. The problem is that his formal resources suffer from some technical difficulties. The least serious of them is the use of 'signal r_a' in the definition of $I(s_a, r_a)$ (Equation (11)): r_a is not a signal but one of the states of the receiver. $I(s_a, r_a)$ should be defined as the amount of information about the state s_a of the source contained in the state r_a of the receiver. It is even more troubling that Dretske uses the same index 'a' to denote the state of the source *and* the state of the receiver, as if there were some special relationship between the elements of certain pairs (s, r). In order to make the definition of the new individual transinformation (Equation (11)) completely general, $I(s_i, r_j)$ should be defined as the amount of information about the state s_i of S received at R through the occurrence of its state r_j:

$$(14) \qquad I(s_i, r_j) = I(s_i) - E(r_j)$$

where $I(s_i)$ and $E(r_j)$ are given by Equations (10) and (12), respectively.

However, we have not yet arrived at the central difficulty. When Dretske's proposal is formally 'cleansed' in this way, its main technical problem becomes manifest. If, as Dretske assumes, $I(s_i, r_j)$ is the 'individual' correlate of the transinformation $I(S, R)$, then $I(S, R)$ must be computed as the average of the $I(s_i, r_j)$. According to the definition of the average of a function of two variables:

$$(15) \qquad I(S, R) = \Sigma_i \Sigma_j p(s_i, r_j) I(s_i, r_j).$$

Replacing Dretske's Equations (10), (12) and (14) into (15):

$$(16) \qquad I(S, R) = \Sigma_i \Sigma_j p(s_i, r_j)[I(s_i) - E(r_j)] = \Sigma_i \Sigma_j p(s_i, r_j) \times [log\, 1/p(s_i) - \Sigma_k p(s_k/r_j)\, log\, 1/p(s_k/r_j).$$

Then:

$$(17) \qquad I(S, R) = \Sigma_i \Sigma_j p(s_i, r_j)\, log\, 1/p(s_i) - \Sigma_i \Sigma_j p(s_i, r_j) \Sigma_k p(s_k/r_j)\, log\, 1/p(s_k/r_j).$$

But, on the other hand, $I(S, R)$ should also be obtained with the standard definitions (2), (5) and (6) which have been accepted by Dretske:

$$(18)\qquad I(S, R) = I(S) - E = \Sigma_i\, p(s_i)\, log\, 1/p(s_i) - \Sigma_i \Sigma_j\, p(r_j, s_i)\, log\, 1/p(s_i/r_j).$$

The trouble is that the right-hand side of Equation (17) is not equivalent to the right-hand side of Equation (18).[6] This means that we cannot accept Dretske's response to those who accuse him of misunderstanding Shannon's theory: his 'interpretation' of the formulas by means of the new quantities is not compatible with the formal structure of the theory.

It might be argued that this is a minor formal detail. However, this point has deep conceptual consequences. When Dretske defines $E(r_j)$, that is, the contribution of r_j to the equivocation E, as a summation over the s_i (Equation (12)), he makes the error of supposing that this individual contribution is a function only of the particular state r_j of the receiver. But the equivocation E is a magnitude that depends essentially on the communication channel. Thus, any individual contribution to E must preserve such a dependence. When we understand this conceptual point, we can retain Dretske's proposal by introducing the appropriate correction in the formalism. In order to perform such a correction, we must define the individual contribution of the pair (s_i, r_j) to the equivocation E as:

$$(19)\qquad E(s_i, r_j) = log\, 1/p(s_i/r_j).$$

With this definition, the average of the $E(s_i, r_j)$ is equal to E:

$$(20)\qquad E = \Sigma_i \Sigma_j\, p(r_j, s_i)\, log\, 1/p(s_i/r_j) = \Sigma_i \Sigma_j\, p(r_j, s_i) E(s_i, r_j).$$

