Causality and intrinsic information

Vitor Silva Tschoepke
vitor.tschoepke@gmail.com

Abstract: This text will discuss the concept of information and its relevance in the study of the nature of the mind. It will analyze a hypothesis that deals with the equivalence between information and causality, which results in information having a double ontological character: "intrinsic" and "extrinsic." A discussion will follow on Integrated Information Theory, which is developed from a variation of this thesis. It will be proposed that this theory does not reach the objective of being an "intrinsic" information theory, precisely because it is an adaptation of classical information theory. For this reason, it neither allows nor avoids the paradoxes of the subject–object distinction, nor shows how the causal evolution of an integrated system can be treated in informational terms; that is, it cannot unite causality and information.

Keywords: integrated information, causality, superposition of states, self-reference.

1) Introduction

The investigation of the nature of mental processes involves the relationship between two categories: information and causality. If an observer is an interpreter of reality, is someone for whom the world, and each of her relationships with it, possesses meaning. However, if he receives information from the world and processes it internally in his brain, in a network of neurons whose communication is given by electrochemical processes, then one neural process transforms into another and another indefinitely, and, therefore, to whom, beyond this continuum, does this information make a difference?

If a brain impulse is a causal event that leads to another in an undefined series, at what descriptive level does a state of the brain cease to be a simple emitter and become a receiver of information? Similarly, if we consider that brain integration, arising from the perspective of individual neurons and structures, identifies patterns of interaction in an architecture that contemplates the brain as a whole, what can we know besides the fact that one momentary configuration is different from another? If the mental process is a succession of states and at the same time is information, how can we correlate one category with another?

If information is a kind of mediation in which one event becomes a type of reading for another, then information will always be something relative to an observer. However, how can an observer be explained in informational terms, that is, how can we explain the transformation of information from "extrinsic" to "intrinsic"? A theory with this proposal will need to present a model of how information emerges in the structure of reality. However, this task can only succeed if consistent premises are adopted on the relationship between causality and information, and in turn on the "extrinsic" and "intrinsic" perspectives.

First, we will discuss a position in which there is a type of equivalence between information and causality—the complementary and coexistent aspects in physical systems. If they are both in essence the same thing, then a descriptive state in informational terms possesses an extrinsic aspect relative to an observer and an intrinsic one relative to itself—an autonomous existence in nature. If information exists in parallel to causality, physically, then every causal event has a kind of primitive cognition. This ontological proposal defended by David Chalmers is the foundation of Giulio Tononi's Integrated Information Theory. We will question this premise, as well as analyze if this theory can explain how cerebral information becomes intrinsic. As it is the starting point of the two theories, it is worth considering a brief summary of the general concepts of information theory.

2) Information Theory

The Mathematical Theory of Communication, or information theory, of Claude Shannon (1948) was developed to address the various theoretical problems that arise from the development of data communication devices and the circumstances that allowed them to become more efficient and economical. The focus of the theory is the encoding and transmission of signals, but it does not address the semantic or conceptual meaning of the message. Like any mathematical theory, it becomes quite complicated in the analysis of details, but this brief explanation will suffice.

Information is measured by a unit called Entropy. It measures the minimum quantity of bits required to represent a space of signal alternatives for data transmission; it is the minimum limit of encoding complexity.

The bit is the most elementary "unit" of transmission, which can be summarized as two states. Within the turning on or off of a lamp, the two possible states can be represented by one bit: 0 and 1. The larger the space of alternatives, the more bits will be needed to encode them. If we have a space of alternatives, such as choosing between "A," "B," "C," or "D," this can be represented by two bits: 00, 01, 10, and 11. Thus, the message "ABACD" would be encoded as "0001001011." A space of eight alternatives can be encoded as three bits: 000, 001, 011, 111, 110, 100, 101, and 010. Thus, the measurement of information is determined by the base 2 logarithm of the number of alternatives (in the binary system), that is, for four alternatives, two bits; for eight alternatives, three bits; and so on.

Considering the degree of certainty or uncertainty of the transmitted event, we have three possibilities. If among all the events, one of them has a probability equal to 1 and the other 0, then the outcome is certain, and the entropy will be zero. If every event has the same probability, then the uncertainty will be maximum. For a set of events of different probabilities, the uncertainty will have a value greater than zero and less than the maximum. Entropy is a logarithmic unit that measures the uncertainty of information.

In the example of the symbols "A," "B," "C," and "D," we initially consider the probability of each as ¼. However, there is a problem. If one symbol is repeated more than the others, the chance of each event occurring must enter into the calculation. Since the theory focuses on the economy of transmission, one can assume that more common symbols are shorter, and the minimum encoding space is proportional to their frequency. If the probabilities of "A," "B," "C," and "D" are ½, ¼, ⅛, and ⅛ respectively, they can be represented by 1, 2, 3, and 3 bits, respectively.

Thus, we arrive at the way of calculating the entropy, the measure of the information, which is the sum total of the results of the minimum number of bits representing each event multiplied by the probability of it appearing. In this case, the result will be ½ x 1 + ¼ x 2 + ⅛ x 3 + ⅛ x 3, which results in 1.75 bits. This will be the average space used to encode each character of the message. Note that four equally probable alternatives use two bits, and so, considering that the frequency of the character is an information-saving factor, on average 0.25 bits will be saved.

An interesting question remains. If a single character is transmitted, such as the letter "I," we have an uncertainty that is determined by the next letter that can succeed it in the formation of a word, which in this case is probably almost the entire alphabet. We can complete the sequence with several options of characters, and for each one, as the word starts to be identified, alternatives of possible letters and words can be eliminated, and the uncertainty decreases. We can go from "I" to "IDENTICAL," "IMPOSSIBLE," and so on. As we reveal the subsequent symbols, we are eliminating alternatives: "IN" ("INTER," "INVERSION," ...), "INF" ("INFLATION," "INFRINGE," ...), "INFO," ("INFORMATION" ...), until we arrive at

the word, with total certainty: "INFORMATION." Thus, the probability of one event will be limited by the previous event, and the next is limited by the current event. This is a conditional distribution, in which the probability p of an event x1 is given by p (x1 | x0 = "I"), which identifies the space of alternatives given that the previous state in this example is "I." These alternatives can be mapped onto a chain of probabilities, with each point having a local probability.

