Section II

Chapter 5

Information

Luciano Floridi

1. Introduction

Information "can be said in many ways", like being (Aristotle, *Metaphysics* Γ.2), and the correlation is probably not accidental. Information, with its cognate concepts like computation, data, communication etc., plays a key role in the ways we have come to understand, model and transform reality. Quite naturally, information has adapted to some of being's ridges.

Because information is a multifaceted and polyvalent concept, the question "what is information?" is misleadingly simple, exactly like "what is being?". As an instance of the Socratic question "ti esti...?", it poses a fundamental and complex problem, intrinsically fascinating and no less challenging than "what is truth?", "what is virtue?" "what is knowledge?" or "what is meaning?". It is not a request for dictionary explorations but an ideal point of intersection of philosophical investigations, whose answers can diverge both because of the conclusions reached and because of the approaches adopted. Approaches to a Socratic question can usually be divided into three broad groups: reductionist, antireductionist and non-reductionist. Theories of information are no exception.

Reductionists support the feasibility of a "unified theory of information" (UTI, see the UTI web site for references), general enough to capture all major kinds information (from Shannon's to Baudrillard's, from genetic to neural), but also sufficiently specific to discriminate between conceptual nuances. They attempt to show that all kinds of information are ultimately reducible conceptually, genetically or genealogically to some *Ur*-concept, mother of all instances. The development of a systematic UTI is a matter of time, patience and intelligent reconstruction. The ultimate UTI will be hierarchical, linear (even if probably branching), inclusive and incompatible with any alternative model.

Reductionist strategies are unlikely to succeed. Several surveys have shown no consensus or even convergence on a single, unified definition of information (see for example Braman 1989, Losee 1997, Machlup 1983, NATO 1974, 1975, 1983, Schrader 1984, Wellisch 1972, Wersig and Neveling 1975). This is hardly surprising. Information is such a powerful and flexible concept and such a complex phenomenon that, as an explicandum, it can be associated with several explanations, depending on the level of abstraction adopted and the cluster of requirements and desiderata orientating a theory. Claude Shannon (1993, 180), for one, was very cautious:

> The word "information" has been given different meanings by various writers in the general field of information theory. It is likely that at least a number of these will prove sufficiently useful in certain applications to deserve further study and permanent recognition. *It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field*.

At the opposite end, antireductionists stress the multifarious nature of the concept of information and of the corresponding phenomena. They defend the radical irreducibility of the different species to a single stem, objecting especially to reductionist attempts to identify Shannon's quantitative concept of information as the required *Ur*-concept and to ground a UTI on the mathematical theory of communication. Antireductionist strategies are essentially negative and can soon become an impasse rather than a solution. They allow specialised analyses of the various concepts of information to develop independently, thus avoiding the vague generalisations and mistaken confusions that may burden UTI strategies. But their fragmented nominalism remains unsatisfactory insofar as it fails to account for the ostensible connections permeating and influencing the various ways in which information *qua* information "can be said". *Connections*, mind, not Wittgensteinian *family resemblances*. The genealogical analogy would only muddy the waters here, giving the superficial impression of having finally solved the difficulty by merely hiding the actual divergences. The die-hard reductionist would still argue that all information concepts descend from the same *family*, whilst the unrepentant antireductionist would still object that we are facing mere *resemblances*, and that the various information concepts truly have different roots.

Non-reductionists seek to escape the dichotomy between reductionism and antireductionism by replacing the reductionist hierarchical model with a distributed network of connected concepts, linked by mutual and dynamic influences not necessarily genetic or genealogical. This "hypertextual analysis" can be centralised in various ways or completely decentralised and perhaps multi-centred.

According to decentralised or multi-centred approaches, there is no key concept of information. More than one concept is equally important, and the "periphery" plays a counterbalancing role. Depending on the orientation, information is seen as interpretation, power, narrative, message or medium, conversation, construction, a commodity and so on,. Thus, philosophers like Baudrillard, Foucault, Lyotard, McLuhan, Rorty and Derrida are united by what they dismiss, if not challenge: the predominance of the factual. For them information is not in, from or about reality. They downplay the aboutness of information and bend its referential thrust into a self-referential circle of hermeneutical communication. Their classic target is Cartesian foundationalism seen as the clearest expression of a hierarchical and authoritarian approach to the genesis, justification and flow of information. Disoriented, they mistake it as the only alternative to their fully decentralised view.

Centralised approaches interpret the various meanings, uses, applications and types of information as a system gravitating around a core notion with theoretical priority. The core notion works as a hermeneutical device that influences, interrelates and helps to access other notions. In metaphysics, Aristotle held a similar view about being, and argued in favour of the primacy of the concept of *substance*. In the philosophy of information, this "substantial" role has long been claimed by *factual* or *epistemically-oriented* semantic information. The basic idea is that, in order to understand what information is, the best thing to do is to start by analysing it in terms of the knowledge it can yield about its reference. The perspective is not without competitors. Weaver (1949), for example, supported a tripartite analysis of information in terms of (1) technical problems concerning the quantification of information and dealt with by Shannon's theory; (2) semantic problems relating to meaning and truth, and (3) what he called "influential" problems concerning the impact and effectiveness of information on human behaviour, which he thought had to play an equally important role. In pragmatic contexts, it is common to privilege a view of information as primarily a resource for decision making processes. One of the tasks of this chapter is to show how in each case the centrality of epistemically-oriented semantic information is presupposed rather than replaced.

We are now well placed to look at the structure of this chapter. In the following pages the question "what is information?" is approached from a non-reductionist and epistemically centralised perspective. In section two, the concept of semantic information is reviewed assuming that factual information is the most important and influential sense in which information *qua* information "can be said". No attempt is made to reduce all other concepts to factual information. Factual information is like the capital of the informational archipelagos, crucially positioned to provide a clear grasp of what information is, and a privileged gateway to other important concepts that are interconnected but not necessarily reducible to a single *Ur*-concept. To show this in practice and to enrich our understanding of what else information may be, we shall look at two neighbouring areas of great importance. Section three summaries the mathematical theory of communication, which studies the statistical behaviour of uninterpreted data, a much impoverished concept of information. Section four outlines some important philosophical programs of research that investigate a more enriched concept of semantic information. Space constraints prevents discussion of several other important concepts of information, but some of them are at least mentioned in the conclusion.

2. Semantic information

In this section, a general definition of semantic information is introduced, followed by a special definition of factually-oriented semantic information. The contents of the section are based on Floridi (2003 and forthcoming a). The approach is loosely connected with the methodology developed in situation logic (see section 3.2).

2.1. Semantic information as content

Information is often used in connection with communication phenomena to refer to objective (in the sense of mind-independent or external, and informee-independent) *semantic contents*. These can be of various size and value, formulated in a range of codes and formats, embedded in physical implementations of different kinds. They can variously be produced, processed, communicated and accessed. The *Cambridge Dictionary of Philosophy*, for example, defines information thus:

> an objective (mind independent) entity. It can be generated or carried by messages (words, sentences) or by other products of cognizers (interpreters) Information can be encoded and transmitted, but the information would exist independently of its encoding or transmission.

Examples of information in this broad sense are this *Guide*, E. A. Poe's *The Raven*, Verlaine's *Song of Autumn*, the Rosetta Stone and the movie *Fahrenheit 451*.

Over the last three decades, many analyses have converged on a General Definition of Information (GDI) as semantic content in terms of *data + meaning* (see Floridi forthcoming a for extended bibliography):

GDI) σ is an instance of information, understood as objective semantic content, if and only if:

GDI.1) σ consists of n *data* (d), for $n \geq 1$;

GDI.2) the data are *well-formed* (wfd);

GDI.3) the wfd are *meaningful* (mwfd = δ).

GDI has become an operational standard especially in fields that treat data and information as reified entities (consider, for example, the now common expressions "data mining" and "information management"). Examples are Information Science; Information Systems Theory, Methodology, Analysis and Design; Information (Systems) Management; Database Design; and Decision Theory. Recently, GDI has begun to influence the philosophy of computing and information (Floridi 1999 and Mingers 1997).

According to GDI, information can consist of different types of data δ. Data can be of four types (Floridi 1999):

δ.1) *primary data.* These are the principal data stored in a database, e.g. a simple array of numbers. They are the data an information-management system is generally designed to convey to the user in the first place.

δ.2) *metadata.* These are secondary indications about the nature of the primary data. They describe properties such as location, format, updating, availability, copyright restrictions, and so forth.

δ.3) *operational data.* These are data regarding usage of the data themselves, the operations of the whole data system and the system's performance.

δ.4) *derivative data.* These are data that can be extracted from δ.1-δ.3, whenever the latter are used as sources in search of patterns, clues or inferential evidence, e.g. for comparative and quantitative analyses (*ideometry*).

GDI indicates that information cannot be dataless, but it does not specify which types of δ constitute information. This *typological neutrality* (TyN) is justified by the fact that, when the apparent absence of data is not reducible to the occurrence of *negative* primary data, what becomes available and qualifies as information is some further non-primary information μ about σ constituted by some non-primary data δ.2-δ.4. For example, if a database query provides an answer, it will provide at least a *negative* answer, e.g. "no documents found". If the database provides no answer, either it fails to provide any data at all, in which case no specific information σ is available, or it can provide some data δ to establish, for example, that it is running in a loop. Likewise, silence, as a reply to a question, could represent negative information, e.g. as implicit assent or denial, or it could carry some non-primary information μ, e.g. the person has not heard the question.

Information cannot be dataless. In the simplest case, it can consist of a single datum (d). A datum is reducible to just a lack of uniformity between two signs. So our definition of a datum (Dd) is:

Dd) d = (x ≠ y)

where the x and the y are two uninterpreted variables.

