Towards a usable notion of information[1]

Fred Truyen, June 2002

Summary: In this article, we will discuss some views on the notion of information, starting from Claude Shannon's communication theory to theories of today that where largely inspired by his influential views. In particular, the theory of Fred Dretske deserves much attention. Dretske sees information as a commodity that can be shared and transmitted through communication. On top of this theory of information, he wants to offer a theory of knowledge understood as "information-caused belief". We will show that Dretske's theory, although fundamentally sound, faces some problems, due to a strong naturalistic stance. These problems are addressed in the theory of John Perry and David Israel, which is based on situation theory. The view of Perry and Israel gives a clearer formal grasp of the constraints involved in identifying information. In their theory, more attention is paid to the context in which information comes to be. This inspires us, in a further move, to the formulation of a more layered model of information, where the different aspects of this contextual embedding are discerned. This more constructivist approach does more justice to the fact that information is always information *to* a processing *agent*, without leaving the basic Shannon/Dretske framework. We conclude with a pragmatic stance, that understands information from the user's point of view.

Shannon's communication theory

In his legendary article "A Mathematical Theory of Communication", Shannon gave birth tot the mathematical discipline now known as "information theory", by offering a sound quantitative approach to information. In fact, Shannon's goal was certainly not to promote the philosophical understanding of the nature of information; rather, as it shows from the title, his aim was to provide a theory of communication. Communication is, as he sees it, the transmission of messages from sender to receiver through a channel. This channel, the medium, can be a carrier wave, a radio wave or a laser beam, anything that can transport a signal.

[1] This is the original author version of the article published in: "In de voetsporen van Jacob van Maerlant: liber amicorum Raf De Keyser: verzameling opstellen over middeleeuwse geschiedenis en geschiedenisdidactiek. Raoul Bauer e.a. (eds.) - Leuven: Universitaire pers, 2002. - p. 596-612.

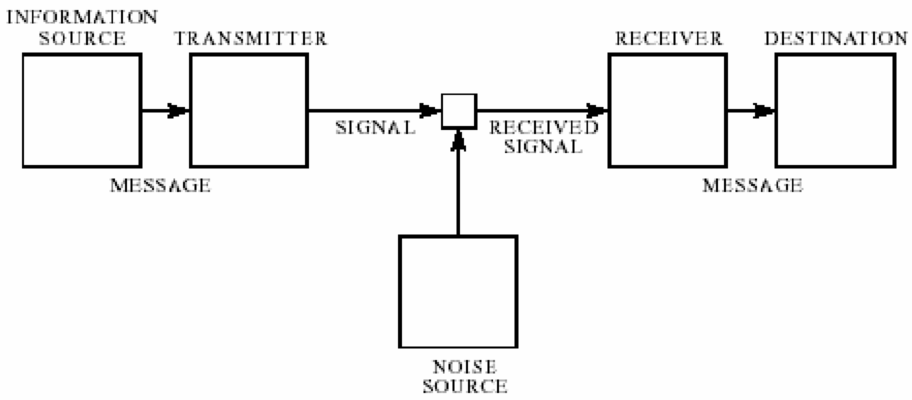


Fig. 1—Schematic diagram of a general communication system.

Shannon gives a very clear account of what he means by the different elements involved, and I think it is justified to quote it:

1. "An ***information source*** which produces a message or sequence of messages to be communicated to the receiving terminal. […]

2. A ***transmitter*** which operates on the message in some way to produce a signal suitable for transmission over the channel. In telephony this operation consists merely of changing sound pressure into a proportional electrical current. […]

3. The ***channel*** is merely the medium used to transmit the signal from transmitter to receiver. It may be a pair of wires, a coaxial cable, a band of radio frequencies, a beam of light, etc.

4. The ***receiver*** ordinarily performs the inverse operation of that done by the transmitter, reconstructing the message from the signal.

5. The ***destination*** is the person (or thing) for whom the message is intended."[2]

The main theory behind this model of communication is best explained through the "20 Questions game", a child's play we all remember. The game consists in trying to guess what someone is thinking about, by asking questions only to be answered by "yes" or "no". The starting situation is one of absolute uncertainty about the object thought of. Through a series of binary choices, one will find out more and more about this object, and discover that it will be easier and easier to predict the outcome of the questions, until finally the right object is singled out.