There is yet another interpretation of uncertainty, which is relative to the lack of clarity of the signal, or the identification between elements of space of states. A larger encoding space gives greater freedom of choice for the emitter and results in greater informational potential; this is the desirable type of uncertainty of the device. However, there is another meaning of uncertainty, which is the *noise* in the transmission. It can be as much the lack of clarity in the signal identification by unforeseen interference as the difficulty in distinctly and unequivocally distinguishing between the corresponding signals.

With these brief considerations in mind, we will continue our study.

3) Intrinsic and extrinsic information

Chalmers (1996) proposes a metaphysical conception that information has two aspects, one relative to the observer and another intrinsic, or phenomenological. In presenting the argument, he first summarizes the idea of information as the selection of an event among a number of alternatives:

> The way I have formalized things here is intended to capture Shannon's idea that information essentially involves a state selected from a number of possibilities (in the relational structure of a space), and also captures the idea that complex information can be built up from simple information (in the combinatorial structure of a space). (Chalmers, 1996, p. 280)

He then goes on to argue that an informational state, rather than being thought of only as an encoding between alternative events of a mechanism, can be taken in parallel with its own physical causality. For this, he uses Bateson's (1972) expression, when referring to the bit and to the information itself, as "the difference that makes a difference." The physical correlations between possible states could thus be thought of as the transmission of information within the physical self—it is information physically realized.

After establishing the correlation between information and the physical, he continues to the connection between information and phenomenology, that is, "the being as something." The evaluation of the differences between experiences, perceptions, sensations, etc., is given by evaluating their similarities and differences. In addition, from the various perceptions, it is possible to find the neurological informational component that corresponds to them.

From this premise, he arrives at the extrapolation: if this evaluation is also an evaluation of (physical) information, then every physical system that alternates between states will have experiences, or states of information can be applied to both directions. He suggests a fundamental principle that information has two aspects, physical and phenomenological:

> States of experience fall directly into information spaces in a natural way. There are natural patterns of similarity and difference between phenomenal states, and these patterns yield the difference structure of an information space. Thus we can see phenomenal states as realizing information states within those spaces. (Chalmers, 1996, p. 284)

He continues:

> The space of relevant possible states here is isomorphic to the space of possible experiences; so we can see the same information state realized both physically and phenomenally. (Chalmers, 1996, p. 285)

If the mind is definable in terms of information processing, then everything possessing information is, in a sense, in possession of a mind, and in this case the physical world is entirely possessed of degrees of consciousness—it is the panpsychism. A brain has many states, but simpler things like thermostats, even though they have a smaller margin of variation, possess associated information. This is the synthesis of the argument:

a) If the mind has experiences, and in the brain there are correlated informational states (alternating between physical states), then information can be positioned as much in terms of one as the other.

b) The mind/brain relationship is the expression of the double nature of the existence of information—intrinsic and extrinsic.

c) If information is in everything, then experience is in everything, from brains to thermostats.

d) The level of complexity of the information space determines more or less complex experiences.

This basic idea has significant repercussions in the scientific world. Although he develops it differently, Tononi (2008), in presenting his Integrated Information Theory (IIT), accepts the same premise, that switching between states results in phenomenological states:

> IIT claims that consciousness is not an all-or-none property, but is graded: specifically, it increases in proportion to a system's repertoire of discriminable states. Strictly speaking, then, the IIT implies that even a binary photodiode is not completely unconscious, but rather enjoys exactly 1 bit of consciousness. Moreover, the photodiode's consciousness has a certain quality to it—the simplest possible quality—that is captured by a single q-arrow of length 1 bit. (Tononi, 2008, p. 236)

Christoph Koch (2012), when expounding in his book on his great appreciation for IIT, also assumes the metaphysical premise that information possesses a phenomenological component:

> David Chalmers is a philosophical defender of information theory's potential for understanding consciousness. His *dual aspect* sketch of consciousness postulates that information has two distinct, inherent and elementary attributes: an extrinsic one and an intrinsic one. The hidden, intrinsic attribute of information is what it *feels* like to be such a system; some modicum of consciousness, some minimal quale is associated with being an information-processing system. This is just the way the universe is. (Koch, 2012, p.124).

The difference is that for these last two authors, what creates the phenomenological aspect is not the simple complexity of the information; it is not the addition of bits, but the level of *integration*, which will be discussed in the following part of the study.

4) Equivalence between causality and information

John Searle (1997b) has presented several objections to the computational approach in the formulation of explanatory models for consciousness, especially due to its dependence on

an external observer. These criticisms can be included in the present discussion because, although the theories analyzed here are not computational, an aspect of the same debate, the definition of information, is in question.

In many areas of scientific research and development, definitions are restricted to one applicable field and become too broad and vague when extrapolated to other areas—they become merely analogies and parallelisms. When it comes to describing the conditions for consciousness to arise in nature, or to describe it scientifically, precise definitions are essential. Without clear definitions, it cannot be said that a brain processes information in the same way as, for example, a computer does. Searle's objections can be summarized in the following points:

a) Both the syntax and the definition of bits are relative to the observer.

b) If the definitions are vague, anything that alternates between states can be a computer, or perform information processing.

c) Considering the current understanding (or lack thereof) on what information means, it is not intrinsic to nature.

In a text critical of Chalmers' theory, the author synthesizes his objection at the epistemological level, attacking not only the conclusions, but the very legitimacy of the investigation:

> To the extent that you make the function and the information specific, they exist only relative to observers and interpreters. Something is a thermostat only to someone who can interpret it or use it as such. The tree rings are information about the age of the tree only to someone capable of interpreting them. If you strip away the observers and interpreters then notions become empty, because now everything has information in it and has some sort of "functional organization" or other. The problem is not just that Chalmers fails to give us any reason to suppose that the specific mechanism by which brains cause consciousness could be explained by "functional organization" and "information", rather he couldn't. As he uses them, these are empty buzzwords. (Searle, 1997a, p. 176)

If this criticism is true, then it does not make sense to treat information as a kind of structural attribute of reality such as electrical fields, matter, or gravity, for it is an interpretation made by nature's own observers. To posit a direct equivalence between information and causality authorizes the model to place an observer where there is none.

