

Philosophical Conceptions of Information

Luciano Floridi^{1,2}

¹ University of Hertfordshire

² University of Oxford

1 Introduction

I love information upon all subjects that come in my way, and especially upon those that are most important.

Thus boldly declares Euphranor, one of the defenders of Christian faith in Berkeley's *Alciphron* (Berkeley, (1732), Dialogue 1, Section 5, Paragraph 6/10). Evidently, information has been an object of philosophical desire for some time, well before the computer revolution, Internet or the dot.com pandemonium (see for example Dunn (2001) and Adams (2003)). Yet what does Euphranor love, exactly? *What is information?* The question has received many answers in different fields. Unsurprisingly, several surveys do not even converge on a single, unified definition of information (see for example Braman 1989, Losee (1997), Machlup and Mansfield (1983), Debons and Cameron (1975), Larson and Debons (1983)).

Information is notoriously a polymorphic phenomenon and a polysemantic concept so, 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. The reader may wish to keep this in mind while reading this article, where some schematic simplifications and interpretative decisions will be inevitable. Claude E. Shannon, 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.* (italics added) (Shannon (1993), p. 180).

Thus, following Shannon, Weaver (1949) 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. And these are only two early examples of the problems raised by any analysis of information.

Indeed, the plethora of different analyses can be confusing. Complaints about misunderstandings and misuses of the very idea of information are frequently expressed, even if to no apparent avail. Sayre (1976), for example, criticised the “laxity in use of the term ‘information’” in Armstrong (1968) (see now Armstrong (1993)) and in Dennett (1969) (see now Dennett (1986)), despite appreciating several other aspects of their work. More recently, Harms (1998) pointed out similar confusions in Chalmers (1996), who

seems to think that the information theoretic notion of information [see section 3, my addition] is a matter of what possible states there are, and how they are related or structured ... rather than of how probabilities are distributed among them (p. 480).

In order to try to avoid similar pitfalls, this article has been organised into three main parts.

Section two attempts to draw a map of the main senses in which one may speak of *semantic information*, and does so by relying on the analysis of the concept of *data* (Fig. 1). Sometimes the several concepts of information organised in the map can be variously coupled together. This should not be taken as necessarily a sign of confusion, for in some philosophers it may be the result of an intentional bridging. The map is not exhaustive and it is there mainly in order to avoid some obvious pitfalls and to narrow the scope of this article, which otherwise could easily turn into a short version of the Encyclopedia Britannica. Its schematism is only a starting point for further research.

After this initial orientation, section three provides a brief introduction to information theory, that is, to the mathematical theory of communication (MTC). MTC deserves a space of its own because it is the quantitative approach to the analysis of information that has been most influential among several philosophers. It provides the necessary background to understand several contemporary theories of semantic information, especially Bar-Hillel and Carnap (1953), Dretske (1981) and Floridi (2004b)).

Section four focuses entirely on the philosophical understanding of semantic information, what Euphranor really loves.

The reader must also be warned that an initial account of semantic information as *meaningful data* will be used as yardstick to outline other approaches. Unfortunately, even such a minimalist account is open to disagreement. In favour of this approach one may say that at least it is less controversial than others. Of course, a conceptual analysis must start somewhere. This often means adopting some working definition of the object under scrutiny. But it is not this commonplace that one needs to emphasize here. The difficulty is rather more daunting. Philosophical work on the concept of (semantic) information is still at that lamentable stage when disagreement affects even the way in which the problems themselves are provisionally phrased and framed. Nothing comparable to the well-polished nature of the Gettier problem is yet available, for example. So the “you are here” signal provided in this article might be placed elsewhere by other philosophers. The whole purpose is to put the concept of semantic information firmly on the philosophical map. Further adjustments will then become possible.

2 An Informational Map

Information is a conceptual labyrinth, and in this section we shall begin to have a look at a general map of one of its regions, with the purpose of placing ourselves squarely in the semantic area. Fig. 1 summarises the main distinctions that are going to be introduced.

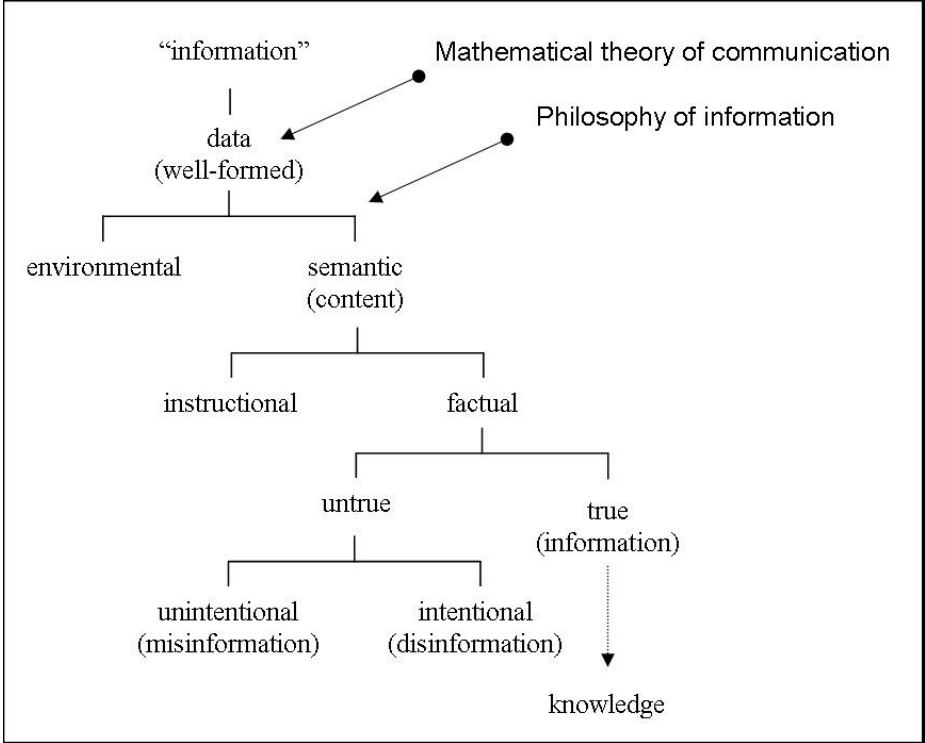


Fig. 1. An informational map

Clearly, percolating through the various points in the map will not make for a linear journey. Using a few basic examples, to illustrate the less obvious steps, will also help to keep our orientation. So let me introduce immediately the one to which we shall return more often.

2.1 An Everyday Example of Information

Monday morning. You turn on the ignition key of your car, but nothing happens: the engine does not even cough. The silence of the engine worries you. Unsurprisingly, you also notice that the red light of the low battery indicator is flashing. After a few more attempts, you give up and ring the garage. You

explain that your husband forgot to switch off the lights of the car last night it is a lie, you did, but you are too ashamed to confess it and now the battery is flat. The mechanic tells you that the instruction manual of your car explains how to use jump leads to start the engine. Luckily, your neighbour has everything you need. You read the manual, look at the illustrations, follow the instructions, solve the problem and finally drive to the office.

This everyday episode will be our “fruit fly”. Although it is simple and intuitive, it provides enough details to illustrate the many ways in which we understand one of our most important resources: *information*.

2.2 The Data-Based Definition of Information

It is common to think of information as consisting of *data*. It certainly helps, if only to a limited extent. For, unfortunately, the nature of data is not well-understood philosophically either, despite the fact that some important past debates - such as the one on the given and the one on sense data - have provided at least some initial insights. There still remains the advantage, however, that the concept of data is less rich, obscure and slippery than that of information, and hence easier to handle. So a data-based definition of information seems to be a good starting point.

Over the last three decades, several analyses in Information Science, in Information Systems Theory, Methodology, Analysis and Design, in Information (Systems) Management, in Database Design and in Decision Theory have adopted a *General Definition of Information* (GDI) in terms of *data + meaning* (see Floridi 2005b) for an extended bibliography). 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”). Recently, GDI has begun to influence the philosophy of computing and information (Floridi (1999) and Mingers (1997)).

A clear way of formulating GDI is as a tripartite definition (Fig. 2):

GDI) σ is an instance of information, understood as 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* ($mwfd = \delta$).

Fig. 2. The General Definition of Information (GDI)

GDI requires a definition of data. This will be provided in the next section. Before, a brief comment on each clause is in order.

According to (GDI.1), data are the stuff of which information is made. We shall see that things can soon get more complicated.

In (GDI.2), “well-formed” means that the data are clustered together correctly, according to the rules (*syntax*) that govern the chosen system, code or language being analysed. Syntax here must be understood broadly (not just linguistically), as what determines the form, construction, composition or

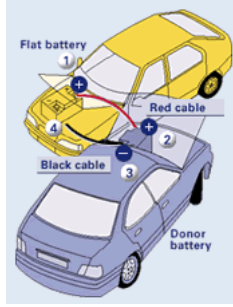


Fig. 3. How to jump start your car ©Copyright Bosh 2005

structuring of something (engineers, film directors, painters, chess players and gardeners speak of syntax in this broad sense). For example, the manual of your car may show (see Fig. 3) a two dimensional picture of the two cars placed one near the other, not one on top of the other.

This pictorial syntax (including the linear perspective that represents space by converging parallel lines) makes the illustrations potentially meaningful to the user. Using the same example, the actual battery needs to be connected to the engine in a correct way to function: this is still syntax, in terms of correct physical architecture of the system (thus a disconnected battery is a syntactic problem). And of course the conversation you carry on with your neighbour follows the grammatical rules of English: this is syntax in the ordinary linguistic sense.

Regarding (GDI.3), this is where semantics finally occurs. “Meaningful” means that the data must comply with the meanings (*semantics*) of the chosen system, code or language in question. However, let us not forget that semantic information is not necessarily linguistic. For example, in the case of the manual of the car, the illustrations are such as to be visually meaningful to the reader.