Now we can correctly rewrite Equation (14) as:

$$(21)\qquad I(s_i, r_j) = I(s_i) - E(s_i, r_j)$$

where the average of the $I(s_i, r_j)$ is the transinformation $I(S, R)$:[7]

$$(22)\qquad I(S, R) = \Sigma_i \Sigma_j\, p(s_i, r_j) I(s_i, r_j) = I(S) - E.$$

This modified version of the formulas allows us to reach Dretske's goal; that is, to adapt the standard theory of information to deal with the information contained in individual messages. We can now return to Dretske's argument. When does the occurrence of the state r_j at the receiver carry the information about the occurrence of the state s_i at the source? The occurrence of the state r_j tells us that s_i has occurred when the amount of information $I(s_i, r_j)$ is equal to the amount of information $I(s_i)$ generated at the source by the occurrence of s_i. This means that there has been no loss of information through the individual communication. In other words, the value of the individual contribution $E(s_i, r_j)$ to the equivocation is zero (Dretske 1981, 55). In fact, according to Equation (21):

$$E(s_i, r_j) = 0 \Rightarrow I(s_i, r_j) = I(s_i).$$

But now, the value of $E(s_i, r_j)$ must be computed with the correct formula (19). At this point it is worth emphasizing again that, contrary to Dretske's assumption, the individual contribution to the equivocation is a function of the communication channel and not only of the receiver. As a consequence, it is not the individual state r_j but the pair (s_i, r_j), with the corresponding conditional probability $p(r_j/s_i)$, what contributes to the equivocation E. This means that we can get completely reliable information about the source even through a very low probability state of the receiver, provided that the channel is appropriately designed.

4. INFORMATIONAL CONTENT

In spite of having taken the formal theory of information as his starting point, Dretske reminds us that Shannon's theory is purely quantitative: it deals only with *amounts* of information, but ignores the content of information relevant for semantic questions. In fact, Shannon claims that: "*[the] 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, 379). But Dretske's main purpose is to formulate a *semantic* theory of information capable of grasping what he considers the nuclear sense of the term 'information': "*A state of affairs contains information about X to just that extent to which a suitable placed observer could learn something about X by consulting it*" (Dretske 1981, 45). In order to reach this goal, he opens the third chapter of his book, 'A Semantic Theory of Information', with the definition of informational content (Dretske 1981, 65):

> A state *r* carries the information that *S* is *F* = The conditional probability of *S*'s being *F*, given *r* (and *k*), is 1 (but, given *k* alone, less than 1).

where *k* stands for what the receiver already knows about the possibilities existing at the source.[8]

Dretske seems to suggest that the semantic character of his theory is captured by this definition to the extent that the concept of informational content takes into account the semantic aspects ignored by Shannon. However, this is not the case: the semantic character of Dretske's proposal does not depend on the concept of informational content. Of course, the definition of this concept cannot be formulated in the context of the original theory of Shannon, but it can be adequately expressed in terms of the new quantities referred to individual messages. In fact, the concept of informational content can be – more precisely – defined as follows:

> A state r_B of the receiver contains the information about the occurrence of the state s_A of the source iff $p(s_A/r_B) = 1$ but $p(s_A) < 1$, given the knowledge of the probability distribution over the possible states of the source.

where s_A stands for S's being F. If the right formulas are used, we can be sure that:

- If $p(s_A) < 1$, then $I(s_A) > 0$ (Equation (10)), that is, there is a positive amount of information generated at the source by the occurrence of s_A.
- If $p(s_A/r_B) = 1$, then $E(s_A, r_B) = 0$ (Equation (19)), that is, the individual contribution of the pair (s_A, r_B) to the equivocation E is zero. And if $E(s_A, r_B) = 0$, then $I(s_A, r_B) = I(s_A)$ (Equation (21)).