Following this critique, we will analyze the relationship between observers and information, and evaluate to what level information requires the presence of an interpreting agent. The possible definitions of information are many, and this list is not intended to be exhaustive. We can list at least four possible meanings for information:

a) The simple causal link – the "communication" between cause and effect.

b) Translation from one level of reality to another by means of a scale, mechanism, or measuring device.

c) Encoding of abstract entities as logical operators ("yes," "no," "or"), letters and numbers, as well as their use in computer programs.

d) Systematic interpretation of an area of reality through a system of concepts and its transmission.

Of all the meanings, "d" is clearly dependent on observers; it is the activity of the intellect to systematize and convey meaningful messages, ranging from a simple chat to the formulation of a theory. In the case of item "c," Searle explains that a computer program can be encoded in a manner equivalent to a mechanism that alternates between physical states, but efficient causality will not be that of the program as an abstract entity. This is because it does not physically exist, from the interactions between subatomic particles, atoms, molecules, and the properties of structures resulting from their aggregations to the macro level, whatever the "program level":

> the system patterns allow us to encode the information into physical characteristics intrinsic to the system, such as voltage levels. And these standards, along with system transducers at *input* and *output*, enable us to use any member of this equivalence class (...). The universality of patterns facilitates the attribution of computational interpretations (...), but the interpretations are not intrinsic to the system. (Searle, 1997b, p.317)

We then move on to items "b" and "a," and verify whether we can preserve the information by eliminating the observer. Item "b" is the case of data transmission and measurement instruments such as telephones, scales, thermometers, and telescopes. To attribute a causal correlation to two events is to bring together two elements to which we often do not have direct access. One only gives hints to the other, and it is necessary to gather evidence that both are related. In these cases, the instrument is designed to allow a correlation between two different types of events, so that one can extract the reality of the other.

In the development of a measuring instrument, all of its planning considers and anticipates the physical correlation between one type of phenomenon and another. The observer assumes the role of tracking the causal link, as there is no exact identifiable link between the two facts, without mixing them in any way. In this sense, the informational aspect—the architecture of the device itself is extrinsic to it—is dependent on the decisions and the interpretation of an observer. The translation by a system of measurement from one level to another requires a pre-selection of what will be measured in the instrument's own construction, and it is the designer who will in anticipation of that give meaning to the information.

The question is, does a device works causally, and is it possible to direct dependent events in order to create an intelligible link? However, for Chalmers, this demonstrates that information is directly related to causality:

> All this is implicit rather than explicit in Shannon's account, where there is no direct treatment of the relationship between physical states and information states, but on a close look it is clear that when information states are individuated, it is the transmittability principle that is doing the work. (Chalmers, 1996, p. 283)

That is, "do the work" is the difference that makes a difference in the transmission. Information is an expression of efficient causality. From this statement, one can conclude that information is dependent on its efficiency, that is, on the "difference that makes a difference." Information is then the presence of a state that is clearly identifiable, recognizable (in a pattern), and transmissible, as it is the unequivocally determinant factor of a result in clear distinction from other possible ones.

However, therein lies a problem in mixing information and causality. For an element or physical property, to make, or not make, a difference to a given result is extrinsic to its constituent properties. As an informational link will be specific among various possible causal links, the information may or may not exist, while there will always be some causality. Aristotle presents one of his four causes as that of movement and rest; *efficiency is causal, but so also is inefficiency.* The insufficiency of conditions to isolate one bit is as causally explainable as their

sufficiency to do so. If the occurrence of each physical result has its own efficiency condition, then something is and is not information depending on the expected occurrence; in one case, it will be information; in another, it will not. However, something cannot both be and at the same time not be a fundamental physical property.

One can understand the "difference that makes a difference" as obtaining the sufficient condition to trigger an event. A difference is identifiable as opposed to something that does not differ, to some non-difference, to some resemblance, but what is the causal power that has the *resemblance*? If difference contrasts with resemblance, and is informative because it is causal, resemblance cannot be causal, but this is absurd. Difference is informative in relation to something that does not differ, but in this case resemblance is as informative as difference. Difference is informative because it is causal, but so too is resemblance, and, therefore, if everything that is causal is information, nothing is. Thus, attributing resemblance and difference requires evaluation by an observer.

The expression "difference that makes a difference," which communicated sufficiently the intentions of its formulator in his essays, does not serve as a definition, and even less as a scientific principle. The expression could be replaced by "clearly identifiable difference," since without the clear distinction of the voltages corresponding to the 0's and 1's, computation would be impracticable. However, this only makes sense in a mechanism designed in advance to consider this difference. The difference alone has no physical meaning.

A physical interpretation of information would be, for example, the sum of the possible states and aspects of an object. A sheet of paper can appear in many ways: smooth, folded, crumpled, etc. The detailed description of these alternative possibilities could even be translated into an encoding in bits. Another way of considering physical information would be the effect of an object on its exterior environment. Thus, the information from a body would come from the possibility of its reconstitution through its influence in the external environment, and this is the basis of the so-called holographic principle. Even if the effect of a body in the environment is dispersed in its surroundings in many subtle levels of interaction, it would theoretically be possible to read these effects, to go back to the original object, and even to reconstitute it. We could also, by the same principle of "micro-reversibility" (Susskind, 2003), reconstitute the order of a deck of cards thrown into the air. It would also be hypothetically possible, with the right adjustments, to burn a newspaper and read the news in the smoke.

A popular theory of Stephen Hawking concludes that when an object such as a sheet of paper falls into a black hole, all possibilities of internal organization of this object disappear, evaporating from reality. Furthermore, any bond between a sheet of paper and its counterpart in the black hole will be essentially untraceable. Due to this, in two senses, information disappears from the universe. However, the conclusion does not presuppose the annulment of the causal link, but rather that it is irreversibly uncertain. Even in this case, information cannot be reduced to a mere causal link—they are not equivalent and interchangeable concepts.

