System Modeling and Information Semantics

Gordana Dodig-Crnkovic

Department of Computer Science and Engineering
Mälardalen University
Västerås, Sweden
gordana.dodig-crnkovic@mdh.se
http://www.idt.mdh.se/personal/gdc/

Abstract. Floridi´s Theory of Strongly Semantic Information defines information as consisting of data and truth in contrast to the standard definition prevailing in empirical sciences in which information is defined as meaningful data. I argue that meaningful data does not necessarily need to be true to constitute information. Partially true information or even completely false information can lead to an outcome adequate and relevant for inquiry. Instead of insisting on the truth of an empirical model, the focus is on basic criteria such as the validity of the model and its appropriateness within a certain well-defined context, as the meaning of the information content of the model is strongly contextual. Even though empirical models could in general only be 'adequate' and not 'true' they may produce results and data from which relevant conclusions could be drawn. If truthlikeness admits of degrees, then the history of inquiry is one of steady progress towards the truth. In that sense models can generate information for improving our knowledge about the empirical world.

1 Introduction

The ubiquitous use of computers has led to the informatisation of society, science, technology and culture in general. Information defines our time, commonly recognized as the Information Era; it is steadily flooding us from innumerable sources. At a fundamental level information can be said to characterize the world itself, for it is through information we gain all our knowledge - and yet we are only beginning to understand its real meaning. [2] Information is to replace matter as the primary stuff of the universe, as von Baeyer [1] suggests; it will provide a new basic unifying framework for describing and predicting reality in the twenty-first century.

As the global informatisation process continues, a theory of information systems (IS) has been found to be necessary. The International Federation for

Information Processing (IFIP) task group "FRamework of Information System COncepts" (FRISCO) has taken a first step in the development of a theory of IS by publishing a report [14]. This report has a solid foundation formed by semiotics and ontology and contains a compendium of core IS concepts and a broad interdisciplinary socio-technical view of IS.

Concurrently, computer based representations, simulations and emulations, modeling, and model-based reasoning are gaining in importance as a consequence of the increasing availability of computer power. Science and technology, businesses and organizations, administration and the financial world all now rely on computer-based models of systems which they find essential in describing, understanding, predicting and controlling their different activities.

This paper addresses the relation between the concepts of reality, model, information and truth.

1.1 System Modeling and Simulation Validation

A system is an entity which maintains its existence through the mutual interaction of its parts. A system exists and operates in time and space. Examples of systems are elementary particles, atoms, molecules, cells, organs, organisms, eco-systems, planets, solar systems, galaxies, universe(s) - in increasing order of complexity. [http://www.systems-thinking.org/systems/systems.htm]

A model may be defined as a simplified description of a system, generally developed for an understanding of, or the analysis, improvement and/or substitution of the system. The model can be used to obtain information about the system and for making predictions about the system behaviour as a result of its activities, relationships and constraints.

A model is the *actual data representation* of the *information at hand* and enables an analysis of the possible effects of changes in the model resulting from changes in the process which it represents, according to our best knowledge. Models are analytical tools (often a system of postulates, data, and inferences presented as a mathematical description of a phenomenon such as an actual system or process) used to assist in generating descriptions and forecasts and to facilitate control.

Experimentation is the backbone of scientific thinking and the sine qua non technique for scientific method. Conducting experiments allows us to go beyond the limits of Aristotelian logic in our investigation of the empirical world.

Simulation is time-dependent, goal-directed experimentation with a dynamic model. When experimentation cannot or should not be performed on a real system, simulation with a dynamic model can be used as a substitute. From a systemic point of view, simulation can be used to find the values of one of the output, input, or state variables of a system; provided that the values of the two other types of the variables are known. Correspondingly, simulation can be used in solving analysis, control, and design problems. The uses of simulation range from gaining insight, testing theories, experimentation with control strategies to prediction of action and performance. [21, 22]

In the concept of simulation as a model-based computational activity, the emphasis is on the *generation of model behavior*. Viewed as a knowledge-generation activity, simulation can be interpreted as model-based experiential knowledge-generation. [23, 24]. Thus, simulation can be combined with other types of knowledge-generation techniques such as optimization, statistical inferencing, reasoning and hypothesis-processing.

