Chapter 2
Information Versus Data

2.1 General Discussion

Similarly to semantics, the relations between information and data are also counterintuitive and somewhat confusing: In the everyday usage the term 'information' is used interchangeably with 'data' and we tend to think that more data leads to more information; and yet, in the context of Shannon's information theory this is not necessarily the case. To see how, consider the Bull by Picasso (Fig. 2.1): From plate to plate there is less data but more Shannonian information (uncertainty) until the final plate which can be a bull, but also a buffalo, a mountain goat, or a gnou (Fig. 2.2). The process is irreversible. This process is called lossy data compression. We find here inverse relations between data and information: data compression is the opposite of information compression—it is information inflation.

2.1.1 On Knowledge

> The question how **knowledge** should be defined is perhaps the most important and difficult [...]. This may seem surprising: at first sight it might be thought that knowledge might be defined as belief which is in agreement with the facts. The trouble is that no one knows what a belief is, no one knows what a fact is, and no one knows what sort of agreement between them would make a belief true. (Bertrand Russell (1926), in "Theory of Knowledge")

The relation "less data → more information" is counterintuitive and might be seen as a result of the fact that Shannon's information is just a technical term with no relations to information as used in everyday language. But note that in some cases it does make intuitive sense as quite often more data leads to more uncertainty, that is, to more Shannonian information. For example, for a person who *knows*—that is, has in memory—one animal only—say, a bull—Picasso's Plate XI would be a bull (i is 0), while for a person who has more knowledge about animals (Fig. 2.2), Plate

H. Haken and J. Portugali, *Information Adaptation: The Interplay Between Shannon Information and Semantic Information in Cognition*,
SpringerBriefs in Complexity, DOI 10.1007/978-3-319-11170-4_2

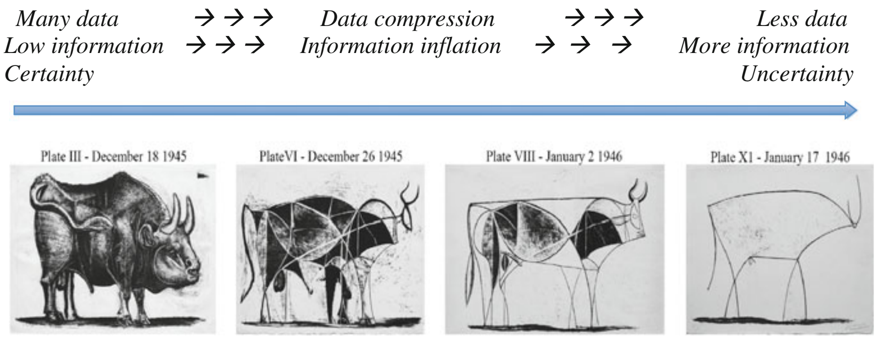

Fig. 2.1 Plates III, VI, VII and XI from the Bull by Picasso and the relations between the data, information and certainty they convey

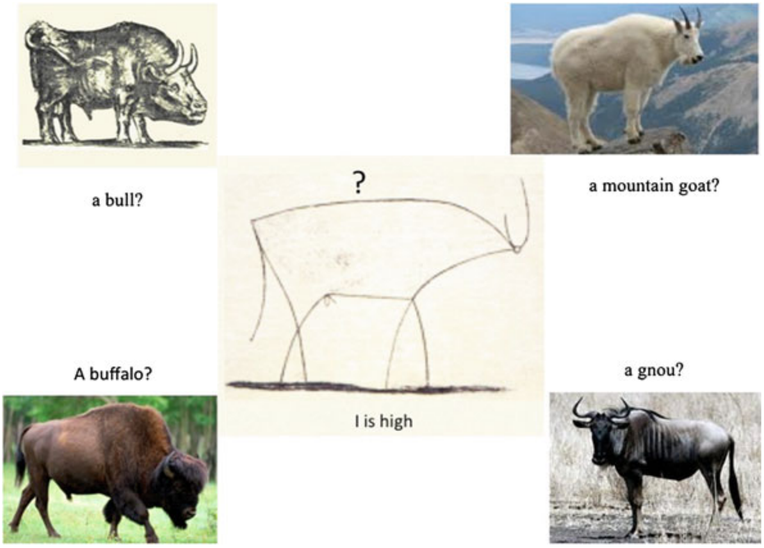

Fig. 2.2 Is the painting at the center a bull?, a mountain goat?, a gnou?, a buffalo?

XI might be several things (i is high). Shannonian information is thus also a measure of more/less knowledge—the number of *meanings* one can attach to a given representation (e.g. Plate XI in Fig. 2.1). Weaver (ibid) suggested that Shannonian information is not only a measure of uncertainty but also of a 'freedom of choice'. To the latter we might add that it is also a measure of expectations, which in the present context mean attractors. So we are driven back to the issue of semantic information.