In other words, the definition says that r_B contains the information about the occurrence of s_A iff the amount of information about s_A received through the occurrence of r_B is equal to the positive amount of information generated by the occurrence of s_A. Dretske tries to express a similar idea when he says:

> if the conditional probability of S's being F (given r) is 1, then the equivocation of this signal must be 0 and (in accordance with formula (1.5)) the signal must carry as much information about S, $I(S, R)$, as is generated by S's being F, $I(S_F)$ (Dretske 1981, 65)

where his formula 1.5 is $I(S, R) = I(S) - E$. The problem is that what Dretske says is wrong: $p(s_A/r_B) = 1$ does not imply that $E = 0$ and $I(S, R) = I(S)$ (see Equation (6)). Why does Dretske use these formulas, which refer to average amounts of information, instead of using the new formulas, which refer to the amount of information contained in *individual* messages, the necessity for which he so strongly argued? The reason is again due to his formal error; that is to say, given his definition of $E(r_B)$ (Equation (12)), $p(s_A/r_B) = 1$ does not make this individual contribution to the equivocation E equal to zero and, therefore, he cannot guarantee that $I(s_A, r_B) = I(s_A)$. Only when the new formulas have been properly corrected, can the idea roughly expressed by Dretske be stated with precision. In short, Dretske's definition of informational content says nothing that cannot be said in terms of the quantitative theory of information conveniently supplemented to deal with individual amounts of information.

5. THE INTENTIONAL CHARACTER OF INFORMATION

For Dretske, the theoretical characterization of a signal's informational content "*enables us to understand the source (the intentionality of natural laws) of the semantic character of information*" (Dretske 1981, 81). As we have seen, the concept of informational content does not capture the semantic dimension of information to the extent that its definition can be expressed in terms of a quantitative theory of information. Nevertheless, the notion of intentionality does have a semantic significance. In fact, from Dretske's point of view, information qualifies as a semantic concept in virtue of the intentionality inherent in its transmission. The ultimate source of this intentionality is the *nomic* character of the regularities on which the transmission of information depends. This means that the channel probabilities $p(r_j/s_i)$ do not represent a set of mere *de facto* correlations, but they are determined by the network of lawful connections between the states of the source and the states of the receiver:

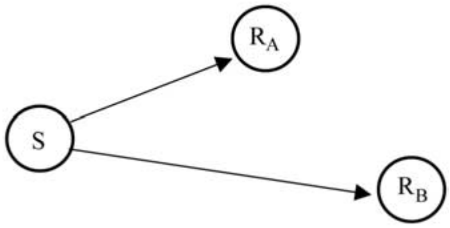
> The conditional probabilities used to compute noise, equivocation, and amount of transmitted information (and therefore the conditional probabilities defining the informational content of the signal) are all determined by the lawful relations that exist between source and signal. Correlations are irrelevant unless these correlations are a symptom of lawful connections (Dretske 1981, 77)

Dretske emphasizes this point because intentionality is what relates information to knowledge. Even if the properties F and G are perfectly correlated – i.e., every F is G and every G is F – this does not guarantee that we can know that 'x is G' by knowing that 'x is F'. If the correlation between F and G is a mere coincidence, x's being F tells us nothing about x's being G. In other words, the mere correlation and even the exceptionless accidental uniformity do not supply knowledge. This fact explains why we are sometimes in a position to know that x is F without being able to tell whether x is G, in spite of the fact that every F is G. Only on the basis of the semantic dimension of information, which relies on the intentionality of natural laws, we can state that "*information is a commodity that, given the right recipient, is capable of yielding knowledge*" (Dretske 1981, 47).

This emphasis on the intentional character of information differs widely from certain perspectives that present the theory of information in a completely syntactic way, with no mention of sources, receivers or messages. For instance, Khinchin (1957) and Reza (1961) conceive information theory as a new chapter of the theory of probability. More recently, Cover and Thomas (1991) define the basic concepts of the theory in terms of random variables and the probability distributions over their possible values. In these syntactic approaches, it is legitimate to define the transinformation between two variables even if there is no nomic relationship between them and their conditional probabilities are computed exclusively by means of mere *de facto* correlations. But purely syntactic approaches to the theory of information, even useful for supplying necessary conditions for an informational analysis of knowledge, fail to provide relevant and close constraints for an informational epistemology.[9] Summing up, Dretske is right when he claims that one of the sources of the semantic character of information is intentionality, and that information inherits its intentional dimension from the lawful regularities on which it depends. But, is this the only element that contributes to the semantic character of Dretske's theory? As we will see, the fact that Dretske's theory qualifies as semantic also depends on a particular interpretation of the very nature of information.

