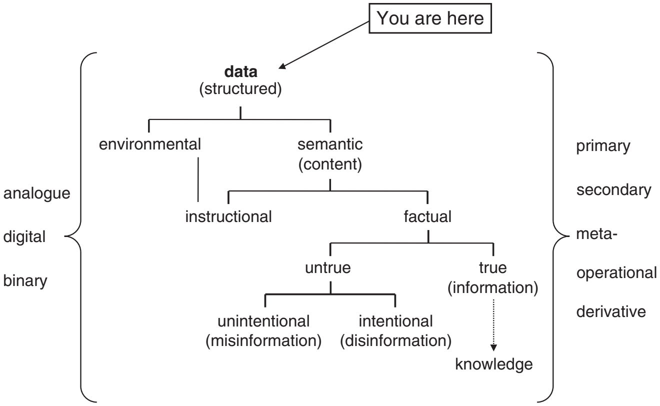
Chapter 2

The language of information

Information is a conceptual labyrinth, and in this chapter we shall look at its general map, with the purpose of finding our bearings. Figure 2 summarizes the main distinctions that are going to be introduced. Some areas will be explored in more depth in the following chapters.

Navigating through the various points in the map will not make for a linear journey, so using a few basic examples to illustrate the less obvious steps will help to keep our orientation. Here is one to which we shall return often.

It is Monday morning. John turns on the ignition key of his car, but nothing happens: the engine does not even cough. The silence of the engine worries him. Looking more carefully, he notices that the low-battery indicator is flashing. After a few more unsuccessful attempts, he gives up and calls the garage. Over the phone, he explains that, last night, his wife forgot to switch off the car's lights – it is a lie, John did, but he is too ashamed to admit it – and now the battery is flat. The mechanic tells John that he should look at the car's operation manual, which explains how to use jump leads to start the engine. Luckily, John's neighbour has everything he needs. He reads the manual, looks at the illustrations, speaks to his



2. A map of information concepts

neighbour, follows the instructions, solves the problem, and finally drives to the office.

This everyday episode will be our 'fruit fly', as it provides enough details to illustrate the many ways in which we understand information. Our first step will now be to define information in terms of data.

The data-based definition of information

Over the past decades, it has become common to adopt a *General Definition of Information* (GDI) in terms of *data* + *meaning*. GDI has become an operational standard, especially in fields that treat data and information as reified entities, that is, stuff that can be manipulated (consider, for example, the now common expressions 'data mining' and 'information management'). A straightforward way of formulating GDI is as a tripartite definition (Table 1):

According to (GDI.1), information is made of data. In (GDI.2), 'well formed' means that the data are rightly put together,

Table 1. The General Definition of Information (GDI)

GDI) σ is an instance of information, understood as semantic content, if and only if:
GDI.1) σ consists of n *data*, for $n \geq 1$;
GDI.2) the data are *well formed*;
GDI.3) the well-formed data are *meaningful*.

according to the rules (*syntax*) that govern the chosen system, code, or language being used. Syntax here must be understood broadly, not just linguistically, as what determines the form, construction, composition, or structuring of something. Engineers, film directors, painters, chess players, and gardeners speak of syntax in this broad sense. In our example, the car's operation manual may show a two-dimensional picture of how to jump-start a car. This pictorial syntax (including the linear perspective that represents space by converging parallel lines) makes the illustration potentially meaningful to the user. Still relying on the same example, the actual battery needs to be connected to the engine in a correct way to function properly: 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 John carries on with his 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. Once again, semantic information is not necessarily linguistic. For example, in the case of the car's operation manual, the illustrations are supposed to be visually meaningful to the reader.