> "Simulation is used to support important policies and decisions. For example, in nuclear fuel waste management systems, simulation is used to study (even several millennia) long-term behavior of nuclear fuel waste. Simulation of safety-critical systems is one of the important application areas of simulation. Currently, simulation is also used in simulation-based acquisition as well as simulation-based prototyping affecting millions of dollars of investments. Simulation has the potential of surpassing its own abilities of being an off-line decision making tool to be also an on-line decision support tool for complex and important problems. The existence of several validation, verification, and accreditation (VV&A) techniques and tools also attests to the importance of the implications of simulation (Davis, 1992)." [23]

Validation: The essence of model validation is the determination of whether or not a model is an appropriate representation of the reality, for a clearly specified goal. There are also other relevant matters to be considered – such as the acceptability of the goal of the study and the experimental conditions. [25]

System modeling, simulation and validation have interesting epistemological aspects that will be addressed in more detail.

2 Semantics

In general, semantics is the study of meaning. Semantics is opposed to syntax, in which case the former is about what something means while the latter represents the formal structure in which something is expressed (Wikipedia, http://en.wikipedia.org/wiki/Semantics)

Semantics thus concerns the relation between the expressions of language and their meanings. But there is a long-standing philosophical dispute concerning the *meaning of meaning*, Gärdenfors [15], with two general traditions in semantics, realistic and cognitivistic. According to the realistic approach to semantics the meaning of an expression is something out there in the world. Realistic semantics comes in two variants: extensional and intensional. In the extensional type of semantics, one begins from a language L, and maps the constituents of L onto a "world". The objective of this kind of semantics is to determine truth conditions for sentences in L. Frege's semantics and Tarski's theory of truth are examples of extensional semantics. In intensional semantics the language L is mapped onto a *set of possible worlds* instead of a single world. The meaning of a sentence is taken to be a proposition which is identified with a set of possible worlds in which the sentence is true. The classic form of this semantics is Kripke's semantics for modal logics.

In a cognitivistic paradigm of semantics the meanings of expressions are mental entities. Semantics is a mapping from the linguistic expressions to cognitive structures. Language itself is seen as part of the cognitive structure, and not as an independent entity. Semantics is thus a relation between mental entities. The external world and truth conditions appear only when the relation between the world and the cognitive structure is considered.

For the sake of completeness we should mention the philosophy of the later Wittgenstein, whose argument in the Philosophical Investigations [30]: "Meaning just is use." is an example of semantic pragmatism, a turn from *mirror-like* to *tool-like* language.

2.1 Information

> "In the beginning there was information. The word came later. The transition was achieved by the development of organisms with the capacity for selectively exploiting this information in order to survive and perpetuate their kind." Dretske [7]

According to Floridi [13] recent surveys have shown no consensus on a single, unified definition of semantic information: "This is hardly surprising. Information is such a powerful and elusive concept that it can be associated with several explanations, depending on the requirements and intentions." He quotes Claude Shannon [27]:

> "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."

As he concludes that information can solely be defined in relation to a well-specified context of application, Floridi analyses only one specific type of information, namely the *alethic*, (pertaining to truth and falsehood) *declarative, objective and semantic (DOS) information*.

The standard work of Carnap and Bar-Hillel [3-5] used inductive logic to define the information content of a statement in a given language in terms of the possible states it rules out. The basic idea is that the more possibilities (possible states of affairs) a sentence rules out, the more informative it is, i.e. information is the elimination of uncertainty. The information content of a statement is thus relative to a language. Evidence, in the form of observation statements, (Carnap's “state descriptions", or Hintikka's “constituents") contains information by virtue of the class of state descriptions the evidence rules out. (*The very essential underlying assumption is that observation statements can be related to experience unambiguously*.)

