

# Notes on “Semantic Conceptions of Information” (Floridi 2015)

Information is a polymorphic phenomenon and a polysemantic concept. Following Shannon, Weaver 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
3. “Influential” problems concerning the impact and effectiveness of information on human behavior

Section 1 draws a map of how one can speak of *semantic information* by relying on the analysis of the concept known as *data*. Section 2 provides an introduction to *information theory* and the *mathematical theory of communication* (MTC). Section 3 analyzes information as semantic content. Section 4 focuses on the philosophical understanding of semantic information.

## 1. An informational map

### 1.1 An everyday example of information

### 1.2 The data-based definition of information

“Information” when defined with MTC is *well-formed data*. These data are classified in a tree structure:

Well-formed = environmental + semantic

Semantic = instructional + factual

Factual = true (information) + untrue

Untrue = intentional (disinformation) + unintentional (misinformation)

**The General Definition of Information (GDI):** *s* is an instance of information, understood as semantic content, if and only if:

(GDI.1) *s* consists of one or more *data*;

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

(GDI.3) the well-formed data in *s* are *meaningful*.

“Data” => substance

“Well-formed” => complies with the syntax/composition/structuring of the system/code/language

“Meaningful” => complies with the semantics of system/code/language

### 1.3 A definition of data

Information cannot be dataless.

**The Diaphoric Definition of Data (DDD):** A datum is a putative fact regarding some difference or lack of uniformity within some context.

DDD can be applied at three levels:

1. data as diaphora *de re*, lack of uniformity in reality, *dedomena*. They are proto-epistemic data, data before interpretation, tears in the fabric of being. Dedomena are never accessed independently of a level of abstraction. We cannot know if they exist, but we can infer that they exist as the external anchors of our information;
2. data as diaphora *de signo*, lacks of uniformity in the perception of at least two physical states;
3. data as diaphora *de dicto*, lacks of uniformity between two symbols.

Depending on your philosophical views, data in (1) may be identical with or what makes possible signals in (2). Signals in (2) are what make possible the coding of symbols in (3).

Semantic information must be encoded, but the choice of *format*, *medium*, and *language* in which it is encoded is often irrelevant. The same semantic information can be analog or digital, printed on paper or viewed on a screen, in words or in pictures.

GDI adheres to taxonomic, typological, ontological, and genetic neutrality.

### 1.4 Taxonomic neutrality

(ON) Data are relata (a relatum is one of a group of related things).

A datum is often considered to be the thing that is exhibiting the anomaly, but really the anomaly can only exist if there is a contrasting uniformity. Thus, data are relata, entities in relation to other entities (i.e. in a *relationship*, both the black dot and the white background at the same time. GDI does not require one to choose whether the data is the dot or the background. The choice is arbitrary, but there is usually a natural preference for one over the other.

The relation in question is *inequality*. This relation is created via a perceptual cut.

### 1.5 Typological neutrality

(TyN) Information can consist of different types of data as relata.

There are different ways to classify data, and the classifications are not mutually exclusive. The same data can play different roles.

(D1) *Primary data*. The data you collected about the actual event, what would be stored in a database.

(D2) *Secondary data*. The data you have due to not having other data, anti-data. If it is silent, you have no noise data, but you do know that it is silent, which is itself a datum. Silence, the absence of noise, may be informative and can be considered secondary information.

(D3) *Metadata*. Data about (usually primary) data you already have. They describe things like location, format, availability, etc. Metainformation is information about information (“‘The battery is flat’ is encoded in English.”).

(D4) *Operational data*. Data about the operations of the data system as a whole, performance stats and internal indicators. Operational information is information about the operations of an information system.

(D5) *Derivative data*. Data that is derived from other data. Perhaps some information is conveyed for a certain purpose, but it contains within itself information that can be used for a different purpose.

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 constituted by non-primary data (D2)-(D5). If a database query provides an answer, it at least provides a negative answer (e.g. “no documents found”). If it provides no answer, it either provides no information (and thus no data) or it provides some non-primary data, such as operational data, which may constitute operational information that explains why it is not providing an answer. Likewise, silence could represent negative primary data (e.g. implicit assent or denial) or it could carry some non-primary data (e.g. the person did not hear you, noise level in room).

## 1.6 Ontological neutrality

(ON) There can be no information without data representation.

This is often interpreted materialistically as stating that information cannot be physically disembodied. It must be carried by something.

