A critical analysis of Floridi's theory of semantic information

Adriaans, P.

[Link to publication](#)

A Critical Analysis of Floridi's Theory of Semantic Information

Pieter Adriaans

In various publications over the past years, Floridi has developed a theory of semantic information as well-formed, meaningful, and truthful data. This theory is more or less orthogonal to the standard entropy-based notions of information known from physics, information theory, and computer science that all define the amount of information in a certain system as a scalar value without any direct semantic implication. In this context the question rises what the exact relation between these various conceptions of information is and whether there is a real need to enrich these mathematically more or less rigid definitions with a less formal notion of semantic information. I investigate various philosophical aspects of the more formal definitions of information in the light of Floridi's theory. The position I defend is that the formal treatment of the notion of information as a general theory of entropy is one of the fundamental achievements of modern science that in itself is a rich source for new philosophical reflection. This makes information theory a competitor of classical epistemology rather than a servant. In this light Floridi's philosophy of information is more a reprise of classical epistemology that only pays lip service to information theory but fails to address the important central questions of philosophy of information. Specifically, I will defend the view that notions that are associated with truth, knowledge, and meaning all can adequately be reconstructed in the context of modern information theory and that consequently there is no need to introduce a concept of semantic information.

1 Introduction

In everyday life we all have an intuitive idea of what information is. It is the stuff that secret services gather about terrorists and it is what we hope to get when we go

P. Adriaans (✉)
University of Amsterdam, IvI, Sciencepark 107, 1089 XG Amsterdam, The Netherlands
e-mail: pietera@science.uva.nl

to an office of the tourist information. So if we ask for a hotel in the center of the city we are happy when we get a concise list of addresses. This is a form of data. They are just signs on a paper without any meaning in themselves. Since we can read we can interpret the signs as a message with a certain structure and since we are beings endowed with general intelligence these data have meaning for us. So even if we have never been to Amsterdam we have a rough expectation of what Hotel Krasnapolsky on the Dam in Amsterdam would look like. This gives meaningful data. We also want this meaningful data to be true. There can be no hotels on the list that are out of business or that have moved to another address. Furthermore, we want it to be adequate, and to the point, so we do not need a book of all hotels in the country. We expect to get true and adequate information from the tourist information because we think it is a reliable source. Good information in general life gives us something to direct our actions. If we base our actions on false information, the outcome might be disastrous, as recent events in Iraq have shown. It is this intuitive every day notion of information that Luciano Floridi (supposedly) tries to capture when he defines the notion of semantic information as "*well-formed, meaningful, and truthful data.*"[1] Knowledge then is defined as: "*relevant semantic information properly accounted for, humans are the only semantic engines and conscious inforgs (informational organisms) in the universe, and the universe is the totality of information.*"[2] In the past decade, Luciano Floridi has developed an impressive and interesting attempt to formulate a coherent theory about information. The quotes above are typical for his style of presentation in which perfectly defendable positions are interleaved with fairly radical statements that are hard to interpret in a sensible way. Here, a reasonable claim about the nature of information, is followed by an unsubstantiated claim about uniqueness of human beings as semantic engines and a completely unverifiable metaphysical claim about the nature of our universe.

In this paper I will focus myself on what I consider to be Floridi's central research program as it is captured in the slogan: information is well-formed, meaningful, and truthful data. Floridi explicitly defends the view that false information is not really information, but pseudo-information, and this is a view that obviously one can sympathize with (Floridi 2005a). We would be disappointed if the tourist office would give us wrong information, and we are appalled to learn that the decision to declare war on Iraq war was partly based on such pseudo-information. We would not accept the excuse that it was still 'real' information albeit 'false' real information. In this sense the intuitions behind Floridi's venture seem perfectly acceptable. Yet, this research program is not without problems. Some are internal to the program others external.

---

[1] Probably the best concise overview of Floridi's philosophical ideas can be found in his article on 'SemanticConceptions of Information' in the Stanford Encyclopedia of Philosophy (Floridi 2005b). I base my interpretation of Floridi's philosophy partly on two manuscripts he was kind enough to send me. One is the manuscript of the second edition of a short introduction called 'Information—a Very Short Introduction'. The other manuscript has the work title 'The philosophy of Information' and is an overview and reworking of most of his publications over the past 10 years.
[2] This concise formulation of the kernel of Floridi's philosophy is quoted from a manuscript that is at the time of writing unpublished. I take the liberty to reproduce it here because it gives such a nice description of his position in a nutshell.

Internally, the program has to explain how knowledge, information, and meaning emerge from data. Indeed these problems are recognized by Floridi in a variety of publications he has mentioned, among others:

The nature of information
The challenge of a unified theory of information
The data-grounding problem: how data acquire their meaning
The semantic problem: how meaningful data acquire their truth value
The ontological status of information
As central problems of his philosophy (see e.g. Floridi 2008).