Carnap and Bar-Hillel have suggested two different measures of information. The first measure of the information content of statement S is called the content measure, cont(S), defined as the complement of the a priori probability of the state of affairs expressed by S

$$\mathrm{cont(S)} = 1 - \mathrm{prob(S)}$$

Content measure is not additive and it violates some natural intuitions about conditional information. Another measure, called the information measure, inf(S) in bits is given by:

$$\mathrm{inf(S)} = \log_2(1/(1-\mathrm{cont(S)})) = -\log_2 \mathrm{prob(S)}$$

prob(S) here again is the probability of the state of affairs expressed by S, not the probability of `S' in some communication channel. According to Bar-Hillel cont(S) measures the substantive information content of sentence S, whereas inf(S) measures the surprise value, or the unexpectedness, of the sentence H.

Although inf satisfies additivity and conditionalisation, it has a following property: If some evidence E is negatively relevant to a statement S, then the information measure of S conditional on E will be greater than the absolute information measure of S. This violates a common intuition that the information of S given E must be less than or equal to the absolute information of S. This is what Floridi [13] calls the Bar-Hillel semantic paradox.

However, a more serious problem with the approach is the linguistic relativity of information, and problems with the Logical Empiricist program that supports it, such as the *theory-ladenness of observation*, Collier [28].

For recent semantic theories such as Dretske [7], Barwise and Perry [6], Devlin [8, 9], see http://www.nu.ac.za/undphil/collier/information/.

The conclusion Floridi draws is that SDI needs to be revised by adding a necessary truth-condition in order to inter alia avoid the Bar-Hillel semantic paradox. Non-declarative meanings of “information”, e.g. referring to graphics, music or information processing taking place in a biological cell or a DNA molecule, such as defined in Marijuán [29] are not considered in this paper, in which objective semantic information is taken to have a declarative or factual value i.e. it is supposed to be correctly qualifiable alethically.

2.2 The Standard Definition of Information

According to Floridi´s account [10-13] the most common definition of DOS information established over the last three decades:

Information = Data + Meaning.

can also be found in Devlin [8, 9], along with a variation on the theme:

Information = Representation + Procedure for Encoding/Decoding.

Or in terms of situation semantics [6]

Information = Representation + Constraint.

"Constraint" is the term used by situation semanticists to refer to the regularities and conventions that enable some configurations of objects to represent or store information.

2.3 Floridi's Theory of Strongly Semantic Information

Floridi [13] contributes to the current debate by criticizing and revising the Standard Definition of declarative, objective and semantic Information (SDI) as meaningful data. The main thesis presented is that meaningful and well-formed data constitute information *only if it also qualifies as truthful*. SDI is criticized for providing insufficient conditions for the definition of information. SDI is in this view incorrect because information is not necessarily true, and misinformation (i.e. false information) is treated as information. Floridi argues strongly against misinformation as a possible source of information or knowledge. As a consequence, SDI should be revised to include a necessary truth-condition.

Floridi's quantitative theory of strongly semantic information is based on truth-values in contrast to the classic quantitative theory of weakly semantic information of Bar-Hillel and Carnap [3] which is based on probability distributions. The classic theory assumes that truth-values supervene on information, yet this principle is found to be too weak and generates the semantic paradox noted by Floridi [13] as mentioned before.

> "A triangle has four sides": according to the classic theory of semantic information, there is more semantic content in this contradiction than in the contingently true statement "the earth has only one moon".

As a remedy, Floridi’s concept of strongly semantic information contains truth from the outset and thus avoids the paradox and is consistent with the common usage of the word information.

A dilemma is apparent here in making a choice between the two definitions of information; the weaker one that accepts meaningful data as information, and the stronger one that claims that information must be true in order to qualify as information. Both approaches will prove to have legitimacy under specific circumstances, and I will try to illuminate why the general definition of information does not explicitly require truth from the data.