(ON.1) There can be no information without physical implementation.

This is the assumption taken in the computational theory of mind. However, if a universe contains only noetic entities (e.g. the conceptual universe), data representations would not necessarily require material implementations. Thus, GDI is also consistent with other assumptions:

(ON.2) *It from bit*. Every item of the physical world has an immaterial source and explanation, an response given to a yes-no question. All physical things are information-theoretic in origin, material objects are secondary manifestations, and the Universe is a participatory simulation.

(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. Information is information, not matter or energy, and materialist theories must admit this.

(ON.4) In fact, what we mean by information—the elementary unit of information—is a difference which makes a difference (i.e. a datum that is meaningful).

### 1.7 Genetic neutrality

(GeN) Data (as relata) can have a semantics independently of any informee.

GeN supports the possibility of information without an informed subject. Meaning is not (or at least not only) in the mind of the user.

This is the weak sense in which (GDI.3) speaks of meaning data being embedded in information-carriers. GeN is NOT the stronger, realist thesis, according to which data could also have their own semantics independently of an intelligent producer/informer. This is *environmental information*.

**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  $G$ .

This is analog information (low battery indicator ( $a$ ) being in a flashing state ( $F$ ) is informative about the battery ( $b$ ) being in a flat state ( $G$ )). One system is an *analogy* for another system.

Environmental may require or involve no semantics at all. It may consist of patterns of correlated data understood as mere differences. Plants, animals, and mechanisms can make practical use of environmental information even in the absence of any (semantic processing of) *meaningful* data.

### 1.8 Summary of the first part

GDI defines information as syntactically well-formed and meaningful data. It has four neutralities: TN, TyN, ON, and GeN).

## 2. Information as data communication

Information can be *encoded*, *transmitted*, and *stored*. Information is also *additive* ( $\text{info } a + \text{info } b = \text{info } a+b$ ) and *non-negative* (the worst case scenario is receiving zero information).

### 2.1 The mathematical theory of communication

MTC is a quantitative approach to information, developed in the field of electrical engineering as a means of solving two fundamental problems: the ultimate level of data compression and the ultimate rate of data transmission. The solutions are respectively the information entropy  $H$  and the channel capacity  $C$ .

**The Shannon-Weaver Model:** Informer (I1) sends a message (written in an alphabet) through Encoding Transmitter (ET). ET creates a signal and sends it through Communication Channel (CC), which is subject to noise. Decoding Receiver (DR) receives the signal from the channel and turns it into a message, which Informee (I2) reads, interprets, and potentially understands. In order for proper communication to occur, I1, ET, DR, and I2 must be familiar with the alphabet of the message.

A *unary device* produces the same response to any question (like Poe’s raven always answering “nevermore”). It produces a zero amount of information because you already know what the response is before even asking the question. It tells you nothing new. Silence is a singular response, so a silent source is an uninformative, unary source. A censored party and a boy crying wolf are both uninformative, unary sources.

A *binary device* can produce two symbols (e.g. a two-sided coin). Before a computation (e.g. a coin toss), the informee is in a state of *data deficit* or *uncertainty*. They do not know the result. The computation produces an amount of information that is a function of the possible outputs and equal in magnitude to the data deficit that it removes.

Consider a more complex system made up of two coins  $A$  and  $B$ . This  $AB$  system has 4 possible outputs, each of which is a separate symbol:  $\langle h, h \rangle$ ,  $\langle h, t \rangle$ ,  $\langle t, h \rangle$ ,  $\langle t, t \rangle$ . The system resolves a data deficit of 4 units by producing an equivalent amount of information.

Given an alphabet of  $N$  equiprobable symbols (outputs), average informativeness per symbol (“uncertainty”) =  $\log_2(N)$  bits of information per symbol. Information can be quantified in terms of “how much uncertainty it resolves.”

The amount of information produced by the biased coin depends on the average informativeness of the  $S$ -length string of  $h$  and  $t$  produced by the coin. The average informativeness of the string depends on the probability of the occurrence of each symbol. The higher the frequency of a symbol in  $S$ , the less information the coin produces when it produces that symbol.

Common events are less informative. Rarer events are more informative. Whether something is common or rare depends on context. For the average information to be high, the context must allow for a large number of unlikely events (i.e. there must be a high level of uncertainty within the context, a wide distribution of probabilities such that there is a great deal of disorder or non-determinism).