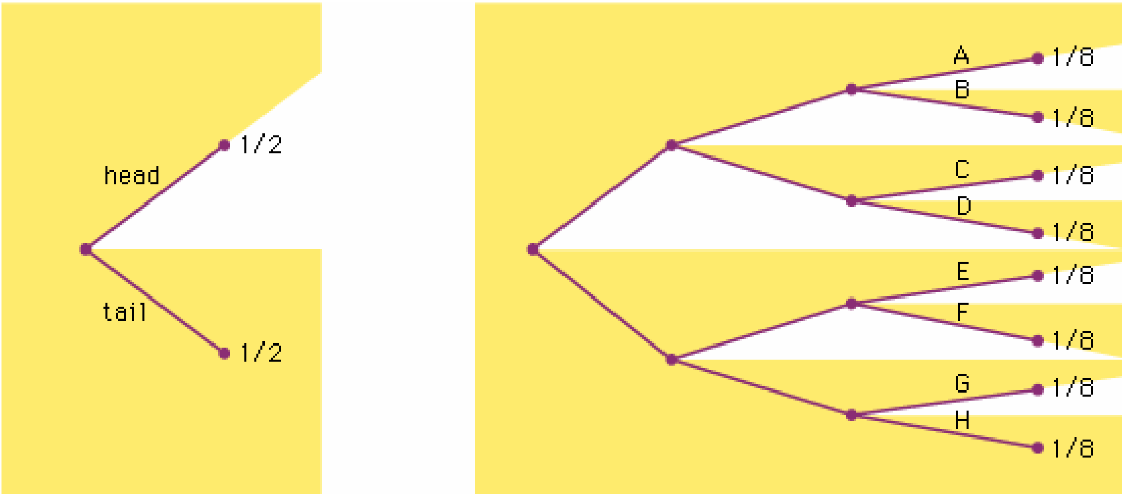
Already Plato understood that the statistically most sound method to perform this game is by trying to divide the realm of unknown objects in 2 quantitatively comparable classes at each question, the so–called "dichotomy" method. It is possible to play the game on-line (http://www.20Q.net) to experience the amazing effectiveness of this basic algorithm. So, when one is in front of a large group of people and has to guess one of them, the first question will probably be like "Is it a boy or a girl?" rather than "Does the person have a green earring?". Anyone understanding the logic of decision-

[2] Shannon 1948: 2

trees will see the exponential nature of this kind of method: with 20 questions, it is easy to single out one out of thousand possibilities.

In this sense, Shannon views information as a commodity that reduces uncertainty, and informing oneself amounts to reducing possibilities. In this way, information can yield knowledge, in the sense of predictability, certainty.

The diagram below shows the eligibility of one out of eight possibilities by three successive choices.



(Left) A binary choice, as in the flip of a coin, in which head and tail each has probability 1/2. (Right) Three successive binary choices, leading to eight possible outcomes, A through H, each with probability 1/8 (see text).

(1) If there are N possibilities, the **measure of information** equals Log_2 N; i.c. Log_2 8 = 3.

This means a sequence of 3 choices predicts any of the eight possibilities.

Transmitting signals through a channel necessarily involves *encoding* the information for that channel. Information will be encoded differently when sent through an optical beam or through a radio wave. This, of course, is the physical encoding. But Shannon rightly introduces encoding as something more abstract: in fact, only “answering by yes or no” in the 20 questions game is already “encoding” the information, in a binary format. Where it is relatively straightforward to see that you can encode non-binary encodings (e.g., decimal encodings) into a sequence of binary ones, Shannon points out that there are a lot of options available in the way one is encoding to binary format.

Compare the following two encoding methods:

Encoding 1 of M using S	
M	S
A	00
B	01
C	10
D	11

Encoding 2 of M using S		
Frequency	M	S
50%	A	0
25%	B	10
12,5%	C	110
12,5%	D	111

Comparison			
Character	Cases	Length ENC 1	Length ENC 2
A	60	120	60
B	30	60	60
C	15	30	45
D	15	30	45
Totals	120	240	210

The difference between the two methods could be proposed as a "standard characterization" of the difference between the 19th Century mathematician, essentially a philosopher and logician, and the 20th Century *engineer*. In the second method, we introduce some *factual* information about languages, *i.e.* the fact that in a language the surprisal value[3] of tokens isn't equally distributed. Even more so if there is no uncertainty then there little point in transmitting from sender to receiver. "If a source can produce only one particular message its entropy is zero, and no channel is required. For example, a computing machine set up to calculate the successive digits of π produces a definite sequence with no chance element. No channel is required to "transmit""[4]

By taking into account these stochastic facts about a language, we can optimize the communication by using the smallest possible token length for the most frequent sign.

If we adapt our formula to take into account the possible differences in *probability* of different signs to calculate the optimal average minimum token length, we get:

(2) Average amount of information[5]: $I(s) = \Sigma p(S_i) \bullet I(S_i)$

The *surprisal* value is the probability that a certain sign occurs within the communication.

Everything that is transmitted and that is not necessary to pass the information is *noise*.[6] It would be preferable to state this in another way: everything that is transmit-

[3] A token with a higher probablity has a lower "surprisal value": its occurence contains *less* information then the occurrence of a less probable token.

[4] Shannon 1948: 18

[5] Dretske 1999: 10

ted and that is not necessary to *obtain* the information is *noise*. Since Shannon's work is at the very origin of the Theory, he tries in first instance to draw a theory on an idealized, *noiseless communication*, to be able to formulate the base maths as clear as possible. Before entering into the details about noise, let us first take a look at the basic insight in Shannon's model.

Entropy

Entropy is a core concept in Shannon's theory: the ***entropy*** is the average minimum encoding length needed to transmit the messages in a way that they can be unequivocally decoded. Maybe a little mesmerizing, this measure of information is in fact also the measure of uncertainty left: the greater the entropy, the greater the uncertainty in the system and hence, the more information *can* be transmitted through the language. Indeed, the larger the optimal minimum message length, the more information you can transmit, or, as Warren Weaver puts it: "the amount of freedom one has to compose messages"[7].