2.3 A Definition of Data

According to GDI, information cannot be dataless but, in the simplest case, it can consist of a single datum (d). Now a datum is reducible to just a lack of uniformity (*diaphora* is the Greek word for “difference), so a general definition of a datum is (Fig. 4):

Depending on philosophical inclinations, the diaphoric definition of data can be applied at three levels:

1. data as diaphora *de re*, that is, as lacks of uniformity in the real world out there. There is no specific name for such “data in the wild”. A possible suggestion is to refer to them as *dedomena* (“data” in Greek; note that our word “data comes from the Latin translation of a work by Euclid entitled *Dedomena*). *Dedomena* are not to be confused with environmental data (see section 2.7.1). They are pure data or proto-epistemic data, that is, data before they are epistemically interpreted. As “fractures in the fabric of being” they can

Dd) datum = x *being distinct* (\neq) from y
 where the x and the y are two uninterpreted variables and the domain is left open to further interpretation.

Fig. 4. The diaphoric definition of data

only be posited as an external anchor of our information, for dedomena are never accessed or elaborated independently of a *level of abstraction* (more on this in section 4.2.2). They can be reconstructed as ontological requirements, like Kant's *noumena* or Locke's *substance*: they are not epistemically experienced but their presence is empirically inferred from (and required by) experience. Of course, no example can be provided, but dedomena are whatever lack of uniformity in the world is the source of (what looks to information systems like us as) as data, e.g. a red light against a dark background. Note that the point here is not to argue for the existence of such pure data in the wild, but to provide a distinction that (in section 2.6) will help to clarify why some philosophers have been able to accept the thesis that there can be no information without data representation while rejecting the thesis that information requires physical implementation;

2. data as diaphora *de signo*, that is, lacks of uniformity between (the perception of) at least two physical states, such as a higher or lower charge in a battery, a variable electrical signal in a telephone conversation, or the dot and the line in the Morse alphabet; and
3. data as diaphora *de dicto*, that is, lacks of uniformity between two symbols, for example the letters A and B in the Latin alphabet.

Depending on one's position with respect to the thesis of ontological neutrality (section 2.6) and the nature of environmental information (section 2.7.1) dedomena in (1) may be either identical with, or what makes possible signals in (2), and signals in (2) are what make possible the coding of symbols in (3).

The dependence of information on the occurrence of differences variously implementable physically, explain why information can so easily be decoupled from its support. The actual *format*, *medium* and *language* in which semantic information is encoded is often irrelevant and hence disregardable. In particular, the same semantic information may be analog or digital, printed on paper or viewed on a screen, in English or in some other language, expressed in words or pictures. Interpretations of this support-independence can vary quite radically. For Dd (see Fig. 4 above) leaves underdetermined

- the classification of the relata (*taxonomic neutrality*);
- the logical type to which the relata belong (*typological 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*).

We shall now look at each form of neutrality in turn.

2.4 Taxonomic Neutrality

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 [black-dot-on-white-sheet] datum itself, together with the fundamental relation of inequality that couples it with the dot. Nothing seems to be a datum *per se*. Rather, being a datum is an external property. So GDI endorses the following thesis:

TaN) a datum is a relational entity.

The slogan is “data are relata”, but GDI is neutral with respect to the identification of data with *specific* relata. In our example, GDI refrains from identifying either the red light or the white background as the datum. To understand why there cannot be “dataless information”, we shall now look at the typological neutrality of GDI.

2.5 Typological Neutrality

According to GDI, information can consist of different types of data as relata (δ). Five classifications are quite common, although the terminology is not yet standard or fixed (but see Floridi (1999)). They are not mutually exclusive, and one should not understand them as rigid: depending on circumstances, on the sort of analysis conducted and on the level of abstraction adopted, the same data may fit different classifications.

- $\delta 1$ *Primary data*. These are the principal data stored e.g. in a database, for example a simple array of numbers. They are the data an information-management system such as the one used in the car to indicate that the battery needs to be charged is generally designed to convey (in the form of information) to the user in the first place. Normally, when speaking of data, and of the corresponding information they constitute, one implicitly assumes that *primary* data/information is what is in question. So, by default, the red light of the low battery indicator flashing is assumed to be an instance of primary data conveying primary information.
- $\delta 2$ *Secondary data*. These are the converse of primary data, constituted by their absence (one could call them anti-data). Recall how you first suspected that the battery was flat: the engine failed to make any of the usual noise. Likewise, in Silver Blaze, Sherlock Holmes solves the case by noting something that has escaped everybody else: the unusual silence of the dog. Clearly, silence may be very informative. This is a peculiarity of information: its absence may also be informative. When it is, the point is stressed by speaking of *secondary* information.
- $\delta 3$ *Metadata*. These are indications about the nature of some other (usually primary) data. They describe properties such as location, format, updating,

availability, usage restrictions, and so forth. Correspondingly, *metainformation* is information about the nature of information. “The battery is flat’ is encoded in English” is a simple example.

- $\delta 4$ *Operational data*. These are data regarding the operations of the whole data system and the systems performance. Correspondingly, *operational information* is information about the dynamics of an information system. Suppose the car has a yellow light that, when flashing, indicates that the car checking system is malfunctioning. The fact that the light is on may indicate that the low battery indicator is not working properly, thus undermining the hypothesis that the battery is flat.
- $\delta 5$ *Derivative data*. These are data that can be extracted from some data whenever the latter are used as indirect sources in search of patterns, clues or inferential evidence about other things than those directly addressed by the data themselves, e.g. for comparative and quantitative analyses (*ideometry*). As it is difficult to define this category precisely, a familiar example may be helpful to convey the point. Credit cards notoriously leave a trail of derivative information. From someones credit card bill, concerning e.g. the purchase of petrol in a certain petrol station, one may derive the information of her whereabouts at a given time. Again, derivative information is not something new. Hume provides a beautiful example in these days of global warming. In the *Essays oral, Political, and Literary* (Part II, Essay 11. Of the Populousness of Ancient Nations, Para. 155/186 mp. 448 gp. 432, see now Hume (1987)) he reports that

It is an observation of LAbbe du Bos, that Italy is warmer at present than it was in ancient times. *The annals of Rome* tell us, says he, that in the year 480 ab U.C. the winter was so severe that it destroyed the trees. [...] Many passages of Horace suppose the streets of Rome full of snow and ice. We should have more certainty with regard to this point, had the ancients known the use of thermometers: But their writers, without intending it, give us information, sufficient to convince us, that the winters are now much more temperate at Rome than formerly.

Hume has just extracted some derivative information from some primary information provided by LAbbe du Bos.

Let us now return to our question: can there be dataless information? GDI does not specify which types of data constitute information. This *typological neutrality* 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 $\delta.2-\delta.5$. For example, if a database query provides an answer, it will provide at least a *negative* answer, e.g. “no documents found”. This is primary negative information. However, if the database provides no answer, either it fails to provide any data at all, in which case no specific information σ is available so the rule “no information without data” still applies or it can provide some data δ to establish, for example, that it is running in a

loop. Likewise, silence, this time as a reply to a question, could represent negative primary information, e.g. as implicit assent or denial, or it could carry some non-primary information μ , e.g. about the fact that the person has not heard the question, or about the amount of noise in the room.

2.6 Ontological Neutrality

By rejecting the possibility of dataless information, GDI also endorses the following modest thesis of ontological neutrality:

ON) no information without data representation.

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

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 information and computing (see Deutsch (1985);(1997) and 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 *dedomena*, such as Leibnizian monads, for example. Indeed, the classic realism debate on the ultimate nature of “being” 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 (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);

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/dedomena 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 loosely 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). The formulation in Mackay (1969) - that is "information is a *distinction* that makes a difference" - predates Batesons but it is slightly different from it in that, by speaking of "distinction" instead of "difference", it has an epistemological rather than an ontological twist. A "difference" (a "distinction") is just a discrete state, namely a datum, and "making a difference" simply means that the datum is "meaningful", at least potentially.

2.7 Genetic Neutrality

Finally, let us consider 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*, a concept sufficiently important to deserve a brief presentation before we close this first part.

Environmental information. One of the most often cited example of environmental information is the series of concentric rings visible in the wood of a cut tree trunk, which may be used to estimate its age. Yet “environmental” information does not need to be *natural*. Going back to our example, when you turned the ignition key, the red light of the low battery indicator flashed. This signal too can be interpreted as an instance of environmental information.

Environmental information is defined relative to an observer (an information agent), who is supposed to have no direct access to pure data in themselves. It requires two systems *a* and *b* to be coupled in such a way that *a*'s being (of type, or in state) *F* is correlated to *b* being (of type, or in state) *G*, thus carrying for the observer the information that *b* is *G* (this analysis is adapted from Barwise and Seligman (1997), who improve on a similar account by Dretske (1981)).

Environmental information: two systems *a* and *b* are coupled in such a way that *a*'s being (of type, or in state) *F* is correlated to *b* being (of type, or in state) *G*, thus carrying for the information agent the information that *b* is *G*.

Fig. 5. Environmental information

The correlation in Fig. 5 is usually *nomio* (it follows some law). It may be engineered as in the case of the low battery indicator (*a*) whose flashing (*F*) is triggered by, and hence it is informative about, the battery (*b*) being flat (*G*). Or it may be natural, as when litmus - a natural colouring matter from lichens - is used as an acid-alkali indicator because it turns red in acid solutions and blue in alkaline solutions. Other typical examples include the correlation between fingerprints and personal identification.

One may be so used to see the low battery indicator flashing as carrying the information that the battery is flat to find it hard to distinguish, with sufficient clarity, between environmental and semantic information. However, it is important to stress that environmental information may require or involve no semantics at all. It may consist of (networks or patterns of) correlated data understood as mere differences or constraining affordances. Plants (e.g., a sunflower), animals (e.g., an amoeba) and mechanisms (e.g., a photocell) are certainly capable of making practical use of environmental information even in the absence of any (semantic processing of) *meaningful* data.

2.8 Summary of the First Part

To summarise, GDI defines information, broadly understood, as 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. Indeed, philosophers have variously interpreted and tuned these four neutralities according to their theoretical needs.

Our next step is to check whether GDI is satisfactory when discussing the most important type of semantic information, namely factual information. Before addressing this issue, however, we need to pause and look at the mathematical theory of communication (MTC).

MTC is not the only successful mathematical approach to the concept of information. *Fisher information* (Frieden (1998)) and the *algorithmic information theory* (Chaitin (1987)) provide two other important examples. However, MTC is certainly the most widely known among philosophers. As such, it has had a profound impact on philosophical analyses of semantic information, to which it has provided both the technical vocabulary and at least the initial conceptual frame of reference. One needs to grasp its main gist if one wishes to make sense of the issuing philosophical debate.

