

# Empirical modeling and information semantics

Gordana Dodig-Crnkovic

Received: 18 January 2006 / Accepted: 29 November 2006 / Published online: 13 March 2007  
© Fondazione Rosselli 2007

**Abstract** This paper investigates the relationship between reality and model, information and truth. It will argue that meaningful data need not be true in order to constitute information. Information to which truth-value cannot be ascribed, partially true information or even false information can lead to an interesting outcome such as technological innovation or scientific breakthrough. In the research process, during the transition between two theoretical frameworks, there is a dynamic mixture of old and new concepts in which truth is not well defined. Instead of veridicity, correctness of a model and its appropriateness within a context are commonly required. Despite empirical models being in general only truthlike, they are nevertheless capable of producing results from which conclusions can be drawn and adequate decisions made.

**Keywords** Modeling · Information · Semantics · Validation and verification · Veridicity · Truthlikeness

## 1 System modeling and simulation: validation and verification

A model is a simplified representation of a complex system or a process developed for its understanding, control and prediction; it resembles the target system in some respects while it differs in other respects that are not considered essential. It follows that a model, which is valid for one objective, may not be valid for another. Models are abstracted or constructed on the

---

G. Dodig-Crnkovic (✉)  
Department of Computer Science and Electronics,  
Mälardalen University, Vasteras, Sweden  
e-mail: gordana.dodig-crnkovic@mdh.se

grounds that they potentially satisfy important constraints of the target domain.

Model-based reasoning is essential for all sciences, particularly for the empirical. It supports conceptual change and facilitates novel insights as demonstrated in Magnani et al. (1999).

When discussing models, two concepts are fundamental: verification and validation; where model verification is the confirmation that the model is constructed as a model specification based on a problem formulation and model validation is the demonstration that the model, within its domain of applicability, is consistent with its objectives.

Consequently, the term “valid” refers to a model that adequately represents a target system in its domain of applicability. Determining whether or not a model is an appropriate representation of reality, for a well-specified goal, is the essence of model validation. The relationship between the model and the physical reality is established by conducting empirical tests. Determining whether or not a model is an appropriate representation of reality, for a well specified goal, is the essence of model validation, but there are other significant factors to be considered such as the relevance of the goal itself, Dodig-Crnkovic 2003

Simulation as a special case of modeling implies time-dependent goal-directed experimentation with a dynamic model. Simulation can be used in analysis, control, and design, Wildberger (2000). It is a tool, which facilitates the gaining of insight, the testing of theories, experimentation with strategies of control, and prediction of performance. In the concept of simulation as a model-based computational activity, the emphasis is on the generation of model behaviour. Simulation can be interpreted as model-based experimental knowledge generation, Ören (2001), and can be combined with different types of knowledge generation techniques such as optimization, statistical inference, reasoning and hypothesis processing.

Questions of interest are to what degree the results of modeling and simulation can be trusted and can they be said to generate reliable information? The former may be answered in a pragmatic way, by asking what would be the alternative to using model-based reasoning, learning and prediction techniques. In the case of weather forecasting, for example, we know that the reliability of the prediction is not extremely high, but it is improving, and it should be compared to a pure guess, which obviously is a less successful prediction method. The output of a model for producing weather forecasts may be seen as information that is probable but not certain (true), yet necessary and useful.

## **2 Construction Process: data → information → knowledge**

Data is generally considered to be a series of disconnected facts and observations. These may be converted to information by analyzing, cross-referring, selecting, sorting, summarizing, or in some way orga-

nizing the data. Patterns of information, in turn, can be worked up into a coherent body of knowledge. Knowledge consists of an organized body of information, such information patterns forming the basis of the kinds of insights and judgments which we call wisdom.

The above conceptualization may be made concrete by a physical analogy (Stonier 1983): consider spinning fleece into yarn, and then weaving yarn into cloth. The fleece can be considered analogous to data, the yarn to information and the cloth to knowledge. Cutting and sewing the cloth into a useful garment is analogous to creating insight and judgment (wisdom). This analogy emphasizes two important points: (1) going from fleece to garment involves, at each step, an input of work, and (2) at each step, this input of work leads to an increase in organization, thereby producing a hierarchy of organization. Stonier (1997).