This is, in the literature on Shannon, seldom well understood: a measure of uncertainty is a measure of the amount of information, whereas information is to be thought of as what yields certainty. This is the core of the "communication" view on information: information comes about in a *transmission.* It is our opinion that Fred Dretske doesn't fully pay tribute to the importance of this property. In our view, there is only information when there is an *exchange process*, a communication process going on. To put it straight: when data gets transmitted from sender to receiver, it becomes information *to* the recipient. As they say in didactics: the whole art is at the receivers' end. Data is being emitted constantly; we are actually surrounded by an enormous amount of signal emitters. Is all this information? Not at all: for these data to *become* information, an active selection process at the receivers' end has to take place. So, information would be a commodity only to be possessed by *agents that want to be informed.*

Noise and equivocation

The concepts of noise and equivocation in Shannon's theory are crucial to later debate in general and to Dretske's position in particular, as we shall see below.

As Weaver puts it, the whole issue with noise is, that it adds to the uncertainty degree of the message. Paradoxically, it doesn't make the information richer, by this means. In fact, "Uncertainty which arises by virtue of freedom of choice on the part of the sender is desirable uncertainty. Uncertainty which arises because of errors or because of the influence of noise is undesirable uncertainty."[8]

---

[6] Dretske 1999: 20

[7] Shannon & Weaver 1949: 13

[8] Shannon & Weaver 1949: 19

Shannon defines this as:

(3) $H(x) - H_y(x) = Hy - H_z(y)$

Where H(x) is the entropy of the source of messages, H(y) the entropy of the received signals, $H_y(x)$ is called the equivocation, or the uncertainty in the source as the signal is known, and $H_z(y)$ is the uncertainty in the received message if the sent signals be known, i.e. the uncertainty caused by noise.

$Hy - H_z(y)$ is the useful received information.

Fred Dretske's naturalist theory of information

In "Knowledge and the flow of information", Fred Dretske sets out a novel theory of knowledge based upon Shannon's theory of communication. In the first chapters, he offers a very clear account of Shannon's theory, as an introduction to Shannon only outclassed by the foreword by Warren Weaver of the 1949 edition of the text.

Dretske restates the technical point about equivocation in defining the *relation* between the information on the sender's side I(s) and the information on the receiver's side I(r). In fact, in his view, a reduction of possibilities takes place when composing a message. This reduction can be more or less independent from the reduction at the receiver's side. Dretske claims the decision mechanism at the sender's side, S, can alter the probability that a particular token will be transmitted. Suppose that you have to choose between 8 names, but you decide to add a procedure, that if the name "Shirley" is chosen, you instead would send "Herman", one of the other 7 names. So the name "Herman" will appear twice as likely than any other name. It is Dretske's view that in this case there is a difference between I(r) and I(s)[9]. The relation between I(s) and I(r) can be named $I_s(r)$.

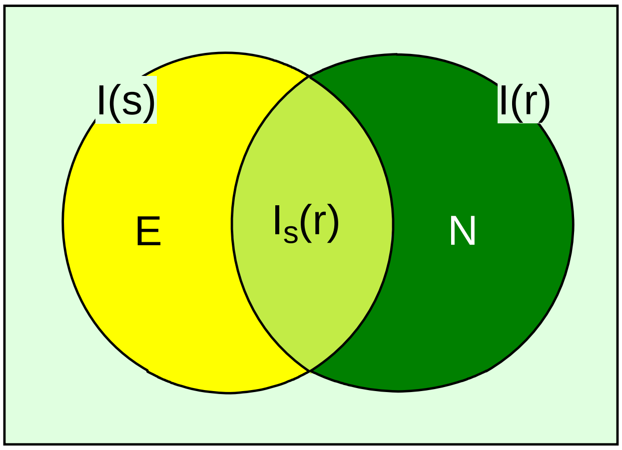
There are two ways of defining $I_s(r)$:

(4) $I_s(r) = I(r) - \text{noise}$

(5) $I_s(r) = I(s) - \text{equivocation}$

In (4) we define the information transmitted from s to r as the information at r without the noise; in (5) we define this information as the information at s without the part that is *not* transmitted to r.

[9] Dretske 1999: 14



The bright side of this theory is that you can see how equivocation and noise are the same at the receiver's side. The negative side is, that it is not clear if the processes at the sender's side are relevant to the communication. Where it is obvious that the probability of the sign at the receiver's end is relevant, this seems far less the case for the "objective" probability at the sender's side. In fact, one could claim that the only thing the example shows is that there is no language with 8 names involved, but one with 7, where the surprisal value of "Herman" is *lower* than that of the other names.

This creates a problem for Dretske's view that in perfect communication, all the information in r is transmitted to s. I think we have reasons to believe you cannot define "all the information at r". We will explain this point when discussing the theory of Perry and Israel.

But first, let us take a look at a terminological matter. It might appear that Dretske's example of the different *information* on the sender's and receiver's side, is a little stretching what Shannon's endavour was all about. Isn't data communication only concerned with the transmission of signals to transfer data? It seems totally irrelevant what kind of decisions are made at the sender's side. The only thing that seems to matter is that that what is sent is also received.