6. THE INTERPRETATION OF INFORMATION

Although Dretske asserts that the communication channel is defined by a network of nomic connections between the states of the source and the states of the receiver, he explicitly claims that a physical link between source and receiver is not necessary for the transmission of information. In this sense, he considers the following case (Dretske 1981, 38–39):



A source S is transmitting information to both receivers R_A and R_B via some physical channel. R_A and R_B are isolated from one another in the sense that there is no physical interaction between them; nevertheless, the correlations between the events occurring at both receivers are not accidental, but they are functions of the common nomic dependence of R_A and

R_B on S. Dretske considers that, even though R_A and R_B are physically isolated from one another, there is an *informational* link between them. According to him, it is correct to say that there is a communication channel between R_A and R_B because *it is possible to learn* something about R_B by looking at R_A and vice versa. Nothing at R_A causes anything at R_B or vice versa; yet R_A contains information about R_B and R_B about R_A. In other words, this is an example of an informational link between two points, in spite of the absence of a physical channel between them. Dretske adds that the receiver R_B may be farther from the source than R_A and, then, the events at R_B may occur later in time than those at R_A. However, this is irrelevant for evaluating the informational relationship between R_A and R_B: even though the events at R_B occur later than those at R_A, R_A carries information about what will happen at R_B. In short:

> from a theoretical point of view [...] the communication channel may be thought of as simply the set of depending relations between S and R. If the statistical relations defining equivocation and noise between S and R are appropriate, then there is a channel between these two points, and information passes between them, even if there is no direct physical link joining S with R (Dretske 1981, 38)

The possibility of informational channels lacking physical interaction shows a new face of the semantic character of Dretske's proposal. In fact, this possibility is at odds with the position adopted in physical sciences, where the existence of an unavoidable link between flow of information and propagation of signals is presupposed. Physicists and engineers unquestioningly assume that the transmission of information between two points of the physical space always needs an information-bearing signal, that is, a physical process propagating from one point to the other. For instance, this is the perspective from which the correlations between spatially separate quantum systems are considered. Let us remember the famous EPR-experiment (Einstein et al. 1935). Two quantum particles interact and then rush off in opposite directions. When they are widely separated from each other, each one of them encounters a measuring apparatus: measurements on one particle can be used, by means of the quantum correlations resulting from the interaction, to generate predictions about the other. A causal account of the correlations between both particles would require action-at-a-distance in the form of causal signals propagating faster than light; but this fact would contradict the central postulate of special relativity. Therefore, the quantum correlations in an EPR-experiment cannot be used to send messages from one subsystem to the other: there is no information flow between the two subsystems (cf., for instance, Hughes 1989. For a detailed discussion on quantum non-local correlations, cf. Berkovitz 1998).

The close link between flow of information and propagation of physical signals leads some authors to advance a step farther, arguing for the physical nature of information. This position is usually expressed by the slogan "no information without representation" in the context of the debate concerning the physical nature of information in computing processes. For instance, Rolf Landauer claims that:

> Information is not a disembodied abstract entity; it is always tied to a physical representation. It is represented by engraving on a stone tablet, a spin, a charge, a hole in a punched card, a mark on a paper, or some other equivalent (Landauer 1996, 188)

According to this view, information is a physical entity that can be generated, accumulated, stored, processed, converted from one form to another, and transmitted from one place to another. It is precisely due to the physical nature of information that the dynamics of its flow is constrained by physical laws and facts: "*Information handling is limited by the laws of physics and the number of parts available in the universe*" (Landauer 1991, 29. Cf. also Bennett and Landauer 1985). The extreme versions of this view conceive information as a physical entity with the same ontological status as energy, and whose essential property is the power of manifesting itself as structure when added to matter (cf. Stonier 1990).