Other problems are more external to his philosophy. Most notably the semantic theory of information is clearly at variance with the standard formal definitions of information that are studied in computer science. I will treat these definitions and their implications in more detail later, but if one takes Shannon's theory as an example then it is clear that Shannon tries to capture the notion of information as a relational concept (Shannon and Weaver 1949). For me as a scientist working in Amsterdam the statement that hotel Krasnapolsky is in the center of Amsterdam contains no information at all. For a tourist this insight might be new and thus 'informative'. The statement that the sun will rise tomorrow contains very little new information. We know this already. The opposite statement, which probably would imply the extinction of life at earth would be very informative. The semantic theory of information prima facie cannot explain this subjective relational aspect of information. It is bound to make information a monolithic static notion that exists independent from any individual observer. Floridi quotes Shannon in this context who states that:

"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." (Shannon 1993, p. 180).

Consequently, Floridi takes this as a free ticket to bring the semantic notion of information in to the discussion. He goes even further, denying that Shannon's Mathematical Theory of Communication (MTC) deals with a proper notion of information at all:

'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".' (Floridi 2005a, b).

There is obviously a circularity in this argument: Shannon's notion of information is not a proper notion because it is not covered by Floridi's definition. There is a tension between Shannon's research program and that of Floridi. They partly contradict each other. These observations suffice as a motivation to study Floridi's program in more depth. In the rest of this paper I will argue that there are at least two relevant research programs in the context of the philosophy of information that both

have deep roots in history: one which I will call the *transcendental* program is rooted in Kant's philosophy, the other that I will call the *empiricist* program finds its roots in the same empiricist Humean school of thought that Kant reacted to in his first Critique. If this distinction is appropriate then clearly Floridi's efforts belong in the transcendental program. His philosophy of information can be seen as an attempt to redefine the popular notion of information in order to revitalize the program of transcendental philosophy. In order to appreciate the situation let's put the notion of semantic information in its proper historical context. But first I will present some preparatory information.

2 The Perspective of a Unified Theory of Entropy-Based Theories of Information

Table 1 shows some highlights of three well-known theories that deal with the notion of entropy (see Adriaans and Van Benthem (2008) for a general introduction in these notions):

- Gibbs entropy: a measure for the amount of 'disorder' S in a closed system of particles in equilibrium in terms of the probability distribution of these energies.
- Shannon information: measures the amount of 'disorder' H(P) in a collection of messages in terms of a probability distribution P over these messages. (Note that the mathematical formulation is equivalent to that of Gibbs entropy). The information content of a message p_i in bits is given by the expression $-\log P(p_i)$, which gives also the optimal code length in bits.
- Kolmogorov complexity: measures the information content $C(x)$ of a binary string x in terms of the length of the shortest program p that computes this string on a universal computer U_t, such that $U_t(p)=x$.

In all of these theories researchers have felt the necessity to formulate a notion of maximal entropy. This is the Boltzmann entropy in Gibbs' theory (which is reached when all energy states have the same probability). The same holds for the maximal entropy in Shannon's theory. In Kolmogorov complexity random strings have maximal entropy (this corresponds to strings that are created by a Shannon system with two equiprobable messages '0' and '1'). Also in these three frameworks one finds the useful notion of the distance between the entropy of the actual system and the maximal entropy. In Gibbs theory this distance is associated with the notion of

Table 1 Well-known theories that deal with the notion of entropy

	Energy in a system of particles (Gibbs)	Information in a system of messages (Shannon)	Complexity of a string (Kolmogorov)
Basic definition	$S = -\sum_i p_i \ln p_i$	$H(P) = -\sum_{i \in A} p_i \log_2 p_i$	$C(x\|y) = \min\{\|p\| : U_i(py) = x\}$ $C(x) = C(x\|\varepsilon)$
Maximal entropy	$S = \ln w$	$H_{\max} = \log n = R$	$C_{\max}(x) = \log_2 x = \|x\|$
'Distance' of maximal Entropy	Free energy $U_{\text{total}} = S_{\max} - S_{\text{act}}$	Absolute redundancy $D = R - r$	Randomness deficiency $\delta(x) = \|x\| - C(x)$

free energy. Shannon defines the equivalent notion of absolute redundancy and in Kolmogorov complexity the idea takes the form of randomness deficiency. Originally, these theories lived in separate domains and the mathematical relations between them were considered to be accidental, but in recent years it has become clear that the connections are pretty deep. The relation between Shannon entropy and Gibbs entropy has been studied for a couple of decades. The role of Kolmogorov complexity in information theory has only recently been appreciated:

> "Gratifyingly, the Kolmogorov complexity K is approximately equal to the Shannon entropy H is the sequence is drawn at random from a distribution that has entropy H. So the tie-in between information theory and Kolmogorov complexity is perfect. Indeed, we consider Kolmogorov complexity to be more fundamental than Shannon entropy. It is the ultimate data compression and leads to a logically consistent procedure for inference" (Cover and Thomas 2006)

The relation between Free energy and randomness deficiency has been investigated in Adriaans (2009).