3 View from the Scientific Window: Truth, Truthlikeness, Information

Plato argued that since knowledge requires certainty, and certainty requires an unchanging subject matter, true knowledge can only be of unchanging forms. It means that there can be no true knowledge of the ever changing physical world, which is the central tenet of modern physics. From the outset, there is no way to reach *absolute certainty* about the physical world, but the problem is more complex than that of absolute certainty.

Science is nevertheless accepted as one of the principal sources of “truth” about the world. It might be instructive to see the view of truth from the scientific perspective. When do we expect to be able to label some information as “true”? Is it possible for a theory, a model or a simulation to be “true”? When do we use the concept of truth and why is it important?

Popper was the first prominent realist philosopher and scientist to adopt a radical fallibilism about science (fallibilism claims that some parts of accepted knowledge could be wrong or flawed) proclaiming at the same time the epistemic superiority of scientific method. Popper was the first philosopher to abandon the idea that science is about truth and take the problem of truthlikeness seriously. In his early work, The Logic of Scientific Discovery, [26] Popper implied that *the only kind of progress an inquiry can make consists in falsification of theories.*

Now how can a succession of falsehoods constitute epistemic progress? Epistemic optimism would mean that if some false hypotheses are closer to the truth than others, if truthlikeness (verisimilitude) admits of degrees, then the history of inquiry may turn out to be one of steady progress towards the goal of truth. [20]

> “While truth is the aim of inquiry, some falsehoods seem to realize this aim better than others. Some truths better realize the aim than other truths. And perhaps even some falsehoods realize the aim better than some truths do. The dichotomy of the class of propositions into truths and falsehoods should thus be supplemented with

> a more fine-grained ordering -- one which classifies propositions according to their closeness to the truth, their degree of truthlikeness or verisimilitude. The problem of truthlikeness is to give an adequate account of the concept and to explore its logical properties and its applications to epistemology and methodology."

Kuipers [16] developed a synthesis of a qualitative, structuralist theory of truth approximation:

> "In this theory, three concepts and two intuitions play a crucial role. The concepts are confirmation, empirical progress, and (more) truthlikeness. The first intuition, the success intuition, amounts to the claim that empirical progress is, as a rule, functional for truth approximation, that is, an empirically more successful theory is, as a rule, more truthlike or closer to the truth, and vice versa. The second intuition, the I&C (idealization and concretization) intuition, is a kind of specification of the first."

According to Kuipers [16-19] the truth approximation is a two-sided affair amounting to achieving 'more true consequences and more correct models', which obviously belongs to scientific common sense.

4 Conclusion

In the standard definition of semantic information commonly used in empirical sciences, information is defined as meaningful data. Floridi in his new Theory of Strongly Semantic Information adds the requirement that standard semantic information should also contain truth in order to avoid the Bar-Hillel logical paradox.

This paper argues that meaningful data need not necessarily be true to constitute information. Partially true information or even completely false information can lead to an outcome adequate and relevant for inquiry, as the history of empirical sciences has shown. If truthlikeness admits of degrees, then the history of inquiry is one of steady progress towards the truth. In that sense models can generate information for improving our knowledge about the empirical world.

References

1. von Baeyer H.C.: Information: The New Language of Science, Weidenfeld and Nicolson (2003)
2. van Benthem J., Adriaans P., ed.: Handbook on the Philosophy of Information, (part of the Gabbay D., Thagard P., and Woods J., Handbook of the philosophy of science, to be published by Elsevier (2005)