3 Information as Data Communication

Some features of information are intuitive. We are used to information being *encoded*, *transmitted* and *stored*. One also expects it to be additive (information a + information b = information $a + b$) and *non-negative*, like other things in life, such as probabilities and interest rates. If you ask a question, the worst scenario is that you receive no answer or a wrong answer, which will leave you with zero new information.

Similar properties of information are quantifiable. They are investigated by the *mathematical theory of communication* (MTC) with the primary aim of devising efficient ways of encoding and transferring data.

The name for this branch of probability theory comes from Shannon's seminal work (Shannon and Weaver (1949)). Shannon pioneered this field and obtained many of its principal results, but he acknowledged the importance of previous work done by other researchers and colleagues 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 second part of the article outlines some of the key ideas behind MTC, with the aim of understanding the relation between MTC and some philosophical theories of semantic information. The reader with no taste for mathematical formulae may wish to go directly to section 3.2, where some conceptual implications of MTC are outlined. The reader interested in knowing more may start by reading Weaver (1949), Pierce (1980), Shannon and Weaver (1949), then Jones

(1979), and finally Cover and Thomas (1991). The latter two are technical texts. Floridi (2003a) provides a simplified analysis oriented to philosophy students.

3.1 The Mathematical Theory of Communication

MTC has its origin in the field of electrical engineering, 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 (how small can a message be, given the same amount of information to be encoded?) and the ultimate rate of data transmission (how fast can data be transmitted over a channel?). The two solutions are the entropy H in equation [9] (see below) and the channel capacity C . The rest of this section illustrates how to get from the problems to the solutions.

To have an intuitive sense of the approach, let us return to our example. Recall the telephone conversation with the mechanic. In Fig. 6, the wife is the *informer*, the mechanic is the *informee*, “the battery is flat” is the (semantic) message (the *informant*), there is a coding and decoding procedure through a natural language (English), a channel of communication (the telephone system) and some possible noise. Informer and informee share the same background knowledge about the collection of usable symbols (technically known as the *alphabet*; in the example this is English).

MTC is concerned with the efficient use of the resources indicated in Fig. 6. Now, the conversation with the mechanic is fairly realistic and hence

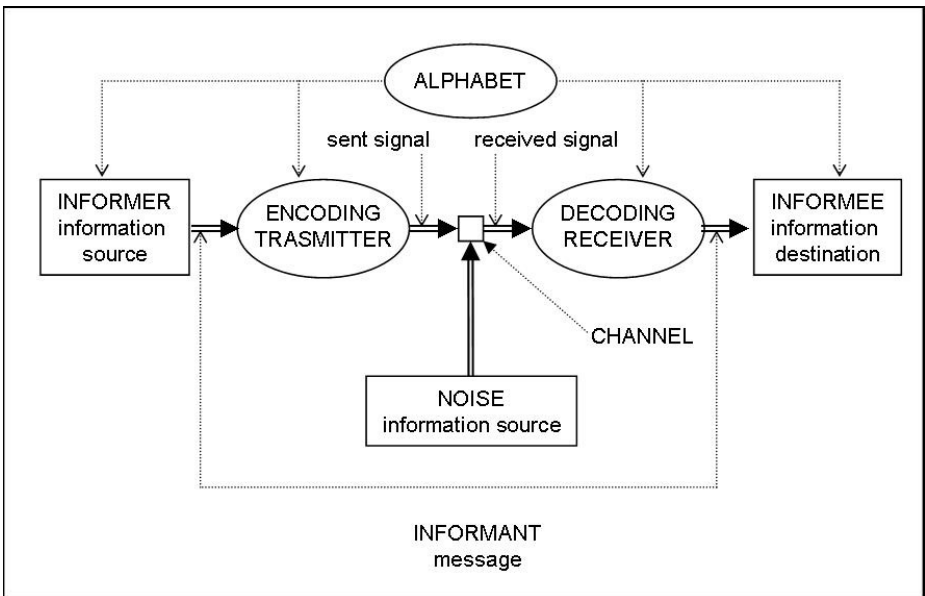


Fig. 6. Communication model (adapted from Shannon and Weaver [1949 rep. 1998])

more difficult to model than a simplified case. We shall return to it later but, in order to introduce MTC, imagine instead a very boring device that can produce only one symbol. Edgar Alan Poe wrote a short story in which a raven can answer only “nevermore” to any question. Poe’s raven is called a *unary device*. Imagine you ring the garage and your call is answered by Poe’s raven. Even at this elementary level, Shannons simple model of communication still applies. It is obvious that the raven (a unary device) produces zero amount of information. Simplifying, we already know the outcome of the communication exchange, so our ignorance (expressed by our question) cannot be decreased. Whatever the informational state of the system is, asking appropriate questions (e.g. “will I be able to make the car start?”, “can you come to fix the car?”) of the raven does not make any difference. Note that, interestingly enough, this is the basis of Platos famous argument in the *Phaedrus* against the value of semantic information provided by written texts:

[Socrates]: Writing, Phaedrus, has this strange quality, and is very like painting; for the creatures of painting stand like living beings, but if one asks them a question, they preserve a solemn silence. And so it is with written words; you might think they spoke as if they had intelligence, but if you question them, wishing to know about their sayings, they always say only one and the same thing [they are unary devices, in our terminology]. And every word, when [275e] once it is written, is bandied about, alike among those who understand and those who have no interest in it, and it knows not to whom to speak or not to speak; when ill-treated or unjustly reviled it always needs its father to help it; for it has no power to protect or help itself.

As Plato well realises a unary source answers every question all the time with only one message, not with silence or message, since silence counts as a message, as we saw in 2.5, when discussing the nature of secondary information. It follows that a completely silent source also qualifies as a unary source. And if silencing a source (censorship) may be a nasty way of making a source uninformative, it is well known that crying wolf is a classic case in which an informative source degrades to the role of uninformative unary device.

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 can be a very misleading term because of the strong epistemological connotations of this term. Remember that the informee can be a simple machine, and psychological, mental or doxastic states are clearly irrelevant.

Once the coin has been tossed, the system produces an amount of 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 now build a slightly more complex system, made of two fair coins A and B . The AB system can produce 4 ordered outputs: $\langle h, h \rangle, \langle h, t \rangle, \langle t, h \rangle, \langle t, t \rangle$. 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 $\langle -, - \rangle$ removes a higher data deficit than the occurrence of a symbol in the A system. In other words, each symbol provides more 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 now 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 a device producing one of two equiprobable symbols (coin A in the example above) plus the amount of information provided by another device producing one of two equiprobable symbols (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. Now, 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 n is the power to which 2 must be raised to give the number n , 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 now rephrase some examples more precisely (Fig. 7) by using equation [1]:

average informativeness per symbol (or “uncertainty”)

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

The basic idea is all in equation [1]. Information can be quantified in terms of decrease in data deficit (Shannon’s “uncertainty”). Unfortunately, real coins are always biased. To calculate how much information they produce one must 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 Shannon’s 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 information produced by the biased coin depends on the average *informativeness* (also known as average *surprisal*, another unfortunate term to

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. 7. Examples of communication devices and their information power

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 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 at all, behaving like the raven or the boy who cries wolf.

So, to calculate the average informativeness of S we need to know how to calculate S and the informativeness of the i^{th} symbol in general. This requires understanding what the probability of the 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:

$$\log(N) = -\log(N^{-1}) = -\log\left(\frac{1}{N}\right) = -\log(P) \quad (2)$$

The value of $\frac{1}{N} = P$ can range from 0 to 1. If Poe's raven is our source, the probability of it saying “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 and, in the worst case scenario, can at most be equal to zero, but never become “negative”. More formally, this means:

$$\sum_{i=1}^N P_i = 1 \quad (3)$$

The sigma notation in [3] is simply a shortcut that indicates that if we add all probabilities values from $i = 1$ to $i = N$ their 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 the i^{th} symbol can be expressed by analogy with $-\log(P)$ in equation [2]:

$$u_i = -\log(P_i) \quad (4)$$

Next, we need to calculate the length of a general string S . Suppose that the biased coin, tossed 10 times, produces the string: $\langle h, h, t, h, h, t, t, h, h, t \rangle$. 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:

$$S = \sum_{i=1}^N S_i \quad (5)$$

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:

$$\frac{\sum_{i=1}^N S_i u_i}{\sum_{i=1}^N S_i} \quad (6)$$

Term [6] can be simplified thus:

$$\sum_{i=1}^N \frac{S_i}{S} u_i \quad (7)$$

Now $\frac{S_i}{S}$ 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 term [7], we have:

$$\sum_{i=1}^N P_i u_i \quad (8)$$

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

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

Equation [9] is Shannon's formula for H = uncertainty, which 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; apparently, Shannon used the letter H because of R.V.L. Hartley's previous work).

Equation [9] indicates that the quantity of 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 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, witness our Monday morning accident. 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 an n -th order Markov source. Consider for example the probability of hearing “n” (followed by the string “ing”) after having received the string of letters “Good mor-” over the phone, when you called the garage. And consider the same example through time, in the case of a child (the son of the mechanic) who is learning how to answer the phone instead of his father. 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. 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 concepts that play a pivotal role both in communication analysis and in memory management 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, as in Fig. 8:

$\langle h, h \rangle = 00$	$\langle h, t \rangle = 01$	$\langle t, h \rangle = 10$	$\langle t, t \rangle = 11$
-----------------------------	-----------------------------	-----------------------------	-----------------------------

Fig. 8. Code 1

Call this Code 1. In Code 1 a message conveys 2 bits of information, as expected. Do not confuse *bits* as *bi*-nary units of information (recall that we decided to use \log_2 also as a matter of convenience) with *bits* as *bi*-nary digits, 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 (Fig. 9):

$\langle h, h \rangle = 0.5$	$\langle h, t \rangle = 0.25$	$\langle t, h \rangle = 0.125$	$\langle t, t \rangle = 0.125$
------------------------------	-------------------------------	--------------------------------	--------------------------------

Fig. 9. A biased system

This biased system produces less information, so by using Code 1 we would be wasting resources. A more efficient Code 2 (Fig. 10) should take into account the symbols probabilities, with the following outcomes:

$\langle h, h \rangle = 0$	0.5×1 binary digit = .5
$\langle h, t \rangle = 10$	0.25×2 binary digits = .5
$\langle t, h \rangle = 110$	0.125×3 binary digits = .375
$\langle t, t \rangle = 111$	0.125×3 binary digits = .375

Fig. 10. Code 2 (Fano Code)

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.