1. To calculate the average informativeness of the string, we must calculate  $S$  and the informativeness of the  $i$ th symbol in general.
2. For the special case of equiprobable symbols, the informativeness of each symbol is  $\log(N) = -\log(1/N) = -\log(P)$ . In general, informativeness of the  $i$ th symbol  $u_i = -\log(P_i)$ . (Boltzmann’s entropy is Shannon’s entropy applied to equiprobable microstates.)

3. For the two-sided coin, the length of  $S$  equals the number of times  $h$  occurs plus the number of times  $t$  occurs. For a general number of symbols,  $S = \sum_{i=1}^N (S_i)$ .
4. Average informativeness for a string of  $S$  symbols is thus  $(\sum_{i=1}^N (S_i \log P_i)) / \sum_{i=1}^N (S_i)$ . This simplifies to  $\sum_{i=1}^N P_i \log P_i$ .
5.  $H = -\sum_{i=1}^N P_i \log P_i$  (bits per symbol). The amount of information produced by a device (RHS) corresponds to the amount of data deficit erased (LHS).

Because 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 (the coin becomes a flip-flop, sequential logic).

The highest quantity of information is produced by a system whose symbols are equiprobable. When a system is biased, you might have one state that occurs much more frequently than others. You would not want to have equal-length encodings in this case; you would want to assign the most frequent state the shortest encoding (e.g. 0) and the less frequent states longer encodings (Fano code).

Redundancy is the difference between the physical representation of a message and the mathematical representation of the same message, which is as compressed as possible while still retaining the original meaning. Redundancy takes up additional resources (i.e. space or time), but it can counteract data loss (*equivocation*) and *noise* during transmission. If your words could be lost or “talked over,” it is worth stating them more than once.

**Shannon’s Fundamental Theorem of the Noiseless Channel:** 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  $C/H - \epsilon$  symbols per second over the channel where  $\epsilon$  is arbitrarily small. It is not possible to transmit at an average rate greater than  $C/H$ .

**Shannon’s Fundamental Theorem for a Discrete 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  $H - C + \epsilon$  where  $\epsilon$  is arbitrarily small. There is no method of encoding which gives an equivocation less than  $H - C$ .

Basically, the transmission rate of semantic information is limited by how complex the encodings are (their average entropy) and by the physical parameters of the channel (namely, its bandwidth, signal power, and noise power). If you are transmitting information faster than the channel can transmit it, some of it will be lost to equivocation.

(Conditional entropy: if every word in the dictionary is represented by a unique symbol, there is much uncertainty as to which word will be chosen; if the first letter of the word is given, there is suddenly much less uncertainty. Markov chains...)

(For a noiseless channel, given a sent message, the received message is certain. The mutual information is equal to the entropies of the received and sent messages; both parties are fully informed. For a noisy channel, the informer can send a correction message (i.e. a redundancy, error-correcting code, checksum) whose entropy is conditional on the received message.)

(Intelligence as the ability to decrease the entropy of systems?)

## 2.2 Conceptual implications of the mathematical theory of communication

How many yes/no questions required to determine what the source is communicating?  $\log(N)$  questions where  $N$  is the number of possible outputs.

The MTC definition of information is not the same as the ordinary definition of information. If a device could randomly send you a single book or every book ever written, you would receive different amounts of bytes of data, but only one bit of information in the MTC sense: the answer to whether you got one book or all of them. One decision was made. An equal distribution of symbols, each chosen randomly, should yield the highest amount of information according to MTC, but such a text would be gibberish.

If information is by definition meaningful, ‘information theory’ is not a good name. The better name is ‘mathematical theory of *data* communication.’ Shannon’s work was strictly in moving symbols from point A to point B; interpretation of those symbols does not factor into the theory.

A proposition can be transformed into a question + the answer “yes.” If you remove the “yes” (dealthicization), you are left with *semantic content* or *unsaturated information*. The datum “yes” works as a key to unlock the information contained in the question. In MTC, information is the amount of details in a signal or message or memory space necessary to saturate the informee’s unsaturated information (i.e. answer the question, decide the truth value of the proposition).

Weaver: Information is not what you *do* say, but what you *could* say; MTC deals with carriers of information, symbols and signals, not with information itself; information is the measure of your freedom of choice in selecting a message.

MTC studies information at the syntactic level, and computers are syntactical devices.