A model or simulation outcome depends essentially on the quality of the input data such as correctness, reliability, sufficiency, relevance and alike. If we see the world itself as informational, modeling implies selection and processing of information of interest. It is the actual data representation of the information at hand which makes possible an analysis of the relationship between changes in the underlying physical process and the changes in the model (observational information versus model-generated information). In order to study the relationship between reality and model, we will first focus on the semantics of information.

Floridi (2004c) suggests a list of the eighteen principal problems of philosophy of information. Among those, the most fundamental is the question: “What is information?”

Searching for the answer, Marijuan (2003) concluded that “Inconsistencies and paradoxes in the conceptualization of information can be found through numerous fields of natural, social and computer science.”

Or, as Floridi (2005) formulates it, “Information is such a powerful and elusive concept that it can be associated with several explanations, depending on the requirements and intentions.” See even the forthcoming Handbook on the philosophy of information, (van Benthem J and Adriaans P eds, <http://www.ilic.uva.nl/HPI/>). In the same spirit, Capurro and Hjørland (2003) analyze the term information, explaining its role as a constructive tool and its theory-dependence as a typical interdisciplinary concept.

On the other hand, Capurro et al. (1999) discuss the possibility of a unified theory of information (UTI), considering this in a cautiously affirmative way. According to the authors, UTI is an expression of the metaphysical quest for a unifying concept of the same fundamental nature as energy and matter. In the unification approach, reality is an information-processing phenomenon. “We would then say: whatever exists can be digitalized. Being is computation.” (ibid) In other words, at a fundamental level, information characterizes the world itself, for it is through information we gain all our knowledge, and yet

we are only beginning to understand its meaning. If information is to be considered the primary stuff of the universe, as information physics and paninformationalism suggest, it will provide a new basic unifying framework for understanding and constructing reality. It is interesting to observe how information can be understood in conjunction with its complementary concept, computation, in a dual-aspect information-computation unified theory, Dodig-Crnkovic (2006).

### 3 The Standard definition of information versus strongly semantic information

A standard definition of information which is assumed to be declarative, objective and semantic is given as data + meaning, (Floridi 2004a, b). In this context Floridi refers to The Cambridge Dictionary of Philosophy definition of information:

an objective (mind independent) entity. It can be generated or carried by messages (words, sentences) or by other products of cognizers (interpreters). Information can be encoded and transmitted, but the information would exist independently of its encoding or transmission.

It is instructive to compare the above formulation with the web <http://www.pespmc1.vub.ac.be/ASC/INFORMATION.html>, Dictionary of Cybernetics and Systems, which offers the following definition of information:

that which reduces uncertainty (Claude Shannon); that which changes us. (Gregory Bateson)

Literally that which forms within, but more adequately: the equivalent of or the capacity of something to perform organizational work, the difference between two forms of organization or between two states of uncertainty before and after a message has been received, but also the degree to which one variable of a system depends on or is constrained by another. E.g., the DNA carries genetic information inasmuch as it organizes or controls the orderly growth of a living organism. A message carries information inasmuch as it conveys something not already known.

In the background there is the most fundamental notion of information, ascribed to a number of authors; “a distinction that makes a difference”, MacKay (1969), or “a difference that makes a difference”, Bateson (1973).

Floridi's (2004a) Outline of a theory of strongly semantic information contributes to the current debate by scrutinizing and revising the standard definition of declarative, objective and semantic information (SDI). The main thesis defended is that meaningful and well-formed data constitute information only if they also qualify as contingently truthful. SDI is criticized for providing insufficient conditions for the definition of information, because

truth-values do not supervene on information. Floridi argues strongly against misinformation as a possible source of information or knowledge. As a remedy, SDI is modified to include a condition about the truth of the data; so that “ $\sigma$  is an instance of declarative objective and semantic information if and only if:

1.  $\sigma$  consists of  $n$  data (d), for  $n \geq 1$ ;
2. the data are well-formed (wfd);
3. the wfd are meaningful (mwfd =  $\delta$ );
4. the  $\delta$  are truthful.”