Given these results we are entitled to formulate the hypothesis that these three notions of entropy actually are different perspectives on a unified theory of entropy. What such a theory would look like is unclear at the moment and the theoretical and philosophical problems around the interpretation of these various notions of entropy and information are formidable. Possible results of such a unified theory of entropy, computation, and information could be:

- A general theory about the interaction between information and computation
- A general theory of induction
- A general frame work to study human cognition and methodology of science
- A theory of non-equilibrium entropy

In this paper, which deals with the interpretation of Floridi's philosophy, I will not treat these issues in any depth. The ambition to develop such a unified theory defines a philosophical research program that is orthogonal to the classical research program of epistemology. It offers various mathematical techniques to select the right model given a set of observations. Philosophy of information studies model selection and probability. Classical epistemology studies the orthogonal notions of truth and justification. This gives us the key to interpret Floridi's philosophy: it adopts selected notions from information theory in a research framework that is strictly classical. In itself there is nothing wrong with that, apart from the fact that information theory as a philosophical research program in the current historical situation seems much more fruitful and promising than classical epistemology.

3 A Short Theoretical Digression

One of the most elegant results of information theory in the past decades is the so-called minimum description length principle (Mitchell 1997; Grünwald 2007). It combines elements from statistics, Shannon information, and Kolmogorov complexity to formulate a strategy for optimal model selection. Suppose we have a set U

of possible models and a set of observations D. What would be the best model to select from U? Below I give the fundamental derivation:

$$\begin{aligned}
M_{\mathrm{map}} &= \arg\max_{M \in U} \frac{P(M)P(D|M)}{P(D)} \\
&= \arg\max_{M \in U} P(M)P(D|M) \\
&= \arg\max_{M \in U} \log P(M) + \log P(D|M) \\
&= \arg\max_{M \in U} -\log P(M) - \log P(D|M)
\end{aligned}$$

In the first line, we use the well-known rule of Bayes to select the model M_{map} that finds the maximum a posteriori probability of the M given the data set D. The operator argmax simply tests all possible models in U and selects the one with the highest score. Since we use the same data set D for each test we simplify the expression in step 2. In step 3 we make a harmless shift from a product to a sum of the logarithms. In step 4 we choose to minimize the negative logarithms instead of maximizing the positive sum. The rationale for this derivation is the insight that the two elements of the equation can be interpreted in the following way:

- log P(M) is the length of the optimal Shannon code for the model, also known as the *Model code*.
- log P(D|M) is the length of the optimal Shannon code for the data given the Model, also known as the *Data-to-Model* code

The derivation in fact tells us that if we want to find the best theory to explain a set of data then we have to find the theory that compresses optimally: the sum of the optimal Shannon coding of the theory in bits plus the optimal Shannon coding of the data given the theory in bits. If we let our possible models range over the class of all finite computer programs then the optimal theory can be expressed in terms of Kolmogorov complexity as:

$$M_{\mathrm{map}} = \arg\min_{M \in U} C(M) + C(D|M)$$

This gives the theoretical optimal solution for theory selection. Now there are a lot of philosophical and theoretical problems with this derivation (Adriaans and Vitányi 2009), but for the moment it suffices to say that, at least from a theoretical point of view, information theory has succeeded in formulating a mathematically sound solution to the general induction problem.

Since Kolmogorov complexity is non-computable we have to rely on approximations to apply this theory in practical situations, but work of Cilibrasi and Vitányi (2006; Vitányi et al. 2008) and Adriaans (2009) has shown that using industrial data compression programs is a viable option for a rich class of problems. The mutual information between two objects for instance can be approximated with industrial compression algorithms using the so-called normalized compression distance (NCD):

$$\mathrm{NCD}(x, y) = \frac{C(xy) - \min(C(x), C(y))}{\max(C(x), C(y))}$$

Here $C(xy)$ indicates the complexity of the concatenation of x and y. The NCD of x and y is 'roughly' 1 when x and y have no information in common and is 'roughly' 0 if $x=y$.

Another fundamental theoretical tool is the Kolmogorov structure function also called the maximum-likelihood estimator:

$$h_x(i) = \min_S\{\log d(S) : x \in S, K(S) \leq i\}$$

This function allows us to select a model S containing data x, with $K(S) \leq i$, where i is a natural number and $K(S)$ is the prefix-free Kolmogorov complexity of S. The expression $h_x(i)$ minimizes the maximal data-to-model length (Li and Vitányi 2008, pg 405). The Kolmogorov structure function allows us to find optimal models of bounded complexity. To give an (simplified) example: suppose the tourist-information service wants to make a model of the city of Amsterdam: e.g., a map for tourists. Any non-information-loss model of Amsterdam would still be extremely complex (probably in the order 10^{26} bits) so this is a bit unpractical. They decide to restrict the size model to 100K bits. The expression $h_A(10^5)$, where A=a non-loss-description-of-Amsterdam, would allow them to decide which map of this size is the best.