In a noiseless channel,  $H$  is a measure of three equivalent quantities:

1. the average amount of information per symbol produced by the informer;

2. the average amount of data deficit (uncertainty) that the informee has before interpretation;
3. the informational potentiality of the source (like a stored energy).

Entropy is a measure of the amount of “mixedupness” in processes and systems bearing energy or information. If a process causes no change in entropy, it is a reversible process. The greater the entropy, the less available the energy.

### 3. Information as semantic content

Information comes in two varieties: factual and instructional. For example, a red flashing light can be translated into semantic content in two senses:

1. as a piece of factual information, representing the fact that the battery is flat; and
2. as a piece of instructional information, conveying the need for a specific action, like recharging or replacing the battery.

#### 3.1 Instructional information

Instructional information is given either *imperatively* (a list of commands) or *conditionally* (a flowchart or inferential procedure). It does not model, describe, or represent a situation or state of affairs *w*. Rather, it is meant to (help to) bring about *w*.

Instructional information includes: stipulations (“let...” or “suppose...”), invitations, orders, instructions, game moves, musical compositions, computer programs. All of these instances have a semantic side.

Instructional information may be related to factual (descriptive) information in performative contexts (e.g. naming a ship, declaring a variable’s type, casting magic spells). But unlike factual information, it cannot be true or false (i.e. it does not qualify alethically).

#### 3.2 Factual information

Factual information is given declaratively and may be true or untrue (i.e. false, given a binary logic). Information is most commonly understood as *true semantic content*, which is necessarily factual. True semantic content is a necessary condition for knowledge.

3 things to cover next: data as constraining affordances, the role played by levels of abstraction in the transformation of constraining affordances into factual information, the relation between factual information and truth.

Data that constitute factual information invite and impede certain constructs. They are in one sense *affordances* and in another sense *constraints* for an information agent. That is, they lead you closer toward inferring certain conclusions



(that might be true or false) and lead you away from other conclusions (that might be true or false). They are answers; it is up to the agent to find its corresponding question.

A level of abstraction (LoA) is a specified 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. LoA is an epistemological concept, not an ontological one. Through an LoA, an information agent accesses a physical or conceptual environment, the system. They mediate the epistemic relation between observer and observed. (It is essentially a universe or class.) Data as constraining affordances are processed semantically at a given LoA and, as a result, are transformed into factual information.

Is some factual content only considered information if it is *true*? This is the question of alethic neutrality for information.

If so:

1. false information (including contradictions) would count as semantic information;
2. necessary truths (including tautologies) would count as semantic information; and
3. “it is true that  $p$ ,” where  $p$  is semantic information, would be an expression of semantic information that is distinct from  $p$ .

Solution to hyperdimensionality: the relation between truth and informativeness in the case of logically equivalent expressions.

“It is true that” could be considered redundant because, strictly speaking, information is not a truth-bearer but already encapsulates truth as truthfulness.

Once information is available, knowledge can be built in terms of *justifiable* or *explainable semantic information*. An agent may use data as resources to construct information and, hence, knowledge.

Whether empirical or conceptual, data make possible only a certain range of information constructs, and not all constructs are made possible equally easily. 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. Information is the result of data modelling; it does not have to model the intrinsic nature of the system analyzed.

When semantic content is false, it is *misinformation*. If the source of misinformation is aware that it is false, it is *disinformation*. The truth-value of information does not necessarily affect its efficacy.

## 4. Philosophical approaches to semantic information

MTC is syntactical. Factual information is semantic. Philosophical approaches consider information as semantic content and thus endorse the centrality of factual content (e.g. “Paris is the capital of France”). The majority of philosophers today agree that MTC provides a rigorous constraint to any further theorizing on the semantics and pragmatics of information. But how strong is the constraint? Does MTC fully describe the rules of semantic-factual information or only partially?

Even though the philosophy of semantic information has become increasingly autonomous from MTC, two connections certainly remain: the communication model and the *Inverse Relationship Principle* (IRP) (i.e. the inverse relationship between the probability of  $p$  and the amount of semantic information carried by  $p$ ).

Just as MTC defines information in terms of probability space distribution, the *probabilistic approach* to semantic information defines the semantic information in  $p$  in terms of logical probability space. The semantic content CONT in  $p$  is measured as the complement of the a priori probability of  $p$ :  $\text{CONT}(p) = 1 - P(p)$ . The informativeness INF of  $p$  is the reciprocal of  $P(p)$ , expressed in bits:  $\text{INF}(p) = \log(1/(1-\text{CONT}(p))) = \log P(p)$ .