The definition of information does not specify a physical law, but an access to a physical state, by means of an earlier one, or by the "exchange" of modification between properties. The correlation it specifies always requires some extra physical law, which has no relationship to the meaning of "information." This is an extra ingredient, added to the possible interpretations of correlations between events.

Information can be measured in degrees, such as, in a more basic way, the possibility of defining a space of variation. If the range of variation cannot even be defined, the information can be elevated to an almost absolute uncertainty. The sand of an hourglass is disordered, but its general shape is determined by the vessel. This portion of sand thrown on the beach, however, would blend in such a way with that of its surroundings that it could never be reunited. To put it another way, for every uncertainty, there is always a greater uncertainty. However,

what about causality: does it make sense to say that it varies in degrees? Are there more or less causal links in nature?

For example, the whole technological effort to create a communication device consists in uniting the two ends of the causal link. Engineers always face two problems, which can be considered as two types of uncertainty:

a) The same original event may have, in addition to the expected effect, other undesired and unanticipated consequences. We can call it uncertainty in the forward direction.

b) The same final event can have several causes, and, even if it is expected, it may have been determined by causes unrelated to the original intention of the designer. It will then be uncertainty in the reverse direction.

If the same event can be determined by several causes, then there will be less information regarding any one of these specific causes. Thus, the information depends on the reliability of the connection between an accurate type of cause and a specific type of effect, linked in a systematic correlation. As much as we have confidence, it is always possible that the same expected event has different causes and that the same action repeated countless times has, on one day, unforeseen consequences. A common concept in these cases is the understanding that information is the degree of trust in a cause-and-effect relationship, but *confidence* (certainty or uncertainty—entropy) will depend on the attribution by an observer—it does not exist by itself as a physical property. Information is a subjective property of an observer; it cannot by itself objectively constitute the observer.

When assembling a device, it is necessary to consider the two uncertainties simultaneously, to isolate the possible consequences of determined processes, and to isolate the possible causes of others. However, we will never have full confidence that a technological device will always work. This is because causality extends and is entangled by networks of relationships that will always remain inscrutable. Every causal link can extend to the last confines of the universe. By uniting the end points of the causal process in an informational reading, we bring only the causal consequences, but not the very causality that existed between them.

Instead of Bateson's motto, we might come up with another definition: "information is the cognitive tension between a known and an unknown state." It is also vague but says more about the scientific and technological problem in question. The difference between a clear and noisy signal, the recognizable individualism between signals, the greater or lesser specificity of the encoding, the confidence in the selected state, as well as the origin of an event among several alternatives are decisive factors in the degree of the known and unknown. It is irrelevant to say that they are causal, since everything is causal in a sense. However, this definition is subjective and requires an external observer to evaluate it as information. There is a parallelism between causality and information, but not an equivalence. The existence of an information theory is the proof that extracting information from causality is a lot of work!

5) Integrated information

Considering all these problems, there is a catch: these objections would only apply to the classical approach, for which information is always extrinsic. The Integrated Information Theory (Tononi, 2012b) proposes defining an "intrinsic" information model. While accepting Chalmers's premise that information as an articulation of states possesses a correlation with the phenomenal or experiential perspective, Tononi proposes that it is not the complexity of an architecture that generates it, but rather the integration. He distinguishes the information in the

intrinsic perspective, permitted by integration, from that dependent on the observer, differentiating his theory from the traditional approach: "Shannon's perspective may be correct only from the extrinsic perspective of an observer" (2012b, p.145). He proposes a *new theory of information*, which deals with a very different field of problems than those of the usual approach.

To explain the concept of integration, the author uses as an example photo-diodes from photographic cameras, which are small receptors of light ordered in rectangular grids, forming true "bit maps." Each of them only works in two states (on or off), according to the light awareness. By combining the states of these small elements, a very large number of images can be formed. However, since each one only functions independently of the others, the repertoire of photographs will be vast, but the total information will be limited by the calculated value of entropy, the sum of the probabilities of the separate "on/off" events.

Furthermore, the author describes a system where information is integrated, in which the units of information are interdependent and influence each other. In this system, the information of the whole will be greater than the sum of the separate units. In the case of a photo-diode, activating it or not is independent of the others; it relies only on the external light that it receives. If it were in an integrated system, the activation or deactivation would depend on a whole internal repercussion of the state of other points and not only on external influence, revealing a pattern of interaction that is beyond them.

An integrated system will have, owing to its combination of patterns, a larger repertoire of differentiated states (that is, it will have more information). Even if our brain is made up of several distinct structures, they can only function together. It is not possible to evaluate the functioning of one cerebral region in isolation from the others; it can have a defined function, but removed from the cerebral context, it possesses no function whatsoever. So, even though brain activity is a set of processes occurring separately, we have a unified conscious experience.

A system in pattern theory changes state due to a scaled correspondence between its states and extrinsic factors. In the model of integration, across the spectrum of states, there are those determined by perceptions and sensory influences (by the impact of the outside world), and there are those determined by their own architecture. Intrinsic information then amounts to the repertoire of informational states that are determined only by the possibilities of internal articulation, that is, only by the patterns of integration. Internal structuring creates many alternative combinations that are different expressions of the "network" itself, irreducible to their components, and, thus, intrinsic. The complexity of reality can be encoded by the vast spectrum of complex states of our integrated brain. In comparing the two sources of states, it is possible to measure the adequacy of the integrated informational space against the reality:

> A working hypothesis is that the quantity of "matching" between the informational relationships inside a complex and the informational structure of the world can be evaluated, at least in principle, by comparing the value of when a complex is exposed to the environment, to the value of when the complex is isolated or "dreaming" (Tononi., 2008, p.240).

There is another aspect of the theory, which is called the "principle of exclusion." Integration is opposed to the dispersion of information in the external environment. Information is integrated and unitary in a specific portion, but is fragmented externally, at the various levels of interaction with the environment. Thus, only the most effective information, in being a result that is not reducible to parts, can count in the "calculation" of the information. More subtle and less effective time scales or degrees of interaction are thereby excluded from the measurement. Only the level or region of maximum integration is of interest to the measurement of a state of complex interaction.