These examples show that modern information theory is much more powerful than a superficial reading of Shannon's results would suggests. Especially in the last decade we have seen many important developments that allow us to study central notions of theory of knowledge with unprecedented mathematical rigor.

4 A Short Historical Digression

The term information as a technical term in philosophy originally was motivated by Plato's theory of ideas or forms. 'In-formare' for Cicero was a technical term associated with process of knowing the world by means of 'planting' forms in one's mind (Adriaans and Van Benthem 2008, pg. 4). In the sixteenth century it emerged in various European languages with a much more colloquial meaning: a man of information simply was a learned man. Still, in the Latin writings of Descartes we find the terms 'informare' and 'forma' in this technical setting.[3] Given the formal nature of this text which sets out to present a geometrical proof of God's existence one can safely say that Descartes uses these terms in their technical sense here, although he never uses the word 'information' in this context. This changes in the works of Locke and Hume. Locke sometimes uses the phrase that our senses 'inform' us about the world and occasionally uses the word 'information'.[4] In the works of Hume the term 'information' finally seems to have

---

[3] Notably in the short formal proof of God's existence in the second answer to Mersenne in the Meditationes de Prima Philosophia "I use the term idea to refer to the **form** of any given thought, immediate perception of which makes me aware of the thought." (Idea nomine intelligo cujuslibet cogitationis **formam** illam, per cujus immediatam perceptionem ipsius ejusdem cogitationis conscious sum.) I call them 'ideas' says Descartes "only in so far as they make a difference to the mind itself when they **inform** that part of the brain." (sed tantum quatenus mentem ipsam in illam cerebri partem conversam **informant**). (Descartes 1641, Ad Secundas Objections, Rationes, dei existentiam & anime distinctionem probantes, more Geometrico dispositae.)

[4] For what **information**, what knowledge, carries this proposition in it, viz. 'Lead is a metal' to a man who knows the complex idea the name lead stands for? (Locke 1689, VIII, 4).

acquired its modern meaning.[5] In the eighteenth century one further sees an explosion of essays that, in the context of the empirical model of the human mind, study the origin of language and Floridi's question 'how meaningless signs (i.e., data) get their meaning'. Among philosophers that have written about this already in the eighteenth century are: Rousseau, Diderot, Vico, Herder, and Monboddo.

Ever since its early conception, the ambition to find a foundation for science as a set of true statements (or true models, pictures, theories etc.) about reality has been the holy grail of philosophy. A problem with this ambition is that people have held a lot of things to be true that later turn out to be untrue (e.g., that the sun circles around the earth, that heavy object fall faster) and even that people have held true beliefs for the wrong reasons. Descartes introduced his famous demon that tries to lead us into false beliefs about mathematical statements. Locke and Hume criticized our ambition to formulate apodictic true statements about causation and identity. This lead Kant to his well-known transcendental program, in which he tried to justify the core of scientific knowledge of his time in the interplay between reason and intuition which produces synthetic a priori true judgments about the structure of reality (e.g., the judgment that space is Euclidian, which later turned out to be untrue). A major breakthrough was the contribution of Frege, who, in his Begriffschrift (Frege 1879), developed a modern theory of the proposition and introduced quantifiers. This lay the groundwork for a mathematic treatment of logic, a challenge that was taken up by Russell and Whitehead. The conception of a proposition frees logic from the Kantian idea that it had to deal with true judgments. The consequent debate has given rise to a plethora of truth theories in the history of twentieth century philosophy (for a good overview, see Haack 1978). Based on contributions of philosophers like Popper, Kuhn, Feyerabend, Lakatosh, Stegmuller, and Sneed in the middle of the twentieth century the common view among scientist is that scientific theories never can claim to be true definitively. What we can only do is try to find and select the best theory that fits the data so far. When new data are gathered, the current theory is either corroborated or, when the data are in conflict with the theory it has to be revised. The best we can reach in science is provisional plausibility. This is effectively the position of mitigated skepticism that is defended by Hume. This methodological position fits perfectly with the recent insights in philosophy of information, notably the theory of general induction that has been initiated by Solomonoff and his theory of algorithmic probability which is a cornerstone of modern information theory. Algorithmic information theory has helped us to formulate the issues around probability, a priori probability, and model selection with much rigor (Adriaans and Vitányi 2009). Floridi ignores these results from information theory in his philosophy of information. Instead he focuses on a pretty elementary interpretation of Shannon's early work from 1949, consequently declares these results irrelevant for his own research agenda and leaves it at that. None of the fundamental results concerning complexity theory, minimum description length theory, Kolmgorov's structure function, identification in the limit, PAC learning, NP-hardness