How data can come to have an assigned meaning and function in a semiotic system like a natural language is one of the hardest questions in semantics, known as the *symbol grounding problem*. Luckily, it can be disregarded here. The only point worth clarifying is that data constituting information can be meaningful independently of an informee. Consider the following example. The Rosetta Stone contains three translations of a single passage, in Egyptian hieroglyphic, Egyptian Demotic, and classical Greek languages. Before its discovery, Egyptian hieroglyphics were already regarded as information, even if their meaning 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 reasonable sense in which one may speak of meaningful data being embedded in information-carriers independently of any informee. It is very different from the stronger thesis, according to which data could also have their own semantics independently of an intelligent *producer/informer*. This is also known as *environmental information*, but, before discussing it, we need to understand much better the nature of data.

Understanding data

A good way to uncover the most fundamental nature of data is by trying to understand what it means to erase, damage, or lose them. Imagine the page of a book written in a language unknown to us. Suppose the data are in the form of pictograms. The regular patterns suggest the compliance with some structural syntax. We have all the data, but we do not know their meaning, hence we have no information yet. Let us now erase half of the pictograms. One may say that we have halved the data as well. If we continue in this process, when we are left with only one pictogram we might be tempted to say that data require, or may be identical with, some sort of representations. But now let us erase that last pictogram too. We are left with a white page, and yet not entirely without data. For the presence of a white page is still a datum, as long as there is a difference between the white page and the page on which something is or could be written. Compare this to the common phenomenon of 'silent assent': silence, or the lack of perceivable data, can be as much a datum as the presence of some noise, exactly like the zeros of a binary system. Recall in our example John's concern when he did not hear any sound coming from his car's engine. That lack of noise was informative. The fact is that a genuine, complete erasure of all data can be achieved only by the elimination of all possible differences. This clarifies why a datum is ultimately reducible to a *lack of uniformity*. Donald MacCrimmon MacKay (1922–1987) highlighted this important point when he wrote that 'information is a distinction that makes a difference'. He was followed by Gregory Bateson (1904–1980), whose slogan is better known, although less accurate: 'In fact, what we mean by information – the elementary unit of information – is a difference which makes a difference'. More formally, according to the *diaphoric interpretation* (*diaphora* is the Greek word for 'difference'), the general definition of a datum is:

> Dd) datum $=_{\text{def.}}$ x being distinct from y, where x and y are two uninterpreted variables and the relation of 'being distinct', as well as the domain, are left open to further interpretation.

This definition of data can be applied in three main ways.

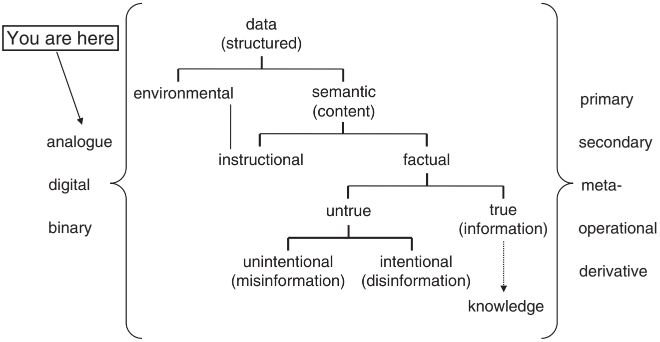
First, data can be lacks of uniformity in the real world. There is no specific name for such 'data in the wild'. One may refer to them as *dedomena*, that is, 'data' in Greek (note that our word 'data' comes from the Latin translation of a work by Euclid entitled *Dedomena*). Dedomena are not to be confused with *environmental information*, which will be discussed later in this chapter. They are pure data, that is, data before they are interpreted or subject to cognitive processing. They are not experienced directly, but their presence is empirically inferred from, and required by, experience, since they are what has to be there in the world for our information to be possible at all. So dedomena are whatever lack of uniformity in the world is the source of (what looks to informational organisms like us as) data, e.g. a red light against a dark background. I shall return to this point in Chapter 5, where we shall see that some researchers have been able to accept the thesis that there can be no information without data while rejecting the thesis that information must have a material nature.

Second, data can be lacks of uniformity between (the perception of) at least two physical states of a system or *signals*. Examples include 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.