Floridi’s concept of strongly semantic information from the outset encapsulates truth and thus can avoid the Bar-Hillel paradox, Floridi (2004a).

It is important to remember that Floridi analyses only one specific type of information, namely the *alethic* (pertaining to truth and falsehood) declarative, objective and semantic information, which is supposed to have definite truth-value. 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 (2004) are not considered.

#### 4 Information, truth and truthlikeness

...by natural selection our mind has adapted itself to the conditions of the external world. It has adopted the geometry most advantageous to the species or, in other words, the most convenient. Geometry is not true, it is advantageous. Poincaré, science and method

Science is accepted as one of the principal sources of “truth” about the physical 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 declare a radical fallibilism about science (claim that accepted knowledge could be wrong or flawed), while at the same time insisting on the epistemic superiority of the scientific method. In his *Logik der Forschung* Popper argues that the only kind of progress an inquiry can make consists in falsification of theories. Popper was the first philosopher to question the idea that science is about truth and instead to consider the problem of truthlikeness as an alternative.

Now the question is: can a succession of falsehoods constitute epistemic progress? This would mean that if some false hypotheses are closer to the truth than others, then the history of inquiry may be seen as steady progress towards the goal of truth Oddie (2001).

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.

On those lines, (Kuipers 1987, 2000, 2002) 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 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.

The conclusion from the scientific methodology point of view is that, at best, we can discuss truthlikeness, but not the truth of a theory. Like Poincaré's geometry, models or theories are in the first place more or less correct and advantageous tools of inquiry.

## 5 Correspondence (static) versus interactive (dynamic) models of information

Information is not a disembodied abstract entity; it is always tied to a physical representation. It is represented by engraving on a stone tablet, a spin, a charge, a hole in a punched card, a mark on paper, or some other equivalent. This ties the handling of information to all the possibilities and restrictions of our real physical world, its laws of physics, and its storehouse of available parts. (Landauer 1996)

In the tradition of Western thought, since the ancient Greeks, information was understood in conjunction with representation. As Zurek (1994) put it: “No information without representation”. In correspondence theory, mind is simply carrying out passive input processing. The transformations from the information in the world into information for the agent are supposed to be causally related.

There are several versions of the correspondence (encoding-decoding) models of representation of information, such as isomorphic correspondence, as in the physical symbol system hypothesis (Newell, Vera and Simon); trained correspondences, as in connectionist models (Rumelhart, McClelland); causal/

nomological (general physical/logical) relationships (Fodor) and representation as function (Godfrey-Smith, Millikan).

In the traditional correspondence framework, the information is caused by some external past event. The problem of this view is to explain what exactly produced the representation in the animal or machine.

Some state or event in a brain or machine that is in informational correspondence with something in the world must in addition have content about what that correspondence is with in order to function as a representation for that system—in order to be a representation for that system. Any such correspondence, for example with this desk, will also be in correspondence (informational, and causal) with the activities the retina, with the light processes, with the quantum processes in the surface of the desk, with the desk last week, with the manufacture of the desk, with the pumping of the oil out of which the desk was manufactured, with the growth and decay of the plants that yielded the oil, with the fusion processes in the sun that stimulated that growth, and so on all the way to the beginning of time, not to mention all the unbounded branches of such informational correspondences. Which one of these relationships is supposed to be the representational one? There are attempts to answer this question too (e.g., Smith 1987), but, again, none that work (Bickhard and Terveen 1995). (Bickhard 2004)

This passage from Bickhard indicates the importance of intentionality in forming representations. Informational content of the world is infinite, and each object is a part of that all-encompassing network of causation and physical interaction. The only way one can explain the fact that the agent extracts (registers) some specific information from the world is the fact that it acts in the world, pursuing different goals, the most basic one being that of survival, and in that way an agent actively chooses particular information of interest.

Pragmatic theory has developed during the last century as an alternative to the correspondence model of representation. (Joas, Rosenthal, Bickhard). Pragmatism suggests interaction as the most appropriate mechanism for understanding information.