---

[5] Two objects, though perfectly resembling each other, and even appearing in the same place at different times, may be numerically different: and as the power, by which one object produces another, is never discoverable merely from their idea, it is evident cause and effect are relations, of which we receive **information** from experience, and not from any abstract reasoning or reflection. (Hume 1739, Part III, section 1.)

etc. have had any impact on Floridi's thinking. In this research program (that is transcendental in flavor) the central notions are not probability and model selection but truth and justification. In this sense Floridi's proposal for a semantic theory of information can be interpreted as a revival of the Kantian ambition to study the structure of true judgments, rather than the development of a mathematical theory about relations between propositions. Also his central research question 'How do data acquire meaning' has a definite Kantian flavor: Ohne Anschaung sind unsere Begriffe leer.

It would not be difficult to reframe Floridi's ideas in terms of modern theories about the relation between syntax and semantics by researchers like Morris, Chomsky, and Montague. Morris (1938) formulated the well-known distinction between: syntax, the study of the combination of signs, semantics, the study of the meaning of signs and pragmatics, the study of the use of signs. In order for a statement to be true, it must be well-formed, e.g., syntactically correct, it must have a proper meaning. So in an example of Chomsky (1957, pg 15) 'Colorless green ideas sleep furiously' we have a sentence that is syntactically correct but semantically inconsistent and therefore we cannot imagine how it could be true (although some philosophers have argued that it is simply false). Montague later worked out this approach in a well-known set of papers in which he tried to model fragments of the English language as a formal system:

> "There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed, I consider it possible to comprehend the syntax and semantics of both kinds of language within a single natural and mathematically precise theory. On this point I differ from a number of philosophers, but I agree, I believe, with Chomsky and his associates." (Montague 1970)

In the past 40 years the philosophical consequences of this work have been extensively studied by philosophers like Partree, van Benthem, Groenendijk and Stokhof, ter Meulen, Thomason etc. (see van Benthem and ter Meulen 1997, for an overview). If one plugs in the definition of semantic information as information as well-formed, meaningful, and truthful data in to this research one would get a potentially rich theory of semantic information. Also the philosophical problems concerning the semantic theory of information could immediately be formulated in a proper rigid framework. It is a bit unfortunate that Floridi tries to develop this whole theoretical building in his writings from scratch.

5 The Problem of a General Definition of Information

In a lot of popular writings one can find the idea that in terms of information theory there are three layers one can distinguish: data, information, and knowledge. A small survey I did among colleagues a couple of years ago revealed that almost none of them took these distinctions to be very clear or valuable, but from the writings of Floridi we learn that there exist 'consensus' on a general definition of information (GDI) as: data + meaning (Floridi 2005a, b). According to Floridi GDI implies a thesis of ontological neutrality: that there can be no information without data representation. One might defend this thesis, but there is from the perspective of

information theory no necessity to do this. In the theory of Kolmogorov complexity all natural numbers carry a certain amount of information that is given by the integer complexity function. The view one wants to defend concerning the ontological status of information thus ultimately is dependent on the view one wants to adopt on the ontological status of (natural) numbers. Natural numbers contain information regardless of their representation. In fact the big-Oh notation has been used in computer science already for decades to abstract away from these representational problems. I will not discuss this issue further here but the definitions introduced by Floridi are certainly not ontologically neutral. Something similar happens in the context of semantics.

6 The Problem of Semantic Information

If we take the estimate of Seth Lloyd (2000) as a starting point, then the total information capacity of the universe is 10^{92} bits. An interesting question is what the actual complexity of the universe is. We have laws of nature, not all configurations are possible, there are patterns, so in the sense of information theory our universe is compressible. Suppose for the sake of argument that our universe is extremely compressible with a factor of 10^{30}. In this case an optimally compressed description of our universe will still contain 10^{62} bits. What is more, since this description is optimally compressed it will be a random string. There will be no patterns and no distinguishable relations between parts of the description. Such a description of our universe will have no sensible meaning. This is an interesting consequence of this information theoretic perspective on our universe. From an information theoretical point of view most facts in our world cannot be explained scientifically.

Suppose some omniscient demon would write an optimally compressed book containing every possible fact that one could know about our universe. This book would have two parts:

- Part one would contain an optimally compressed description of the Laws of Nature
- Part two would contain a factual optimally compressed description of our world given the laws of nature

Since the text of the book is optimally compressed there would be no structure left: the ultimate description of our universe in this sense would consist of two random strings. In fact the two strings are simply two indexes that describe our specific universe in the hypothetical set of all possible universes that could exist. What is more since the two strings do not contain any mutual information we could easily change facts in part two without affecting the consistency of the whole. Any random edit of the Part two description would also lead to a viable description of a universe.