If informational states allow "encoding," then their wider repertoire signifies more possibilities for phenomenological states. To return to the idea of information, a state is

significant only in the midst of a space of differentiated states. The more informational a space is, the greater its spectrum of differentiable states becomes, the more exclusive these are compared to others, and the more informative they will be. Effects of disturbance or noise, which hinder the distinction between states, limit their informational character. That is, as a state must be a well-defined "grain" in the midst of a repertoire of possible states, any superposition of states signifies *increased uncertainty.*

However, the adoption of this principle shows the difficulty in moving away from the pattern-theory approach.[1] Although discussing a different concept of information, the author continues to apply the classic form of thinking, while attempting to transpose it to another domain. In an attempt to present a theory that explains how information can cease to be extrinsic and become intrinsic, he adopts two premises originating from the paradigm that identifies information in causality. The first is the idea we have already encountered in Chalmers, according to which a state is *intrinsically* informative simply by being within a space of states. The adoption of this interpretation brings with it the problem of mixing the subject and the object of the information. The second is that if information is comparable to causality, then the information transmitted in time will correspond to the causal succession. The evolution or integration of information in time will occur as the system evolves in a less uncertain sense, and it is this process that "unites causality and information." These two problems will be discussed below.

5.1) Subject and object of the information

The theory stems from three propositions implicitly assumed when dealing with an information system consisting of a space of states. The informative character of the state will then include a type of "cognition" that it will naturally have in respect to i) its place within the space of states; ii) its representation of reality (or what it "encodes" from reality, for example, one bit for the alternation between light and dark); and iii) its role, not as a mediator, but rather as the subject of the information. This results in the correlation that the author makes between the discrimination or selection of alternatives and their intrinsic aspect. The objection that follows is conceptual; it is relative to the model and its understanding of information. Consider these two propositions of the theory:

a) an experience (state of) exists only in relation to the repertoire of possible experiences:

> "For the IIT, as long as a system has the right internal architecture and forms a complex capable of discriminating a large number of internal states, it would be highly conscious." (Tononi, 2008).

---

[1]   There are objections to the theory (Cerullo, 2015) regarding the generic character of the definition of integration. Depending on what is defined as the limits of a system, integration can be applied to anything, and many things can have properties that are not reducible to their component elements. Any physical system can have internal states determined only by combinations of its elements, as opposed to others determined by external forces, and thus every system will have some kind of intrinsic perspective. There is no doubt that the brain is a richly integrated organ and that this characteristic is decisive for its functioning as opposed to other biological structures. However, this does not mean that we should regard simple integration as a sufficient condition to explain its nature. Presumably, there are several alternative models of integrated systems, and the question is precisely how the brain integrates, which is the point at which IIT presents problems. Many things can be considered integrated, such as microorganisms, atomic structures, etc., and some integrated things, like the brain, generate attributes like the mind. Thus, the issue is not simply that there is integration, but what integration occurs in each case.

b) the more information each state possesses the better it is at differentiating, or excluding others:

> Consciousness is exclusive: each experience excludes all others – at any given time there is only one experience having its full content, rather than a superposition of multiple partial experiences; (Oizumi, Albantakis, Tononi, 2014)

These two statements sound contradictory, but are simultaneously defended by the author, in their respective order, by the following:

a') a greater repertoire of states means a greater possibility of the "encoding" of reality, which signifies more information.

b') an experience must be a state well defined and well differentiated from others, which is implied in its informative character. A disturbance in differentiation is equivalent to noise.

Although the theory deals with integrated information, the meaning of the adopted information is exactly the same as that of the original theory—it is the reduction of uncertainty about a state, within a space of states. However, the distinction between the subject and object of the information, which is well established in the original theory, becomes critical here. This confusion between subject and object will lead to several paradoxes.

The existence of alternative states of a structure, even integrated, is not sufficient to define intrinsic information. Even though integration allows for more states, it still assumes the same reasoning as an extrinsic information model, since by the principle of exclusion, each state is only informative when interspersed with the others.

Even when a state is significant within the repertoire of states, it can still only be an expression of the alternate structure. How can one state can receive information from the others, thus perceiving its place in the space of states, if one of the conditions to be informative is that it differs from them, and exists in order to exclude others? Not being in a simultaneous confrontation, in a superposition, one state cannot have meaning in relation to the others, but only express itself contrary to them. Thus, it will need to be simultaneously distinct and the same. It is not possible for a state to be intrinsically informative under these two conditions (being distinct, to reach its degree of distinction), as they result in opposite consequences.

It is a problem similar to defining information as the "difference that makes a difference." A difference can only be assessed by something that is different, and by something that is similar, which confers on it the degree of difference. A state excluded from others is neither different nor similar—only in a simultaneous confrontation of any differences and similarities does difference become a significant property.

There is a difference between possessing a larger repertoire of states—due only to the possible combinations of their internal elements—and having the ability to distinguish between them. Owing to the principle of exclusion, the system cannot even *discriminate* between the alternatives. This is the extrapolation that the model does not allow. One state cannot contain in itself the representation of its degree of difference or similarity with the others, without raising itself abstractly above the system of alternatives in some type of superposition.

For the state to intrinsically possess its place in the space of states, it will need an abstract synthesis of the repertoire and of its place among the alternatives. However, this is precisely the explanation that the model lacks. This is only possible through an internal representation of the structure itself.

If the system alternates between states always opening one repertoire in relation to a past repertoire (disregarding whether or not this is the case of a brain), then this process needs

to be part of the internal content of the state, independent of the repertoire, or in other words, a general property must be "intrinsic" to a particular state. Only then will it be more than a mere (internal) contingency between one state and another. If having an intrinsic perspective means being self-referential, the system must deal with abstractions and generalizations; that is, it needs representative properties. For the author (Oizumi, Albantakis, Tononi, 2014), self-reference is a kind of internal self-restraint, a limitation of a system to its possibilities of articulation. However, in this case, the model is not capable of being self-referential because it is not representative. If a system or module only switches between one state and another, it expresses the potentialities of this structure, but its states will be neither representative nor self-referential.

