How can data acquire meaning in the informational framework? A review of Floridi's attempt at grounding data

Luciano Floridi develops philosophy of information as a separate branch of philosophy, with the method of levels of abstraction[1] and a comprehensive argument for why philosophy of information should be considered a field of its own. He suggests that we should think of philosophy as "conceptual engineering"[2] (Floridi, 2011, p. 11) - something that seeks out conceptual problems and proposes solutions to them. The field has risen together with the technological developments since 1950s[3], raising new issues and providing new avenues for old philosophical problems to be explored (Floridi, 2011, p. 18). Today issues that stem from technological advancements are many and in different areas. Online piracy is a great example - never before was it possible to obtain something owned by someone else with the original owner still having the property. It has created a plethora of debates in this area alone. Yet, it is not just that technological improvements let us explore the world in new ways, scientific progress has created new conceptual frameworks, such as scientific realism. The concept of information is now considered as one of the most fundamental. An analogous conceptual transformation can be seen in how our understanding of the atom has changed from being that indivisible particle that makes reality, to being granulated into many smaller particles. If knowledge was an atom, information is a quark. Like all fundamental concepts, information

[1] There is a difference between levels of abstraction and 'levelism' as an approach to interpretation (Floridi p. 47). A simple example of levels of abstraction is consideration of cars. To evaluate a car, different levels of abstraction are appropriate. 'car for racing' observables include things like top speed and acceleration, 'family car' observables include things like efficiency, number of seats and the size of a boot. Different observables are relevant for different levels of abstraction.

[2] Term originally coined by Richard Creath.

[3] Initial steps towards philosophy of information can be said to come from philosophically minded scientists and engineers, rather than philosophers, with works such as Wiener's Cybernetics or Shannon's Mathematical Theory of Communication.

invites discussion and avoids a simple definition. In this essay, I will explore the assumptions that have to be made in order to explain how we define information. By assumptions, I mean the idea that information can be defined by a reductionist principle of splitting it into certain types of data. In particular, the focus is on the debate of how data, a building block of information, acquires meaning.

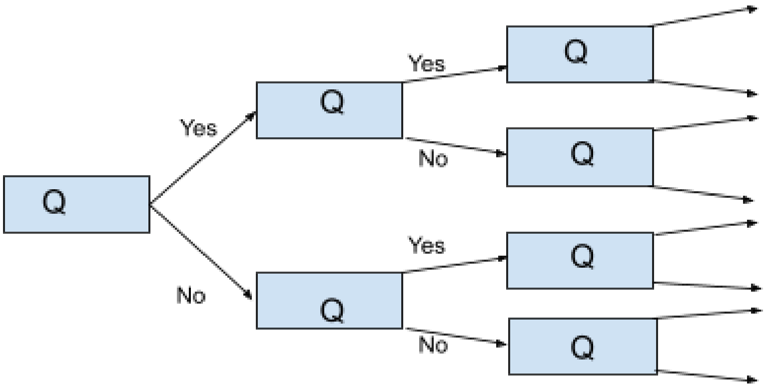
We will begin by highlighting a distinction most prevalent in the theory of information - semantic and syntactic types of information. It is important to understand the basics of these approaches to see how important conception of data is. The syntactic approach is that of probabilities and computation. It is not interested in looking at the meaning, relevance or interpretation of information, but at how it appears and what is needed to produce it (Lombardi, 2016, p. 31). The semantic approach tries to evaluate the necessity, meaning and truthfulness of information. Both of these approaches have a place in trying to answer the question 'what is information' and in conjunction provide more questions, such as whether there is a hierarchy between the semantic and syntactic approach? In an attempt to answer this question, we will look at the concept of information, data and the symbol grounding problem. In particular, we will see how Floridi's attempt to ground symbols, that is, connect semantics with syntax, satisfies the so called Z condition only under specific conditions and also opens further questions.

Syntactic approach

To begin with, let us start with physics and information. Following Floridi, we will work under the realist assumption - reality is mind-independent with objects that provide a structure for it (Harshman, 2016 p. 7). From an epistemological perspective information is assumed to be a property of an interaction between an observer and a system. Conversely, the ontological view suggests that information is a real thing - a property of a system that exists independently of the observer. In an extreme ontological view, matter and energy are properties of information. Life, in this view, is just a sub-system: information gathering and using system (Harshman, 2016 p. 8).