It is the ambition of science to explain and to predict. Ideally, we gather data, formulate a theory, test the theory on new data, and update the theory if necessary. An essential element of the scientific method is that this process must be open for public inspection. The data must be publicly available and the experiments must in principle be repeatable. This implies that a lot of individual facts and events are not

open for scientific validation. I have for instance a nice happy early childhood memory: I was sitting behind my mother while she was riding her bike in the town of Middelburg. This event probably took place somewhere in 1958. I do not know why this particular event got stored in my brain, but what I do know is that the validity of this memory as it occurred to me as an individual can never be validated scientifically. With enough funding, scientist could try to reconstruct what the town of Middelburg looked like 50 years ago, whether my mother had a bike etc. But, no matter how much information they gather, the specific individual event as I perceived it can never be reconstructed. Most things that are meaningful for individual people, their memories, events that overcome them, their loves, their believes are fundamentally closed for scientific inquiry. It is exactly this dimension that make our lives meaningful. I call this dimension of factual data in which our existence is embedded, with a term I derive from Heidegger, the dimension of facticity. It is quite surprising that this notion of facticity that is closely related to Sartre's notion of the absurd, and is supposed to belong to the more obscure phenomena in Heidegger's philosophy, emerges in the context of a philosophy of information (Heidegger 1977; Gadamer 1975). It is possible to develop a theory of informational esthetics (Adriaans 2009) on the basis of this formal notion of facticity (with the central insight that good artists optimize facticity).

These observations serve to illustrate the fact that semantics is not something that stands aside from information theory and can be plugged in at will, as happens in Floridi theory of semantic information. Theory of information inherently implies a treatment of the notion of meaning. The observation that the theory of Shannon does not explain the phenomenon, which is correct, does not imply that we have to develop a special of semantic information in order to repair this deficit. It is true that Shannon information measures the amount of information in terms of a scalar value. For a certain observer the messages: 'John passed his exam' and 'It will rain tomorrow' can contain the same amount of information measured in bits, but this does not imply that they contain the same information. Much in the same way a book and a meal both can both cost 20 Euros, without the implication that one can eat a book or read a meal. Modern information theory has much more powerful techniques to analyze meaning. Conditional probabilities can perfectly be used to analyze the validity of definitions for instance: if P(John is a bachelor | John is an unmarried male) = P(John is an unmarried male | John is a bachelor) =1 this indicates that the messages 'John is a bachelor' and 'John is an unmarried male' in the context of this specific set of messages not only carry the same amount of information, but have the same meaning. Thus a formal treatment of semantic information in the context of classical information theory seems possible (Adriaans 2009).

7 The Problem of Truth

A central role in Floridi's philosophy is played by the so-called Gettier problems (Gettier 1963): suppose that we say that an agent S 'knows that p' if and only if:

1. p is true
2. S believes that p

3. S is justified in believing that p. Sometimes a fourth condition is added
4. *S*'s belief that p is not inferred from any falsehood (Steup 2006)

The problem is that there are clear situations in which all four conditions are met but where we are reluctant to say that S knows p. The following example is due to Goldman (1976) and cited in Steup (2006):

> Suppose there is a county in the Midwest with the following peculiar feature. The landscape next to the road leading through that county is peppered with barn-facades: structures that from the road look exactly like barns. Observation from any other viewpoint would immediately reveal these structures to be fakes: devices erected for the purpose of fooling unsuspecting motorists into believing in the presence of barns. Suppose Henry is driving along the road that leads through Barn County. Naturally, he will on numerous occasions form a false belief in the presence of a barn. Since Henry has no reason to suspect that he is the victim of organized deception, these beliefs are justified. Now suppose further that, on one of those occasions when he believes there is a barn over there, he happens to be looking at the one and only real barn in the county. This time, his belief is justified and true. But since its truth is the result of luck, it is exceedingly plausible to judge that Henry's belief is not an instance of knowledge. Yet condition (iv) is met in this case. His belief is clearly not the result of any inference from a falsehood.

Clearly such counter examples form a problem for a philosophical program that defines knowledge as: "*relevant semantic information properly accounted for*". Consequently, Floridi goes to great length to analyze the Gettier problems (Floridi 2004). By means of a clever reduction of the problem to the 'coordinated attack' problem he is able to defend the position that the Gettier problem is logically unsolvable. This result then becomes a cornerstone of his philosophical system which forces him to formulate a 'non-doxastic foundation of knowledge' that denies the prima facie plausible principle that "knowledge that p implies a belief that p" (Floridi 2006). The project then evolves via a network theory of account in to a refutation of digital ontology and a defense of Informational structural realism. But these excursions are less relevant here because they are more or less directly implied by the original definition of semantic information.