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* (data received but unwanted). 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 the manual of your car includes both verbal explanations and pictures to convey (slightly redundantly) the same information.

We are now ready to understand Shannons two fundamental theorems. Suppose the 2-coins biased system AB produces the following message:

$$\langle t, h \rangle \langle h, h \rangle \langle t, t \rangle \langle h, t \rangle \langle h, t \rangle$$

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). Shannons fundamental theorem of the noiseless channel states that there is no method of encoding which gives an equivocation less than $H - C$, as explained in Fig. 11.

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 $C/H - \varepsilon$ 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 and Weaver (1949 rep. 1998), 59.

Fig. 11. Shannon’s fundamental theorem of the noiseless channel

Let a discrete channel have the capacity C and a discrete source the entropy per second H . If $H \leq C$ 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 $H - C + \varepsilon$ where ε is arbitrarily small. There is no method of encoding which gives an equivocation less than $H - C$.

Shannon and Weaver (1949 rep. 1998), 71.

Fig. 12. Shannon's fundamental theorem for a discrete channel

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, as explained in Fig. 12.

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. They are limiting results in information theory that constrain any conceptual analysis of semantic information. They are thus comparable to Gdel's Turing's and Church's theorems in logic and computation. With our message finally sent, we may close this section and return to a more philosophical approach.

3.2 Conceptual Implications of the Mathematical Theory of Communication

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 information is by considering the number of yes/no questions required to determine what the source is communicating. One question is sufficient to determine the output of a fair coin, which therefore is said to produce 1 bit of information. A 2-fair-coins system produces 4 ordered outputs: $\langle h, h \rangle$, $\langle h, t \rangle$, $\langle t, h \rangle$, $\langle t, t \rangle$, and therefore requires at least two questions, each output containing 2 bits of information, and so on. This *erotetic* (the Greek word for "question") analysis clarifies two important points.

First, MTC is not a theory of information in the ordinary sense of the word. In MTC, information has an entirely technical meaning. Consider some examples. According to MTC, two equiprobable "yes"'s contain the same quantity of information, no matter whether their corresponding questions are "have the lights of your car been left switched on for too long, without recharging the battery?" or "would you marry me?". If we knew that a device could send us, with equal probabilities, either this article or the whole *Stanford Encyclopedia of Philosophy*, by receiving one or the other we would receive very different amounts of bytes of data but actually only one bit of information in the MTC sense of the word. On June 1

1944, the BBC broadcasted a line from Verlaine's *Song of Autumn*: "Les sanglots longs des violons de Automne". 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 Shannons sense of the word, but failed to understand (or believe in) the crucial importance of those two small bits of data. As for ourselves, we were not surprised to conclude in the previous section that the maximum amount of information (again, in the MTC sense of the word) is produced by a text where each character is equally distributed, that is by a perfectly random sequence. According to MTC, the classic monkey randomly pressing typewriter keys is indeed producing a lot of information.

Second, since MTC is a theory of information without meaning (not in the sense of meaningless, but in the sense of not yet meaningful), and since we have seen that [information - meaning = data], "mathematical theory of data communication" is a far more appropriate description of this branch of probability theory than "information theory". This is not a mere question of labels. Information, as semantic content (more on this shortly), can also be described erotetically as *data + queries*. Imagine a piece of (propositional) 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], such as [does the earth have only one moon? + yes]. Subtract the "yes" - which is at most 1 bit of information, in the equiprobable case of a yes or no answer - and you are left with virtually all the semantic content, fully de-aethicised (from *aletheia*, the Greek word for truth; the query is neither true nor false). To use a Fregean expression, *semantic content* is *unsaturated information*, where the latter is semantic information that has been "eroteticised" and from which a quantity of information has been subtracted equal to $\log P(\text{yes})$, with P being the probability of the yes-answer.

The datum "yes" works as a key to unlock the information contained in the query. MTC studies the codification and transmission of information by treating it as data keys, that is, as the amount of details in a signal or message or memory space necessary to saturate the informees unsaturated information. As Weaver (1949) remarked

the word information relates not so much to what you do say, as to what you could say. The mathematical theory of communication 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 (p.12).

Since MTC deals not with semantic information itself but with the data that constitute it, that is, with messages comprising 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.

H is also known in MTC as *entropy*. It seems we owe this confusing label to John von Newman, who recommend it to Shannon:

You should call it entropy for two reasons: first, the function is already in use in thermodynamics under the same name; second, and more importantly, most people don't know what entropy really is, and if you use the word *entropy* in an argument you will win every time (quoted by Golan (2002)).

Von Newman proved to be right on both accounts, unfortunately.

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

- a) the average amount of information per symbol produced by the informer, or
- b) the corresponding average amount of data deficit (Shannons 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: I do not know whether it is head or tail. Use two fair coins and my deficit doubles, but use the raven, and my deficit becomes null. My empty glass (point (b) above) is an exact measure of your capacity to fill it (point (a) above). Of course, it makes sense to talk of information as quantified by H only if one can specify the probability distribution.

Regarding (c), MTC treats 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, that is, it is easily associated to interpretations, as I used to try to explain to my parents when I was young). Entropy is 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 information in Shannon sense, and hence it causes a smaller data deficit, which can be close to zero (remember the raven). By contrast, 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 Information as Semantic Content

We have seen that, when data are well-formed and meaningful, the result is also known as *semantic content* (Bar-Hillel and Carnap (1953; Bar-Hillel (1964). Information, understood as semantic content, comes in two main varieties: factual and instructional. In our example, one may translate the red light flashing into semantic content in two senses:

- a) as a piece of factual information, representing the fact that the battery is flat; and
- b) as a piece of instructional information, conveying the need for a specific action, e.g. the re-charging or replacing of the flat battery.

In this third part of the article we shall be concerned primarily with (a), so it is better to clear the ground by considering (b) first. It is the last detour in our journey.

4.1 Instructional Information

Instructional information is a type of semantic content. An instruction booklet, for example, provides instructional information, either imperatively - in the form of a recipe: first do this, then do that - or conditionally, in the form of some inferential procedure: if such and such is the case do this, otherwise do that.

Instructional information is not about a situation, a fact, or a state of affairs w and does not model, or describe or represent w . Rather, it is meant to (help to) bring about w . For example, when the mechanic tells one over the phone to connect a charged battery to the flat battery of ones car, the information one receives is not factual, but instructional.

There are many plausible contexts in which a stipulation (“let the value of $x = 3$ ” or “suppose we discover the bones of a unicorn”), an invitation (“you are cordially invited to the college party”), an order (“close the window!”), an instruction (“to open the box turn the key”), a game move (“1.e2-e4 c7-c5” at the beginning of a chess game) may be correctly qualified as kinds of instructional information. The printed score of a musical composition or the digital files of a program may also be counted as typical cases of instructional information.

All these instances of information have a semantic side: they have to be at least potentially meaningful (interpretable) to count as information. Moreover, instructional information may be related to factual (descriptive) information in

performative contexts, such as christening - e.g. “this ship is now called *HMS The Informer*” - or programming - e.g. as when deciding the type of a variable. The two types of semantic information (instructional and factual) may also come together in magic spells, where semantic representations of x may be (wrongly) supposed to provide some instructional power and control over x . Nevertheless, as a test, one should remember that instructional information does not qualify alethically (cannot be correctly qualified as true or false). In the example, it would be silly to ask whether the information “only use batteries with the same rated voltage” is true. Stipulations, invitations, orders, instructions, game moves, and software cannot be true or false. As Wittgenstein remarks “The way music speaks. Do not forget that a poem, even though it is composed in the language of information, is not used in the language-game of giving information.” (Zettel, §160, see Wittgenstein (1981)).

4.2 Factual Information

In the language game that Wittgenstein seems to have in mind, the notion of “semantic information” is intended in a declarative or factual mode. Factual information may be true or untrue (false, in case one adopts a binary logic). *True semantic content* is the most common sense in which information seems to be understood (Floridi (2004b)). It is also one of the most important, since information as true semantic content is a necessary condition for knowledge. Some elaboration is in order, and in the following sub-sections we shall briefly look at the concept of data as constraining affordances, at the role played by levels of abstraction in the transformation of constraining affordances into factual information, and finally at the relation between factual information and truth.

4.2.1 Constraining Affordances

The data that constitute factual information allow or invite certain constructs (they are *affordances* for the information agent that can take advantage of them) and resist or impede some others (they are *constraints* for the same agent), depending on the interaction with, and the nature of, the information agent that processes them. For example, the red light flashing repetitively and the engine not starting allow you (or any other information agent like you) to construct the information that (a) the battery is flat, while making it more difficult to you (or any other information agent like you) to construct the information that (b) there is a short circuit affecting the proper functioning of the low battery indicator, where the engine fails to start because there is no petrol in the tank, a fact not reported by the relevant indicator which is affected by the same short circuit. This is the sense in which data are *constraining affordances* for (an information agent responsible for) the elaboration of factual information.

4.2.2 Levels of Abstraction

In section 2.3, we saw that the concept of pure data in themselves (*dedomena*) is an abstraction, like Kant’s noumena or Locke’s substance. The point made

was that data are never accessed and elaborated (by an information agent) independently of a *level of abstraction* (LoA). The time has come to clarify what a LoA is.

A LoA is a specific set of typed variables, intuitively representable as an interface, which establishes the scope and type of data that will be available as a resource for the generation of information (Floridi and Sanders (forthcoming)). This concept of LoA is purely epistemological, and it should not be confused with other forms of “levellism” that are more or less explicitly based on an ontological commitment concerning the intrinsic architecture, syntax or structure of the system discussed (Dennett (1971), Marr (1982), Newell (1982), Simon (1969), see now Simon (1996); Poli (2001) provides a reconstruction of ontological levellism; more recently, Craver (2004) has analysed ontological levellism, especially in biology and cognitive science, see also Craver (forthcoming)). Ontological levellism has come under increasing attack. Heil (2003) and Schaffer (2003) have seriously and convincingly questioned its plausibility. However, epistemological levellism is flourishing, especially in computer science (Roever et al. (1998), Hoare and Jifeng (1998)), where it is regularly used to satisfy the requirement that systems constructed in levels (in order to tame their complexity) function correctly.