Whichever view we adopt at this point, we are considering the syntax of information, because we either consider it as a physical thing in the world (ontological view), or as a property of interaction between physical things (epistemological view). Both are, at least initially, embedded in physical structures. This leads to a better understanding of what is the structure of information, if it can be considered an object or merely structures between objects. I will highlight two approaches here, but the list of different interpretations is vast and the views are not necessarily mutually exclusive (Floridi, 2011, p. 31). The present essay will use descriptions, examples and illustrations to explain the ideas discussed and will include sources to approaches that use more of logical and mathematical analysis. A question to keep in mind is whether there is a hierarchy between semantic and syntactic approaches to information. What structure information has and how we perceive it goes side by side with the question of what it means for something to be information. For example, one way to think of information is in terms of probability. The more likely something is, the less information it has and vice versa (Lombardi, 2016, p. 33). Such a proposition is clearly subjective and requires interpretation of information itself, because information is then merely a measurement tool, used to quantify probabilities, but it also presupposes our ability to make sense of some data, otherwise no such probabilities could be constructed at all.

Claude Shannon's non-semantic notion of information is one of the major contributors to a syntax based view of information. After a brief description, the present essay shall show its limitations. The advantages of such a view will appear later in the essay. Shannon's mathematical model of communication considers information from a purely syntactic point of view (Shannon, 1948, p. 379). Shannon's approach is concerned with "the fundamental problem of communication" (Piccinini, 2016, p. 26). He tried to show that there is a way to talk about information exchange between the sender and receiver without having any knowledge of what the content entails. Basically, his insight was that we are informed of a symbol being selected and also gain the information that other symbols have not been selected. As Piccinini and Scarantino have put it: "information is in effect the reduction of uncertainty generated by the occurrence of one of a set of alternative possible outcomes" (Piccinini, 2016, p.26). It is achieved by using the concept of information as being the resource needed to identify one outcome out of many alternatives (Lombardi, 2016, p. 31). Ralph Hartley (1928) introduced information in the technical sense when considering actions with equal outcomes, while Shannon expanded beyond that. The question here is practical, how do we measure information? The answer provided by this approach is counting information by bits, where one bit is "the amount of information obtained when one of two equally likely alternatives is specified" (Lombardi, 2016, p. 31). In other words, the general approach to information here is to continually ask questions about a given system with only two possible answers until the event is found, then measure how many questions were necessary. The picture below helps to illustrate why such measurement is algorithmic - each new layer adds a degree of possibilities. So if we have 3 different spices and want to know which have been added to the mix, we ask: was the first one added? Yes/No, was the second added? Yes/No. It will take 3 questions, so 2^3 is 8 - that is the amount of alternative possibilities. What is counted as an information bit is each question, and since the question has yes/no answers, the lowest amount of information required to find out which spices have been added is 3 bits.



Shannon's theory is a mathematical model of communication, so the notion of information there is, in a sense, secondary. The notion of information is understood as an event to be recorded to pass from the source to the receiver. In other words: "the significant aspect is that the actual message is one selected from a set of possible messages" (Shannon, 1948, p. 379). A serious issue with this approach can be seen when we consider that the amount of information can differ based on our knowledge of a system that the information resides in. In other words, the probability of events is arbitrary. We can devise a system of possible alternatives for any event and measure the amount of information that is transmitted when the message is received, but such information is simply a counting of possible alternatives. It may be useful in physics and engineering, but it is counting data rather than information and data interpretation is already presupposed. Presuposed in a sense that its measure of quantity is assumed to be knowable without interpreting its meaning. In essence, it is an approach that uses a vague definition of information based on the idea of quantifying it. It works when addressing engineering problems, because they have a clear purpose. However, the qualitative aspect of information is completely ignored. That is, a predefined conception of what a data point means, will determine how it is quantified. This approach is so quantitatively powerful, because the information that is being calculated is always comprised of data points that are already understood. An example Shannon gives is with letters A B and C making a sequence, the interpretation of each one of them being a separate symbol is how the calculation begins. Yet it could easily be the case that different symbols represent the same data. For example, they are all letters of the English alphabet. The differentiation of symbols depends on what we want them to mean. Is there a way to assume the most appropriate interpretation of data?