There are several important differences between the interactive model of representation and standard correspondence models. Interactive explanation is future-oriented; based on the fact that the agent is concerned with anticipated future potentialities of interaction. So the actions are oriented internally to the system, which optimize its internal outcome, while the environment constitutes resources for the agent. Correspondence with the environment is only a part of the interactive relation: representation emerges in the anticipatory interactive processes in (natural or artificial) agents, who are pursuing their goals while communicating with the environment.

In the contemporary fields of artificial intelligence, cognition, cognitive robotics, consciousness, language and interface design, interactive models are becoming more and more prominent. This is in parallel with the new inter-

active computing paradigm (Wegner, Goldin), and new approaches to logic (dialogic logic, game-theoretic approaches to logic), see Dodig-Crnkovic (2006).

## 6 Conclusion

There are two major approaches to the individuation of scientific theories that have been called syntactic and semantic. We prefer to call them the linguistic and non-linguistic conceptions. On the linguistic view, also known as the received view, theories are identified with (pieces of) languages. On the non-linguistic view, theories are identified with extralinguistic structures, known as models. We would like to distinguish between strong and weak formulations of each approach. On the strong version of the linguistic approach, theories are identified with certain formal-syntactic calculi, whereas on a weaker reading, theories are merely analyzed as collections of claims or propositions. Correspondingly, the strong semantic approach identifies theories with families of models, whereas on a weaker reading the semantic conception merely shifts analytical focus, and the burden of representation, from language to models. Hendry and Psillos (2004)

Here we can refer to Laudan's Methodological Naturalism, in Psillos (1997) formulation:

- All normative claims are instrumental: methodological rules link up aims with methods, which will bring them about, and recommend what action is more likely to achieve one's favoured aim.
- The soundness of methodological rules depends on whether they lead to successful action, and their justification is a function of their effectiveness in bringing about their aims. A sound methodological rule represents our "best strategy" for reaching a certain aim.

In the actual process of discovery and in model building, information is the fundamental entity. During the process, information is transformed and it changes its place and physical form. Depending on context, it also changes its meaning, (Barwise and Perry 1983).

When dealing with empirical information we meet the fact that the real world never perfectly conforms to the ideal abstract structures of the model (Plato's stance). Ideal atoms might be represented by ideal spheres. Real atoms have neither perfect shape nor sharp boundaries. In the physical world of technological artefacts and empirical scientific research, situations are rare in which models can be sharply divided into true and false. However, it is often possible to conventionally set the limits for different outcomes that we can label as "acceptable/non-acceptable" which in turn can be translated in terms of "true/false" if we agree to use the term truth in this very specific sense.



There are cases in the history of science in which false information/knowledge (false for us here and now) has led to the production of true information/knowledge (true for us here and now). A classical example is serendipity, making unexpected discoveries by accident. The pre-condition for the discovery of new scientific “truths” (where the term “true” is used in its limited sense to mean “true to our best knowledge”) is not that we start with a critical mass of absolutely true information, but that in continuous interaction (feedback coupling of learning process) with the empirical world we refine our set of (partial) truths. With good reason, truth is not an operative term for scientists.

Christopher Columbus had, for the most part, incorrect information about his proposed journey to India. He never saw India, but he made a great discovery. The “discovery” of America was not incidental; it was a result of a combination of many favourable historical preconditions combined with both true and false information about the state of affairs. Similar discoveries are constant occurrences in science.

“Yet libraries are full of ‘false knowledge’”, as Floridi (2004d) rightly points out. Nevertheless we need all that “false knowledge”. Should we throw away all books containing false information, and all newspapers containing misinformation, what would be left? And what would our information and knowledge about the real world look like?

In the standard (general) definition of semantic information commonly used in empirical sciences, information is defined as meaningful data. Floridi in his Theory of Strongly Semantic Information adds the requirement that standard semantic information should also contain truth in order to avoid the logical paradox of Bar-Hillel’s semantic theory. 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. Instead of insisting on the veridicity of the empirical model, we should focus on such basic criteria as the validity of the model and its appropriateness within a given context. Models and theories are seen as instruments of epistemic progress, which enable us to learn from the interaction with the empirical world, changing the world, our understanding of it and our epistemic tools in a dynamic process of meaning production.