The dependence of information on the occurrence of syntactically well-formed data, and of data on the occurrence of differences variously implementable physically, explain why information can be decoupled from its support. Interpretations of this support-independence vary radically because Dd leaves underdetermined not only the logical type to which the relata belong (see TyN), but also *the classification* of the relata (*taxonomic neutrality*), *the kind of support* required for the implementation of their inequality (*ontological neutrality*) and the dependence of their semantics on a producer (*genetic neutrality*).

Consider the *taxonomic neutrality* (TaN) first. A datum is usually classified as the entity exhibiting the anomaly, often because the latter is perceptually more conspicuous or less redundant than the background conditions. However, the relation of inequality is binary and symmetric. A white sheet of paper is not just the necessary background condition for the occurrence of a black dot as a datum, it is a constitutive part of the datum itself, together with the fundamental relation of inequality that couples it with the dot. Nothing is a datum *per se*. Being a datum is an external property. GDI endorses the following thesis:

TaN) a datum is a relational entity.

No data without relata, but GDI is neutral with respect to the identification of data with specific relata. In our example, GDI refrains from identifying either the black dot or the white sheet of paper as the datum.

Understood as relational entities, data are *constraining affordances*, exploitable by a system as input of adequate queries that correctly semanticise them to produce information as output. In short, information as content can also be described erotetically as *data + queries* (Floridi, 1999). I shall return to this definition in section 3.2.

Consider now the *ontological neutrality* (ON). By rejecting the possibility of dataless information, GDI endorses the following modest thesis:

ON) no information without data representation.

Following Landauer and Bennett 1985 and Landauer 1987, 1991 and 1996, ON is often interpreted materialistically, as advocating the impossibility of physically disembodied information, through the equation "representation = physical implementation":

ON.1) no information without physical implementation.

ON.1 is an inevitable assumption when working on the physics of computation, since computer science must necessarily take into account the physical properties and limits of the data carriers. Thus, the debate on ON.1 has flourished especially in the context of the philosophy of quantum computing (see Landauer 1991, Deutsch 1985, 1997; Di Vincenzo and Loss 1998; Steane 1998 provides a review). ON.1 is also the ontological assumption behind the Physical Symbol System Hypothesis in AI and Cognitive Science (Newell and Simon 1976). But ON, and hence GDI, does not specify whether, ultimately, the occurrence of every discrete state necessarily requires a *material* implementation of the data representations. Arguably, environments in which all entities, properties and processes are ultimately noetic (e.g. Berkeley, Spinoza), or in which the material or extended universe has a noetic or non-extended matrix as its ontological foundation (e.g. Pythagoras, Plato, Descartes, Leibniz, Fichte, Hegel), seem perfectly capable of upholding ON without necessarily embracing ON.1. The relata in Dd could be monads, for example. Indeed, the classic realism debate can be reconstructed in terms of the possible interpretations of ON.

All this explains why GDI is also consistent with two other popular slogans this time favourable to the proto-physical nature of information and hence completely antithetic to ON.1:

ON.2) "*It from bit*. Otherwise put, every "it" —every particle, every field of force, even the space-time continuum itself—derives its function, its meaning, its very existence entirely—even if in some contexts indirectly—from the apparatus-elicited answers to yes-or-no questions, binary choices, *bits*. "It from bit" symbolizes the idea that every item of the physical world has at bottom—a very deep bottom, in most instances—an immaterial source and explanation; that which we call reality arises in the last analysis from the posing of yes-no questions and the registering of equipment-evoked responses; in short, that all things physical are information-theoretic in origin and that this is a *participatory universe*." (Wheeler 1990, 5);

and

ON.3) "[information is] a name for the content of what is exchanged with the outer world as we adjust to it, and make our adjustment felt upon it." (Wiener 1954, 17). "Information is information, not matter or energy. No materialism which does not admit this can survive at the present day" (Wiener 1961, 132).

ON.2 endorses an information-theoretic, metaphysical monism: the universe's essential nature is digital, being fundamentally composed of information as data instead of matter or energy, with material objects as a complex secondary manifestation (a similar position has been defended more recently in physics by Frieden 1998, whose work is based on a Platonist perspective). ON.2 may but does not have to endorse a computational view of information processes. ON.3 advocates a more pluralistic approach along similar lines. Both are compatible with GDI.

A final comment concerning GDI.3 can be introduced by discussing a fourth slogan:

ON.4) "In fact, what we mean by information - the elementary unit of information - is a difference which makes a difference". (Bateson 1973, 428).

ON.4 is one of the earliest and most popular formulations of GDI (see for example Franklin 1995, 34 and Chalmers 1996, 281; note that the formulation in MacKay 1969, that is "information is a *distinction* that makes a difference", predates Bateson's and, although less memorable, is more accurate). A "difference" is just a discrete state (that is, a datum), and "making a difference" simply means that the datum is "meaningful", at least potentially.

Finally, let us considers the semantic nature of the data. How data can come to have an assigned meaning and function in a semiotic system in the first place is one of the hardest problems in semantics. Luckily, the point in question here is not *how* but *whether* data constituting information as semantic content can be meaningful *independently* of an informee. The *genetic neutrality* (GeN) supported by GDI states that:

GeN) δ can have a semantics *independently* of any informee.

Before the discovery of the Rosetta Stone, Egyptian hieroglyphics were already regarded as information, even if their semantics was beyond the comprehension of any interpreter. The discovery of an interface between Greek and Egyptian did not affect the semantics of the hieroglyphics but only its accessibility. This is the weak, conditional-counterfactual sense in which GDI.3 speaks of meaningful data being embedded in information-carriers informee-independently. GeN supports the possibility of *information without an informed subject*, to adapt a Popperian phrase. Meaning is not (at least not only) in the mind of the user. GeN is to be distinguished from the stronger, realist thesis, supported for example by Dretske (1981), according to which data could also have their own semantics independently of an intelligent *producer/informer*. This is also known as *environmental information*, and a typical example given are the concentric rings visible in the wood of a cut tree trunk, which may be used to estimate the age of the plant.

To summarise, GDI defines information broadly understood as semantic content comprised of syntactically well-formed and meaningful data. Its four types of neutrality (TyN, TaN, ON and GeN) represent an obvious advantage, as they make GDI perfectly scalable to more complex cases and reasonably flexible in terms of applicability and compatibility. The next question is whether GDI is satisfactory when discussing the most important type of semantic information, namely factual information.

2.2. Semantic information as factual information

We have seen that semantic information is usually associated with communication. Within this context, the most important type of semantic information is *factual information*, which tells to the informee something *about* something else, for example where a place is, what the time is, whether lunch is ready or that penguins are birds. Factual information has a declarative (Kant's judicial) nature, is satisfactorily interpretable in terms of first-order, classic predicate logic, is correctly qualifiable alethically and can be appropriately analysed in the following form "*a*'s being (of type) *F* carries the information that *b* is *G*" (Dretske 1981, Barwise and Seligman 1997).

Does GDI provide a definition of factual information? Some philosophers (Barwise and Seligman 1997, Dretske 1981, Floridi 2003 and forthcoming a, Grice 1989) have argued that it does not, because otherwise false information would have to count as a type of factual information, and there are no convincing reasons to believe it does, whilst there are compelling reasons to believe that it does not (for a detailed analysis see Floridi forthcoming a). As Dretske and Grice have put it: "[…] *false* information and *mis*-information are not kinds of information – any more than decoy ducks and rubber ducks are kinds of ducks" (Dretske 1981, 45) and "False information is not an inferior kind of information; it just is not information" (Grice 1989, 371). Let us see the problem in more detail.

The difficulty lies here with yet another important neutrality in GDI. GDI makes no comment on the truthfulness of data that may comprise information (*alethic neutrality* AN):

AN) meaningful and well-formed data qualify as information, no matter whether they represent or convey a truth or a falsehood or have no alethic value at all.

Verlaine's *Song of Autumn* counts as information even if it does not make sense to ask whether it is true or false, and so does every sentence in *Old Moore's Almanac*, no matter how downright false. Information as purely semantic content is completely decoupled from any alethic consideration (Colburn 2000 and Fox 1983 can be read as defending this perspective). However, if GDI is taken to define also factual information, then

a) false information about the world (including contradictions), i.e. *misinformation*, becomes a genuine type of factual information;

b) tautologies qualify as factual information;

c) "it is true that *p*" where *p* can be replaced by any instance of genuine factual information, is no longer a redundant expression, e.g. "it is true" in the conjunction "'the earth is round' qualifies as information *and* it is true" cannot be eliminated without semantic loss; and finally

d) it becomes impossible to erase factual information semantically (we shall be more and more informed about x, no matter what the truth value of our data about x is).

None of these consequences is ultimately defensible, and their rejection forces a revision of GDI. "False" in "false information" is used attributively, not predicatively As in the case of a false constable, false information is not factual information that is false, but not factual information at all. So "false information" is, like "false evidence", not an oxymoron, but a way of specifying that the informational contents in question do not conform to the situation they purport to map, and so fail to qualify as factual information. Well-formed and meaningful data may be of poor quality. Data that are incorrect (vitiated by errors or inconsistencies), imprecise (precision is a measure of the repeatability of the collected data) or inaccurate (accuracy refers to how close the average data value is to the actual value) are still data and may be recoverable. But, if the are not truthful, they can only amount to semantic content at best and misinformation at worst.

The special definition of information (SDI) needs to include a fourth condition about the positive alethic nature of the data in question:

SDI) σ is an instance of factual information if and only if:

SDI.1) σ consists of n *data* (d), for n ≥ 1;

SDI.2) the data are *well-formed* (wfd);

SDI.3) the wfd are *meaningful* (mwfd = δ);

SDI.4) the δ are *truthful*.