To recapitulate: interpretation of data is through difference. As MacKay[4] and later Bateson[5] highlight, information is a difference that makes a difference. A datum is "ultimately reducible to a lack of uniformity" (Floridi, 2011, p. 85). Such reduction is hardly satisfactory and has several issues in its interpretation: It can be understood as data 'out-there' in the real world (dedomena), as data between signals or as data between symbols (Floridi, 2011, pp. 85-86). There can be different variations between those interpretations which either support a syntactic or semantic model of information. However, since meaning is contained in form, although not necessarily as a representation, the crucial question is how that meaning emerges.[6]

---

[4] MacKay, Donald, *Information, Mechanism and Meaning*, London: MIT Press, 1969.
[5] Bateson, Gregory, *Steps To An Ecology Of Mind,* New York: Balantine Books, 1973.
[6] To say that meaning is contained in form, simply means that we cannot communicate it without form, not that it is bound by form.

Data is therefore a necessary but insufficient condition of information. General Definition of Information[7] is that information is well-formed meaningful data (Floridi, 2011, pp. 83-84). Floridi builds upon that definition to add the veridicality thesis which says that for data to become information it also has to be truthful; truthful in the sense that information encapsulates truth (Floridi, 2011, p. 106). However, if information is defined as a well-formed meaningful and truthful data, the most basic problem to be solved is the symbol grounding problem (SGP) (Floridi, 2011, p. 34). The problem refers to the difficulty in figuring out how any system could internally, but not innately, construct and elaborate meaning from symbols. In simple words - how is meaning from symbols possible? This is a fundamental problem that has to be addressed and in the process we will see how meaning can be constructed from data and thus interpreted as information. To do so, an answer to SGP has to satisfy the Z condition (Floridi, 2011, p. 137). Z condition has three main tenets. No innate semantic resources can be presupposed, no external semantic resources can be assumed to come from outside the system and that the system can have its own capacities and resources to ground symbols. What this essay is trying to find out is whether Floridi's solution manages to adhere to those conditions. Several attempts have been made to solve the symbol grounding problem, however, they all fail the Z condition that Floridi specifies. We will look at Floridi's attempt at solving SGP next.

Praxical solution

[7] General Definition of Information is a definition used by Floridi, to showcase a definition that is used by several prominent theorists that work at analysing information. Such as Davis and Olson, 1985, p. 200 or Lucey, 1991, p. 5.

Two main components make up Floridi's attempt to solve the SGP. First is what he calls 'action-based semantics' and then 'two-machine artificial agents'. The idea of action-based semantics is that initially an artificial agent would have meaningful symbols only as internal states of the agent that directly correlate with actions the agent takes. There is no purpose or goal orientation at this stage. In other words, actions that are taken are the meaning of the syntax within the artificial agent's inner workings. The grounding of action as a symbol is performed as a consequence of the action, without any consideration of its goal or purpose. One may wonder whether it is still a symbol that is being grounded, or whether it is being grounded at all. The solution is made clearer with the introduction of two-machine artificial agent, which is actually a two-agent system that is based on metaprogramming architecture (Floridi, 2011, p. 166). Such a system has in effect two agents that operate at different levels of abstraction. Such programming architecture allows a program to decide its architecture during the runtime, based on data and possible actions. We are using 'program' as agent here only to show that we are talking about 'simple systems' - issue here relates to the debate between natural and artificial systems, but is beyond the scope of this essay.[8] So the artificial agent of such design, would have an agent1 who operates at object level, interacting with the external environment. The data and computation of the first agent then allows the agent2 to compute as well, and those computations have an impact on agent1 and its computation. In other words, the grounding of symbols is done within the system, using only actions it can perform, without innate or external resources. A view that the symbol grounding problem is solved is also shared by Luc Steels[9]. An example sometimes given of a

[8] For more see: Vassie K. and Morlino G. 'Natural and Artificial Systems: Compare, Model or Engineer?'.