Finally, data can be lacks of uniformity between two *symbols*, for example the letters B and P in the Latin alphabet.

Depending on one's interpretation, the *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 syntactically well-formed data, and of data on the occurrence of differences variously implementable physically, explain why information can



3. Analogue, digital, and binary data

so easily be decoupled from its support. The actual *format*, *medium*, and *language* in which data, and hence information, are encoded is often irrelevant and disregardable. In particular, the same data/information may be printed on paper or viewed on a screen, codified in English or in some other language, expressed in symbols or pictures, be analogue or digital. The last distinction is the most important and deserves some clarification.

Analogue versus digital data

Analogue data and the systems that encode, store, process, or transmit them vary continuously. For example, vinyl records are analogue because they store mechanical, continuous data that correspond to the recorded sounds. On the contrary, digital data and the related systems vary discretely between different states, e.g. on/off or high/low voltage. For example, compact discs are digital because they store sounds by transforming them as series of pits (indentations) and lands (the areas between pits). They *encode* rather than just *record* information.

Our understanding of the universe is firmly based not only on digital, discrete, or grainy ideas – the natural numbers, the heads or tails of a coin, the days of the week, the goals scored by a football team, and so forth – but also on many analogue, continuous, or smooth ideas – the intensity of a pain or pleasure, the real numbers, continuous functions, differential equations, waves, force fields, the continuum of time. Computers are usually seen as digital or discrete information systems, but this is not entirely correct, for two reasons. As Turing himself remarked,

> The digital computers [. . .] may be classified amongst the 'discrete state machines', these are the machines which move by sudden jumps or clicks from one quite definite state to another. These states are sufficiently different for the possibility of confusion between them to be ignored. Strictly speaking there are no such machines.

> Everything really moves continuously. But there are many kinds of machine, which can profitably be thought of as being discrete state machines.

And there are analogue computers. These perform calculations through the interaction of continuously varying physical phenomena, such as the shadow cast by the gnomon on the dial of a sundial, the approximately regular flow of sand in an hourglass or of water in a water clock, and the mathematically constant swing of a pendulum. Clearly, it is not the use of a specific substance or reliance on a specific physical phenomenon that makes an information system analogue, but the fact that its operations are directly determined by the measurement of continuous, physical transformations of whatever solid, liquid, or gaseous matter is employed. There are analogue computers that use continuously varying voltages and a Turing machine (the logically idealized model of our personal computers) is a digital computer but may not be electrical. Given their physical nature, analogue computers operate in real time (i.e. time corresponding to time in the real world) and therefore can be used to monitor and control events as they happen, in a 1:1 relation between the time of the event and the time of computation (think of the hourglass). However, because of their nature, analogue computers cannot be general-purpose machines but can only perform as necessarily specialized devices. The advantage is that analogue data are highly resilient: a vinyl record can be played again and again, even if it is scratched.

Binary data

Digital data are also called binary data because they are usually encoded by means of combinations of only two symbols called *bits* (*b*inary dig*its*), as strings of 0s and 1s comparable to the dots and dashes in the Morse code. For example, in binary notation the number three is written 11 (see Table 2). Since the value of any position in a binary number increases by the power of 2 (doubles) with each move from right to left (i.e.. . . 16, 8, 4, 2, 1; note that it could have been 1, 2, 4, 8, 16, and so on, but the binary system pays due homage to the Arabic language and moves from right to left) 11 means $(1 \times 2) + (1 \times 1)$, which adds up to three in the decimal system. Likewise, if one calculates the binary version of 6, equivalent to $(1 \times 4) + (1 \times 2) + (0 \times 1)$ one can see that it can only be 110.

Table 2. Decimal and binary notations of positive integers

		Decimal Notation			
	. . .	$10^3 = 1000$	$10^2 = 100$	$10^1 = 10$	$10^0 = 1$
one apple					1
two apples					2
. . .					
six apples					6
. . .					
thirteen apples				1	3
. . .					
		Binary Notation			
	. . .	$2^3 = 8$	$2^2 = 4$	$2^1 = 2$	$2^0 = 1$
one apple					1
two apples				1	0
. . .					
six apples			1	1	0
. . .					
thirteen apples		1	1	0	1
. . .					