Factual information encapsulates truthfulness, which does not contingently supervene on, but is necessarily embedded in it. And since information is "said primarily in factual ways", to put it in Aristotelian terms, false information can be dismissed as no factual information at all, although it can still count as information in the sense of semantic content.

3. The mathematical theory of communication

Some features of information are intuitively quantitative. Information can be *encoded*, *stored* and *transmitted*. We also expect it to be *additive* and *non-negative*. Similar properties of information are investigated by the *mathematical theory of communication* (MTC) with the primary aim of devising efficient ways of encoding and transferring data.

MTC is not the only successful mathematical approach to information theory, but it certainly is the best and most widely known, and the one that has had the most profound impact on philosophical analyses. The name for this branch of probability theory comes from Shannon's seminal work (Shannon 1948, now Shannon and Weaver 1998). Shannon pioneered this field and obtained many of its principal results, but he acknowledged the importance of previous work done by other researches at Bell laboratories, most notably Nyquist and Hartley (see Cherry 1978 and Mabon 1975). After Shannon, MTC became known as *information theory*, an appealing but unfortunate label, which continues to cause endless misunderstandings.

Shannon came to regret its widespread popularity, and we shall avoid using it in this context.

This section outlines some of the key ideas behind MTC, with the aim of understanding the relation between MTC and the philosophy of information. The reader with no taste for mathematical formulae may wish to go directly to section 3.2, where some implications of MTC are discussed. The reader interested in knowing more can start by reading Weaver 1949 and Shannon 1993b, then Schneider 2000, Pierce 1980 and Jones 1979 and finally Cover and Thomas 1991.

3.1. The quantification of raw information

MTC has its origin in the field of electrical communication, as the study of communication limits. It develops a quantitative approach to information as a means to answer two fundamental problems: the ultimate level of data compression and the ultimate rate of data transmission. The two solutions are the entropy H in equation [9] and the channel capacity C. The rest of this section illustrates how to get from the problems to the solutions.

Imagine a very boring device that can produce only one symbol, like Poe's raven, who can answer only "nevermore". This is called a unary device. Even at this elementary level, Shannon's simple model of communication applies (see Fig. 1).

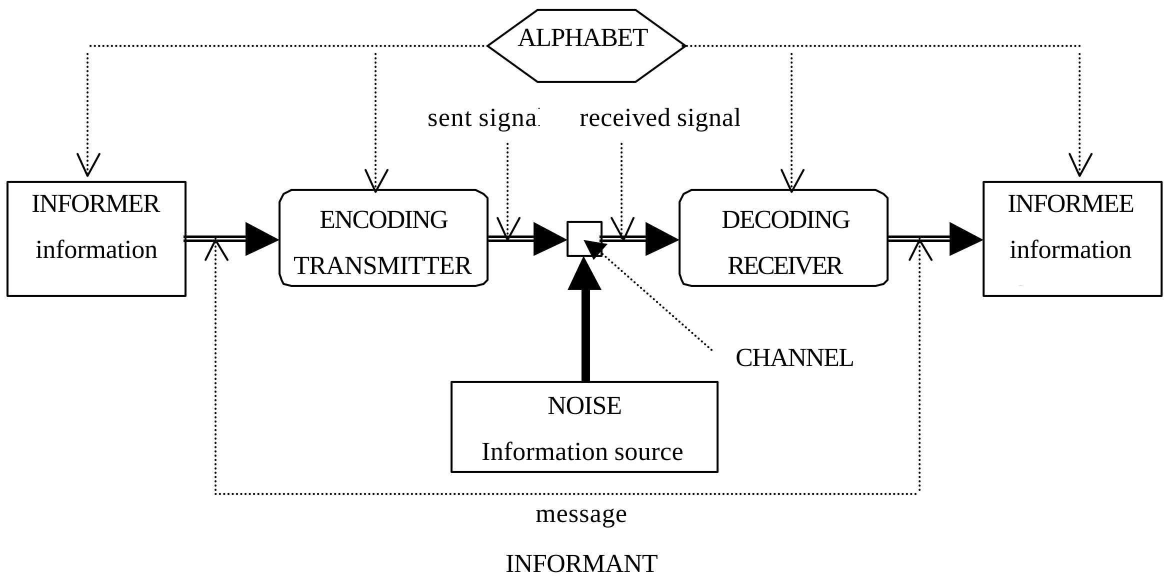


Fig. 1 Communication model (adapted from Shannon 1948, 1998)

The raven is the informer, we are the informee, “nevermore” is the message (the informant), there is a coding and decoding procedure through English, a channel of communication and some possible noise.

Informer and informee share the same background knowledge about the collection of usable symbols (the *alphabet*). Given this a priori knowledge, it is obvious that a unary device produces zero amount of information. Simplifying, we already know the outcome so our ignorance cannot be decreased. Whatever the informational state of the system, asking appropriate questions to the raven does not make any difference. Note that a unary source answers every question all the time with only one symbol, not with silence or symbol, since silence counts as a signal, as we saw in 2.1. A completely silent source also qualifies as a unary source.

Consider now a binary device that can produce two symbols, like a fair coin *A* with its two equiprobable symbols {*h*, *t*}; or, as Matthew 5:37 suggests, “Let your communication be Yea, yea; Nay, nay: for whatsoever is more than these cometh of evil”. Before the coin is tossed, the informee (for example a computer) is in a state of *data deficit* greater than zero: the informee does not “know” which symbol the device will actually produce. Shannon used the technical term “uncertainty” to refer to data deficit. In a non-mathematical context this is a misleading term because of its strongly semantic connotations. Recall that the informee con be a very simple machine, and psychological, mental, doxastic or epistemic states are clearly irrelevant. Once the coin has been tossed, the system produces an amount of raw information that is a function of the possible outputs, in this case 2 equiprobable symbols, and equal to the data deficit that it removes.

Let us build a slightly more complex system, made of two fair coins *A* and *B*. The *AB* system can produce 4 ordered outputs: <*h*, *h*>, <*h*, *t*>, <*t*, *h*>, <*t*, *t*>. It generates a data deficit of 4 units, each couple counting as a symbol in the source alphabet. In the *AB* system, the occurrence of each symbol removes a higher data deficit than the occurrence of a symbol in the *A* system. In other words, each symbol contains more raw information. Adding an extra coin would produce a 8 units of data deficit, further increasing the amount of information carried by each symbol in the *ABC* system, and so on.

We are ready to generalise the examples. Call the number of possible symbols *N*. For $N = 1$, the amount of information produced by a unary device is 0. For $N = 2$, by producing an equiprobable symbol, the device delivers 1 unit of information. And for N = 4, by producing an equiprobable symbol the device delivers the sum of the amount of information provided by coin A plus the amount of information provided by coin B, that is 2 units of information, although the total number of symbols is obtained by multiplying A's symbols by B's symbols. Our information measure should be a continuous and monotonic function of the probability of the symbols. The most efficient way of satisfying these requirements is by using the logarithm to the base 2 of the number of possible symbols (the logarithm to the base 2 of a number is the power to which 2 must be raised to give the number, for example $\log_2 8 = 3$, since $2^3 = 8$). Logarithms have the useful property of turning multiplication of symbols into addition of information units. By taking the logarithm to the base 2 (henceforth log simply means $\log_2$) we have the further advantage of expressing the units in bits. The base is partly a matter of convention, like using centimetres instead of inches, partly a matter of convenience, since it is useful when dealing with digital devices that use binary codes to represent data. Given an alphabet of N equiprobable symbols, we can rephrase some examples more precisely (Fig. 2) by using equation [1]:

$$\log_2(N) = \text{bits of information per symbol} \qquad [1]$$

Device	Alphabet	Bits of information per symbol
Poe's raven (unary)	1 symbol	log(1) = 0
1 coin (binary)	2 equiprobable symbols	log(2) = 1
2 coins	4 equiprobable symbols	log(4) = 2
1 die	6 equiprobable symbols	log(6) = 2.58
3 coins	8 equiprobable symbols	log(8) = 3

Fig. 2

The basic idea is all in equation [1]. Raw information can be quantified in terms of decrease in data deficit (uncertainty). Unfortunately, real coins are always biased. To calculate how much information they produce one needs to rely on the frequency of the occurrences of symbols in a finite series of tosses, or on their probabilities, if the tosses are supposed to go on indefinitely. Compared to a fair coin, a slightly biased coin must produce less than 1 bit of information, but still more than 0. The raven produced no information at all because the occurrence of a string S of "nevermore" was not *informative* (not *surprising*, to use a more intuitive, but psychologistic vocabulary), and that is because the *probability* of the occurrence of "nevermore" was maximum, so overly predictable. Likewise, the amount of raw information produced by the biased coin depends on the average *informativeness* (also known as average *surprisal*, another unfortunate term to refer to the average statistical rarity) of the string *S* of *h* and *t* produced by the coin. The average informativeness of the resulting string *S* depends on the *probability* of the occurrence of each symbol. The higher the frequency of a symbol in *S*, the less raw information is being produced by the coin, up to the point when the coin is so biased to produce always the same symbol and stops being informative, behaving like the raven. So, to calculate the average informativeness of *S* we need to know how to calculate *S* and the informativeness of a i^{th} symbol in general. This requires understanding what the probability of a i^{th} symbol (P_i) to occur is.