But how should  $P(p)$  be interpreted? According to Bar-Hillel and Carnap [1953], it is the outcome of a logical construction of atomic statements according to a chosen formal language. This would require a strict correspondance between observational and formal language. In Dretske, the probability refers to the observed state of affairs:  $I(s) = -\log P(s) = \log P(p)$ .

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 the descriptions of the relevant possible states of the universe that are excluded by  $p$ .

In the *systemic approach*, 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 (i.e. given a certain context). Information tracks possible transitions in a system’s state space under normal conditions. Typically, this approach requires some presence of information already immanent in the environment (environmental information).

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.

These 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.

#### 4.1 The Bar-Hillel-Carnap Paradox

Because tautologies have a probability of 1, most philosophers agree that they carry no information. But, by the same reasoning, one could argue that contradictions, which have a probability of 0, would contain the highest amount of semantic information. This is not true; at a certain point, the statement “implodes,” becoming “too informative to be true.”

**Bar-Hillel-Carnap Paradox (BCP):** It is strange that a self-contradictory sentence would be considered as carrying the most semantic information. However, semantic information here does not imply truth. A self-contradictory sentence says a lot, so much that it cannot be true. Its constraints are too strict, its scope too specific, for it to be true in any situation that may arise in the system.

*Theory of weakly semantic information*  $\Rightarrow$  truth values play no role in it.

BCP is an unfortunate fact that must be accepted for weakly semantic information theories. However, there are strategies that redefine the quantitative measurement of semantic information in the case of contradiction:

1. assign to all inconsistent cases the same, infinite information value (similar to an economic approach);
2. eliminate all inconsistent cases a priori (a syntactic approach like MTC);
3. assign to all inconsistent cases the same zero information value (similar to the *strongly semantic approach*).

#### 4.2 The strongly semantic approach to information

Hypothesis: BCP implies that something is wrong with theories of weakly semantic information. A semantically stronger approach, according to which information encapsulates truth, can avoid the paradox.

MTC identifies the quantity of information associated with the occurrence of a signal (i.e. a realizable event) with the elimination of possibilities (reduction of uncertainty) represented by that signal. MTC does not exhibit BCP, so one could argue that an MTC-like semantic theory based in alethic and discrepancy values rather than probabilities would avoid BCP as well.

According to Floridi, semantic-factual information is defined, in terms of data space, as well-formed, meaningful, and truthful data. This further constrains the probabilistic approach by requiring first a qualification of the content as truthful. After this qualification, the quantity of semantic information in  $p$  is calculated in terms of distance of  $p$  from the situation  $w$  that  $p$  is supposed to model.

Situation ( $w$ ): there will be three guests for dinner tonight. Possible models: there may or may not be some guests for dinner tonight (T), there will be some guests tonight (V), there will be three guests tonight (P).

The *degree of informativeness* of T is zero because T is a tautology. V performs better, and P has the maximum degree of informativeness because it is a fully accurate, precise, and contingent truth. Generalizing, the more distant some semantic-factual information  $s$  is from its target  $w$ , the larger the number of situations to which it applies, the lower its degree of informativeness becomes (consider the genus-species taxonomy: closer to root => more widely applicable => fewer details). A tautology is a true  $s$  that is the most distant from  $w$ .

Given some  $s$  and  $w$ , there exists some distance  $d$  between them.  $d(T) = 1$ ,  $d(P) = 0$ , and let us say that  $d(V) = 0.25$ . We now need a formula to calculate the degree of informativeness  $i$  of  $s$  in relation to  $d(s)$ :  $i(s) = 1 - d(s)^2$ .

$d(s)$  ranges from -1 (contradiction) to 1 (tautology), and  $i(s)$  ranges from 0 to 1. However, semantic information must be truthful => its  $d(s)$  must be on the interval  $[0,1]$ .

Max quantity of semantic information  $a = \int(0,1,i(s),dx)$ .

Amount of vacuous information  $b = \int(0,d,i(s),dx)$ .

Amount of semantic information  $c(s) = a - b$ .

T is so distant from  $w$  as to contain only vacuous information. T contains as much vacuous information as P contains relevant information.