These considerations show that it is possible to agree about the formal theory of information and even about some interpretative issues but, in spite of this fact, to dissent when the very nature of information is considered. Information may be conceived as a *physical entity*, whose essential feature is its capacity to be *generated* at one point of the physical space and transmitted to another point. In this case, the capability of providing knowledge is not a central issue since the transmission of information can be also used for *control* purposes, for instance, for controlling a device at the receiver end by modifying the state of the source. This view requires an information-bearing signal that can be modified at the transmitter end in order to carry information to the receiver end. If there is no physical link between the source A and the receiver B, we cannot control the states at A to send information to B and, therefore, it is not possible to define an information channel between them. But information can also be conceived as a *semantic item*, whose essential property is its capability of providing *knowledge*. From this viewpoint, the possibility of controlling the states at a point A to send information to a point B is not a necessary condition for defining an information channel between A and B: the only requirement is the possibility of knowing the state at A by looking at B. This means that we can distinguish two different concepts of information, which are relevant for different purposes. The *semantic* concept can provide fruitful insights in cognitive and semantic studies.[10] The *physical* concept is useful in communication theory, where the main problem consists in optimizing the transmission of information by means of physical signals whose energy and bandwidth is constrained by technological and economic limitations.

When Dretske presents the fundamental ideas of Shannon's theory in the first chapter of his book, 'Communication Theory', he does not distinguish between the two concepts of information and ignores the fact that the possibility of information channels with no physical signals would be absolutely rejected in physical sciences. This shows that the semantic character of Dretske's theory does not arise from supplementing Shannon's communication theory with a semantic dimension, but it is presupposed from the very beginning when the 'nuclear' sense of the term 'information' is proposed as what all scientists and technologists have in mind. Let us make the point in a different way. Dretske's semantic view and the physical view of information share the assumption that the channel probabilities are defined by lawful regularities and not by mere *de facto* correlations. However, according to the physical view, information is not a semantic item but a physical entity. Therefore, the semantic character of Dretske's theory cannot rely on (or, at least, not exclusively on) the intentionality of these underlying regularities. Dretske's theory qualifies as semantic because it takes a position about the very nature of information "*as something capable of yielding knowledge*" (Dretske 1981, 45).

7. CONCLUSIONS

Our ordinary language includes the word 'information' in a variety of contexts, as if we all precisely knew what information is. Moreover, the explosion in telecommunications and computer sciences endows the concept of information with a scientific prestige that makes supposedly unnecessary any further explanation. This apparent self-evidence hides disagreements about the sense of the term 'information' and even about the interpretation of the concept.

The attempt to use the concept of information to elucidate central notions in the theory of knowledge is a very valuable proposal. Dretske is right when he claims that a semantic theory of information is necessary for this purpose since only the semantic dimension of the theory can relate information with knowledge. In the present paper we tried to stress some points that are not sufficiently clear in Dretske's book. In particular, we showed that the semantic character of Dretske's theory does not rely on the concept of informational content to the extent that such a concept can be defined by means of the quantitative theory of Shannon *adequately* adapted to deal with individual events. On the other hand, we argued that the intentionality of the natural laws on which the transmission channel is defined is not the ultimate source of the semantic dimension of Dretske's theory, since intentionality does not distinguish between a semantic concept and a physical concept of information. The semantic dimension of Dretske's theory results from a particular interpretation of the nature of information, which differs widely from the interpretation usually adopted in physical sciences.