The probability P_i of the i^{th} symbol can be "extracted" from equation [1], where it is embedded in log(*N*), a special case in which the symbols are equiprobable. Using some elementary properties of the logarithmic function we have:

$$\log(N) = -\log(N^{-1}) = -\log(\frac{1}{N}) = -\log(P) \qquad [2]$$

The value of $1/N = P$ can range from 0 to 1. If the raven is our source, the probability of "good morning" is 0. In the case of the coin, $P(h) + P(t) = 1$, no matter how biased the coin is. Probability is like a cake that gets sliced more and more thinly depending on the number of guests, but never grows beyond its original size. More formally:

$$\sum_{i=1}^{N} P_i = 1 \qquad [3]$$

The sigma notation simply means that if we add all probabilities values from $i = 1$ to $i = N$ the sum is equal to 1.

We can now be precise about the raven: "nevermore" is not informative at all because $P_{nevermore} = 1$. Clearly, the lower the probability of occurrence of a symbol, the higher is the informativeness of its actual occurrence. The informativeness *u* of a i^{th} symbol can be expressed by analogy with $-\log(P)$ in equation [2]:

$$u_i = -\log(P_i) \qquad [4]$$

Next, we need to calculate the length of a general string *S*. Suppose that the biased coin, tossed 10 times, produces the string: <*h, h, t, h, h, t, t, h, h, t*>. The (length of the) string *S* (in our case equal to 10) is equal to the number of times the *h* type of symbol occurs added to the numbers of times the *t* type of symbol occurs. Generalising for *i* types of symbols:

$$S = \sum_{i=1}^{N} S_i \qquad [5]$$

Putting together equations [4] and [5] we see that the average informativeness for a string of *S* symbols is the sum of the informativeness of each symbol divided by the sum of all symbols:

$$\frac{\sum_{i=1}^{N} S_i u_i}{\sum_{i=1}^{N} S_i} \qquad [6]$$

Formula [6] can be simplified thus:

$$\sum_{i=1}^{N} \frac{S_i}{S} u_i \qquad [7]$$

Now S_i/S is the frequency with which the i^{th} symbol occurs in *S* when *S* is finite. If the length of *S* is left undetermined (as long as one wishes), then the frequency of the i^{th} symbol becomes its probability P_i. So, further generalising formula [7] we have:

$$\sum_{i=1}^{N} P_i \, u_i \qquad [8]$$

Finally, by using equation [4] we can substitute for u_i and obtain

$$H = -\sum_{i=1}^{N} P_i \log P_i \text{ (bits per symbol)} \qquad [9]$$

Equation [9] is Shannon's formula for *H* = uncertainty, what we have called *data deficit* (actually, Shannon's original formula includes a positive constant *K* which amounts to a choice of a unit of measure, bits in our case; Shannon used the letter *H* because of R.V.L. Hartley's previous work). Equation [9] indicates that the quantity of raw information produced by a device corresponds to the amount of data deficit erased. It is a function of the average informativeness of the (potentially unlimited) string of symbols produced by the device. It is easy to prove that, if symbols are equiprobable, [9] reduces to [1] and that the highest quantity of raw information is produced by a system whose symbols are equiprobable (compare the fair coin to the biased one).

To arrive at [9] we have used some very simple examples: a raven and a handful of coins. Things in life are far more complex. For example, we have assumed that the strings of symbols are *ergodic*: the probability distribution for the occurrences of each symbol is assumed to be stable through time and independently of the selection of a certain string. Our raven and coins are *discrete* and *zero-memory sources*. The successive symbols they produce are statistically independent. But in real life occurrences of symbols are often interdependent. Sources can be non-ergodic and have a memory. Symbols can be continuous, and the occurrence of one symbol may depend upon a finite number n of preceding symbols, in which case the string is known as a Markov chain and the source a n*-th* order Markov source. Consider for example the probability of being sent an "e" before or after having received the string "welcom". And consider the same example through time, in the case of a child learning how to spell English words. In brief, MTC develops the previous analysis to cover a whole variety of more complex cases. We shall stop here, however, because in the rest of this section we need to concentrate on other central aspects of MTC.

The quantitative approach just sketched plays a fundamental role in coding theory (hence in cryptography) and in data storage and transmission techniques. Recall that MTC is primarily a study of the properties of a channel of communication and of codes that can efficiently encipher data into recordable and transmittable signals. Since data can be distributed either in terms of here/there or now/then, diachronic communication and synchronic analysis of a memory can be based on the same principles and concepts (our coin becomes a bistable circuit or flip-flop, for example), two of which are so important to deserve a brief explanation: *redundancy* and *noise*.

Consider our *AB* system. Each symbol occurs with 0.25 probability. A simple way of encoding its symbols is to associate each of them with two digits:

<*h*, *h*> = 00

<*h*, *t*> = 01

<*t*, *h*> = 10

<*t*, *t*> = 11

Call this Code 1. In Code 1 a message conveys 2 bits of information, as expected. Do not confuse *bits* as *bi*-nary uni*ts* of information (recall that we decided to use $\log_2$ also as a matter of convenience) with *bits* as *bi*-nary digi*ts*, which is what a 2-symbols system like a CD-ROM uses to encode a message. Suppose now that the *AB* system is biased, and that the four symbols occur with the following probabilities:

<*h*, *h*> = 0.5

<*h*, *t*> = 0.25

<*t*, *h*> = 0.125

<*t*, *t*> = 0.125

This system produces less information, so by using Code 1 we would be wasting resources. A more efficient Code 2 should take into account the symbols' probabilities, with the following outcomes:

<*h*, *h*> = 0	0.5 × 1 binary digit = .5
<*h*, *t*> = 10	0.25 × 2 binary digits = .5
<*t*, *h*> = 110	0.125 × 3 binary digits = .375
<*t*, *t*> = 111	0.125 × 3 binary digits = .375

In Code 2, known as Fano Code, a message conveys 1.75 bits of information. One can prove that, given that probability distribution, no other coding system will do better than Fano Code. On the other hand, in real life a good codification is also modestly redundant. *Redundancy* refers to the difference between the physical representation of a message and the mathematical representation of the same message that uses no more bits than necessary. *Compression* procedures work by reducing data redundancy, but redundancy is not always a bad thing, for it can help to counteract *equivocation* (data sent but never received) and *noise* (received but unwanted data). A message + noise contains more data than the original message by itself, but the aim of a communication process is *fidelity*, the accurate transfer of the original message from sender to receiver, not data increase. We are more likely to reconstruct a message correctly at the end of the transmission if some degree of redundancy counterbalances the inevitable noise and equivocation introduced by the physical process of communication and the environment. Noise extends the informee's freedom of choice in selecting a message, but it is an undesirable freedom and some redundancy can help to limit it. That is why, in a crowded pub, you shout your orders twice and add some gestures.

We are now ready to understand Shannon's two fundamental theorems. Suppose the 2-coins biased system produces the following message: <*t*, *h*> <*h*, *h*> <*t*, *t*> <*h*, *t*><*h*, *t*>. Using Fano Code we obtain: 11001111010. The next step is to send this string through a channel. Channels have different transmission rates (*C*), calculated in terms of bits per second (bps). Shannon's fundamental theorem of the noiseless channel states that

> Let a source have entropy H (bits per symbol) and a channel have a capacity C (bits per second). Then it is possible to encode the output of the source in such a way as to transmit at the average rate of C/H – ε symbols per second over the channel where ε is arbitrarily small. It is not possible to transmit at an average rate greater than C/H. (Shannon 1998, 59).

In other words, if you devise a good code you can transmit symbols over a noiseless channel at an average rate as close to *C*/*H* as one may wish, but, no matter how clever the coding is, that average can never exceed *C*/*H*. We have already seen that the task is made more difficult by the inevitable presence of noise. However, the fundamental theorem for a discrete channel with noise comes to our rescue:

> Let a discrete channel have the capacity C and a discrete source the entropy per second H. If H ≤ C there exists a coding system such that the output of the source can be transmitted over the channel with an arbitrarily small frequency of errors (or an arbitrarily small equivocation). If H > C it is possible to encode the source so that the equivocation is less than H – C + ε where ε is arbitrarily small. There is no method of encoding which gives an equivocation less than H – C. (Shannon 1998, 71)

Roughly, if the channel can transmit as much or more information than the source can produce, then one can devise an efficient way to code and transmit messages with as small an error probability as desired These two fundamental theorems are among Shannon's greatest achievements. And with our message finally sent, we may close this section.

3.2. Some conceptual implications of MTC

For the mathematical theory of communication (MTC) information is only a selection of one symbol from a set of possible symbols, so a simple way of grasping how MTC quantifies raw information is by considering the number of yes/no questions required to guess what the source is communicating. One question is sufficient to guess the output of a fair coin, which therefore produces 1 bit of information. A 2-fair-coins system produces 4 ordered outputs: <*h*, *h*>, <*h*, *t*>, <*t*, *h*>, <*t*, *t*> and therefore requires two questions, each output containing 2 bits of information, and so on. This erotetic analysis clarifies two important points.

First, MTC is not a theory of information in the ordinary sense of the word. The expression "raw information" has been used to stress the fact that in MTC information has an entirely technical meaning. Consider some examples. Two equiprobable "yes" contain the same quantity of raw information, no matter whether their corresponding questions are "would you like some tea?" or "would you marry me?”. If we knew that a device could send us with equal probabilities either the movie *Fahrenheit 451* or this whole *Guide*, by receiving one or the other we would receive many bytes of data but only one bit of raw information. On June 1 1944, the BBC broadcasted a line from Verlaine’s *Song of Autumn*: “Les sanglots longs des violons de Autumne”. The message contained almost 1 bit of information, an increasingly likely “yes” to the question whether the D-Day invasion was imminent. The BBC then broadcasted the second line “Blessent mon coeur d'une longueur monotone”. Another almost meaningless string of letters, but almost another bit of information, since it was the other long-expected “yes” to the question whether the invasion was to take place immediately. German intelligence knew about the code, intercepted those messages and even notified Berlin, but the high command failed to alert the Seventh Army Corps stationed in Normandy. Hitler had all the information in Shannon’s sense of the word, but failed to understand the real meaning and importance of those two small bits of data. As for ourselves, we were not surprised to conclude that the maximum amount of raw information is produced by a text where each character is equally distributed, that is by a perfectly random sequence.