Let's carefully look into this remark. It is very important to make a clear distinction between signals, data and information. In electronics, "a signal is an electric current or electromagnetic field used to convey data from one place to another"[10]. Data is "information that has been translated into a form that is more convenient to move or process"[11] In telecommunications, data is often understood as digitally encoded information. By *transmitting* signals, you *transfer* data to *convey* information. And there's the hitch: Dretske overlooks the fact that in complex systems, and there are a lot of them around, the majority of signals transmissions is only involved in transferring data, without the direct intention to convey information. Under these conditions, you may choose the signals that fits best to transmit, the only thing that counts is that you get the same data at the receiver's end as at the sender's side.

---

[10] http://searchnetworking.techtarget.com/sDefinition/0,,sid7_gci212986,00.html

[11] http://searchstorage.techtarget.com/sDefinition/0,,sid5_gci211894,00.html

In networking, both the overwhelmingly used TCP/IP 4-layered network protocol stack as the more idealised OSI 7-layered Reference model make clear distinctions between lower-level datalink and transport layers and the higher level application layers (level 4 in TCP/IP; levels 5-7 in OSI)[12]. It seems as though the information where Dretske is alluding at in his example is something that is to be situated at these higher levels of interchange, the communication between people *using* these kind of telecommunication networks. Once an information has been encoded in a digital format for an electronic carrier, these *data* can be stored, shared, copied, transmitted in many different ways, where they will be often reprocessed and re-encoded (e.g. compressed) or even encoded for another transport (e.g. to switch from an ethernet network to a fiber-optic network). These are all mindless, automated processes taking place in full compliance with Shannon's theory, and the original "information" is never changed in between.

By asking the question whether the decision process at the sender's side is relevant, Dretske seems to take an "intentional stance", which is rather unexpected considering his naturalism.

But of course, this doesn't change the main point that what is encoded in the first place is the information, and that information is the alpha and gamma of the whole activity. No one would claim that the World-Wide-Web is simply about moving data around; its whole essence is to be an information system. To speak only about signals, or as we put it, data, seems to overlook the fact that there are no such things as mere data, that in fact, in the information chain, processed information becomes the data for another information process. This is very apparent in *sensor fusion* where data from a radar and an electro-optical passive receiver are processed to become useful information, and then in a second move reprocessed to build an integrated representation.

There is even more to it, according to Dretske, and that leads us to the core of his interesting vision: where one can easily describe signal transmission as a causal chain of events, there is not necessarily a *causal* link between the information sent and the information received. Instead, there is an *informational* link between signal sent and signal received.

Dretske warns against misinterpreting this theory as being about the transfer of *meaning* across the communication channel.[13] There is a distinction between what a signal means and the information it carries. Information is for Dretske something that can yield knowledge. Depending on what you know already, you could get other information out of a signal, more or less, but sometimes only remotely related to its meaning. If we accept this view, Dretske holds that communication theory can offer us two interesting laws about information understood that way:

(6) $I(s_a) = \log 1/p(s_a)$

the amount of information generated by a particular state of affairs s_a, and

---

[12] See e.g. http://www.rfc-editor.org/rfc/rfc1122.txt and http://www.iso.ch/iso/en/CatalogueListPage.CatalogueList?ICS1=35&ICS2=100&ICS3=01

[13] Dretske 1999: 42 f.

(7) $I_s(r_a) = I(s_a) - E(r_a)$

the amount of information carried by a particular signal r_a about a state of affairs s_a. Dretske tries to defend these measures as useful, albeit "just to make comparisons".[14]

In Dretske's view, Shannon's theory offers a *communication* condition for any valid theory of information. It amounts to the following definition[15].

CC: If a signal carries the information that *s* is *F*, then it must be the case that:

a. The signal carries as much information about *s* as would be generated by *s*'s being *F*

b. *s* is *F*.

For Dretkse, more is needed however to come to a genuine theory of information. The quantitative approach in Shannon's theory doesn't provide a total account of what we understand under the notion "information". To develop such an account, it is necessary to work out a *semantic* theory of information. It is easy to see where the problem lies with the quantitive approach in CC. If you have a red square *s*, it could be the case that the amount of information needed to transmit that *s* is red is equal to the amount needed to transmit that *s* is square. If you where transmitting that *s* is red, then under CC this would be sufficient to having transmitted that *s* is square, since both conditions are met[16]. For a *semantic* theory, one needs a further condition, assuring that the *right* information is transmitted:

c. The quantity of information the signal carries about *s* is (or includes) that quantity generated by *s*'s being *F* (and not, say, by *s*'s being *G*)

These three conditions combined offer an account of informational content, according to Dretske:

> *Informational content*: A signal *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* is what the receiver already knows.

It is our view that the awkward way Dretske formulates this definition of informational content betrays a deeper problem: the fact that Dretske, driven by an all to naturalistic approach, wants to see information as an objective commodity *transmitted* by the signals. The definition of informational content however shows that one cannot make abstraction of the way the data are *processed* by the receiver to *obtain* informa-

[14] Dretske 1999: 54

[15] Dretske 1999: 63-64

[16] Dretske 1999: 64

tion. Depending on his foreknowledge, *different* information comes about. Information is not given as such, but comes to be in a communication where *agents* are involved that have information needs and process the data accordingly. It is not necessary to try to embed all these “action constraints” into a static definition of information. We will show in our discussion of the theory of Perry and Israel below, that it is also impossible to do so.