Through a LoA, an information agent (the observer) accesses a physical or conceptual environment, the system. LoAs are not necessarily hierarchical and they are comparable. They are interfaces that mediate the epistemic relation between the observed and the observer. Consider, for example, a motion detector (Fig. 13). In the past, motion detectors caused an alarm whenever a movement was registered within the range of the sensor, including the swinging of a tree branch (object a in Fig. 13). The old LoA1 consisted of a single typed variable, which may be labelled MOVEMENT. Nowadays, when a PIR (passive infrared) motion detector registers some movement, it also monitors the presence of an infrared signal, so the entity detected has to be something that also emits infrared radiation usually perceived as heat before the sensor activates the alarm. The new LoA2 consists of two typed variables: MOVEMENT and INFRARED RADIATION. Clearly, your car (object b in Fig. 13) leaving your house is present for both LoAs; but for the new LoA2, which is more finely grained, the branch of the tree swinging in the garden is absent. Likewise, a stone in the garden (object c in Fig. 13) is absent for both the new and the old LoA, since it satisfies no typed variable of either one.

The method of LoA is an efficient way of making explicit and managing the ontological commitment of a theory. In our case, “the battery is what provides electricity to the car” is a typical example of information elaborated at a drivers LoA. An engineers LoA may output something like “12-volt lead-acid battery is made up of six cells, each cell producing approximately 2.1 volts”, and an economists LoA may suggest that “a good quality car battery will cost between \$50 and \$100 and, if properly maintained, it should last five years or more”.

Data as constraining affordances - answers waiting for the relevant questions - are transformed into factual information by being processed semantically at a given LoA (alternatively: the relevant question is associated to the right answer

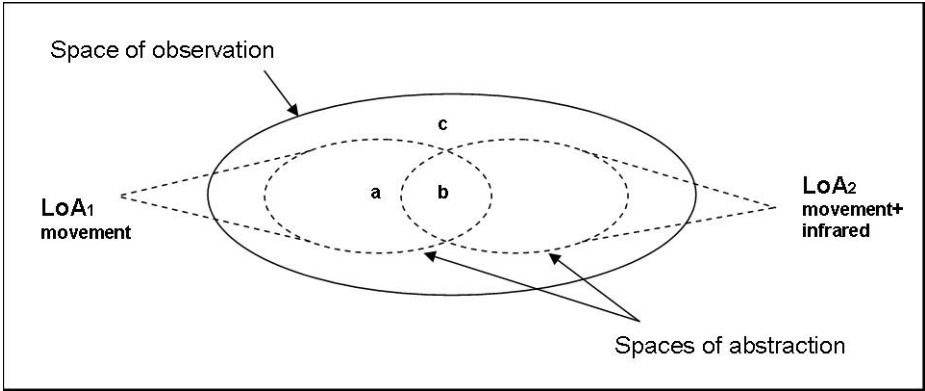


Fig. 13. An example of Levels of Abstraction

at a given LoA). Once data as constraining affordances have been elaborated into factual information at a given LoA, the next question is whether truth values supervene on factual information.

4.2.3 Information and Truth

Does some factual content qualify as information only if it is true? Defenders of the alethic neutrality of semantic information (Fetzer (2004) and Dodig-Crnkovic (2005), who criticises Floridi (2004b); Colburn (2000), Fox (1983), and, among, situation theorists, Devlin (1991)) argue that meaningful and well-formed data already qualify as information, no matter whether they represent or convey a truth or a falsehood or indeed have no alethic value at all. Opponents, on the other hand, object that “[...] *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 that “false information is not an inferior kind of information; it just is not information” (Grice (1989), 371; other philosophers who accept a truth-based definition of semantic information are Barwise and Seligman (1997) and Graham (1999)). The result is a definition of factual semantic information as well-formed, meaningful and truthful data (defended in Floridi (2004b), (2005b)), where “truthful” is only a stylistic choice to be preferred to “true” because it enables one to say that a map conveys factual information insofar as it is truthful.

Once again, the debate is not about a mere definition, but concerns the possible consequences of the alethic neutrality thesis, three of which can be outlined here, whereas a fourth requires a longer analysis and will be discussed in section 5.1.

If the thesis “meaningful and well-formed data already qualify as information” is correct then

- i) false information (including contradictions) would count as a genuine type of semantic information, not as pseudo-information;
- ii) all necessary truths (including tautologies) would qualify as information (on this issue see Bremer (2003)); and

- iii) “it is true that p ” - where p is a variable that can be replaced by any instance of genuine semantic information - would not be a redundant expression; for example, “it is true” in the conjunction “‘the earth is round’ qualifies as information and it is true” could not be eliminated without semantic loss.

All these new issues are grafted to some old branches of the philosophical tree. Whether false information is a genuine type of information has important repercussions on any philosophy and pragmatics of communication.

The question about the informative nature (or lack thereof) of necessary truths, tautologies, equations or identity statements is an old one, as it runs through Hume, Kant, Frege and Wittgenstein. The latter, for example, interestingly remarked:

Another expression akin to those we have just considered is this: ‘Here it is; take it or leave it!’ And this again is akin to a kind of introductory statement which we sometimes make before remarking on certain alternatives, as when we say: ‘It either rains or it doesn’t rain; if it rains we’ll stay in my room, if it doesn’t . . .’. The first part of this sentence is no piece of information (just as ‘Take it or leave it’ is no order). Instead of, ‘It either rains or it doesn’t rain’ we could have said, ‘Consider the two cases . . .’. Our expression underlines these cases, presents them to your attention (*The Blue and Brown Books*, The Brown Book, II, p. 161, see Wittgenstein (1960)).

The solution of the problem of hyperintensionality (how one can draw a semantic distinction between expressions that are supposed to have the same meaning according to a particular theory of meaning that is usually model-theoretic or modal in character) depends on how one can make sense of the relation between truth and informativeness in the case of logically equivalent expressions.

Finally, the possibly redundant qualification of information as true is also linked with the critique of the deflationary theories of truth (DTT), since one could accept a deflationary T-schema as perfectly correct, while rejecting the explanatory adequacy of DTT. “It is true that” in “it is true that p ” could be redundant in view of the fact that there cannot be factual information that is not true, but DTT could mistake this linguistic or conceptual redundancy for unqualified dispensability. “It is true that” could be redundant because, strictly speaking, information is not a truth-bearer but already encapsulates truth as truthfulness. Thus, DTT may be satisfactory as theories of truth-ascriptions while being inadequate as theories of truthfulness.

Once information is available, knowledge can be built in terms of *justifiable* or *explainable semantic information*. An information agent knows that the battery is flat not by merely guessing rightly, but because e.g. it perceives that the red light of the low battery indicator flashing and/or that the engine does not start. In this sense, information provides the basis of any further scientific investigation. Note, however, that the fact that data may count as *resources* for (i.e. inputs an agent can use to construct) information, and hence for knowledge, rather than *sources*, may lead to constructionist arguments against mimetic theories that

interpret information as some sort of picture of the world. The point requires some elaboration.

Whether empirical or conceptual, data make possible only a certain range of information constructs, and not all constructs are made possible equally easily. An analogy may help here. Suppose one has to build a shelter. The design and complexity of the shelter may vary, but there is a limited range of “realistic” possibilities, determined by the nature of the available resources and constraints (size, building materials, location, weather, physical and biological environment, working force, technical skills, purposes, security, time constraints, etc.). Not any shelter can be built. And the type of shelter that will be built more often will be the one that is more likely to take close-to-optimal advantage of the available resources and constraints. The same applies to data. Data are at the same time the resources and constraints that make possible the construction of information. The best information is that better tuned to the constraining affordances available. Thus informational coherence and adequacy do not necessarily entail nor support naive or direct realism, or a correspondence theory of truth as this is ordinarily presented. Ultimately, information is the result of a process of data modelling; it does not have to represent or photograph or portray or photocopy, or map or show or uncover or monitor or . . . the intrinsic nature of the system analysed, no more than an igloo describes the intrinsic nature of snow or the Parthenon indicates the real properties of stones.

When *semantic content is false*, this is a case of *misinformation* (Fox (1983)). And if the source of misinformation is aware of its nature, one may speak of *disinformation*, as when one says to the mechanic “my husband forgot to turn the lights on”. Disinformation and misinformation are ethically censurable but may be successful in achieving their purpose: tell the mechanic that your husband left the lights on last night, and he will still be able to provide you with the right advice. Likewise, information may still fail to be successful; just imagine telling the mechanic that your car is out of order.

5 Philosophical Approaches to Semantic Information

What is the relation between MTC and the sort of semantic information that we have called factual? 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, reliability, usefulness or interpretation of information, but only in the level of detail and frequency in the uninterpreted data, being these symbols, signals or messages. Philosophical approaches differ from MTC in two main respects.

First, they 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 can semantic information be generated and flow?”, “how is information related to error,

truth and knowledge?”, “when is information useful?”. Wittgenstein, for example, remarks that “One is inclined to say: ‘Either it is raining, or it isn’t - how I know, how the information has reached me, is another matter.’ But then let us put the question like this: What do I call information that it is raining’? (Or have I only information of this information too?) And what gives this ‘information’ the character of information about something? Doesn’t the form of our expression mislead us here? For isn’t it a misleading metaphor to say: “My eyes give me the information that there is a chair over there”? (*Philosophical Investigations*, I. §356, see now Wittgenstein (2001)).

Second, philosophical theories of semantic information also seek to connect it to other relevant concepts of information and more complex forms of epistemic, mental and doxastic phenomena. For instance, Dretske (1981) and Barwise and Seligman (1997) attempt to ground information, understood as factual semantic contents, on environmental information. The approach is also known as the *naturalization of information*. A similar point can be made about Putnam’s twin earths argument, the externalization of semantics and teleosemantics.

Philosophical analyses usually adopt a propositional orientation and an epistemic outlook, endorsing, often implicitly, the prevalence or centrality of factual information within the map outlined in Fig. 1. They tend to base their analyses on cases such as “Paris is the capital of France” or “The Bodleian Library is in Oxford”. How relevant is MTC to similar researches?

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). According to MacKay, information is linked to an increase in knowledge on the receiver’s side:

Suppose we begin by asking ourselves what we mean by information.