I agree with Floridi that the Gettier problem is fundamental and I also agree that as a problem in epistemology it is unsolvable. As a philosopher of information I am less impressed by the directions his project takes after these observations, since specifically within the framework of the theory of information a completely different analysis of the Gettier problem is possible. In this view, information theory is not so much a servant of classical epistemology, but more a competitor. Within the context of information theory, the problem of *founding knowledge as true justified belief* is replaced by the problem of *selecting the optimal model that fits the observations*. As we have seen in recent years, research in information theory has come up with various approaches to formulate optimal strategies for model selection (Grünwald 2007; Adriaans 2008; Adriaans and Vitányi 2009). So let us see what Information theory has to offer in terms of an interpretation of the Gettier paradox.

First of all observe that Information theory is not dealing with the correct interpretation of the modal operator 'Knows'. It simply gives rules for selecting 'the best model'. If somebody wants to add under certain conditions the propositional attitude "I know that" to a statement 'p' that has been chosen as the best model explaining the observations so far, he or she is free to do that. According to this interpretation of information theory this extra label 'I know that p is true' does not really add anything to the statement 'p' by the same agent. In this sense it is comparable to Ramsey's redundancy theory of truth. In this context, a general informational framework for induction emerges:

- Given a set of non-contradictory facts (observations or data) D there is always an infinite number of finite theories/models that explain these data. Let's call this domain U.
- The more data we have the less theories/models M fit the data (Note that an infinite set of finite objects still can have an infinite number of proper subsets). Gathering more data will decrease the number of theories that fit the data, the viable theories will be less dense in the total space of possible theories although the total amount is still infinite.
- Given an appropriate fitness measure (e.g., randomness deficiency, minimum description length) the theories/models can be ordered in terms of their plausibility, probability, appropriateness (note that such a measure not necessarily needs to be computable, it can also be approximated by resource bounded procedures). The optimal theory at any moment is given by:

$$M_{\text{map}} = \arg \min_{M \in U} C(M) + C(D|M)$$

which is uncomputable but can be approximated. What kind of cognitive compression mechanisms take place within Henry's brain is unclear at this moment, but it is reasonable to suppose that data compression is a basic capacity that underlies human cognition (Wolff 2006; Chater and Vitányi 2003).

- Given these insights the best we can do is: codify our observations, codify our theories, compute the plausibility scores, and select the theory with the best score so far. In other words, when confronted with his false or unjustified beliefs, Henry can draw on fundamental results of information theory to justify his shifts in beliefs by saying that more complex data allow him to formulate more complex theories (i.e. the theory that he is passing the set of a movie). Henry is only in trouble when a traditional Kantian philosopher asks him to formulate his beliefs in terms of apodictic knowledge that cannot be retracted.

If we apply this framework to the barn-façade problem then the following situation arises. In principle Henry has access to an infinity of possible explanations of his observations. Some are really farfetched: he might be mislead by a Cartesian demon, he might be a solipsist, he might be dreaming, he might be playing the main character in some sort of Trueman show. Others are more realistic: he might be passing the set of a movie or an exhibition of the national contest of barn-façade designers. The best theory however that he could select in this case is, given his general experience as a human being, that when he sees something that looks like a

barn it really is a barn. Now there is always some room for doubt. Such a judgment is never absolutely true. If Henry leaves his car, inspects the barns from close by and finds out that there is really only a façade, he will update his theory. The crux of the argument is that with the formal approach of information theory Henry can perfectly justify this shift from one theory to another, regardless of the fact whether he is using the operator 'to know'. It is clear that when one reserves the propositional attitude "to know that _" only for models that fit the observations with absolute certainty, as seems to be the ambition behind Floridi's concept of semantic information, it can never be used in the context of model selection based on insights of information theory. This is a fact of life that we have to live with, regardless what philosophical ambitions we have.

The propositional attitudes "to know that p" and "to believe that p" seem to be dialogical markers that indicate to what extend we want to base our actions on the truth of proposition p. We tend to use the operator "to know that p" when we claim general validity of the proposition p, we use the weaker operator "to believe that p" we are personally ready to base our actions on the truth of p without any intention to ask the same from others. In the interpretation the statement "John knows that p" implies the statement "John believes that p". The crux of the matter is that information theory helps us to select the most plausible model regardless of our epistemological interpretation of the situation. It is exactly in this sense that the approach of information theory is orthogonal to the classical epistemological program.

This is not to say that the information theoretical research program is without problems. There are a number of problems. To name a few:

- How can we find good approximations for the computational routines that govern model selection?
- How can the general experience of human agents be codified in terms of probability distributions over possible observations?
- Which hidden logical, physical, biological and cultural bias govern our cognitive capacities?
- What is the cognitive relevance of these distributions?

These are some of the central philosophical research questions of philosophy of information.