The antonym of lossy data compression is *lossless data compression* in which a given data representing a given entity can be compressed without loosing the specific meaning of that entity (e.g. Plates III–VII in Fig. 2.1). This is possible when

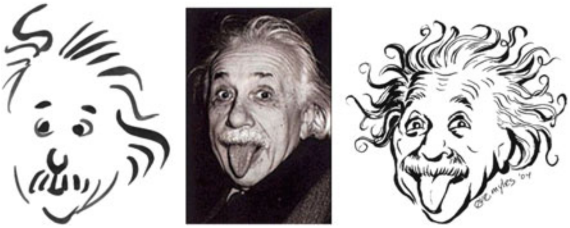

Fig. 2.3 Two caricatures of Einstein: Left by Stoyan L and right by E Myles

we deal with redundant data; the process is reversible—the original data can be recovered. Here too more data doesn't necessarily lead to more information. Lossless data compression can be exemplified by reference to caricatures that with few lines convey the face of a person (e.g. Einstein in Fig. 2.3, left).

This is also the idea behind the 'principle of parsimony' and famous phrases such as:

> *Less is more* (Andreas del Sarto 1855)
> *It is futile to do with more things that which can be done with fewer.* Occam's Razor (1288–1348)
> *I apologise for the length of this letter, but I didn't have time to write a shorter one*. Blaise Pascale (1623–1662).
> *Everything should be made as simple as possible, but no simpler*. Albert Einstein (1879–1955)

The notions 'information compression' and 'data compression' might often oppose each other, as we've just illustrated, and yet in the literature they often appear interchangeably. To avoid confusion we suggest leaving the term 'data compression' as it is, and instead of 'information compression', use *information deflation*. The antonym of the latter would be *information inflation*. Thus in the case of Picasso's Bull, each sequential drawing conveys less data but more information, more uncertainty and more possibilities. What we have here is thus a sequential process of information inflation implemented by means of data compression.

The discussions on lossy/lossless data compression refer to situations in which a given amount of data can or cannot be recovered. But the cognitive process of information adaptation by inflation/deflation goes beyond that. As we illustrate below in Chaps. 3 and 4, in some cases of information adaptation the brain adds data that doesn't exist in the stimulus, while in other cases the brain ignores data that does exist in the stimulus.

The above distinction between data and information feeds back to the discussion about the relations between Shannonian and semantic information, namely, it illustrates in a different way, how semantics enters in disguise into the definition of Shannonian information: Data in itself, that is, "pure data" with no meaning at all, has no Shannonian information. In order for data to become quantitative Shannonian information, semantic information is required in order to distinguish in a meaningful way between the different items that constitute the data, for example, to (pattern) recognize the items.

2.2 Mathematical Formulation. Some Basic General Concepts

Mathematically we start from the notion of "representation" as a set (or string) of symbols (or data), where we assume for clarity that these are numbers, e.g. in a binary system, 0 and 1. Thus a specific string, e.g. (0, 1, 1, 0, 1) is a representation. In this way, any picture can be encoded by a string of symbols: Introduce pixels and put: white <−>0, black <−>1.

This enables us, in a first step, to ignore "meaning", and deal with the processing of *representations*, e.g. by a human or some computer. How to interpret "processing"? It means in each actual case the transformation of a specific representation into another representation in one of the following manners:

(a) *one-to-one*
(b) *many-to-one (or fewer)*
(c) *one-to-many*

Compression of representation has only to do with (a).

A specific representation is replaced by a different shorter representation (string). "Lossless" is now defined as *reversible*: the original representation can be fully restored. (There are deep questions related to the Turing machine concerning the reversibility of this process, however!). "Lossy" compression is *irreversible*, the original representation cannot be fully restored or not at all.

Interestingly, these two different processes were discussed in computer theory: Rolf Landauer (1961) of IBM studied the "lossy" case which is realized in all computers and amounts to doing away bits, where pro bit one kT is produced (where k is Boltzmann's constant and T the absolute temperature), i.e. a heating up of the computer (which actually can be much larger than kT). Charles Bennett (1982), also of IBM, showed that a *lossless* computer is also feasible, at least in principle. Our brain produces heat!

Probably one may define redundancy in several ways, e.g. in the context of engineering. In the present context we would suggest: A specific representation is redundant if it can be *reversibly* compressed. (Redundancy may also mean that a set of representations can be replaced by a single representation provided the latter serves the same purpose(s) as the original set.)

We can now refer to information in the Shannon sense. Here we are dealing with a *whole set* of representations. Let P_j be the probability of the "event": occurrence of a representation labeled by j, then

$$i_j = -\log_2 P_j$$

is the information of that representation.

If $P_j = 1$ (centainty!)
$i_j = 0$
If $P_j \approx 0$ (great improbability because of the possibility of other representations)
i_j very large.