[9] Steels Luc, 2008. He also does a great job at clarifying the differences in how conception of symbols is used in philosophy and computer science, thus making a distinction between c-symbols as understood in computer science and m-symbols as understood philosophically as meaning-symbols. This distinction and a consequent argument for the solving of the symbol grounding problem is interesting and relevant, but is beyond the scope of this essay.

working two-level agent interpretation can be a detection device (object level agent) and movement device (meta-level agent) coupled together to make a robot. A detection device can be a light sensor, which gives energy to the movement device when light is detected, and does not when light is not detected.

Agency and Action

An important issue in such a conception of finding meaning in action is whether it is justified to conceptualise such artificial agent as not having innate or external resources. Action based semantics (AbS) assumes that meaning is generated by action and that symbols are records of actions. If the symbols are recorded internally and the action of agent1 has some immediate consequence for the agent2 at the lower level of abstraction (that the agent2 operates on), then it can as a consequence also induce action from agent1. A certain recursive loop based on outside stimuli is then possible. But that is all it is, a recursive loop of events within an open system. One issue in such conception is that action already implies purpose. If we assume that an agent acted in some way, we assume agency, that is, a behaviour with a goal of some kind. Floridi is careful to note that "the purpose of the action has no direct influence in the generation of the meaning" (Floridi, 2011, p. 165). However, without a purpose, an action is merely an event - it does not have agency. The underlying question here is what can we call an agent? For example, a sun's generation of light and heat is a recursive loop of nuclear fusion where atoms are in a sense "molded" together and create so much energy that it molds other atoms together. By the logic of action based semantics and two-machine artificial agents, the sun is such an agent. In fact, it is a very complex agent from this perspective, because multiple layers of it have an immediate effect on other layers.

For example, gravity forces a fusion reaction at the sun's core, where two protons make a hydrogen atom, a positron and a neutrino, they in turn combine further to make more and more complex reactions that make more complex atomic structures and release energy. The Helium-4 atom that is created after several fusion reactions, has less mass than the two hydrogen atoms that started the reaction, so the difference in mass is converted into energy and expended as light. Eventually this will weaken the gravitational pull and the sun will begin expanding. If we adjust our perspective of time, the object level agent here is gravitational pull that started the reaction and the fusion reaction then is the meta level which "acts" based on the data of gravitational pull and feeds back data to the object level agent by expending energy outwards.

I would argue that practically any energy expending system can be understood this way, simply because the levels of abstraction that are used to determine the operations of the two agents have to be semantically selected. In other words, we do not find out how meaning is constructed from syntax, but merely construct syntax that has a recursive loop to maintain and change its structure. Reasoning from this side of the argument, we can see that unless actions/events are taken for granted as events that 'just happen', semantics cannot be ignored.

Furthermore, since AbS is one of the main tools Floridi uses to ground the symbols (data), concept of action requires clarification. As K. Beliecka says: "in AbS, it is impossible to distinguish a successful action from a non-successful one" (Beliecka, 2015, p. 82). I would add that if, as Floridi says, the actions are to be considered without a goal or a purpose, the conception of successful action does not make sense, because a failed action will still be an action that the two-machine model will be able to ground as a symbol. In that sense, all actions are successful. The issue that is being highlighted here, is that our interpretation of action is irrelevant.

Also, another possible interpretation is to say that what we are talking about are the changes in the environment that affect the agent1, which in turn affects the agent2. In this case, how do we even differentiate agent as such? There is no action, it is merely a cause and effect relationship within the environment that both 'agents' belong to. Can we assume that the same change in the environment will have the same affect every time? Not without conditions that are pre-established and thus fail the z condition. Some of those conditions would include things like considering a process in isolation and not differentiating between slight differences. A presupposed Evolutionary Local Selection Algorithm, although only meant to represent a general framework, in fact presupposes an intrinsic capacity of reinforcement, as would any other similar framework. With it, Floridi is able to construct a scenario where two-machine artificial system could indeed overcome the symbol grounding problem and construct semantic content from data. However, symbols in this system are grounded in a way that leaves little room for agency. Working under the assumption of a mind-independent reality, the mind as consciousness and intentionality is put into question.