8 Conclusion

Let's return to the example of the tourist information that we started with. Although initially the idea emerged that information only was real information when it was syntactically well formed, meaningful, and truthful, it is now clear that, although these things are be pragmatically true, they do not force us to develop a specific theory of semantic information. The adequacy and truthfulness of models, the notion of meaning, the well-formedness of syntactic structures all can well be explained using the rich toolbox of modern information theory with such diverse tools as:

- The minimum description length principle
- Kolmogorov's structure function and
- Normalized compression distance

This makes the notion of semantic information from the perspective of information theory somewhat superfluous. One cannot forbid philosophers to study it, but it would be somewhat of a caricature to present such research under the ambitious label of '*the* philosophy of information'. The risk is that the real questions that are in the heart of philosophy of information get less attention than they deserve. Instead of studying semantic information I urge the students of philosophy of information to direct their endeavors to such problems as:

– The nature of various probability distributions that dominate logical, physical, biological end cultural domains
– The interaction between information and computation
– The approximation of various compression measures
– The study of cognition and learning as data compression

References

Adriaans, P. W. (2008). The philosophy of learning, the cooperative computational universe. In P. W. Adriaans & J. F. A. K. van Benthem (Eds.), *Handbook of Philosophy of Information*. Amsterdam: Elsevier.

Adriaans, P. W. (2009). Between Order and Chaos: The Quest for Meaningful Information, *Theory of Computing Systems*, Volume 45, Issue 4 (July 2009), Special Issue: Computation and Logic in the Real World; Guest Editors: S. Barry Cooper, Elvira Mayordomo and Andrea Sorbi, 650–674.

Adriaans, P. W., & van Benthem, J. F. A. K. (Eds.), (2008). Handbook of Philosophy of Information. Amsterdam: Elsevier

Adriaans, P. W., & Vitányi, P. (2009). Approximation of the two-part MDL code, Comput. Sci. Dept., Univ. of Amsterdam, Amsterdam. *Information Theory, IEEE Transactions on, 55*(1), 444–457.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences, 7*(1), 19–22.

Chomsky, N. (1957). *Syntactic Structures*. The Hague/Paris: Mouton.

Cilibrasi, R., & Vitányi, P. M. B. (2006). Similarity of objects and the meaning of words. *TAMC, 2006*, 21–45.

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information theory*. New York: Wiley.

Descartes, R. (1641). Meditationes de prima philosophia, in qua dei existentia et anime immortalitas demonstratur, Paris.

Floridi, L. (2004). On the logical unsolvability of the Gettier problem. *Synthese*, Volume 142, Number 1.

Floridi, L. (2005a). Is Semantic Information Meaningful Data? *Philosophy and Phenomenological Research,* Vol. LXX, No. 2.

Floridi, L. (2005b). Semantic conceptions of information. *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/information-semantic/.

Floridi, L. (2006). The Logic of Being Informed. *Logique & Analyse* 196.

Floridi, L. (2008). Trends in the Philosophy of information. In P. W. Adriaans & J. F. A. K. van Benthem (Eds.), *Handbook of Philosophy of Information*. New York: Elseviers.

Frege, G. (1879). *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle.

Gadamer, H.-G. (1975). *Wahrheit und Methode*. Tübingen.

Gettier, E. L. (1963). Is justified true belief Knowledge? *Analysis, 23*, 121–23.

Goldman, A. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy, 73*, 771–791.

Grünwald, P. D. (2007). *The Minimum Description Length Principle*. Cambridge: MIT Press.

Haack, S. (1978). *Philosophy of Logics*. Cambridge: Cambridge University Press.

Heidegger, M. (1977). *Sein und Zeit*. Tübingen: Max Niemeyer Verlag.

Hume, D. (1739). *A treatise of human nature*.

Li, M., & Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications*, Third Edition, Springer Verlag.

Locke, J. (1689). *Essay concerning human understanding*.

Lloyd, S. (2000). Ultimate physical limits to computation. *Nature, 406*(8), 1047–1054.

Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.

Montague, R. (1970). Universal grammar. *Theoria, 36*(1970), 373–398. reprinted in Thomason, 1974.

Morris, C. W. (1938). *Foundations of the Theory of Signs*.

Shannon, C. E. (1993). *Collected Papers*. In: N. J. A. Sloane, A. D. Wyner (Eds.), New York: IEEE Press.

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication.* Univ. of Illinois Press.

Steup, M. (2006). The analysis of knowledge, *Stanford encyclopedia of philosophy*, http://plato.stanford.edu/entries/knowledge-analysis/

van Benthem, J. F. A. K., & ter Meulen, A. G. B.(eds.), (1997). *Handbook of Logic and Language*. Elsevier/MIT Press.

Vitányi, P. M. B., Balbach, F. J., Cilibrasi, R., & Li, M. (2008). Normalized Information Distance, *CoRR* abs/0809.2553.

Wolff, J. G. (2006). Unifying *computing and cognition, The SP Theory and its Applications*, CognitionResearch.org.uk.