Roughly speaking, we say that we have gained information when we know something now that we didn’t know before; when ‘what we know’ has changed. (Mackay (1969), p. 10).

Around the same years, 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, any philosophical theory of semantic-factual information is supposed to be *very strongly* constrained, perhaps even overdetermined, by MTC, somewhat as mechanical engineering is by Newtonian physics. Weavers optimistic interpretation of Shannons work is a typical example.

At the other extreme, any philosophical theory of semantic-factual information is supposed to be *only weakly* constrained, perhaps even completely underdetermined, by MTC, somewhat as 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 1950s generated earlier philosophical enthusiasm that has gradually cooled down through the decades. Historically, philosophical theories of semantic-factual information have moved from “very strongly constrained” to “only weakly constrained”. Recently, we find positions that carefully appreciate MTC for what it can provide in terms of a robust and well-developed statistical theory of correlations between states of different systems (the sender and the receiver) according to their probabilities. This can have important consequences in mathematically-friendly contexts, such as some approaches to naturalised epistemology (Harms (1998)) or scientific explanation (Badino (2004)).

Although the philosophy of semantic information has become increasingly autonomous from MTC, two important connections have remained stable between MTC and even the most recent philosophical accounts:

1. the communication model, explained in section 3.1 (see Fig. 6); and
2. what Barwise labelled the “Inverse Relationship Principle” (IRP).

The communication model has remained virtually unchallenged, even if nowadays theoretical accounts are more likely to consider as basic cases multiagent and distributed systems interacting in parallel, rather than individual agents related by simple, sequential channels of communication. In this respect, the philosophy of information (Floridi (2002);(2004a)) is less Cartesian than “social”.

IRP refers to the inverse relation between the probability of p - which may range over sentences of a given language (as in Bar-Hillel and Carnap) or events, situations or possible worlds (as in Dretske) - and the amount of semantic information carried by p (recall that Poe’s raven, as a unary source provides no information because its answers are entirely predictable). It states that information goes hand in hand with unpredictability. Popper (1935) is often credited as the first philosopher to have advocated IRP explicitly. However, systematic attempts to develop a formal calculus involving it were made only after Shannons breakthrough.

We have seen that 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 (1953) (see also Bar-Hillel (1964)) and further developed by Kemeny (1953), Smokler (1966), Hintikka and Suppes (1970) and Dretske (1981). 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 :

$$CONT(p) = 1 - P(p) \tag{10}$$

$CONT$ 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 $P(p) = 1 - CONT(p)$:

$$INF(p) = \log \frac{1}{1 - CONT(p)} = -\log P(p) \quad (11)$$

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 (1953), 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 the observed states of affairs (s), that is:

$$I(s) = -\log P(s) \quad (12)$$

The *modal approach* further 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, it is less ontologically demanding than the modal approach, since it assumes a clearly limited domain of application. It is also 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 theorists require some presence of information already immanent in the environment (*environmental information*), as nomic regularities or constraints. This "semantic externalism" can be controversial.

The *inferential approach* defines information in terms of entailment space: information depends on valid inference relative to an information agent's theory or epistemic state.

Each of the previous extensionalist 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. Wittgenstein addressed this distinction in his *Remarks on the Philosophy of Psychology* I §817 (Wittgenstein (1980))

The important insight is that there is a language-game [Wittgenstein seems to have in mind here the information game we have already encountered above] in which I produce information automatically, information which can be treated by other people quite as they treat non-automatic information - only here there will be no question of any 'lying' - information which I myself may receive like that of a third person. The

‘automatic’ statement, report etc. might also be called an ‘oracle’. - But of course that means that the oracle must not avail itself of the words ‘I believe ...’.

5.1 The Bar-Hillel-Carnap Paradox

Insofar as they subscribe to the Inverse Relationship Principle, the extensionalist approaches outlined in the previous section can be affected by what has been defined, with a little hyperbole, as the *Bar-Hillel-Carnap Paradox* (BCP, Floridi (2004b)).

In a nutshell, we have seen that, following IRP, the less probable or possible p is the more semantic information p is assumed to be carrying. This explains why most philosophers agree that tautologies convey no information at all, for their probability or possibility is 1. But it also leads one to consider contradictions - which describe impossible states, or whose probability is 0 - as the sort of messages that contain the highest amount of semantic information. It is a slippery slope. Make a statement less and less likely and you gradually increase its informational content, but at certain point the statement “implodes” (in the quotation below, it becomes “too informative to be true”).

Bar-Hillel and Carnap (1953) were among the first to make explicit this *prima facie* counterintuitive inequality. Note how their careful wording betrays the desire to defuse the problem:

BCP)

It might perhaps, at first, seem strange that a self-contradictory sentence, hence one which no ideal receiver would accept, is regarded as carrying with it the most inclusive information. It should, however, be emphasized that semantic information is here not meant as implying truth. A false sentence which happens to say much is thereby highly informative in our sense. Whether the information it carries is true or false, scientifically valuable or not, and so forth, does not concern us. A self-contradictory sentence asserts too much; it is too informative to be true (p. 229).

Since its formulation, BCP has been recognised as an unfortunate, yet perfectly correct and logical inevitable consequence of any quantitative *theory of weakly semantic information* (TWSI; “weakly” because truth values play no role in it). As a consequence, the problem has often been either ignored or tolerated (Bar-Hillel and Carnap (1953)) as the price of an otherwise valuable approach. Sometimes, however, attempts have been made to circumscribe its counterintuitive consequences. This has happened especially in Information Systems Theory (Winder et al. (1997)) - where consistency is an essential constraint that must remain satisfied for a database to preserve data integrity - and in Decision Theory, where inconsistent information is obviously of no use to a decision maker.

In these cases, whenever there are no possible models that satisfy a statement or a theory, instead of assigning to it the maximum quantity of semantic information, three strategies have been suggested:

1. assigning to all inconsistent cases the same, infinite information value (Lozinski (1994)). This is in line with an economic approach, which defines x as impossible if and only if x has an infinite price;
2. eliminating all inconsistent cases a priori from consideration, as impossible outcomes in decision-making (Jeffrey (1990)). This is in line with the syntactic approach developed by MTC;
3. assigning to all inconsistent cases the same zero information value (Mingers (1997), Aisbett and Gibbon (1999)).

The latter approach is close to the *strongly semantic approach*, to which we shall now turn.

5.2 The Strongly Semantic Approach to Information

The general hypothesis is that BCP indicates that something has gone essentially amiss with TWSI. TWSI is based on a semantic principle that is too weak, namely that truth-values are independent of semantic information. A semantically stronger approach, according to which information encapsulates truth, can avoid the paradox and is more in line with the ordinary conception of what generally counts as factual information, as we have seen in section 4.2.3. MTC already provides some initial reassurance. MTC identifies the quantity of information associated with, or generated by, the occurrence of a signal (an event or the realisation of a state of affairs) with the elimination of possibilities (reduction in uncertainty) represented by that signal (event or state of affairs). In MTC, no counterintuitive inequality comparable to BCP occurs, and the line of argument is that, as in the case of MTC, a *theory of strongly semantic information* (TSSI), based on alethic and discrepancy values rather than probabilities, can also successfully avoid BCP (Floridi (2004b); (2005b), see Bremer and Cohnitz (2004) chap. 2 for an overview). The idea is to define semantic-factual information in terms of data space, as well-formed, meaningful and truthful data. This constrains the probabilistic approach introduced above, by requiring first a qualification of the content as truthful. Once the content is so qualified, the quantity of semantic information in p is calculated in terms of distance of p from the situation/resource w that p is supposed to model. Total distance is equivalent to a p true in all cases (all possible worlds or probability 1), including w and hence minimally informative, whereas maximum closeness is equivalent to the precise modelling of w at the agreed level of abstraction.

Suppose there will be exactly three guests for dinner tonight. This is our situation w . Imagine we 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 some semantic-factual information 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 now 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 of a Cartesian diagram the values of ϑ given a specific σ and a corresponding target w . In our example, we know that $\vartheta(T) = 1$ and $\vartheta(P) = 0$. For the sake of simplicity, let us assume that $\vartheta(V) = 0.25$ (see Floridi (2004b) on how to calculate ϑ values). We now need a formula to calculate the *degree of informativeness* ι of σ in relation to $\vartheta(\sigma)$. It can be shown that the most elegant solution is provided by the complement of the square value of $\vartheta(\sigma)$, that is $y = 1 - x^2$. Using the symbols just introduced, we have:

$$\iota(\sigma) = 1 - \vartheta(\sigma)^2 \tag{13}$$

Fig. 14 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).