If we think about the linguistic aspect, the representation occurs through the use of symbols, and these express concepts. A symbol[2] is applied to something when the object in the world fits into the classification of the concept; that is, to represent is to compare a concrete and singular element or object with a general and abstract definition, valid for an entire class of elements. The activity of representing parts of the general, and without this pre-existence of the general defining character, there is no way to designate an object precisely. In this sense, a representative property is one to which an element places itself in parallel to its generalization.

Withdrawing from the linguistic meaning, representation becomes the very act of thought that considers objects specific to their immediate reality and confronts their generalized perspectives. Without this act, linguistic representation itself would be impossible. Representation, as an activity of thought, is the abstract extension of the general from the singular, of the absent from the present, of the space from one of its positions, of the time from the current moment.

However, a referential property requires another step; the property needs to be self-inclusive. All representation is only seen in terms of the other possible representations, when the representation as a property becomes an element of a particular representation. Thus, self-reference is a further step of mental activity, when the abstract process as such accompanies its content. To think of something as a specific object in relation to a general category is to abstract; every abstraction is possible within the others, and the generalization of these same abstractive capacities will accompany every particular abstraction.

Thus, the model, because it is not representational, does not show how the structure can discriminate between states as it does not explain how it can make an abstraction of itself as a space of states, independent of movement to one or the other. Therefore, the informative character will always be outside the particular state, even if it is the result of an independent articulation of external influences, or irreducible to its parts.

5.2) Temporal integration

The model proposes that the system is also integrated in time, in succession; the integration contemplates the past and the future of the system, and, therefore, it unites the causality and the information. If information is to some extent causal, then it is possible to "physically" describe the evolution *as informational*. The evolution of the system in time is

---

[2] It is not possible to represent it by singular or concrete objects, because one object can be as different or as similar to another by any aspect one thinks about. One cannot represent "sand" with a handful of sand. To use another concrete element to isolate one aspect does not help as this also has its own collection of aspects. That is why signs, notes, and images have such limited communicative power and are only effective within their specific contexts.

determined by informational criteria, which causally determine what the next state will be, and what the previous state was, given the present state.

> Qualia space (Q) is a space where each axis represents a possible state of the complex, each point is a probability distribution of its states, and arrows between points represent the informational relationships among its elements generated by causal mechanisms (connections). (Tononi, 2008)

The integrated system as a whole has a vast repertoire of possible states. Since it does not make sense for it to jump randomly from one to the other, there must be an order that directs it to the most-likely states, ignoring the others. Within each group of most-likely states, if they all have an equal probability, the uncertainty will be high, and the system will be indefinite about its future. More certainty means that there are, among the candidates, one that is more likely than others, and, thus, the system will always pass between them in this direction. If the information on the previous state causes the system to lean toward one state, then the certainty will be greater, and this will be the "chosen" future state. The author proposes that this is integration and the "intrinsic" perspective in the temporal sense:

> A central postulate concerns information: from the intrinsic perspective of a system, a mechanism in a state generates information only if it has both selective causes and selective effects within the system – that is, the mechanism must constitute "a difference that makes a difference within the system". This intrinsic, causal notion of information can be assessed by examining the cause-effect repertoire (CER) specified by a mechanism in a state – the set of past system states that could have been the causes of its present state and the set of future system states that could have been its effects. If a mechanism in a state does not specify either selective causes or selective effects (for example by lacking inputs or outputs), then the mechanism does not generate any cause-effect information (CEI) within the system. Ontologically, the information postulate claims that, from the intrinsic perspective of a system, only differences that make a difference within the system exist. (Tononi. 2012a)

The situation is similar to that discussed in the presentation of information theory, in which each element has a conditional distribution relationship with the previous one. If we have the first letter "I," (in Portuguese) it will never precede the "space" character, but this can occur for another vowel or consonant. As we complete the word, there will be a repertoire of alternatives from the previous step and another repertoire from the following step. We are eliminating unlikely states and concentrating on the most likely, and, in that sense, the conditional distribution will say something about the past and the future of the system. A similar model can also describe the moves of a chess game, their combined uncertainties, possibilities with each new move, and the possibility of stepping back through the previous moves. For each new situation in the game, there are those moves that do not result in any tactical advantage, that lose time or advantage, and there are a few that have a true strategic meaning, and thus the uncertainty is less with them. The arrangement of the pieces of the board also allows us to draw conclusions about the past and the future.

The problem at this point is the extrapolation of information to causality, since it starts from the premise that successive conditional relationships will express the causal sequence. However, how does the mapping of states by a large chain of probabilities make the past and the future intrinsic to the state? The author argues that there is temporal integration based on the information principle, like "the difference that makes a difference," because in this way an event can be treated in both causal and informational terms, and then an informational selection criterion will result in the causality being transmitted between states. However, this equivalence cannot be made.

Even if the information is being read from a device that functions causally, it has a cost that is also causal. Imagine the construction of a scale, a thermometer, or a telescope. These devices require a number of technical decisions and interventions, such as materials,

calibrations, tests, experience, etc., in addition to the machinery to manufacture the final product. The ideal thermometer, scale, or telescope results from the filtering of all unwanted causal links, which could impede accurate measurements. Information such as the correlation between two categories of events—in a way that the occurrence of one presents a sufficient degree of reliability as to the occurrence of another—is the filtering by one criterion of a causal link from several. A measuring instrument is only informative by what is causally filtered from it.