Second, since MTC is a theory of information without meaning, and information – meaning = data, *mathematical theory of data communication* is a far more appropriate description than *information theory*. In section 2.1 we saw that information as semantic content can also be described erotetically as *data + queries*. Imagine a piece of information such as “the earth has only one moon”. It is easy to polarise almost all its semantic content by transforming it into a query + binary answer: “does the earth have only one moon? + yes”. Subtract the “yes” and you are left with virtually all the semantic content, fully de-alethicised (the query is neither true nor false). The datum “yes” works as a key to unlock the information contained in the query. MTC studies the codification and transmission of raw information by treating it as data keys, as the amount of details in a signal or message or memory space necessary to unlock the informee’s knowledge. As Weaver (1949, 12) remarked “the word information relates not so much to what you do say, as to what you could say. MTC deals with the carriers of information, symbols and signals, not with information itself. That is, information is the measure of your freedom of choice when you select a message”.

Since MTC deals not with information itself but with the carriers of information, that is messages constituted by uninterpreted symbols encoded in well-formed strings of signals, it is commonly described as a study of information at the *syntactic* level. MTC can be successfully applied in ICT (information and communication technologies) because computers are syntactical devices. What remains to be clarified is how H in equation [9] should be interpreted.

Assuming the ideal case of a noiseless channel of communication, H is a measure of three equivalent quantities:

a) the average amount of raw information per symbol produced by the informer, or
b) the corresponding average amount of data deficit (Shannon's "uncertainty") that the informee has before the inspection of the output of the informer, or
c) the corresponding informational potentiality of the same source, that is, its *informational entropy*.

H can equally indicate (a) or (b) because, by selecting a particular alphabet, the informer automatically creates a data deficit (uncertainty) in the informee, which then can be satisfied (resolved) in various degrees by the *informant*. Recall the erotetic game. If you use a single fair coin, I immediately find myself in a 1 bit deficit predicament. Use two fair coins and my deficit doubles, but use the raven, and my deficit becomes null. My empty glass is an exact measure of your capacity to fill it. Of course, it makes sense to talk of raw information as quantified by H only if one can specify the probability distribution.

Regarding (c), MTC treats raw information like a physical quantity, such as mass or energy, and the closeness between equation [9] and the formulation of the concept of entropy in statistical mechanics was already discussed by Shannon. The informational and the thermodynamic concept of entropy are related through the concepts of probability and *randomness* ("randomness" is better than "disorder" since the former is a syntactical concept whereas the latter has a strongly semantic value), entropy being a measure of the amount of "mixedupness" in processes and systems bearing energy or information. Entropy can also be seen as an indicator of reversibility: if there is no change of entropy then the process is reversible. A highly structured, perfectly organised message contains a lower degree of entropy or randomness, less raw information and causes a smaller data deficit, consider the raven. The higher the potential randomness of the symbols in the alphabet, the more bits of information can be produced by the device. Entropy assumes its maximum value in the extreme case of uniform distribution. Which is to say that a glass of water with a cube of ice contains less entropy than the glass of water once the cube has melted, and a biased coin has less entropy than a fair coin. In thermodynamics, we know that the greater the entropy, the less available the energy. This means that high entropy corresponds to high energy deficit, but so does entropy in MTC: higher values of *H* correspond to higher quantities of data deficit.

4. Some philosophical approaches to semantic information

The mathematical theory of communication approaches information as a physical phenomenon. Its central question is whether and how much uninterpreted data can be encoded and transmitted efficiently by means of a given alphabet and through a given channel. MTC is not interested in the meaning, aboutness, relevance, usefulness or interpretation of information, but only in the level of detail and frequency in the uninterpreted data, being these symbols, signals or messages. On the other hand, philosophical approaches seek to give an account of information as semantic content, investigating questions like "how can something count as information? and why?", "how can something carry information about something else?", "how is information related to error, truth and knowledge?", "when is information useful?". Philosohers usually adopt a propositional orientation and an epistemic outlook, endorsing, often implicitly, the prevalence of the factual (they analyse examples like "The Bodleian library is in Oxford"). How relevant is MTC to similar analyses?

In the past, some research programs tried to elaborate information theories *alternative* to MTC, with the aim of incorporating the semantic dimension. Donald M. MacKay (1969) proposed a quantitative theory of qualitative information that has interesting connections with situation logic (see below), whereas Doede Nauta (1972) developed a semiotic-cybernetic approach. Nowadays, few philosophers follow these lines of research. The majority agrees that MTC provides a rigorous constraint to any further theorising on all the semantic and pragmatic aspects of information. The disagreement concerns the crucial issue of the *strength* of the constraint. At one extreme of the spectrum, a theory of semantic information is supposed to be *very strongly* constrained, perhaps even overdetermined, by MTC, somewhat like mechanical engineering is by Newtonian physics. Weaver's interpretation of Shannon's work is a typical example. At the other extreme, a theory is supposed to be *only weakly* constrained, perhaps even completely underdetermined, by MTC, somewhat like tennis is constrained by Newtonian physics, that is in the most uninteresting, inconsequential and hence disregardable sense (see for example Sloman

1978 and Thagard 1990). The emergence of MTC in the fifties generated earlier philosophical enthusiasm that has gradually cooled down through the decades. Historically, philosophical theories of semantic information have moved from "very strongly constrained" to "only weakly constrained", becoming increasingly autonomous from MTC (for a review, see Floridi forthcoming b).

Popper (1935) is often credited as the first philosopher to have advocated the inverse relation between the probability of *p* and the amount of semantic information carried by *p*. However, systematic attempts to develop a formal calculus were made only after Shannon's breakthrough. MTC defines information in terms of probability space distribution. Along similar lines, the *probabilistic approach* to semantic information defines the semantic information in *p* in terms of logical probability space and the inverse relation between information and the probability of *p*. This approach was initially suggested by Bar-Hillel and Carnap (Bar-Hillel and Carnap 1953, Bar-Hillel 1964) and further developed by Hintikka (especially Hintikka and Suppes 1970) and Dretske 1981 (on Dretske's approach see also chapters 17 and 18). The details are complex but the original idea is simple. The semantic content (CONT) in *p* is measured as the complement of the a priori probability of *p*:

$$\text{CONT}(p) = 1 - P(p) \qquad [10]$$

CONT does not satisfy the two requirements of additivity and conditionalization, which are satisfied by another measure, the informativeness (INF) of *p*, which is calculated, following equations [9] and [10], as the reciprocal of $P(p)$, expressed in bits, where $P(p) = 1 - \text{CONT}(p)$:

$$\text{INF}(p) = \log \frac{1}{1-\text{cont}} = -\log P(p) \qquad [11]$$

Things are complicated by the fact that the concept of probability employed in equations [10] and [11] is subject to different interpretations. In Bar-Hillel and Carnap the probability distribution is the outcome of a logical construction of atomic statements according to a chosen formal language. This introduces a problematic reliance on a strict correspondence between observational and formal language. In Dretske, the solution is to make probability values refer to states of affairs (*s*) of the world observed:

$$\text{I}(s) = -\log P(s) \qquad [12]$$

The *modal approach* modifies the probabilistic approach by defining semantic information in terms of modal space and in/consistency. The information conveyed by *p* becomes the set of all possible worlds or (more cautiously) the set of all the descriptions of the relevant possible states of the universe that are excluded by *p*. The *systemic approach*, developed especially in situation logic (Barwise and Perry 1983, Israel and Perry 1990, Devlin 1991; Barwise and Seligman 1997 provide a foundation for a general theory of information flow) also defines information in terms of states space and consistency. However, is less ontologically demanding than the modal approach, since it assumes a clearly limited domain of application, and it is compatible with Dretske's probabilistic approach, although it does not require a probability measure on sets of states. The informational content of *p* is not determined a priori, through a calculus of possible states allowed by a representational language, but in terms of factual content that *p* carries with respect to a given situation. Information tracks possible transitions in a system's states space under normal conditions. Both Dretske and situation theories require some presence of information already immanent in the environment (*environmental information*), as nomic regularities or constraints. This "semantic externalism" can be controversial both epistemologically and ontologically. Finally, the *inferential approach* defines information in terms of entailment space: information depends on valid inference relative to a person's theory or epistemic state.

Most approaches close to MTC assume the principle of *alethic neutrality*, and run into the difficulties outlined in 2.2 (Dretske and Barwise are important exceptions; Devlin rejects truthfulness as a necessary condition). As a result, the *semantic approach* (Floridi 2003 and forthcoming a) adopts SDI and defines factual information in terms of data space.