3. Carnap, R., Bar-Hillel, Y.: An Outline of Semantic Information. In Bar-Hillel, Y. (Ed.) Language and Information: Selected Essays on Their Theory and Application. Addison-Wesley, Reading, Mass. (1964)
4. Bar-Hillel, Y.: Language and Information, Reading, MA, London, Addison-Wesley (1964)
5. Carnap, R.: Logical Foundations of Probability. University of Chicago Press, Chicago (1950)
6. Barwise, J. and Perry J.: Situations and Attitudes. MIT Press, Cambridge, MA (1983)
7. Dretske, F.: Knowledge and the Flow of Information, MIT Press Cambridge, MA (1981)
8. Devlin K.: InfoSense: Turning Information into Knowledge, Freeman (1999).
9. Devlin K.: Logic and Information, Cambridge University Press (1995)
10. Floridi, L.: Blackwell Guide to the Philosophy of Computing and Information, Blackwell Publishers, Inc., Cambridge, MA (2003)
11. Floridi. L.: Afterword - LIS as Applied Philosophy of Information: a Reappraisal, Library Trends, 52(3) (2004)
12. Floridi, L. (forthcoming): Is Information Meaningful Data?, Philosophy and Phenomenological Research, http://www.wolfson.ox.ac.uk/floridi/papers.htm.
13. Floridi, L.: Outline of a Theory of Strongly Semantic Information" Minds and Machines, 14.2, pp. 197-222 (2004)
14. FRISCO Report http://www.liacs.nl/~verrynst/fri-full-7.pdf
15. Gardenfors, P.: Conceptual Spaces--The Geometry of Thought, The MIT Press, Cambridge, Massachusetts (2000)
16. Kuipers, T.A.F.: From instrumentalism to constructive realism: on some relations between confirmation, empirical progress, and truth approximation - Dordrecht Kluwer Academic (2000)
17. Kuipers, T.A.F. (ed.) What is closer-to-the-truth? A parade of approaches to truthlikeness, Poznan Studies in the Philosophy of the Sciences and the Humanities, Volume 10, Amsterdam: Rodopi (1987)
18. Kuipers, T.A.F.: Inductive Aspects of Confirmation, Information, and Content. To appear in the Schilpp-volume The Philosophy of Jaakko Hintikka (2002)
19. Kuipers, T.A.F.: Empirical and conceptual idealization and concretization. The case of truth approximation, to appear in (English and Polish editions of) Liber Amicorum for Leszek Nowak
20. Oddie, G.: Truthlikeness, The Stanford Encyclopedia of Philosophy, E.N. Zalta (ed.), http://plato.stanford.edu/archives/fall2001/entries/truthlikeness/ (2001)
21. Wildberger, A.M.: AI & Simulation, Simulation, 74 (2000)
22. Denning, P.J.: How We Will Learn. In: P.J. Denning and R.M. Metcalfe (eds.) Beyond Calculation – The Next Fifty Years of Computation. Copernicus (Springer Verlag). New York (1997)
23. Ören, T.I.: Impact of Data on Simulation: From Early Practices to Federated and Agent-Directed Simulations. In: A. Heemink et al. (eds.) Proc. of EUROSIM 2001, June 26-29, Delft, the Netherlands (2001)
24. Ören, T.I.: Responsibility, Ethics, and Simulation. Transactions of the SCS, San Diego, CA., 17:4, (2000)
25. Davis, P.K.: Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations, RAND, R-4249-ACQ. (1992)

26. Popper, K.: The Logic of Scientific Discovery. Tenth (revised) edn, Unwin Hyman Ltd, London. Translation of *Logik der Forschung* (1980)
27. Shannon C.E.: The lattice theory of information, in Symposium on Information Theory (Report of Proceedings) (W. Jackson, ed.), (London), pp. 80, The Ministry of Supply, 1950. Held in the Lecture Theatre of the Royal Society, Burlington House, September (1950)
28. Collier, J.D.: Intrinsic Information In Philip Hanson (ed) Information, Language and Cognition: Vancouver Studies in Cognitive Science, Vol. 1, University of Oxford Press (1990)
29. Marijuan, P. C.: Foundations of Information Science: Selected papers from FIS 2002. Entropy 5, 214-219 (2003)
30. Wittgenstein, L.: Philosophical Investigations. The German text, with a revised English translation, 3rd ed., Blackwell Publishing Ltd., 1953, 1958, (2001)