The filtering of excess system causality will not be in the result, precisely because the information is a causal simplification. The work of constructing a measuring instrument removes complexity. Information is a simplification of a complex causality, the reduction of several dimensions of complexity to just one. However, the "work" of simplification does not enter into informational reading. The author applies this same reasoning to the discussion on the principle of exclusion:

> To understand the motivation behind the exclusion postulate as applied to a mechanism, consider neuron with several strong synapses and many weak synapses (...). From the intrinsic perspective of the neuron, any combination of synapses could be a potential cause of firing, including "strong synapses", "strong synapses plus some weak synapses", and so on, eventually including the potential cause "all synapses", "all synapses plus stray glutamate receptors", "all synapses plus stray glutamate receptors plus cosmic rays affecting membrane channels", and so on, rapidly escalating to infinite regress. The exclusion postulate requires, first, that only one cause exists. This requirement represents a causal version of Occam's razor, saying in essence that "causes should not be multiplied beyond necessity", i.e. that causal superposition is not allowed. (Tononi, 2014)

In fact, information is a type of exclusion. An informational reading will select specific causal conditions to the exclusion of others. The causal world consists of things that change each other, and these changes occur in a continuum that extends to the confines of the universe. A system that extends that far can have no intrinsic perspective. In this case, there would be only objects and no subjects. If one part of reality is linked to any other, an infinite causal chain is created in which information about one state of things cannot be isolated, since it neither delimits nor restricts its equivalents to the source and the receiver. The source is everything, and the receiver is in everything. If everything is information, then nothing is information. This interpretation is in agreement with the original information theory.

From the perspective of the neuron, this makes sense, but with an integrated structure, *the exclusion is no longer applicable*. Information is the reduction of complexity, but integration is essentially complex and by definition cannot be simplified; it consists of a multiplicity of causes and effects that cannot be decoupled. Every trigger or event of association or dissociation in positions of the network is caused by several competing events of interactions that it is not possible to isolate or treat as exclusively decisive. Network integration means the integration of causality itself. Thus, information is essentially mixed with causality, or it is the expression of a complete causality as an indivisible block, so causal filtering or simplification becomes impossible.

Every link will be entangled in others so that it is not possible to perform any filtering, and, thus, no informational reading will adequately express reality. Isolating a criterion of simplification or exclusion from all others may be more complex than the network itself, and thus it cannot be handled. The cost of uniting causality and information will be to put an end to the information.

If one still tries to equate causality and information, it is presumed that information will contain (intrinsically) the excluded causality, which is precisely what defines its informative character. If a causal process is informative in relation to what it excludes, the excluded links cannot by definition constitute what remains. If information is a reduction of dimensions of

complexity, an informational reading of integration will be at the cost of the integration itself. So it can only be causality itself *en bloc*, not an informational reading that determines the succession between states.

If the difference is determined by one criterion to the exclusion of others, the criteria that define the difference are also causal, but in the result they are not included in the measure of the difference. If uncertainty is what unites the stages, then there is at most a causal reduction of what is happening in the system. Each state will have the result of the difference,[3] but it will not have the process of differentiation, and this is what matters in the intrinsic perspective.

However, another interpretation may be adopted. The division of causes would be contrary to the idea of integration and would result in dispersed information about the world. Thus, integration reduces the complexity of a state to a unique informational "arrow," in successive states of *full integration*. However, integration does not necessarily presuppose its absolute character, reducing a state to a single vector leading to singular cause-effect pairs, but rather it can be interpreted as competing causes being paired in a unifying perspective.

If integration were to be full, and its complexity then reducible to a single plane of a causal direction, then the experience we would have would be at best an indistinct mass of information. Take the case of perception: at the same time, someone can sense their legs, arms, hearing, and vision. Yet, in an integrated perspective, we can only perceive hearing, vision, and touch separately, because they distinguish themselves in this integration in a concurrent way.

If we accept that integrated systems determine complex interactions and a varied spectrum of independent states, then uncertainty is a decisive factor in the evolution of states if the motives for the modifications are extrinsic. Whenever information is treated as uncertainty, it is treated as something extrinsic. The intrinsic is independent of contingencies, while uncertainty is an extrinsic measure. Essentially, the evolution of an intrinsic information system cannot be described by a relationship between uncertainties. Uncertainty measures at most the relationship of the intrinsic to the extrinsic, but says nothing about the nature of the intrinsic.

6) Integration and superposition

If information in the traditional sense is the uncertainty about one state in relation to others, what is the use of the definition of integrated information if it follows the same line of thought? If the concept of integrated information means something, as opposed to simple information, should this concept not imply a greater level of communication between the potentialities of the structure? Rather than showing itself as a state of an object alternate to others, is it not an element in itself, distinct from the alternative partial expressions, elevating itself to a perspective that contemplates them? Thus, would it not be reflected in the transition of time, and the expression of the succession, as opposed to consisting of alternate states? Could not the differential of integration be the possibility that the evolution of the system is represented in its micro-relations, returning to it in an extended causal dynamic? As information is "greater than the sum of the parts," rather than expressing itself as one or another state differentiated from the others, could it not be precisely the *superposition* of these states?

We alternate between perceptual states because the succession of the perceptive experience accompanies the structure of reality—this is what imposes on us each time a perceived aspect of each situation to the exclusion of the others. However, any alternation or

---

[3] In this proposition "Based on the information postulate, a mechanism in a state (s) generates information from the intrinsic perspective of a system only if it both detects differences in the past states of the system and it makes a difference to its future states. (Tononi, 2012a)", the detection of difference is the informational filter.

confirmation of the degree of similarity or difference between experiences can only occur in the continuous competition between internal states. Memory is competitive, since the successive stages do not disappear but continue to determine the "tone" of the present state. This may be literally true; to prove it, just listen to the *C* note of the piano preceded by a *B* (prolonging the pause as much as you want) and then hear the same *C* after a *Db*. Out of a system of competing directions, we have only one structure that moves between potentialities, and any difference or similarity has no meaning.

The accompaniment of a succession of experiences only makes sense to the extent that one is confronted with another in a superposition, when causality becomes isolated from dispersed random links and bends over on itself. The superposition state described in this way not only establishes the inseparable correlation between events, but also the reconstitution of events to each state. It is the pairing between the cause and the effect, accumulated in a progression, in a continuum; it is not the mere contingency between one state and another, or between a state and a dispersed event that modifies it. It is the reconstitution of the causal link that determined the state; that is, the internal situation of the structure contains its own reconstitution; it says more about its reality. In an informational reading, such as in the use of a measuring instrument, the security of the causal link is precisely the question: how many different situations can determine the same reading, and how many similar situations can determine different readings?