Suppose there will be exactly three guests for dinner tonight. This is our situation *w*. Imagine that you are told that

T) there may or may not be some guests for dinner tonight; or

V) there will be some guests tonight; or

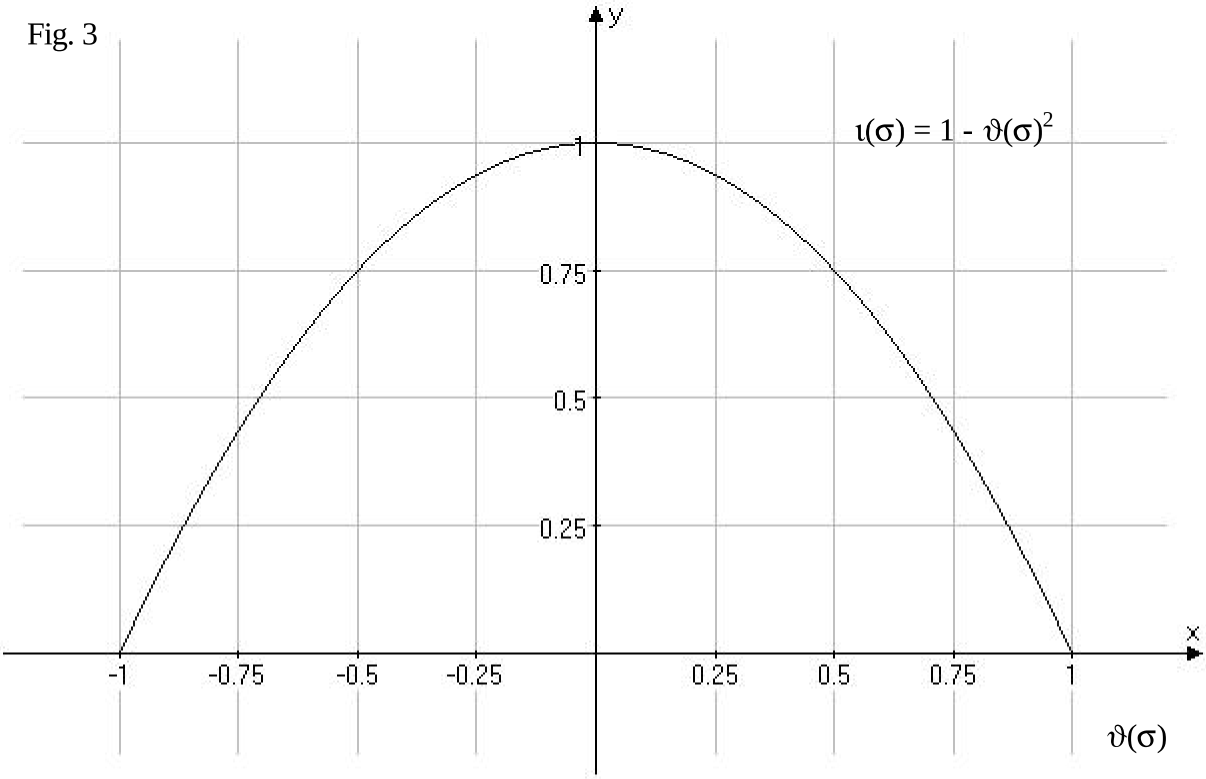
P) there will be three guests tonight.

The *degree of informativeness* of T is zero because, as a tautology, T applies both to *w* and to $\neg w$. V performs better, and P has the maximum degree of informativeness because, as a fully accurate, precise and contingent truth, it "zeros in" on its target *w*. Generalising, the more distant a true σ is from its target *w*, the larger is the number of situations to which it applies, the lower its degree of informativeness becomes. A tautology is a true σ that is most "distant" from the world. Let us use the letter ϑ to refer to the distance between a true σ and *w*. Using the more precise vocabulary of situation logic, ϑ indicates the degree of support offered by *w* to σ. We can now map on the x axis the values of ϑ given a specific σ and a corresponding target *w*. In our example, we know that ϑ(T) = 1 and ϑ(P) = 0. For the sake of simplicity, let us assume that ϑ(V) = 0.25 (see Floridi 2003 on how to calculates ϑ values). We now need a formula to calculate the *degree of informativeness* ι of σ in relation to ϑ(σ). It can be shown that the most elegant solution is provided by the complement of the square value of ϑ(σ), that is $y = 1 - x^2$. Using our symbols we have:

$$\iota(\sigma) = 1 - \vartheta(\sigma)^2 \quad [13]$$

Fig. 3 shows the graph generated by equation [13] when we include also negative values of distance for false σ (ϑ ranges from –1 = contradiction to 1 = tautology).

Fig. 3

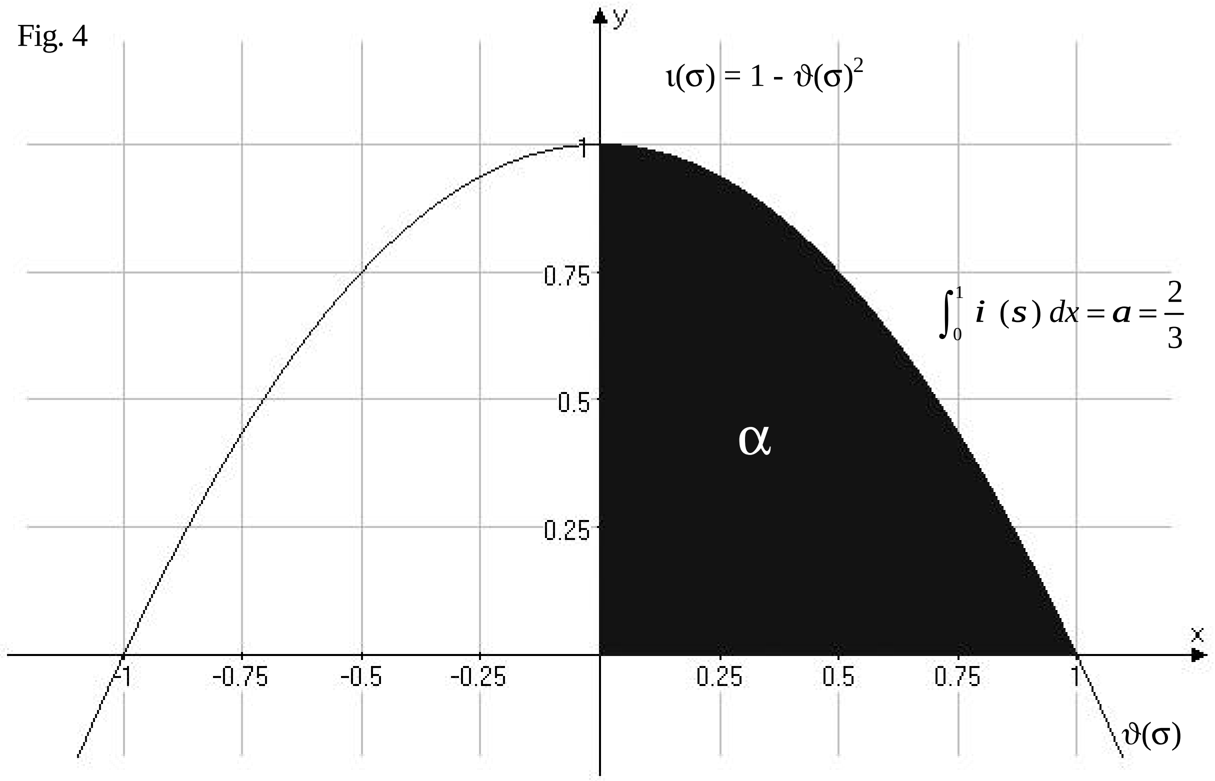


If σ has a very high degree of informativeness ι (very low ϑ) we want to be able to say that it contains a large quantity of semantic information and, vice versa, the lower the degree of informativeness of σ is, the smaller the quantity of semantic information conveyed by σ should be. To calculate the quantity of semantic information contained in σ relative to ι(σ) we need to calculate the area delimited by equation [13], that is the definite integral of the function ι(σ) on the interval [0, 1]. As we know, the maximum quantity of semantic information (call it α) is carried by P, whose ϑ = 0. This is equivalent to the whole area delimited by the curve. Generalising to σ we have:

$$\int_0^1 \iota(\sigma)\, dx = \alpha = \frac{2}{3} \qquad [14]$$

Fig. 4 shows the graph generated by equation [14]. The shaded area is the maximum amount of semantic information α carried by σ.

Fig. 4



An interesting property of equation [14] is that, if we express it in bits, we have

$$\log\frac{2}{3} = \log 2 - \log 3 = 1\,bit - 1\,trit = 1\,sbit \qquad [15]$$

A trit is one base-3 digit and represents the amount of information conveyed by a selection among one of three equiprobable outcomes. It is linearly equivalent to log 3. The term *sbit* (*semantic bit*) indicates our unit of maximum semantic information α, concerning a given situation *w*, that can be conveyed by σ with ϑ = 0.