Another example that highlights the problem of interpreting meaning through action was offered by MacKay. When we receive information, such as that it is raining outside, it requires no action from us. However, if someone were to come inside, we might ask them about it or if we were to go out, we would be ready to take an umbrella. "What has been affected by your understanding of the message is not necessarily what you do - as some behaviourists have suggested - but rather what you would be ready to do if given (relevant) circumstances arose" (MacKay, 1969, p. 22). Floridi's theory may have a way to overcome this difficulty by once again referring to the levels of abstraction. For information to affect our understanding, we assume some process in our brain happens - some neurons fire. In the action based semantics, that is how the data would be grounded, not in an immediate action from object level agent, but in the meta level agent of neuron activity in the brain. That 'record' of information, the action taken by brains' neurons, is what would in turn change the behaviour of object level agent and make them take an umbrella. On the other hand, the real difficulty lies in avoiding any reliance on innatism or externalism - on satisfying the Z condition. Floridi's attempt fails because structure is necessary to ground the symbol, be it a structure of neurons firing in a particular way, or light sensor affecting robot's movement.

Satisfying the Z condition

When Floridi considers other approaches to the symbol grounding problem, he notes that they considered meaning and symbol as two aspects of the same data (Floridi, 2011 p. 180). Praxical method that Floridi introduces suggests we think of meaning and symbol as two kinds of independent data (Floridi, 2011, p. 180). Considering that Floridi defines datum as "x being distinct from y" (Floridi, 2011, p. 85) or simply data as difference, we have to think of how meaning and symbol as independent data can exist. One interpretation of this is that the two-machine agents each interpret data at different levels of abstraction. Indeed, as Floridi notes, pure data in themselves cannot be accessed without adopting a level of abstraction (Floridi, 2011, p. 87). If we have to adopt different levels of abstraction for meaning and symbol of data, we are already interpreting it. K. Beliecka points out that interpretation does not necessarily violate the Z condition, because such adoption of different levels of abstraction is interpreted by the outside observer, rather than the system itself. The system may well be 'unaware' of any interpretation accuring (Beleicka, 2015, p. 83). However, this is not enough, for we are not considering the dedomena (data before interpretation), yet presuppose its existence based on the adopted definition of data as difference.

So there are two possible options. The two-machine agents interpret data as meaning innately, which fails the z condition. The argument here is that action that has an effect to initiate another action has meaning innately. The receiving of action from agent1 has to have a structure that is capable of receiving it. We may not call it interpretation, but the mere fact of it means there are some prerequisites that the agent2 meets, which are innate to it. The acceptance of that fact means that there is a minimal innate structure in anything we would consider as an agent. Another option is to say that data being interpreted is understood as a symbol in both cases, but at different levels of abstraction. This means that in effect, meaning is a datum of a symbol that is considered at a different level of abstraction. There is no object of a datum, as datum itself is the difference, whether it is a dedomena or difference between symbols or difference between signals - the difference that exists can be considered in a plethora of ways and which interpretation we pick to represent as meaningful depends on what effect we assume it to have. The symbols that affect the environment are symbols that acquire meaning.

However, a question well formulated by William J. Rapaport: "what sense does it make to say that syntax is sufficient for semantics?" (Rapaport, 1995, p. 49) now needs to be addressed here as well, although I will rephrase it to suit our needs. If syntax is enough for data to have meaning, how is it acquired? An answer is already here, we just need to drag it out of hiding - meaning is acquired through difference. As Rapaport puts it: "one understands something relative to one's understanding of another thing" or "gets used to it" (Rapaport, 1995, p. 53). Data is interpreted - formed into information, through difference between symbols or signals that create difference themselves. As I said before, a datum is not an object, but a difference. Interpreting it is a process of that difference generating other differences - generating new data points. The data that is recursive, that is, that repeats itself, can be interpreted as a pattern and we reach a different level of abstraction. There we interpret patterns and Shannon's non-semantic communication theory, for example, is applicable - the enclosed system is a set level of abstraction and information amount is dependent on recursiveness of data. That is, repeated appearance of difference that can be interpreted as data. In terms of Floridi’s action-based semantics, meanings of symbols generated by the artificial agent are simply internal states of that artificial agent, which in turn generate action (Floridi, 2011, p. 164). Yet they are meanings only for an outside observer, interpretation need not occur within the system (Beliecka, 1995, p. 83). So putting it all together, we find that in Floridi’s action-based semantics approach, the symbol grounding can only ground symbols without failing the z condition for an outside observer. For the agent, the symbol grounding is an environmental effect, a syntactic move that an agent has no ‘agency’ on.