In recent times, a new and vitally important field of research has emerged around the concept of information. This field, which has been referred to as '*the philosophy of information*' (cf. Floridi 2002), is concerned with the critical investigation of the conceptual nature and basic principles of information. The philosophy of information presents itself as a new paradigm, which not only endorses new areas of inquiry, but also provides innovative methodologies to address traditional problems from novel perspectives. In particular, information-theoretic methods has been applied in the theory of knowledge to extend our understanding of human cognitive abilities. Therefore, the fact that this work of Dretske does not follow the line of mainstream epistemology does not mean that it has to be conceived as an isolated proposal. Dretske's informational approach to the theory of knowledge should be viewed as a manifestation of that new and increasingly fertile trend in philosophical inquiry.

ACKNOWLEDGEMENTS

I am very grateful to Fred Kronz and Silvio Chibeni for their encouragement and stimulating comments. I am also grateful to the anonymous referees whose comments greatly improved this final version. This paper was supported by grants of CONICET and of Secretaría de Estado de Educación y Universidades de España.

NOTES

1 Here we work with discrete situations, but the definitions can be extended to the continuous case (cf. Cover and Thomas 1991, 224–225).

2 One bit is the amount of information obtained when one of two equally likely alternatives is specified. The choice of a logarithmic base amounts to a choice of a unit for measuring information. In his original paper, Shannon (1948, 349) discusses the reason for the choice of a logarithmic function and, in particular, of the logarithm to the base 2 for measuring information. If the natural logarithm is used, the resulting unit of information is called '*nat*' – a contraction of *natural unit*. If the logarithm to base *10* is used, then the unit of information is the *Hartley*. The existence of different units for measuring information

shows the difference between the amount of information associated with an event and the number of binary symbols necessary to codify the event.

[3] Shannon's Second Theorem demonstrates that the channel capacity is the maximum rate at which we can send information over the channel and recover the information at the receiver with a vanishingly low probability of error (cf., for instance, Abramson 1963, 165–182).

[4] Whereas mainstream epistemology has been largely concerned with a doxastic rather than an informational analysis of knowledge, some authors coming from physical sciences find a close link between knowledge and information and assume that information modifies the state of knowledge of those who receive it. This is the case of D. A. Bell, who defines the measure of information in terms of knowledge: "*[information] is measured as a difference between the state of knowledge of the recipient before and after the communication of information*" (Bell 1957, 7).

[5] Dretske uses $Is(r)$ for the transinformation and $Is(r_a)$ for the new individual transinformation. We have adapted Dretske's terminology in order to bring it closer to the most usual terminology in this field.

[6] Perhaps Dretske's mistake is the result of misusing the subindices of the summations in (16).

[7] Replacing the new Equations (19) and (21) into (22): $I(S, R) = \Sigma_i \Sigma_j p(s_i, r_j)[I(s_i) - E(s_i, r_j)] = \Sigma_i \Sigma_j p(s_i, r_j)[log\, 1/p(s_i) - log\, 1/p(s_i/r_j)] = \Sigma_i \Sigma_j p(s_i, r_j)\ log\, 1/p(s_i) - \Sigma_i \Sigma_j p(s_i, r_j)\ log\, 1/p(s_i/r_j)$. But since $p(s_i, r_j) = p(s_i)p(r_j/s_i)$ (Equation (8)) and $\Sigma_j p(r_j/s_i) = 1$, then $I(S, R) = \Sigma_i \Sigma_j p(s_i)p(r_j/s_i)\ log\, 1/p(s_i) - \Sigma_i \Sigma_j p(s_i, r_j)\ log\, 1/p(s_i/r_j) = \Sigma_i p(s_i)\ log\, 1/p(s_i)\Sigma_j p(r_j/s_i) - \Sigma_i \Sigma_j p(s_i, r_j)\ log\, 1/p(s_i/r_j) = \Sigma_i p(s_i)\ log\, 1/p(s_i) - \Sigma_i \Sigma_j p(s_i, r_j)\ log\, 1/p(s_i/r_j)$. This shows that I(S, R) has the right form expressed by (18).