A *bit* is the smallest unit of information, nothing more than the presence or absence of a signal, a 0 or a 1. A series of 8 bits forms a *byte* (*by* eigh*t*), and by combining bytes it becomes possible to generate a table of 256 (2^8) characters. Each character of data can then be stored as a pattern of 8 bits. The most widely used binary code is known as ASCII (American Standard Code for Information Interchange), which relies on only 7 bits out of 8 and therefore consists of a table of 128 (2^7) characters. Here is how a computer spells 'GOD' in binary: 010001110100111101000100 (Table 3):

Table 3. Example of binary encoding

G	off = 0	on = 1	off = 0	off = 0	off = 0	on = 1	on = 1	on = 1
O	off = 0	on = 1	off = 0	off = 0	on = 1	on = 1	on = 1	on = 1
D	off = 0	on = 1	off = 0	off = 0	off = 0	on = 1	off = 0	off = 0

Quantities of bytes are then calculated according to the binary system:

- 1 Kilobyte (KB) = 2^{10} = 1,024 bytes
- 1 Megabyte (MB) = 2^{20} = 1,048,576 bytes
- 1 Gigabyte (GB) = 2^{30} = 1,073,741,824 bytes
- 1 Terabyte (TB) = 2^{40} = 1,099,511,627,776 bytes

and so forth.

This is why the precise size of the random-access memory (RAM) of a computer, for example, is never a round number.

The binary system of data encoding has at least three advantages. First, bits can equally well be represented semantically (meaning True/False), logico-mathematically (standing for 1/0), and physically (transistor = On/Off; switch = Open/Closed; electric circuit = High/Low voltage; disc or tape = Magnetized/

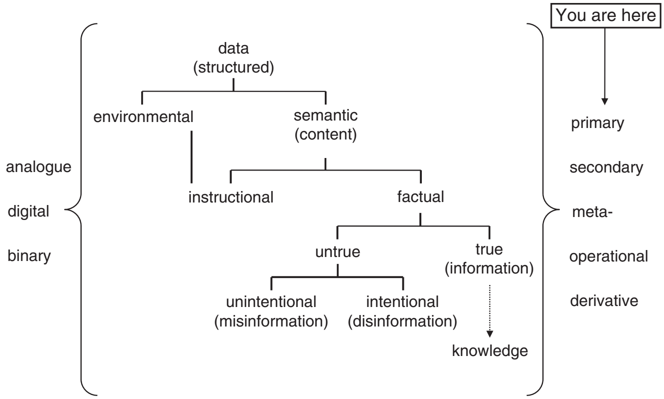
Unmagnetized; CD = presence/absence of pits, etc.), and hence provide the common ground where semantics, mathematical logic, and the physics and engineering of circuits and information theory can converge.

This means (second advantage) that it is possible to construct machines that can recognize bits physically, behave logically on the basis of such recognition, and therefore manipulate data in ways which we find meaningful. This is a crucial fact. The only glimpse of intelligence everyone is ready to attribute to a computer uncontroversially concerns the capacity of its devices and circuits to discriminate between binary data. If a computer can be said to perceive anything at all, this is the difference between a high and a low voltage according to which its circuits are then programmed to behave. The odd thing is that this may be somewhat true of biological systems as well, as we shall see in Chapter 6.

Finally, since digital data normally have only two states, such *discrete variation* means that a computer will hardly ever get confused about what needs to be processed, unlike an analogue machine, which can often perform unsatisfactorily or imprecisely. Above all, a digital machine can recognize if some data are incomplete and hence recover, through mathematical calculations, data that may have got lost if there is something literally odd about the quantity of bits it is handling.