**Acknowledgments** This article is a revised version of the paper presented at the International Conference Model-Based Reasoning in Science and Engineering (MBR04), held at the University of Pavia, Italy (16–18 December 2004) and chaired by Lorenzo Magnani. I wish to thank Lorenzo Magnani for organizing MBR 2004, with great efficiency and unadulterated enthusiasm. I also wish to thank Lorenzo Magnani for organizing E-CAP 2004, the conference that created so much interest for the field of Computing and Philosophy.

## References

- Barwise J, Perry J (1983) *Situations and attitudes*. MIT, Cambridge
- Bateson G (1973) *Steps to an ecology of mind*, Paladin. Frogmore, St Albans

- Capurro R, Hjørland B (2003) The concept of information In: Annual Review of Information Science and Technology (ARIST) (ed) Blaise Cronin, Information Today, Medford
- Capurro R, Fleissner P, Hofkirchner W (1999) Is a unified theory of information feasible? A dialogue In: Hofkirchner W (ed) The quest for a unified theory of information. Proceedings of the second international conference on the foundations of information science, Gordon & Breach, New York
- Dodig-Crnkovic G (2003) Shifting the paradigm of the philosophy of science: the philosophy of information and a new renaissance, *Mind Mach* 13(4):521–536
- Dodig-Crnkovic G (2006) Investigations into information semantics and ethics of computing, Mälardalen University Press, Sweden
- Floridi L (2004a) Information. In: Floridi L (ed) The Blackwell guide to the philosophy of computing and information. Blackwell, Oxford
- Floridi L (2004b) Outline of a theory of strongly semantic information. *Mind Mach* 14(2):197–222
- Floridi L (2004c) Open problems in the philosophy of information, *metaphilosophy*, vol 35 issue 4
- Floridi L (2004d) Afterword—LIS as applied philosophy of information: a reappraisal. *Libr Trends* 52(3)
- Floridi L (2005) Is information meaningful data? *Philos Phenomenol Res* 70(2):351–370
- Hendry RF, Psillos S (2004) How to do things with theories: an interactive view of language and models in science. In: Paprzycka K, Przybysz P (eds) *Idealization and concretization*. Rodopi, Amsterdam
- Kuipers TAF (1987) ed What is closer-to-the-truth? A parade of approaches to truthlikeness, *Poznan Studies in the philosophy of the sciences and the humanities*, vol 10. Rodopi, Amsterdam
- Kuipers TAF (2000) From instrumentalism to constructive realism: on some relations between confirmation, empirical progress, and truth approximation. Kluwer, Dordrecht
- Kuipers TAF (2002) Inductive aspects of confirmation, information, and content, to appear in the *Schilpp-volume The Philosophy of Jaakko Hintikka*
- MacKay DM (1969) *Information, mechanism and meaning*. MIT, Cambridge
- Magnani L, Nersessian NJ, Thagard P (1999) (eds) *Model-based reasoning. Scientific discovery*. Kluwer, New York
- Marijuan PC (2003) Foundations of information science: selected papers from FIS 2002. *Entropy* 5:214–219
- Marijuán PC (2004) Information and life: towards a biological understanding of informational phenomena. *TripleC* 2(1): 6–19, ISSN 1726-670X, <http://www.tripleC.uti.at>
- Oddie G (2001) Truthlikeness, *The Stanford Encyclopedia of Philosophy*. In: Zalta EN (ed) <http://www.plato.stanford.edu/archives/fall2001/entries/truthlikeness/>
- Stonier T (1997) *Information and meaning. An evolutionary perspective*. Springer, Heidelberg
- Wildberger AM (2000) AI & simulation. *Simulation*, 74
- Ören TI (2001) Impact of data on simulation: from early practices to federated and agent-directed simulations. In: A. Heemink et al (eds) *Proceedings of EUROSIM 2001, Delft, 26–29 June 2001*
- Zurek WH (1994) Decoherence and the existential interpretation of quantum theory. In: Grassberger P, Nadal J-P (eds) *From statistical physics to statistical inference and back*. Plenum, Dordrecht pp 341–350