Consider now V, "there will be some guests tonight". V can be analysed as a (reasonably finite) string of disjunctions, that is V = ["there will be one guest tonight" or "there will be two guests tonight" or … "there will be n guests tonight"], where n is the reasonable limit we wish to consider (things are more complex than this, but here we only need to grasp the general principle). Only one of the descriptions in V will be fully accurate. This means that V also contains some (perhaps much) information that is simply irrelevant or redundant. We shall refer to this "informational waste" in V as to the vacuous information in V. The amount of vacuous information (call it β) in V is also a function of the distance ϑ of V from *w*, or more generally

$$\int_0^{J} i\ (s)\, dx = b \qquad [16]$$

Since ϑ(V) = 0.25, we have

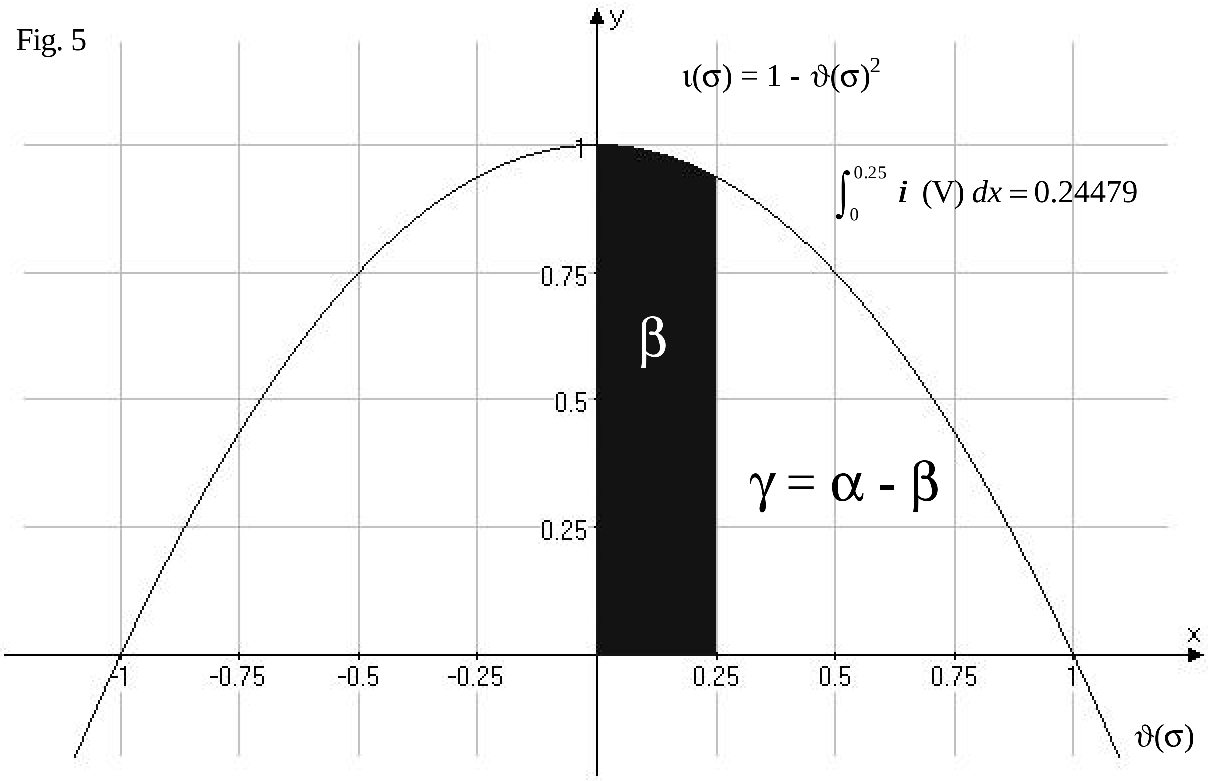
$$\int_0^{0.25} i\ (\mathrm{V})\, dx = 0.24479 \qquad [17]$$

Fig. 5 shows the graph generated by equation [17]. The shaded area is the amount of vacuous information β in V. Clearly, the amount of semantic information in V is simply the difference between α (the maximum amount of information that can be carried in principle by σ) and β (the amount of vacuous information actually carried by σ), that is the clear area in the graph of Fig. 5. More generally, and expressed in bits, the amount of semantic information γ in σ is:

$$\gamma(\sigma) = \log(\alpha - \beta) \qquad [18]$$

Note the similarity between [14] and [16]. When ϑ(σ) = 1, that is, when the distance between σ and *w* is maximum, then α = β and γ(σ) = 0. This is what happens when we consider T. T is so distant from *w* to contain only vacuous information. In other words, T contains as much vacuous information as P contains relevant information, namely 1 sbit.

Fig. 5



A final comment, before closing this section. Each of the previous extentionalist approaches can be given an intentionalist interpretation by considering the relevant space as a doxastic space, in which information is seen as a reduction in the degree of personal uncertainty given a state of knowledge of the informee.

5. Conclusion

In this chapter we have been able to visit only a few interesting places. The connoisseur might be disappointed and the supporter of some local interests appalled. To try to appease both and to whet the appetite of the beginner here is a list of some very important concepts of information that have not been discussed:

informational complexity (Kolmogorov and Chaitin, among others), a measure of the complexity of a string of data defined in terms of the length of the shortest binary program required to compute that string. Note that Shannon's H can be considered a special case of Kolmogorov complexity K, since $H \approx K$ if the sequence is drawn at random from a probability distribution with entropy = H;

instructional information (imagine a recipe, an algorithm or an order), a crucial concept in fields like computer science, genetics, biochemistry, neuroscience, cognitive science and AI (chapters 2 and 3);

pragmatic information, central in any theory addressing the question of how much information a certain informant carries for an informee in a given doxastic state and within a specific informational environment. This includes *useful information*, a key concept in economics, information management theory and decision theory, where characteristics such as relevance, timeliness, updatedness, usefulness, cost, significance and so forth are crucial (chapter 23);

valuable information in ethical contexts (see chapter 6 and Floridi, forthcoming d);

environmental information, that is the possible location and nature of information in the world (Dretske 1981 and chapters 12-14);

physical information and the relation between being and information (see Leff and Rex 1990 and chapters 12-14);

biological information (see chapter 16). The biologically minded reader will notice that the 4 symbols in the AB system we built in section 3.1 could be adenine, guanine, cytosine and thymine, the four bases whose order in the molecular chain of DNA or RNA codes genetic information.

The nature of these and other information concepts, the analysis of their interrelations and of their possible dependence on MTC, and the investigation of their usefulness and influence in the discussion of philosophical problems are some of the crucial issues that a philosophy of information needs to address. There is clearly plenty of very interesting and important work to do.

Acknowledgements

I am very grateful to Mark Bedau, John Collier, Phil Fraundorf, Ken Herold, James Fetzer and Kia Nobre for their very valuable comments on earlier drafts.

References

Bar-Hillel, Y. 1964, *Language and Information* (Reading, Mass.; London: Addison Wesley). (collection of influential essays on semantic information, graduate level) .

Bar-Hillel, Y. and Carnap, R. 1953, "An Outline of a Theory of Semantic Information", rep. in Bar-Hillel 1964, pp. 221-274 (one of the first and most influential attempts to develop a quantitative analysis of semantic information, graduate level).

Barwise, J. and J. Seligman 1997, *Information flow: the logic of distributed systems.* (Cambridge: Cambridge University Press) (innovative approach to information flow, graduate level, but the modular structure contains some very accessible chapters).

Barwise, J. and Perry, J. 1983, *Situations and Attitudes* (Cambridge, Ma.: MIT Press) (influential text in situation logic, graduate level).

Bateson, G. 1973, *Steps to an Ecology of Mind* (Paladin. Frogmore, St. Albans) (the beginning of the ecological approach to mind and information, undergraduate level).

Braman, S. 1989, "Defining Information". *Telecommunications Policy* 13, 233-242.

Chalmers, D. J. 1997, *The Conscious Mind: in Search of a Fundamental Theory* (Oxford: Oxford University Press) (accessible to undergraduates).

Cherry, C. 1978, *On Human Communication* 3rd ed. (Cambridge, Ma: MIT Press) (clear and accessible introduction to communication theory, old but still valuable).

Colburn, T. R. 2000, "Information, Thought, and Knowledge", *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics* (Orlando FL, July 23-26, 2000), vol. X, 467-471 (analyse the standard definition of knowledge as justified true belief from an informational perspective, very accessible).

Cover, T. and J. A. Thomas 1991, *Elements of information theory* (New York: Chichester, Wiley) (standard textbook in the field, requires a solid mathematical background, graduate level only, see Jones 1979 for a more accessible text, or Pierce 1980).

Deutsch, D. 1985, "Quantum theory, the Church-Turing Principle and the Universal Quantum Computer", *Proceedings of the Royal Society*, 400, 97-117 (information and computation in quantum computing, requires a solid mathematical background, graduate level only).

Deutsch, D. 1997, *The Fabric of Reality* (London: Penguin) (on the ontological implications of quantum physics, advanced undergraduate level).

Devlin, K. 1991, *Logic and Information* (Cambridge: Cambridge University Press) (reviews and improves on situation logic, undergraduate level).

Di Vincenzo, D. P. and Loss, D. 1998, "Quantum Information is Physical", *Superlattices and Microstructures* 23, 419-432, special issue on the occasion of Rolf Landauer's 70th birthday (also available at http://xxx.lanl.gov/abs/cond-mat/9710259) (reviews the debate on the physical aspects of information, graduate level).

Dretske, F. 1981, *Knowledge and the Flow of Information* (Cambridge, Ma: MIT Press, rep. Stanford: CSLI, 1999) (classic informational analysis of knowledge, advanced undergraduate level).

Floridi, L. (2003), "Outline of a Theory of Strongly Semantic Information", forthconming in *Minds and Machines*. Preprint available at http://www.wolfson.ox.ac.uk/~floridi/papers.htm (develops a truth-based approach to semantic information, graduate level)

Floridi, L. (forthcoming a), "Is Semantic Information Meaningful Data?". Preprint available at http://www.wolfson.ox.ac.uk/~floridi/papers.htm (defines semantic information as well-formed, meaningful and truthful data, graduate level)

Floridi, L. (forthcoming b), "Information, Semantic Conceptions of", *Stanford Encyclopedia of Philosophy*. (reviews philosophical conceptions of semantic information, undergraduate level)

Floridi, L. (forthcoming c) "On the Intrinsic Value of Information Objects and the Infosphere". Preprint available at http://www.wolfson.ox.ac.uk/~floridi/papers.htm (develops an ethical approach to information environments, undergraduate level)

Floridi, L. 1999, *Philosophy and Computing – An Introduction* (London – New York: Routledge) (textbook that complements this *Guide*, elementary undergraduate level)

Fox, C. J. 1983, *Information and Misinformation – An Investigation of the Notions of Information, Misinformation, Informing, and Misinforming* (Westport, Conn.: Greenwood Press) (analysis of information based on information science, undergraduate level).