Information i is defined as average

$$i = \sum_j P_j i_j = -\sum_j P_j \log_2 P_j.$$

In this case, it will be better to speak of information entropy to make a distinction with respect to the information of a single representation.

Now we can define information *deflation* as lowering the (numerical) value of i, and information *inflation* as increase of i. This can be directly related to data processing.

2.2.1 *Information Deflation*

The number of representations is reduced (case (b) above). For illustration: When there are originally N representations with $P_j = 1/N$, then $i = \ln N$ means large uncertainty for large N, whereas reduction to 1 representation: $P_k = 1$, all other $P_j = 0, j \neq k$ implies information i = 0

i.e. complete certainty.

Or, more generally, reduction of high information (entropy) to zero information, means a move from uncertainty –> certainty. This process is irreversible because the final representation has several originators.

2.2.1.1 Information Inflation

Here one representation is replaced by many representations (case (c) above). This clearly means increase of information (entropy). Whether this process is reversible depends on additional knowledge: If there is only one original representation, the process is reversible. On the other hand, one or several representations could also stem from the transformation of different representations: then (c) is irreversible.

2.3 Data, Information and Meaning. How Are These Related?

We will try to formalize this problem on the basis of our former interpretation of meaning. Meaning does not result from data per se but only in the context of a receiver, which is a complex system in the sense noted above, namely, that it has memory. Such a receiver has in its memory $N + 1$ states, $k = 0, 1, \ldots, N$, which in a task of pattern recognition of faces, for instance, "mean" "face 1" or "face 2" etc., while $k = 0$ means no effect on the receiver.

The reader will note that here we do not distinguish between semantic and pragmatic information. This implies that we adopt a strictly operational point of view. We leave it open, however, how to observe the effect on the receiver. (This effect could be a specific reaction of the receiver (person/machine) or the storage in memory.)

We distinguish the representations by a label j. The length of the corresponding j is denoted by Lj In the context of our contribution we consider the case that a representation represents a "pattern" (e.g. a picture, see below, or a string of letters etc.).

Then *pattern recognition* is formalized by

> Pattern j represented by representation j
> causes receiver to acquire state k with the conditional probability $p(k|j)$ for $k = 1, \ldots, N$ or $p(0|j)$, i.e. no pattern is recognized.

The probabilities are normalized so that $\sum_k p(k|j) = 1$.

Here, we do not discuss how to determine the probability distribution experimentally.

Lossless compression of a representation is now defined by $p(k|j) = p(k|j')$ for the case of two different representations j and j' with respective lengths $L_{j'} < L_j$.

Equivalently, we may speak of *meaning conservation*, if the same meaning k is conveyed by compressed representations or in view of our above statements, by reduced redundancy.

A special case and of particular interest is $p(k|j) = 1$, where a specific state k (a specific meaning!) is attached to the given representation j with certainty. Or in other words, pattern j has been "recognized".

Lossy compression of a representation means $p(k|j') < p(k|j)$for the case of j, j' with $L_{j'} < L_j$.

An example is

$$\begin{array}{ll} p(k|j) = 1 & \text{for a fixed } k \text{ and } j, \text{ but} \\ p(k|j') = \frac{1}{2} & j' \neq j \\ p(k'|j') = \frac{1}{2} & \text{for a fixed } k, k' \text{ and } j' \end{array}$$

The transition from j to j' increases the number of possible interpretations: i.e. information inflation.

Application to Picasso's bull: To bring out the essentials: take the number of black pixels as measure of length L_j. The index j = 1, 2, 3, 4 corresponds to the plates III, VI, VII, XI. Then L_j decreases from left to right in the picture sequence.

Left picture, $j = 1$, if $k = k_0$ means "bull", then $p(k_0|1) = 1$

Right picture, $j = 4$.

Lossless compression of representation: if

$$p(k_0|4 = 1)$$

This is the required feature of a caricature!

Lossy compression of representation: if (for instance)

$p(k_0|4) = \frac{1}{2}$ and $p(k_1|4) = \frac{1}{2}$, where k_1 means deer.

In this case we deal with increased uncertainty, information inflation, partial loss of meaning of representation 4. Complete loss of meaning of representation j' happens, if for

$$\begin{aligned} p(k|j') &= 0 \quad \text{for}\, k = 1, .., \ \text{N, and} \\ p(k|j') &= 1 \quad \text{for}\, k = 0, \ \text{i.e. no effect on receiver.} \end{aligned}$$

Categorization can now be formalized along these lines as well as the process of improving pattern recognition (e.g. by increasing the length L_j, i.e. by providing more and more details of a picture).

http://www.springer.com/978-3-319-11169-8

Information Adaptation: The Interplay Between Shannon Information and Semantic Information in Cognition
Haken, H.; Portugali, J.
2015, XIV, 90 p. 49 illus., 12 illus. in color., Softcover
ISBN: 978-3-319-11169-8