[8] According to Dretske (1981, 66), the informational content of a state is being expressed in the form '*S* is *F*', where the letter *S* is understood to be an indexical or demonstrative element referring to some item at the source. As Barwise and Seligman (1997) note, this way of talking about information may seem ill-suited for the description of the information carried by *events*. Nevertheless, these authors point out that it is possible to remain within Dretske's scheme in the case of events by talking of the information carried by *e*'s being *E*, where e is an event token and *E* is a type of event.

[9] The syntactic approach has, nevertheless, its own advantages since, by turning information into a syntactic concept, it makes Shannon's theory applicable to very different fields such as statistical physics, computer sciences and statistical inference; communication becomes only one of these many applications.

[10] Dretske is not the only author who uses a semantic concept of information. For example, Barwise and Seligman (1997) also adopt a semantic approach under the assumption that there is a close connection between information and knowledge, and that epistemology should be based on a theory of information. However, whereas Dretske uses a probabilistic notion to account for information (following Shannon's theory), Barwise and Seligman propose a logical perspective based on the notion of 'local logic': the local logic of a system is a model of the regularities that support information flow within the system, as well as the exceptions to these regularities.

REFERENCES

Abramson, N.: 1963, *Information Theory and Coding*, McGraw-Hill, New York.

Barwise, J. and J. Seligman: 1997, *Information Flow. The Logic of Distributed Systems*, Cambridge University Press, Cambridge.

Bell, D. A.: 1957, *Information Theory and its Engineering Applications*, Pitman & Sons, London.

Bennett, C. H. and R. Landauer: 1985, 'The Fundamental Physical Limits of Computation', *Scientific American* **July**, 48–56.

Berkovitz, J.: 1998, 'Aspects of Quantum No-Locality I: Superluminal Signalling, Action-at-a-Distance, Non-Separability and Holism', *Studies in History and Philosophy of Modern Physics* **29**, 183–222. 'Aspects of Quantum No-Locality II: Superluminal Causation and Relativity', *Studies in History and Philosophy of Modern Physics* **29**, 509–545.

Cover, T. and J. A. Thomas: 1991, *Elements of Information Theory*, John Wiley & Sons, New York.

Dretske, F.: 1981, *Knowledge & the Flow of Information*, M.I.T. Press, Cambridge Mass. (Reprinted: Dretske, F.: 1999, *Knowledge & the Flow of Information*, CSLI, Stanford).

Einstein, A., B. Podolsky and N. Rosen: 1935, 'Can Quantum-Mechanical Description of Physical Reality be Considered Complete?', *Physical Review* **47**, 777–780.

Floridi, L.: 2002, 'What is the Philosophy of Information?', *Metaphilosophy* **33**, 123–145.

Hughes, R. I. G.: 1989, *The Structure and Interpretation of Quantum Mechanics*, Harvard University Press, Cambridge MA.

Khinchin, A. I.: 1957, *Mathematical Foundations of Information Theory*, Dover Publications, New York.

Landauer, R.: 1991, 'Information is Physical', *Physics Today* **44**, 23–29.

Landauer, R.: 1996, 'The Physical Nature of Information', *Physics Letters A* **217**, 188–193.

Reza, F. M.: 1961, *Introduction to Information Theory*, Mc.Graw-Hill, New York.

Shannon, C.: 1948, 'The Mathematical Theory of Communication', *Bell System Technical Journal* **27**, 379–423 (Reprinted: Shannon, C. and W. Weaver: 1998, *The Mathematical Theory of Communication*, with a foreword by R. E. Blahut and B. Hajek, University of Illinois Press, Urbana and Chicago).

Stonier, T.: 1990, *Information and the Internal Structure of the Universe*, Springer-Verlag, London.

CONICET-Universidad Nacional de Quilmes
Instituto de Estudios sobre la Ciencia y la Technología
Rivadavia 2328, 6° Derecha, 1034
Ciudad de Buenos Aires
Argentina
E-mail: olimpiafilo@arnet.com.ar