Types of data/information

Information can consist of different types of data. Five classifications are quite common, although the terminology is not yet standard or fixed. 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 perspective adopted, the same data may fit different classifications.



4. Types of data/information

Primary data

These are the principal data stored in a database, for example a simple array of numbers in a spreadsheet, or a string of zeroes and ones. They are the data an information-management system – such as the one indicating that the car's battery needs to be recharged – is generally designed to convey to the user in the first place, in the form of information. 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, not some secrete message for a spy.

Secondary data

These are the converse of primary data, constituted by their absence. Recall how John first suspected that the battery was flat: the engine failed to make any noise, thus providing secondary information about the flat battery. 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*.

Metadata

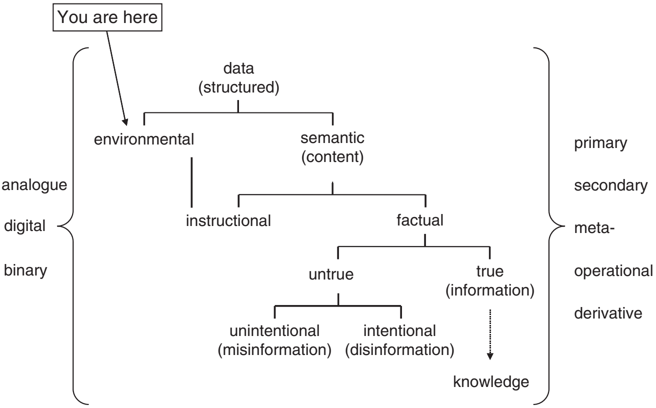
These are indications about the nature of some other (usually primary) data. They describe properties such as location, format, updating, availability, usage restrictions, and so forth. Correspondingly, *metainformation* is information about the nature of information. The copyright note on the car's operation manual is a simple example.

Operational data

These are data regarding the operations of the whole data system and the system's 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 yellow light is on may indicate that the low-battery indicator (the red light flashing) is not working properly, thus undermining the hypothesis that the battery is flat.

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, for example for comparative and quantitative analyses. As it is difficult to define this category precisely, let me rely on our familiar example. Credit cards notoriously leave a trail of derivative information. From John's credit card bill, concerning his purchase of petrol in a specific petrol station, one may obtain the derivative information of his whereabouts at a given time.



5. Environmental data/information

We are now ready to look at environmental information.

Environmental information

We speak of *environmental information* when we wish to stress the possibility that data might be meaningful independently of an intelligent *producer/informer*. One of the most often cited examples 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. Viewers of *CSI: Crime Scene Investigation*, the crime television series, will also be well acquainted with bullet trajectories, blood spray patterns, organ damages, fingerprints, and other similar evidence. Yet 'environmental' information does not need to be *natural*. Going back to our example, when John turned the ignition key, the red light of the low-battery indicator flashed. This *engineered* signal too can be interpreted as an instance of environmental information. The latter is normally defined relative to an observer (an informational organism or informee), who relies on it instead of having direct access to the original data in themselves. It follows that environmental information requires two systems, let us call them *a* and *b*, which are linked in such a way that the fact that *a* has a particular feature *F* is correlated to the fact that *b* has a particular feature *G*, so that this connection between the two features tells the observer that *b* is *G*. In short:

Table 4. Environmental information

Environmental information $=_{\text{def.}}$ two systems *a* and *b* 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 of *a* the information that *b* is *G*.

The correlation in Table 4 follows some law or rule. A *natural* example is provided by litmus. This is a biological colouring matter from lichens that is used as an acid/alkali indicator because it turns red in acid solutions and blue in alkaline solutions. Following the definition of environmental information, we can see that litmus (*a*) and the tested solution (*b*) are coupled in such a way that litmus turning red (*a* being in state *F*) is correlated to the solution being acid (*b* being of type *G*), thus carrying the information for the observer of the litmus (*a*) that the solution is acid (*b* is *G*). Our car example provides an *engineered* case: the low-battery indicator (*a*) flashing (*F*) is triggered by, and hence is informative about, the battery (*b*) being flat (*G*).