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

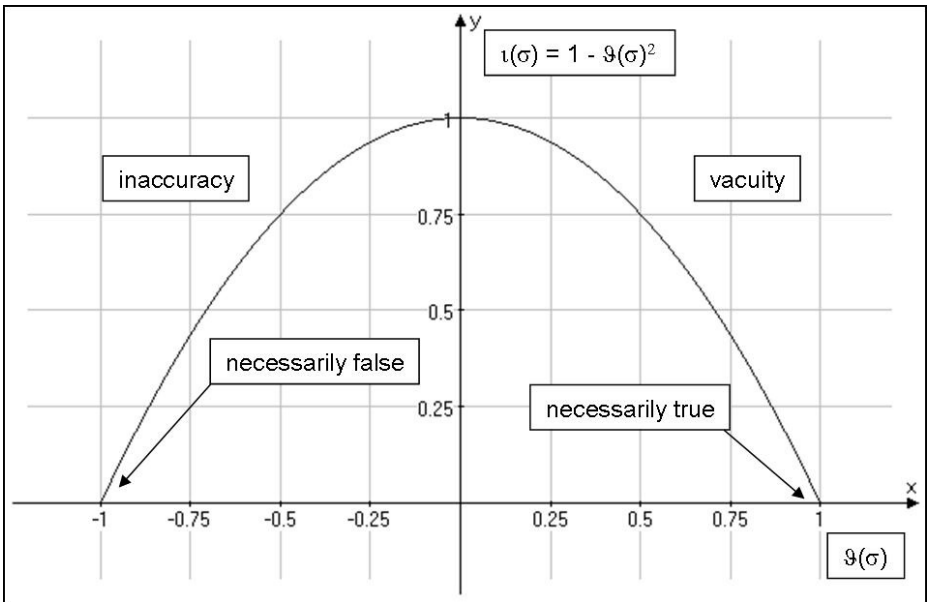


Fig. 14. Degree of informativeness

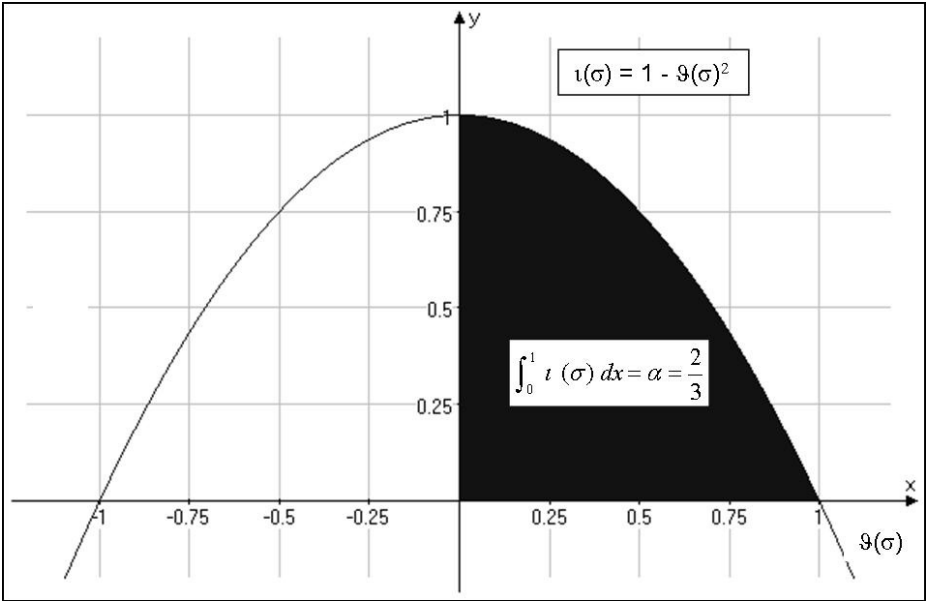


Fig. 15. Maximum amount of semantic information α carried by σ

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 $l(\sigma)$ we need to calculate the area delimited by equation [13], that is, the definite integral of the function $l(\sigma)$ on the interval $[0, 1]$. As we know, the maximum quantity of semantic information (call it α) is carried by P, whose $\vartheta = 0$. This is equivalent to the whole area delimited by the curve. Generalising to σ we have:

$$\int_0^1 l(\sigma) dx = \alpha = \frac{2}{3} \tag{14}$$

Fig. 15 shows the graph generated by equation [14]. The shaded area is the maximum amount of semantic information α carried by σ .

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 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

$$\int_0^{\vartheta} \iota(\sigma) dx = \beta \tag{15}$$

Since $\vartheta(V) = 0.25$, we have

$$\int_0^{0.25} \iota(\sigma) dx = 0.24479 \tag{16}$$

Fig. 16 shows the graph generated by equation [16]. 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. 16. More generally, and expressed in bits, the amount of semantic information γ in σ is:

$$\gamma(\sigma) = \log(\alpha - \beta) \tag{17}$$

Note the similarity between [14] and [15]. When $\vartheta(\sigma) = 1$, that is, when the distance between σ and w is maximum, then $\alpha = \beta$ and $\gamma(\sigma) = 0$. This is what happens when we consider T . T is so distant from w as to contain only vacuous information. In other words, T contains as much vacuous information as P contains relevant information.

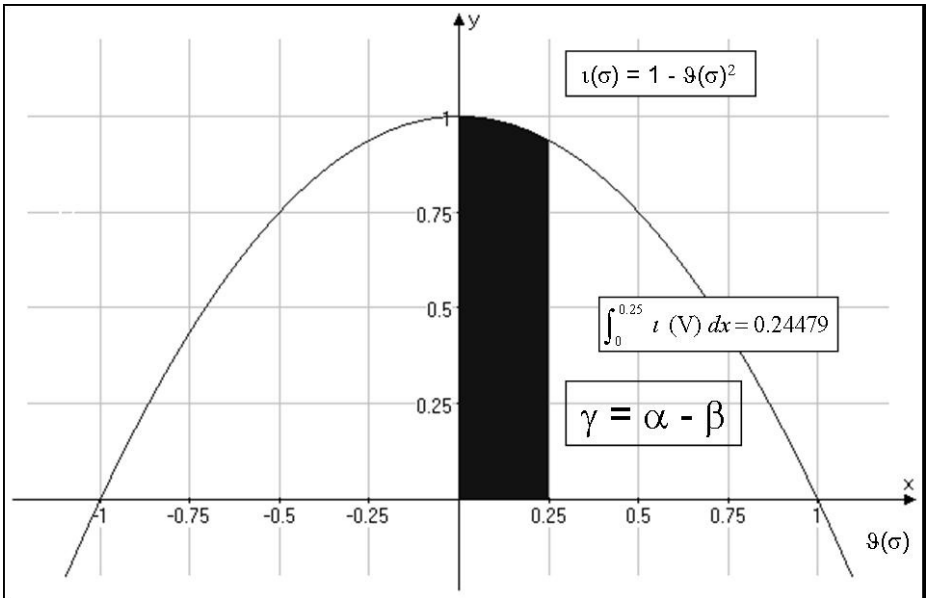


Fig. 16. Amount of semantic information γ carried by σ

6 Conclusion

Philosophical theories of semantic information have recently contributed to a new area of research in itself, the philosophy of information (Adams (2003), Floridi (2002), (2003b), (2004a)). The two special issue volumes of *Minds and Machines* on the philosophy of information (Floridi (2003c)) provide an overview of the scope and depth of current work in the field. Information seems to have become a key concept to unlock several philosophical problems. “The most valuable commodity I know of is information”, boldly declares Gordon Gekko in Oliver Stones *Wall Street* (1987). Euphranor would probably have concurred. The problem is that we still have to agree about what information is exactly.

Acknowledgements

This article is based on Floridi (2005a) and I am grateful to Ed Zalta and the Editorial Board of the *Stanford Encyclopedia of Philosophy* for permission to reproduce the original text. I benefited enormously from many insightful editorial comments by Fred Kroon, Jerry Seligman and Ed Zalta on previous drafts. I am also very grateful to several colleagues and friends for their helpful suggestions and conversations on previous drafts and past papers on which this article is based. They are responsible only for the improvements not for any remaining mistake: Frederick R. Adams, Mark Bedau, John Collier, Ian C. Denger, Michael Dunn, Roger Brownsword, Timothy Colburn, James Fetzer, Phil Fraundorf, Gian Maria Greco, Ken Herold, Bernard Katz, Philipp Keller, Gianluca Paronitti, Jeff Sanders, Sebastian Sequoiah-Grayson, Janet D. Sisson, Giovanni Sommaruga, Ernest Sosa, J. L. Speranza and Matteo Turilli. Finally, I would like to thank all the participants to the Workshop on information theories, and the organizers, Juerg Kohlas and Giovanni Sommaruga.

References

- Adams, F.: The Informational Turn in Philosophy *Minds and Machines* 13(4), 471–501 (2003)
- Aisbett, J., Gibbon, G.: A practical measure of the information in a logical theory. *Journal of Experimental and Theoretical Artificial Intelligence* 11(2), 201–218 (1999)
- Armstrong, D.M.: *A Materialist Theory of the Mind*. London Routledge and Kegan Paul (1968)
- Armstrong, D.M.: *A Materialist Theory of the Mind*, 2nd edn. London Routledge (1993)
- Badino, M.: An Application of Information Theory to the Problem of the Scientific Experiment *Synthese* 140, 355–389 (2004)
- Bar-Hillel, Y.: *Language and Information: Selected Essays on Their Theory and Application*. Addison-Wesley, Reading (1964)
- Bar-Hillel, Y., Carnap, R.: An Outline of a Theory of Semantic Information (1953); repr. in Bar-Hillel, 221–274 (1964)

- Barwise, J., Seligman, J.: *Information Flow: The Logic of Distributed Systems*. Cambridge University Press, Cambridge (1997)
- Bateson, G.: *Steps to an Ecology of Mind*, Frogmore, St. Albans, Paladin (1973)
- Berkeley, G.: *Alciphron, Or the Minute Philosopher*, 1732, Edinburgh 1948-57 Thomas Nelson
- Braman, S.: *Defining Information Telecommunications Policy* 13, 233–242 (1989)
- Bremer, M.E.: *Do Logical Truths Carry Information?* *Minds and Machines* 13(4), 567–575 (2003)
- Bremer, M., Cohnitz, D.: *Information and Information Flow - an Introduction*. Ontos Verlag, Frankfurt, Lancaster (2004)
- Chaitin, G.J.: *Algorithmic Information Theory*. Cambridge University Press, Cambridge (1987)
- Chalmers, D.J.: *The Conscious Mind: In Search of a Fundamental Theory*. Oxford Univ. Press, New York (1996)
- Cherry, C.: *On Human Communication: A Review, a Survey, and a Criticism*, 3rd edn. MIT Press, Cambridge (1978)
- Colburn, T.R.: *Philosophy and Computer Science*. M.E. Sharpe, Armonk (2000)
- Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. Wiley, Chichester (1991)
- Craver, C.F.: *A Field Guide to Levels Proceedings and Addresses of the American Philosophical Association* 77(3) (2004)
- Craver, C.F.: *Explaining the Brain: A Mechanist's Approach* (forthcoming)
- Debons, A., Cameron, W.J. (eds.): *Perspectives in Information Science: Proceedings of the Nato Advanced Study Institute on Perspectives in Information Science*, Aberystwyth, Wales, Uk, August 13-24, 1973. Leiden, Noordhoff (1975)
- Dennett, D.C.: *Content and Consciousness*. Routledge and Kegan Paul, London (1969)
- Dennett, D.C.: *Intentional Systems*. *The Journal of Philosophy* 68, 87–106 (1971)
- Dennett, D.C.: *Content and Consciousness*, 2nd edn. Routledge and Kegan Paul, London (1986)
- Deutsch, D.: *Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer Proceedings of the Royal Society*. 400, 97–117 (1985)
- Deutsch, D.: *The Fabric of Reality*. Penguin, London (1997)
- Devlin, K.J.: *Logic and Information*. Cambridge University Press, Cambridge (1991)
- Di Vincenzo, D.P., Loss, D.: *Quantum Information Is Physical, Superlattices and Microstructures - Special issue on the occasion of Rolf Landauer's 70th birthday* 23, 419–432 (1998)
- Dodig-Crnkovic, G.: *System Modeling and Information Semantics*. In: Bubenko, J., Eriksson, O., Fernlund, H., Lind, M. (eds.) *Proceedings of the Fifth Promote IT Conference*, Borlänge, Sweden, Studentlitteratur, Lund (2005)
- Dretske, F.I.: *Knowledge and the Flow of Information*. Blackwell, Oxford (1981); CSLI Publications, Stanford (reprinted, 1999)
- Dunn, J.M.: *The Concept of Information and the Development of Non-Classical Logics Non-Classical Approaches in the Transition from Traditional to Modern Logic*, Stelzner, W. (ed.): pp. 423–448. de Gruyter, Berlin
- Fetzer, J.H.: *Information, Misinformation, and Disinformation Minds and Machines* 14(2), 223–229 (2004)
- Floridi, L.: *Philosophy and Computing: An Introduction*. Routledge, London (1999)
- Floridi, L.: *What Is the Philosophy of Information?* *Metaphilosophy* 33(1-2), 123–145 (2002)
- Floridi, L. (ed.): *Information in The Blackwell Guide to the Philosophy of Computing and Information*, pp. 40–61. Blackwell, Oxford (2003a)