Information and situation theory

In this sense, the embedding of an information theory in a situation theory has interesting advantages. We find a formulation of this theory in the article “What is Information?” by John Perry and David Israel[17]. Perry and Israel formulate the following principles.

A. Facts carry information.

B. The informational content of a fact is a true proposition.

C. The information a fact carries is relative to a constraint.

D. The information a fact carries is not an intrinsic property of it.

E. The informational content of a fact can concern remote things and situations.

F. Informational content can be specific; the propositions that are informational contents can be about objects that are not part of the indicating fact.

G. Indicating facts contain such information only relative to connecting facts; the information is incremental, given those facts.

H. Many different facts, involving variations in objects, properties, relations and spatiotemporal locations, can indicate one and the same informational content—relative to the same or different constraints.

I. Information can be stored and transmitted in a variety of forms.

J. Having information is good; creatures whose behavior is guided or controlled by information (by their information carrying states) are more likely to succeed than those which are not so guided.

In many ways, these 10 principles help to promote Dretske’s theory to a genuine *information* theory, into a layer superseding the datacommunication layer where Shannon laid the foundations. Already in the first statement, a strong departure is made from the naturalist point of view that *objects* carry information.

Information is rather something carried by *facts*. Of course, intelligent action is only possible because situations involve one another in a way governed by regularities or

---

[17] Perry 1990: 3

constraints. It is important to see that the information that a fact carries depends on other facts; it is therefore not an intrinsic property of it.

Perry and Israel define information as:

> Let *C* be some constraint. The fact σ carries the pure information that P relative to C iff
>
> 1. $C = \langle\langle$ Involves, *T*, *T'*; 1 $\rangle\rangle$
> 2. For any anchor f such that $\sigma = cond(T)[f]$, P = the proposition that $\exists a_1, \ldots a_n(con(T')[f])$.

Grossly stated, a particular arrangement in the data of the x-ray *indicates* something to someone. It carries the fact that some dog's leg is broken. This is the *informational content*. But, depending on other facts, e.g. that the x-ray is known to be taken from the dog Jackie, it carries also the information that *Jackie's* leg is broken. This information is called *incremental information*, which is context-dependent. Since these contexts change, one can't get a grasp on the totality of incremental information involved. Perry and Israel discuss an example involving an x-ray of a partcular dog Jackie's broken leg. More specifically, lets take a look at the statement "The fact that the x-ray has such-and-such a pattern indicates that Jackie has a broken leg".

> "As we noted above, Jackie is not a part of the x-ray, and it does not seem that her having a broken leg could be constantly conjoined with x-rays exhibiting the pattern that the vet recognizes. So how can the informational content of the x-ray have her as a constituent?
>
> We shall call the sort of information reported, e.g., in (3) *incremental* information. The conception is most easily grasped if we think of what the x-ray indicates *to* the vet. If she does not know which dog the x-ray is of, it simply indicates that some dog has been x-rayed that has a broken leg. We call this the *pure* information. But if she knows that Jackie was x-rayed, then the pattern on the x-ray indicates to her the additional or incremental information that Jackie has a broken leg. This is the incremental information carried by the x-ray, given the fact that the x-ray is of Jackie."[18]

We think this is a strong indication against the position that somehow all of "the information" can be *transmitted* in a communication. It is rather the case that one can *obtain* the information *that P* through a transmission, if one *has* the appropriate incremental information. Perry and Israel do not give sufficient tools to distinguish in any context what they call "pure" information from "incremental" information. It seems as though the notion of "pure" information is rather a very *abstract, idealized* notion of information. Information gathering is a highly dynamic process, and it is difficult to conceive of living beings that are not continuously exchanging information. Drawing the line between "pure" and "incremental" seems a daunting task!

Anyway, the theory of Perry and Israel, by acknowledging the notion of incremental information that contributes to the information without being transmitted, clearly make an advance over Dretske's original theory from our point of view. On the other hand, they even more explicitly limit the theory of information to the very cognitive

---

[18] Perry 1990: 4

stance of viewing the informational content as truly propositional. We have already objected against this view, by mentioning the fact that a lot of information hosted by the internet consists of music, images and the like. But then, it could be argued that even an image or a soundtrack is essentially a collection of facts. Lacking specific arguments, though, I personally don't feel comfortable with this view. I feel a lot for an "informational theory of knowledge", but would prefer a less epistemic account of information.

A layered information model

Dretske is right to claim that Shannon's communication theory is not a genuine theory of information, since, as was already stipulated in Warren Weaver's preface to the text in 1949, it doesn't offer a view on the *meaning* of information. But when he states that it does usefully give us a measure for the amount of information transmitted, there remain some doubts. Still, the challenging view on knowledge as "information caused belief" deserves much attention.

There is somewhat of a paradox in Dretske's endeavour to build a theory of knowledge and meaning, obviously higher-level concepts, out of Shannon's basic theory of communication. When one looks at current developments in the study field of information theory[19] (and it is this discipline that truly holds Shannon's legacy, rather than communication science), one can find, besides studies in cryptography, data compression, error-handling, and probability, also the subdomain called "information retrieval". This noble science has offered us the insights behind search-engines such as Google or AltaVista. It starts with a fundamental assumption justifying its existence as a separate research field, *i.e.* the *difference* between "data retrieval" and "information retrieval".