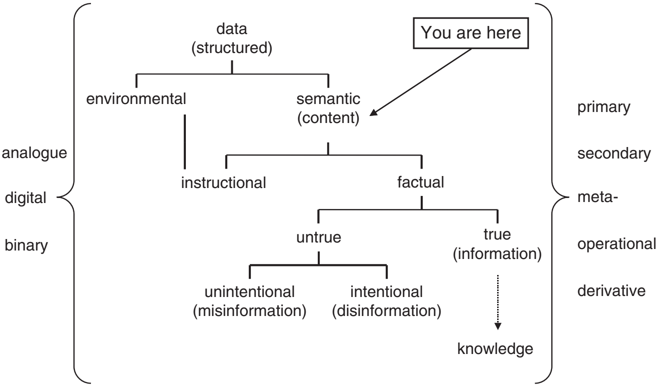
We may be so used to seeing the low-battery indicator flashing as carrying the information that the battery is flat that we may find it hard to distinguish, with sufficient clarity, between environmental and semantic information: the red light flashing *means* that the battery is low. However, it is 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 physical differences. Plants, animals, and mechanisms – e.g., a sunflower, an amoeba, or a photocell – are certainly capable of making practical use of environmental information even in the absence of any semantic processing of *meaningful* data (see Chapter 6).

Information as semantic content

When data are well formed and meaningful, the result is also known as *semantic content*. Information, understood as semantic content, comes in two main varieties: *instructional* and *factual*. In our example, one may translate the red light flashing into semantic content in two senses:

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

Chapter 4 will be primarily about (b), so this chapter ends with a discussion of (a).



6. Information as semantic content

Instructional information can be a type of environmental information or of semantic content, depending on whether meaning is a required feature. For example, the logic gates in the motherboard of a computer merely channel the electric voltage, which we may then interpret in terms of instructional information (logic instructions), such as ‘if . . . then’. In this case, there is no semantics involved at the level of the gates. The car’s operation manual, on the contrary, provides *semantic* 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.

Whether environmental or semantic, 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 (contribute to) bring about *w*. Compare the difference between ‘the water in the kettle has just boiled’, which is an instance of factual semantic information, and the process caused by the steam when it heats up the bimetallic strip enough to break the circuit of electricity flowing through the element inside the kettle, which might be interpreted in terms of instructional information. In our example, when the mechanic tells John, over the phone, to connect a charged battery to the flat battery of his car, the information John receives is not factual, but instructional. We shall return to environmental instructional information in Chapter 6, when discussing biological information. Here, let us concentrate on the semantic aspects.

There are many plausible contexts in which a stipulation (‘let the value of × be 3’ or ‘suppose we genetically engineer 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 semantic 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. Such semantic instances of instructional information have to be at least potentially meaningful (interpretable) to count as information. Finally, there are performative contexts in which we do things with words, 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). In these cases, factual (descriptive) information acquires an instructional value.

As readers of Harry Potter might suspect, the two types of semantic information (instructional and factual) may come together in magic spells, where semantic representations of *x* may be supposed to provide some instructional power and control over *x*, wrongly in real life, rightly in Harry Potter's adventures. Nevertheless, as a test, one should remember that instructional information 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. Likewise, stipulations, invitations, orders, instructions, game moves, and software cannot be true or false.

Semantic information is often supposed to be *declarative* or *factual*. Factual information such as a train timetable, a bank account statement, a medical report, a note saying that tomorrow the library will not be open, and so forth, may be sensibly qualified as true or false. *Factual semantic content* is therefore the most common way in which information is understood and also one of the most important, since information as true semantic content is a necessary condition for knowledge. Because of this key role, Chapter 4 is entirely dedicated to it. Before dealing with it, however, we need to complete our exploration of the concepts of information that require neither meaning nor truth. This is the task of the next chapter, dedicated to the mathematical theory of communication, also known as information theory.