Floridi’s attempt seems to deal with the question of how symbols become meaningful in a very pragmatic way. Understanding philosophy as conceptual-engineering seems very fitting in this framework. Because of its pragmatic nature, Floridi’s philosophy of information has an impressive explanatory power within its own framework, however, it is not without problems. In this essay, I focused on one problem in particular - under the definition of information as a well-formed, meaningful and truthful data, the difficulty of understanding and applying the concept of data was highlighted, the question of symbol(data) grounding was explored with the focus on syntax and semantics - where does one begin and another end? We saw that under some interpretations, the syntax has a hierarchy over semantics, and semantics can only be seen as a derivation of syntax. Floridi’s use of concepts such as action and agent were also criticised for a lack of clarity and this lead into the question of whether Floridi's attempt to solve the symbol grounding problem satisfied the Z condition. The answer generates a new difficulty in conceptualising and evaluating the individual's perspective within the system and the perspective of those outside a system being considered - something to be discussed in a future essay.

Bibliography:

Bateson, G., 1973. *Steps To An Ecology Of Mind,* New York: Balantine Books.

Bielecka K., 2015. 'Why Taddeo and Floridi did not solve the symbol grounding problem', *Journal of Experimental & Theoretical Artificial Intelligence*, volume: 27, issue: 1, pp. 79-93, DOI: 10.1080/0952813X.2014.940138.

Creath R., 1990. *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work: Edited and with an introduction by Richard Creath.* USA: University of California Press.

Davis G. B. and Olson, M. H., 1985. *Management Information Systems: Conceptual Foundations, Structure, and Development*, 2nd edition, New York: McGraw Hill.

Floridi, L., 2011. *The Philosophy of Information*. Oxford: Oxford University Press.

Harshman N. L., 2016. ‘Physics and information’ in Floridi, L. ed. *The Routledge of Philosophy of Information.* London: Routledge. pp. 7-15.

Hartley R., 1928. ‘Transmission of information’ in *The Bell System Technical Journal,* volume: 7, issue: 3, pp. 535 - 563. https://ieeexplore.ieee.org/document/6769394

Lucey T., 1991. *Management Information Systems,* 6th edition, London: DP Publications.

Lombardi O., 2016. ‘Mathematical theory of information (Shannon)’ in Floridi, L. ed. *The Routledge of Philosophy of Information.* London: Routledge. pp. 30-37.

MacKay, D., 1969. *Information, Mechanism and Meaning*, London: MIT Press.

Piccini G. and Scarantino A., 2016. ‘Computation and information’ in Floridi, L. ed. *The Routledge of philosophy of information.* London: Routledge. pp. 23-30.

Rapaport, W., 1995. ‘Understanding understanding: syntactic semantics and computational cognition’ in *Philosophical Perspectives*, volume: 9, pp. 49-88, www.jstor.org/stable/2214212.

Shannon, C., 1948. ‘A mathematical theory of communication’ in *The Bell System Technical Journal,* volume: 27, issue: 3, pp. 379 - 423. https://ieeexplore.ieee.org/document/6773024.

Steels, L., 2008. ‘The symbol grounding problem has been solved. So what’s next?’ in *Symbols and Embodiment: Debates on Meaning and Cognition,* pp. 223-244. Oxford: Oxford University Press, pp. 223-244.

Vassie, K. and Morlino, G., 2012. ‘Natural and Artificial Systems: Compare, Model or Engineer?’ in *From Animals To Animats 12*, Edited by Balkenius C., Hallam J., Ziemke T., Publisher: Springer. pp. 1-11.

Wiener, N., 1948. *Cybernetics: Or, Control And Communication In The Animal And The Machine*, New York : Paris: Wiley; Hermann et Cie.