Data retrieval	Information retrieval
structured data	larger scale inconsistently structured data
query determined by data structure	query determined by information need
exact match	relevance
success iff: all and only matching records are retrieved	success iff: the most relevant information is selected

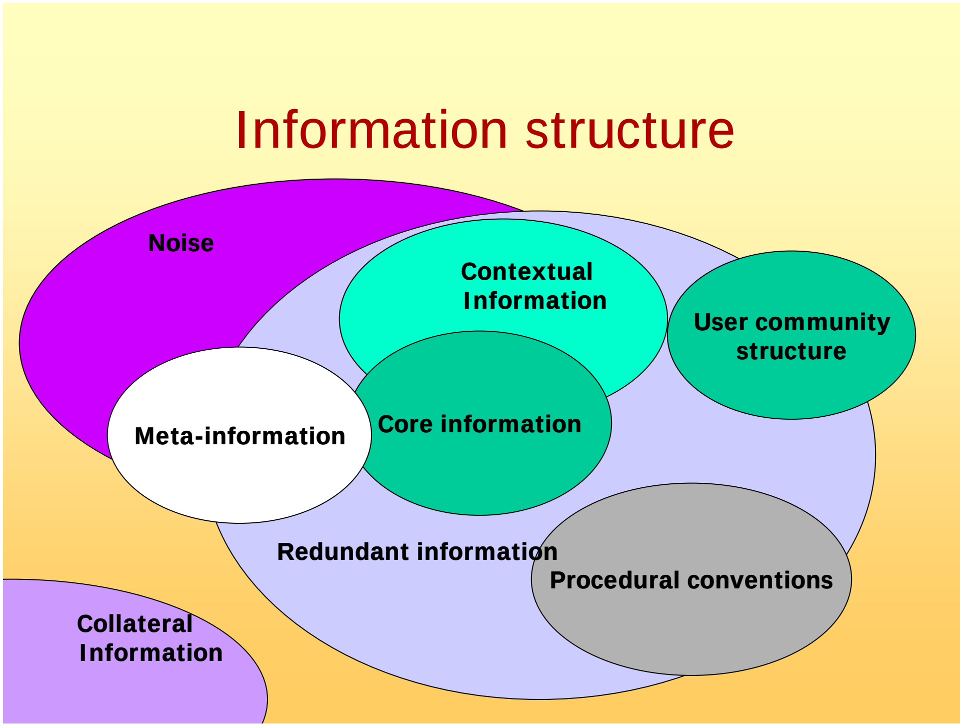
If procedures of data retrieval are always at the bottom line in information systems, at a higher, application-oriented level, we will use principles of information retrieval. The reason is simply the sheer amount of data to be processed in an average real-life

[19] see: http://www.itsoc.org/

information context. In fact, just as in low-level perception, where billions of photons hit the retina but fail to be processed because it isn't needed to perceive what is to be noticed, at a higher, more cognitive level, we need to select small chunks of new information out of terabytes of already known, or otherwise, useless information.

A good understanding of information interchange needs to appreciate aspects of scale, processing power, levels of aggregation and efficiency constraints. Without being some teleologist, describing information processes without reference to their *use* by actual information processing *systems* and the level on which they operate, is in my view nonsense. Information is always the result of a process.

To break with a tradition of an all too much speculative approach to the concept of information and in an attempt to stay more true to the spirit of Shannon's work as an *engineer*, let us take some hindsight, and take an example out of current, nowadays practice, in a time where there actually exist people whose business cards call them "knowledge managers", "information brokers" and even "truth engineers". Anyway, we can consider "information" at a higher level, not in the data link layer described by Shannon but in the application layer, where "information" is actually "used". For the sake of simplicity, I will use the word "information" in the way Dretske does, as a commodity which *resides* in information "carriers": it is at *this* level that this makes most sense.



When designing hypermedia systems at the Maerlant Centre, we use a "home-grown" information model to analyze the content of the projects[20]. When translating an historical file to a computer interface, we have to separate content items in different categories according to their use in the interface. Usually, we get the content in the

[20] It is true, however, that the center holds on to a social constructivist doctrine for didactics.

form of a written essay, structured as an article in a scientific journal. To adapt this kind of document to a hypertext sequence of computer screens is no mean feat. Various levels of adaptation impose themselves. It can best be described as a reformatting of the document, or, better perhaps, as adapting the document to another "format".

First of all, we migrate cross-referencing elements in the text to navigational elements in the interface. You seldom read things like "On the other hand", "however", "After having discussed these preliminary matters, let's now focus our attention on ..." in a computer interface. These "binding paragraphs", the "glue" of an exposé, are translated into buttons, arrows, scrollbars and the like.