If integration is what defines intrinsic character, then integration itself must be expressed and imprinted on the constitution of the state. Information can only be integrated through integrated causality. Integration by itself is like a space without bodies; it is only a precondition; it is a *form*. It is expressed only in succession; it is then that it turns the obtaining of information into part of the information. This is not a quantitative relationship, and it is not about the complexity of the structure.

The integration will then define a form common to all modifications and alternative states. The unity of consciousness, for example, is a property that accompanies every particular state and is an expression of the integrated form. Therefore, it can be considered that the unit is *formal* for being an expression of it. However, this is not a property that a state will or will not have; every state will express itself as a unit; this property will run through all internal variation, and there is nothing probabilistic about it. We can define memory, in simple and "phenomenological" language, as a kind of accumulation of subtle influences that progressively weaken and are superimposed by others, but which give a general notion of collectivity and unity to an action. Even more, we can verify the change of our internal states because we have simultaneously a sense of permanence, which accompanies and reunites all these states. Every tendency for one state or another will be within this unifying perspective; without it, we would have a Humean consciousness.

If it is a formal property, it is an anticipation of the property that a state will have in the continuity of time; that is, the state possesses in its instantaneity a perspective on the continuum, on the temporal succession; in itself, it has an extended causal perspective. Using Aristotle's distinctions of cause, we find that the cause of movement reveals the formal cause. If a formal property is part of the content of the state, and it accompanies every specific causal unfolding, then the content of each state is placed beyond its temporal location and has a reference to this permanent property. Every specific moment, every particular state will follow this dynamic.

Integration is the element that gives the succession of alternative states an ordering and a sense of unity. The shape of the structure reverberates in the form of the state. If unification is a consequence of structure, then this is the form of each state, and it is determined by the structure and this form; it will follow every eventual link, and it will be a permanent element. This unifying perspective expresses itself at every stage of succession; it is the unification of states in which this unification is present, and thus it becomes the content of its own act. The

structure will be expressed in each state, and the confrontation between states will result in the confrontation of the structure with itself. In the continuous confrontation between the momentary and the permanent, the self-referential step is determined. A state is only informative through confrontation with its generalization, and in this sense, a referential property will be a type of informational transcendence.

7) Conclusion

We initially discussed Chalmers's thesis that information possesses an autonomous existence in nature, constituting itself as "intrinsic," by expressing a rudimentary cognition of a system in a state in relation to the spectrum of possible states of its structure. Then we highlighted Searle's objection that information depends on an observer to interpret the alternations between states as significant. It was then proposed, in agreement with it, that causality and information are essentially distinct properties, and that the latter depends on the perspective of the observer, since it relies on the degree of access to unknown states, and uncertainty is the measure of that unfamiliarity. The anticipation of an observation from one perspective or another will predetermine an expected object or aspect, whether or not the information depends on the focus given by this anticipation. Since everything is causal, "being the difference that makes a difference" depends on the perspective by which the viewer sees the model. Thus, information as uncertainty is a subjective property.

Giulio Tononi's Theory of Integrated Information was also discussed, and it was proposed that the theory fails to demonstrate how the information of the brain becomes intrinsic, since its formulation, paradoxes, and contradictions arise precisely because it is an adaptation of classical information theory. The theory does not adequately address the subject–object distinction by requiring that distinct and alternating states be simultaneously significant in relation to the same space of states. In order to be able to deal with spectra of states in relation to particular states, it would need to be able to deal with this generalized perspective; that is, it would have to be representative.

The theory cannot unite information and causality by proposing that the uncertainty between successive repertoires of alternatives expresses the causality between states. In attempting to integrate past and future, this model only repeats the conditional distribution of the classical approach. This is flawed for two reasons. The first is that information will always be a dimensional reduction of a higher level of complexity, but the integration is essentially complex and not reducible in that sense. There is in integration an accumulation of levels of interaction and competing paths so interconnected that the multiplicity of causes cannot be reduced to a single informational cause without eliminating the integration of the informational reading. If integration is complex and concurrent, obtain information from integration by measuring the difference between states is not possible. However, if we have full and reducible integration into a unique cause, this ends with the distinction of perception, and the state information becomes an indistinct mass, which eliminates the information itself. The second reason is that the uncertainty in relating the stages is only an attribute of extrinsic elements, and an intrinsic information system should not be explained only by uncertainties, or even by the allusion to uncertainty.

Next, it was proposed that the intrinsic perspective of an integrated model is more coherent in a model of superposition of states. This modification may allow us to avoid the paradoxical aspects of the theory discussed above and provide a greater correspondence with mental properties such as memory and self-reference.

8) References

ARISTÓTELES. *Metafísica*, Bauru, SP: Edipro. 2006.

BATESON, Gregory. *Steps to an ecology of mind*. Jason Aronson Inc. 1972.

CERULLO, Michael A. *The problem with Phi: A critique of integrated information theory*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4574706/#pcbi.1004286.ref008. 2015.

CHALMERS, David John. *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press. 1996.

KOCH, Christoph. *Consciousness: Confessions of a romantic reductionist*. MIT Press. 2012.

OIZUMI, M., ALBANTAKIS, L., & TONONI, G. From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *Computational Biology*. 5(10): 1–25. 2014.

SEARLE, John. *The mystery of consciousness*. New York: The New York Review of Books. 1997.

___ *Redescoberta da Mente*. São Paulo: Martins Fontes.1997.

SHANNON, Claude E. A mathematical theory of communication, *The Bell System Technical Journal*. Vol. 27, pp. 379–423, 623–656, July, October, 1948.

SUSSKIND, Leonard. Black holes and the information paradox. *Scientific American*. 2003.

TONONI, Giulio. Consciousness as integrated information: A provisional manifesto. *Biol. Bull*. 215: 216–242. 2008.

___ Integrated information theory of consciousness: An updated account. *Arch Ital Biol.* 150(4): 293–329. 2012.

___ *PHI: A voyage from the brain to the soul*. Pantheon Books. New York, 2012.