Franklin, S. 1995, *Artificial Minds* (Cambridge, Mass.: The MIT Press) (undergraduate level).

Frieden, B. R. 1998, *Physics from Fisher Information: a Unification* (Cambridge: Cambridge University Press) (controversial attempt to provide an interpretation of physics in terms of information, requires a solid background in mathematics, graduate level only) .

Grice, P. 1989, *Studies in the Way of Words* (Cambridge Mass.: Harvard University Press) (collection of Grice's influential works, advanced undergraduate level).

Hanson, P. 1990 (ed.), *Information, Language and Cognition* (Vancouver: University of British Columbia Press) (important collection of essays, most at graduate level).

Hintikka, J. and P. Suppes 1970, *Information and inference* (Dordrecht: Reidel) (important collection of philosophical essays on information theory, graduate level).

Israel, D. and Perry J. 1990, "What is Information?" in Hanson 1990, pp. 1-19 (analyses information on the basis of situation logic, graduate level).

Jones, D. S. 1979, *Elementary information theory* (Oxford: Clarendon Press) (brief textbook on information theory, less mathematical than Cover and Thomas 1991, but still more demanding than Pierce 1980).

Landauer, R. 1987, "Computation: A Fundamental Physical View", *Physica Scripta* 35, 88-95 (graduate level only).

Landauer, R. 1991, "Information is Physical"", *Physics Today* 44, 23-29 (graduate level only).

Landauer, R. 1996, "The Physical Nature of Information" *Physics Letter* (A 217), 188 (graduate level only).

Landauer, R. and Bennett, C. H. 1985, "The Fundamental Physical Limits of Computation", *Scientific American* (July), 48-56 (a more accessible presentation of the view that information requires a physical implementation, undergraduate level).

Leff, H. S. and A. F. Rex 1990, *Maxwell's demon : entropy, information, and computing* (Bristol: Hilger) (collection of essays on this classic problem, graduate level).

Losee, R. M. 1997, “A Discipline Independent Definition of Information”, *Journal of the American Society for Information Science*, 48.3, 254-269.

Mabon, P. C. 1975, *Mission Communications: The Story of Bell Laboratories*. (very readable account of the people and the discoveries that made information theory possible, undergraduate level)

Machlup, F. 1983, “Semantic Quirks in Studies of Information”, in Machlup, F. and Mansfield, U. eds. (1983). *The Study of Information: Interdisciplinary Messages*, pp. 641-671 (New York: John Wiley).

MacKay, D. M. 1969, *Information, Mechanism and Meaning* (Cambridge, Ma.: MIT Press) (develops an alternative view of information to Shannon’s, graduate level) .

Mingers, J. 1997, “The Nature of Information and its Relationship to Meaning”, in R. L. Winder *et al.*, *Philosophical Aspects of Information Systems* (London: Taylor and Francis), pp. 73-84 (analyses information from a system theory perspective, advanced undergraduate level).

NATO 1974, Advanced Study Institute in Information Science, Champion, 1972. *Information Science: Search for Identity*, ed. by A. Debons (New York: Marcel Dekker).

NATO 1975, Advanced Study Institute in Information Science, Aberystwyth, 1974. *Perspectives in Information Science*, ed. by A. Debons and W. J. Cameron. (Leiden: Noordhoff).

NATO 1983, Advanced Study Institute in Information Science, Crete, 1978. *Information Science in Action: Systems Design*, ed. by A. Debons and A. G. Larson. (Boston: Martinus Nijhoff).

Nauta, D. 1972, *The meaning of information* (The Hague: Mouton) (reviews various analyses of information, advanced undergraduate level).

Newell, A. and Simon, H. A. 1976, “Computer Science as Empirical Inquiry: Symbols and Search” *Communications of the ACM*, 19 (March), 113-126 (the classic paper presenting the Physical Symbol System Hypothesis in AI and Cognitive Science, graduate level).

Pierce, J. R. 1980, *An introduction to information theory : symbols, signals and noise* (New York, Dover Publications) (old but still very valuable introduction to information theory for the non-mathematician, undergraduate level).

Popper, K. R. 1935, *Logik der Forschung*: *zur Erkenntnistheorie der modernen Naturwissenschaft* (Wien: J. Springer), Eng. tr. *The logic of scientific discovery* (London: Hutchinson, 1959) (Popper's classic text, graduate level).

Schrader, A. 1984, "In Search of a Name: Information Science and its Conceptual Antecedents", *Library and Information Science Research* 6, 227-271.

Schneider, T. 2000, "Information Theory Primer – With an Appendix on Logarithms", version 2.48, postscript version ftp://ftp.ncifcrf.gov/pub/delila/primer.ps, web version http://www.lecb.ncifcrf.gov/~toms/paper/primer/ (a very clear and simple introduction that can also be consulted for further clarification about the mathematics involved, undergraduate level).

Shannon, C. E. 1993a, *Collected Papers*, ed. by N. J. A. Sloane and A. D. Wyner (Los Alamos, Ca: IEEE Computer Society Press) (mostly graduate level only).

Shannon, C. E. 1993b, the article on information theory, *Encyclopedia Britannica*, reprinted in his *Collected Papers*, pp. 212-220 (an accessible presentation of information theory by its founding father, undergraduate level).

Shannon, C. E. and W. Weaver 1998 (orig. 1948), *The mathematical theory of communication*, with a foreword by R. E. Blahut and B. Hajek (Urbana and Chicago, Ill.: University of Illinois Press) (the classic text in information theory, graduate level; Shannon's text is also available on the web, see below).

Sloman A. 1978, *The Computer Revolution in Philosophy* (Atlantic Highlands: Humanities Press) (one of the earliest and most insightful discussions of the informational/computation turn in philosophy, most chapters undergraduate level).

Steane, A. M. 1998 "Quantum Computing", *Reports on Progress in Physics*, 61, 117-173 (a review, graduate level, also available online at http://xxx.lanl.gov/abs/quant-ph/9708022)

Thagard, P. R. 1990, "Comment: Concepts of Information", in Hanson 1990.

Weaver, W. 1949, "The Mathematics of Communication." *Scientific American* 181.1, 11-15 (very accessible introduction to Shannon's theory, undergraduate level).

Wellisch, H. 1972, "From Information Science to Informatics", *Journal of Librarianship* 4, 157-187.

Wersig, G. and Neveling, U. 1975, "The Phenomena of Interest to Information Science", *Information Scientist* 9, 127-140.

Wheeler, J. A. 1990, "Information, Physics, Quantum: The Search for Links" in W. H. Zureck (ed.) *Complexity, Entropy, and the Physics of Information* (Redwood City, Cal.: Addison Wesley) (introduces the It from Bit hypothesis, graduate level).

Wiener, N. 1954, *The Human Use of Human Beings: Cybernetics and Society,* 2nd ed. (London), reissued in 1989 with a new introduction by Steve J. Heims (London: Free Association) (a very early discussions of the ethical and social implications of the computer revolution, undergraduate level).

Wiener, N. 1961, *Cybernetics or Control and Communication in the Animal and the Machine*, 2nd ed. (Cambridge, Mass.: MIT Press) (the foundation of cybernetics, graduate level).

Further readings

Brillouin, L. 1962, *Science and information theory* (New York: Academic Press) (discusses the relation between information theory and physics, requires a solid mathematical background, graduate level).

Buckland, M. 1991, "Information as Thing", *Journal of the American Society of Information Science* (42.5), 351-360 (discusses the concept of information as hypostatised entity, undergraduate level).

Campbell, J. 1983, *Grammatical man*: *information, entropy, language, and life.* (London: Allan Lane) (introductory, undergraduate level).

Checkland, P. B. and Scholes, J. 1990, *Soft Systems Methodology in Action* (New York: John Wiley & Sons) (standard reference for the data + meaning analysis of information, undergraduate level).

Newman, J. 2001, "Some Observations on the Semantics of 'Information'" *Information Systems Frontiers* 3.2, pp. 155-167 (very useful and accessible review of some approaches to information theory and the analysis of semantic information, undergraduate level).

Rényi, A. 1987, *A diary on information theory* (New York, N.Y.: Chichester: Wiley) (introduces information theory and discusses some of its conceptual implications, graduate level).

Siegfried, T. 2000, *The bit and the pendulum*: *from quantum computing to M theory--the new physics of information* (New York, N.Y.: Chichester, Wiley) (very accessible account of some of the problems in the physics of information, undergraduate level).

Szaniawski K. 1984, *On Science, Inference, Information and Decision Making, Selected Essays in the Philosophy of Science*, ed. by A. Chmielewski and J. Wolenski (Dordrecht: Kluwer, 1998) (contains several essays on the philosophy of information, often undergraduate level).

Some web resources

There are many useful resources freely available on the web, the following have been used in writing this chapter:

Feldman D., *A Brief Tutorial on Information Theory, Excess Entropy and Statistical Complexity*: http://hornacek.coa.edu/dave/Tutorial/index.html

Fraundorf P., *Information-Physics on the Web*: http://newton.umsl.edu/infophys/infophys.html

Introduction to Information Theory, by Lucent Technologies Bell Labs Innovation: http://www.lucent.com/minds/infotheory/

MacKay J. C., *A Short Course in Information Theory*: http://www.inference.phy.cam.ac.uk/mackay/info-theory/course.html

Schneider T. 2000, *Information Theory Primer – With an Appendix on Logarithms*: http://www-lmmb.ncifcrf.gov/~toms/paper/primer/index.html, (a very clear and accessible introduction, undergraduate level).

UTI, the Unified Theory of Information website, contains documents, and links about the development of UTI: http://kaneda.iguw.tuwien.ac.at/uti/uti4/index.html.