The ***core information*** itself, then, reduced to its bare bones, is often restyled into a text where verbs tend to be replaced by concepts, names, and attributes. "Rockox lived in Antwerp ..." becomes "Life" "Antwerp" "Intellectual Center" etc. This, of course, enhances our *grasp* on the content. Besides that, the hypertext environment offers the possibility of *layering* the information, i.e., moving contextual information out of focus. In this way, the novel information on the screen can be put in better evidence. It is quite awkward, indeed, to have something like this: "Nicolaas Rockox met with a good deal of humanist intellectual leadership in his days. Humanism is a ... ". Most, but not all, readers of a CD-ROM on Nicolaas Rockox actually *know* this kind of trivialities about Humanism. So, this kind of ***foreknowledge or redundant information*** is left out, and a hyperlink is put under the word "humanist". But the same holds for the opposite, ***contextual information*** proper to the core nucleus (often indexical information), and that can be offered through another hyperlink: under "his days" we could link to a chronology carefully delimiting the context we are constrained to. Of course, this contextual information changes according to time and place: a hyperlink pointing to the actual President of The United States points to someone else now than it did 2 years ago. This kind of continuous re-embedding of core information is something undergraduate students have difficulties to cope with. They even seldom wonder that the "Duke of Flanders" in 1302 wasn't really the Duke of what now is called Flanders; something quite essential as the later world harbor of Antwerp was simply missing is his inventory! This constant recontextualisation of core information is a very interesting point in today's dynamic information systems, and force us to radically different views on knowledge and judgment.

But the analysis doesn't stop there. It becomes more interesting, if one tries to detach from the information lump those parts that indicate the ***user community*** that uses the information. Information for scholars can be differently presented than information for undergraduate students. Books often contain that kind of information, on their covers, in the series title, in the foreword by the editor. A genuine hypertext system will try to acquire knowledge on its user community, and the structure of it. Intelligent, dynamic systems even use this information to offer a differential reading. A research database, e.g., could indicate to a presumed novice that 80% of the scholars using the tool choose for option three at a certain point. More and more, stochastic methods are used that enable interfaces to interact in an appropriate way with the user. Intelligent training software adjusts the average question level to the performance of the trainee. If we pay attention to it, we discover how much user community information is hidden in many kinds of documents and information sources.

Next, a good document should offer some indication as to *how* it should be used. This kind of ***procedural information*** is embedded in a lot of publications. For example, Classicists use a wealth of solid reference works, like dictionaries, prosopographies, citation indices etc., all with very precise directives on how they should be used. In a computer interface, we try to separate this information neatly from the "pure content", to enhance usability, and to be able to adapt the procedures to the different user groups with their possibly different levels of skill.

Maybe best known and understood at the moment, is the difference between all the core information and possible ***meta-information*** of the document: who wrote it? For what purpose? How does it relate to other documents (is it, e.g., number three in a series dedicated to the subject)? This kind of information can be added all along the road, and doesn't have to be provided at the source. In our knowledge organization, it was the job of librarians, being supplemented more and more by the new-generation "cybrarians", that will classify, order, present and store information.

At this moment, it is a good time to look back at the diagram; you will have noticed that it isn't a tree, with hierarchical branches, nor a Venn-diagram, with intersecting spheres, and not a flow-chart either. Rather, such diagrams are often called "concept maps". Of course: in the documents *we* receive to work with, all these kinds of information are intertwined, making it sometimes very difficult to produce a correct analysis. In the interfaces we produce, there will be a ***network*** of hyperlinks that expresses these various kinds of information and their relations.

A special case is what could be called "***collateral information***". A lot of information travels along the road, but isn't included in the network. This can be for many reasons. One, different causal chains lead to different, independent information networks which somehow relate to common information sources. Another, it could be no one ever noticed that the information could be interrelated. New ***usage contexts*** could suddenly tie together formerly independent information chains, like DNA research and Archaeology, to name one example. Of course, from this "definition" follows that collateral information doesn't fit in the model. But on the other hand, collateral information can at any time become ***incremental information.*** This illustrates the advantage of a "highly educated" information processing agent, who has a lot of knowledge about other information chains. In teaching contexts, since the whole idea of a constructivist approach is to "hook in" new information into the student's network of foreknowledge, you will often be happy to take whatever possible link you can find. The flexibility a teacher has in "slotting" suddenly surprising information chunks into the information framework he is supposed to explain is an important skill, which is poorly documented in constructivist literature. That's why a solid hypertext network of information content should offer ***gateways*** *to* or entry points *for* collateral information.

Noise, last but not least, is always present in real life information systems. In the case of a computer interface for a multimedia application, it comes in unused corners of the screen, in superfluous functionality, in unused code. More interesting, a specific kind of noise is generated when the interface requires more *actions* than *decisions*. If you want to look up an explanation for a word in the interface, and it takes you 4 steps to go there, this will be a less optimal situation then when you achieve this in fewer steps. Elegant design requires the reduction of noise.

A more constructivist approach

Anyhow, very patent from this intuitive taxonomy of "kinds" of information involved in the practice of multimedia engineering, is the hidden assumption that information is a commodity for users, i.e. ***agents***. I see no compelling reason to limit this to ***human agents***. For the very straightforward reason that information is also processed and *used* by non-human agents, that can be animals, or even plants, that have to grow when light signals are given, or web servers, robots, computers whatsoever.