- Floridi, L.: Two Approaches to the Philosophy of Information. *Minds and Machines* 13(4), 459–469 (2003)
- Floridi, L.: Open Problems in the Philosophy of Information. *Metaphilosophy* 35(4), 554–582 (2004)
- Floridi, L.: Outline of a Theory of Strongly Semantic Information. *Minds and Machines* 14(2), 197–222 (2004)
- Floridi, L.: Information, Semantic Conceptions Of, *Stanford Encyclopedia of Philosophy*, Zalta, E.N. (ed.) (2005), <http://plato.stanford.edu/entries/information-semantic/>
- Floridi, L.: Is Information Meaningful Data? *Philosophy and Phenomenological Research* 70(2), 351–370 (2005)
- Floridi, L. (ed.): *Minds and Machines* (Special issue of the journal on The Philosophy of Information) (2003)
- Floridi, L., Sanders, J.W.: *Levellism and the Method of Abstraction* (forthcoming)
- Fox, C.J.: *Information and Misinformation: An Investigation of the Notions of Information, Misinformation, Informing, and Misinforming*. Greenwood Press, Westport (1983)
- Franklin, S.: *Artificial Minds*. MIT Press, Cambridge (1995)
- Frieden, B.R.: *Physics from Fisher Information: A Unification*. Cambridge University Press, Cambridge (1998)
- Frieden, B.R.: *Science from Fisher Information: A Unification*, 2nd edn. Cambridge University Press, Cambridge (2004)
- Golan, A.: Information and Entropy Econometrics - Editor's View. *Journal of Econometrics* 107(1-2), 1–15 (2002)
- Graham, G.: *The Internet: A Philosophical Inquiry*. Routledge, London (1999)
- Grice, H.P.: *Studies in the Way of Words*. Harvard University Press, Cambridge (1989)
- Hanson, P.P. (ed.): *Information, language, and cognition*. University of British Columbia Press, Vancouver (1990)
- Harms, W.F.: The Use of Information Theory in Epistemology. *Philosophy of Science* 65(3), 472–501 (1998)
- Heil, J.: Levels of Reality, *Ratio* 16(3), 205–221 (2003)
- Hintikka, J., Suppes, P. (eds.): *Information and Inference*. Reidel, Dordrecht (1970)
- Hoare, C.A.R., Jifeng, H.: *Unifying Theories of Programming*. Prentice-Hall, London (1998)
- Hume, D.: *Essays, Moral, Political, and Literary* 1987, Indianapolis, Liberty Classics. Edited and with a foreword, notes, and glossary by Eugene F. Miller; with an apparatus of variant readings from the, edition by T.H. Green and T.H. Grose. Based on the 1777 edn. originally published as v. 1 of *Essays and treatises on several subjects* (1987)
- Jeffrey, R.C.: *The Logic of Decision*, 2nd edn. University of Chicago Press, Chicago (1990)
- Jones, D.S.: *Elementary Information Theory*. Clarendon Press, Oxford (1979)
- Kemeny, J.: A Logical Measure Function. *Journal of Symbolic Logic* 18, 289–308 (1953)
- Landauer, R.: *Computation: A Fundamental Physical View*. *Physica Scripta* 35, 88–95 (1987)
- Landauer, R.: Information Is Physical. *Physics Today* 44, 23–29 (1991)
- Landauer, R.: The Physical Nature of Information. *Physics Letter A* 217, 188 (1996)
- Landauer, R., Bennett, C.H.: The Fundamental Physical Limits of Computation. *Scientific American*, 48–56 (July 1985)

- Larson, A.G., Debons, A.: *Information Science in Action: System Design*. In: *Proceedings of the Nato Advanced Study Institute on Information Science, Crete, Greece, August 1-11*, M. Nijhoff, The Hague (1983)
- Losee, R.M.: *A Discipline Independent Definition of Information*. *Journal of the American Society for Information Science* 48(3), 254–269 (1997)
- Lozinskii, E.: *Information and evidence in logic systems*. *Journal of Experimental and Theoretical Artificial Intelligence* 6, 163–193 (1994)
- Mabon, P.C.: *Mission Communications: The Story of Bell Laboratories*. Bell Telephone Laboratories, Murray Hill (1975)
- Machlup, F., Mansfield, U. (eds.): *The Study of Information: Interdisciplinary Messages*. Wiley, Chichester (1983)
- MacKay, D.M.: *Information, Mechanism and Meaning*. MIT Press, Cambridge (1969)
- Marr, D.: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, San Francisco (1982)
- Mingers, J.: *The Nature of Information and Its Relationship to Meaning*. In: Winder, R.L., et al. (eds.) *Philosophical Aspects of Information Systems*, pp. 73–84. Taylor and Francis, London (1997)
- Nauta, D.: *The Meaning of Information*. The Hague, Mouton (1972)
- Newell, A.: *The Knowledge Level*. *Artificial Intelligence* 18, 87–127 (1982)
- Newell, A., Simon, H.A.: *Computer Science as Empirical Inquiry: Symbols and Search*. *Communications of the ACM* 19, 113–126 (1976)
- Pierce, J.R.: *An Introduction to Information Theory: Symbols, Signals and Noise*, 2nd edn. Dover Publications, New York (1980)
- Poli, R.: *The Basic Problem of the Theory of Levels of Reality*. *Axiomathes* 12, 261–283 (2001)
- Popper, K.R.: *Logik Der Forschung: Zur Erkenntnistheorie Der Modernen Naturwissenschaft, 1935*, Wien, J. Springer. Eng. tr. *The Logic of Scientific Discovery*, London, Hutchinson (1959)
- Roever, W.P.d., Engelhardt, K., Buth, K.-H.: *Data Refinement: Model-Oriented Proof Methods and Their Comparison*. Cambridge University Press, Cambridge (1998)
- Sayre, K.M.: *Cybernetics and the Philosophy of Mind*. Routledge and Kegan Paul, London (1976)
- Schaffer, J.: *Is There a Fundamental Level?* *Nous* 37(3), 498–517 (2003)
- Shannon, C.E.: *Collected Papers*. In: Sloane, N.J.A., Wyner, A.D. (eds.). IEEE Press, New York (1993)
- Shannon, C. E., Weaver, W.: *The Mathematical Theory of Communication*, 1949 rep., Urbana, University of Illinois Press. Foreword by Richard E. Blahut and Bruce Hajek (1998)
- Simon, H.A.: *The Sciences of the Artificial*, 1st edn. MIT Press, Cambridge (1969)
- Simon, H.A.: *The Sciences of the Artificial*, 3rd edn. MIT Press, Cambridge (1996)
- Sloman, A.: *The Computer Revolution in Philosophy: Philosophy, Science and Models of Mind*, Hassocks, Harvester (1978)
- Smokler, H.: *Informational Content: A Problem of Definition*. *The Journal of Philosophy* 63(8), 201–211 (1966)
- Steane, A.M.: *Quantum Computing*. *Reports on Progress in Physics* 61, 117–173 (1998)
- Thagard, P.R.: *Comment: Concepts of Information in 1990*
- Weaver, W.: *The Mathematics of Communication*. *Scientific American* 181(1), 11–15 (1949)
- Wheeler, J.A.: *Information, Physics, Quantum: The Search for Links, Complexity, Entropy, and the Physics of Information*. In: Zureck, W.H. (ed.). Addison Wesley, Redwood City (1990)

- Wiener, N.: *The Human Use of Human Beings: Cybernetics and Society*, 1st edn. (1954); 2nd edn., London, reissued with a new introduction by Steve J. Heims, London, Free Association (1989)
- Wiener, N.: *Cybernetics or Control and Communication in the Animal and the Machine*, 2nd edn. MIT Press, Cambridge (1961)
- Winder, R.L., Probert, S.K., Beeson, I.A.: *Philosophical Aspects of Information Systems*. Taylor & Francis, London (1997)
- Wittgenstein, L.: *Preliminary Studies for the Philosophical Investigations: Generally Known as the Blue and Brown Books*, 2nd edn. Basil Blackwell, Oxford (1960)
- Wittgenstein, L.: *Remarks on the Philosophy of Psychology*, vol. 2, University of Chicago Press, Chicago. Basil Blackwell, Oxford (1980); Edited by G. E. M. Anscombe and G. H. von Wright; translated by G. E. M. Anscombe; vol. 2 edited by G.H. von Wright and H. Nyman; translated by C.G. Luckhardt and A.E. Aue
- Wittgenstein, L.: *Zettel*, 2nd edn. Blackwell, Oxford (1981); Edited by G.E.M. Anscombe and G.H. von Wright; translated by G.E.M. Anscombe
- Wittgenstein, L.: *Philosophical Investigations: The German Text with a Revised English Translation*, 3rd edn. Blackwell, Oxford (2001); Translated by G.E.M. Anscombe. Incorporates final revisions made by Elizabeth Anscombe to her English edition. Some typesetting errors have been corrected, and the text has been repaginated