It is important to note, in our example, that what is new information to the *undergraduate student*, is *redundant* to the scholar. The scholar has a different internal information network, which makes him "hook in" the data in a different way. Reading the same text, he will be able to inform himself in a different way. In this sense, another communication occurs, from the same information source to a different system at the receiver's side. Although it seems that information *resides* in the book or CD-ROM, in fact it only becomes information again and again in each access, each *use* of it, each *query* made to it. This is not necessarily a "skeptic" constructivist position, claiming there is no objectivity involved. On the contrary, *information processing systems, and their capabilities* are very "objective" contexts. The only claim is that it doesn't make much sense to speak of "information" *outside* this reality of interacting agents.

A pragmatic stance: usage contexts

It is this usage context that defines *what* is becoming *information* at a certain stage in the process. On March 30-31, 1990, Belgian F16 pilots did a sighting of what is called a "UFO"[21]. Suppose one wants to take such an event seriously. In that case, it could be an idea to look at radar data of civilian and military air traffic control centers, to see if in those *data* it isn't possible to find *information* on the event. In opposition to a fighter radar, that tries to identify any flying object meeting certain possible threat criteria (dimension, speed, direction), the civilian systems focus on monitoring civilian air traffic. Civil airplanes do have a transponder, which helps air traffic control to locate and identify the aircraft. Other signals are mostly suppressed by the sensor fusion systems that will generate a representation of the airspace for the traffic controller.

What we are looking for, then, is if in the *noise* of the air traffic communication we can find "abnormal regularities" that can yield information related to the UFO sighting. Suppose we come to the "information **that** a flight pattern is discernible", and we call this Q.

Is the *information* a hidden treasure *in* the radar data? A commodity, waiting to be discovered but essentially there? Is Q in those data? This, I think, is naturalism beyond the call of duty. I think that there is a finite, enumerable amount of data present at any level of analysis (molecular, atomic, electronic, ...). But, it might be the case that you could obtain greater amounts of information *out of it*.

"Swimming in the pool" is indeed, as I have tried to exemplify by showing how informational products are actually designed, the way we learn to handle information. I

[21] see e.g. http://www.qtm.net/~geibdan/belgium/belsum.html

cannot sufficiently stress how wrong the image of *transferring* a portion of information from a sender to a happy receiver is. We are nowhere "empty barrels" waiting to be filled with information. Instead, *information-capable agents* are constantly exchanging data with their environment. These data are processed, prioritized, and integrated into their internal information network. This internal network enables capable interaction and enhances survivabilility. When I read a book, I am not taking the information out of the book and pouring it in an internal vase (the kind of "perfect communication" mentioned by Dretske). On the contrary, my internal representation space is already largely filled. What I need to do, for reasons of economy, is *adjust* and *rearrange* the internal network to more adequately cope with possible new situations. The more mature this network is, the more communication-capable, the faster and more efficient will be an interchange with another human, computer system or book.

Any sound information theory should start from the basic fact that there is already too much possible information coming your way. Not perfect transmission, as it is conceived of in data retrieval theories, but cost-efficient handling, selection and response are the criteria for information interchange.

Conclusion

In this paper, we started with a discussion of Shannon's powerful theory of communication. It provides a healthy base for an understanding of the notion of information. We learn that information is a commodity that can be quantitatively measured. On this foundation, Fred Dretkse is capable of proposing a novel theory of knowledge as "information caused belief", an epistemological theory fit for the information age.

On the other hand, we must be aware of a very strong naturalistic stance in Dretkse's view. Fully understanding the notion of information requires at the least an apprehension of the role of processing by a user.

It is in this field that the situation theory of Perry and Israel offers some clear advantages, as we have shown. Depending on incremental information, that is not transmitted in a communication, different information can be obtained through that communication. This is more the rule than the exception.

A genuine theory of information, therefore, should consider information both as the input and the result of processing by intelligent agents that continuously try to adjust their beliefs to changing situations. Of these agents, it is reasonable to say that they *have* information. This information, as Perry and Israel indicate, is not only a commodity, it is a *good*.

Bibliography

BAEZA-YATES, R. — RIBEIRO-NETO, B., *Modern Information Retrieval*, Addison-Wesley, New York, 1999.

DENNETT, D., *Kinds of Minds. Towards an understanding of* Consciousness, Phoenix, London, 1996.

DRETSKE, F., *Knowledge and the flow of Information*, CSLI Publications, Stanford, 1999 (1981)

DRETSKE, F., *Perception, Knowledge and Belief. Selected Essays*, Cambridge University Press, Cambridge, 2000

ISRAEL, D. — PERRY, J., "What is Information?", *Information, Language and Cognition*, edited by Philip Hanson, Vancouver: University of British Columbia Press, 1990: 1-19.

ISRAEL, D. — PERRY, J., "Information and Architecture", in: *Situation Theory and Its Applications, vol. 2* edited by Jon Barwise, Jean Mark Gawron, Gordon Plotkin and Syun Tutiya, 147–60. Stanford University: Center for the Study of Language and Information, 1991.

PIERCE, J.R., *An Introduction to Information Theory*. *Symbols, Signals and Noise*, Dover Publications, New York, 1980 (1961).

SHANNON, C., "A Mathematical Theory of Communication", in: *Bell System Technical Journal,* vol. 27, pp. 379-423 and 623-656, July and October, 1948.

SHANNON, C. — WEAVER, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1963 (1949).

SPARCK JONES, K. — WILLETT, P. (eds.), *Readings in Information Retrieval*, Morgan Kaufmann, San Fransisco, 1997.