triple**C** cognition
communication
co-operation

An integrated framework for information, communication and knowledge definitions

Roberto Gejman

Computer Science Departament, Engineering Faculty, Pontificia Universidad Católica de Chile, Campus San Joaquín, Santiago, CHILE. E-mail: rgejman@deloitte.com

Abstract: *An integrated set of definitions and distinctions in the information, communication and knowledge field is proposed. It is argued that more attention must be given to records, usually confused with information. Descriptions are shown to be the fundamental facts behind information. Information is defined as an abstract concept, free of material or syntactic constraints. Information is observer free but it is dependent on the shared ontological background of communities. The world may be characterized as being structured by only six components. Knowledge is the capacity to act effectively and it is not information. Information quantity and its relation with the entropy of physical systems are shown to be more ambiguous and less important than they are usually thought of. It is argued that information studies should move on from these old concerns to confront the vital information challenges in this globalized information society with information superabundance.*

The current dialogue about Information, Communication and Knowledge (which I will term the "ICK" domain or problem) has been for a long time focusing on its basic definitions without being able to reach a consensus or, at least, a small limited number of approaches, to settle such fundamental set of questions. This paper hopes to be a contribution to resolve such discussions by proposing clear cut definitions, rooted on primitive concepts that either are free of discussion or are new distinctions already implicit in our common sense.

1. The information – communication - knowledge foundations

The proposed definitions for information, communication and knowledge (in short ICK) need a few previous or foundational definitions or assumptions that are explained and presented in this section. The definitions to be presented are:

- The world
- Agents and communities
- The world texture
- The shared ontological background
- Descriptions
- Records

1.1. The world

The first definition is that there is a real world, including physical and abstract entities. Examples of physical entities are stones and chairs. Examples of abstract entities are the UK's economy, the Waterloo battle, the couple made up by Mr. and Mrs. Obama and the partnership conformed by John, Patrick and Abraham to invest in real estate.

1.2. Agents and communities

Some of the world entities are autonomous agents, objects whose future state cannot be easily predicted by other agents from an

examination of their current state and the prevailing world laws. Agents may include human beings, animals and robots. Our discussion is mostly interesting and makes sense for agent communities, whether they be anthills, herds, flocks, people groups or even heterogeneous communities whose members are drawn from two or more species. It is the need for social coordination and for individual and social effectiveness (success) that causes the emergence of the ICK phenomena. This is similar to saying that language makes sense or is interesting mainly in communal settings.

1.3. The world texture

It is important to define how the world is structured, in the "eyes" of the agents. It has been verified by more than 40 years of successes in IT systems design that only six types of entities, elements or objects are needed to characterize most relevant aspects of the world (Chen, 1976). These are as follows:

- Objects or systems such as people, products, customers, warehouses and vehicles.
- Attributes pertaining or associated to those objects such as weight, color, size, temperature, price and cost and other so called values or properties.
- Relations, structured by two or more associated objects, as for example "spouse of"; "employee of"; "affiliated to"; "part of"; "used for"; "owned by"; "partners in"; etc. (Relations are used here exactly as they are defined in set theory, as subsets of Cartesian products of two or more sets).
- Events, transactions or actions, such as a sale, a car accident, a marriage, a divorce, a birth, etc. These may also be described as state changes.
- Time, the dimension ordering the events over a linear infinite sequence.
- Laws that constrain the configurations that the above elements may be in such as "If A lays on top of B, then B must be below A"; "If X is running faster than Y, and Y is running faster than Z, then X is running faster than Z"; "if J is K's mother then K is

J's child" and so forth. (All true natural laws are included here).

It is important to state that any specific relation instance and every event instance is also an object. They may have attributes and they may be related to other objects. For example the couple formed by John and Mary, which is an instance of a relation, may have attributes such as the date of their marriage and whether it was a civil marriage or a religious marriage. It may participate as an object in further relations such as the one relating it to the place where the marriage took place and also to the "objects" Patricia and Kevin, their children. It is also very important to clarify that objects include both material things and also abstract objects, such as corporations, laws, projects, games, investments, intentions, goals, months, banks, nations, the weather, etc. In fact, all relations and events are abstract objects.

Obviously, objects, relations and events may be grouped into classes or types, which by themselves are also objects. The members of a class are usually called class instances.

Countless databases have been designed and built on these six concepts and operate successfully in the everyday life of almost any organization in the world, providing plenty of evidence that these basic building elements are powerful enough to describe most aspects of interest about the world and that they are deeply entrenched in people's common sense.

1.4. The shared ontological background

The world as described above, including the six types of components, (objects, attributes, events, time, relations and laws), is built and structured by the community members as they encounter the (real) world, interact between them and coordinate their actions. They are continually performing distinction operations, that is, they pick pieces of the world that are useful to single out, provide sufficiently precise definitions of their boundaries, properties, usages and relationships and eventually define names for them. (Maturana, H., Varela, F., 1992)

I must make clear here that this is not a negation of the existence of the real world, of the down to earth physical facts. It is just a statement about the impossibility of any agent to grasp the world in its entirety, in all levels of detail simultaneously and the need to refer to it by recognizable chunks. People, for example, must be able to say "Please, sit on that chair" and not be forced to say "Please seat on the set formed by that atom, and that atom, and that atom, and that atom …" This point of view could be traced to a social constructionism (Berger and Luckmann, 1966).

The results of all distinction operations are accumulated in a big abstract set, a big dictionary or thesaurus, which is shared by the community members, being an important component of the community's culture. Let us call this set the "Shared Ontological Background" or SOB. The SOB includes then all objects, attributes, events, the time, relations and laws; as well as the lawful interactions between them. The SOB reflects the community's agreement on how to look at the world and on how the world is structured.

Of course, as the community goes on existing and developing, the SOB grows and changes and so does also the community language or languages. As the ontological background includes not only its elements but also the names and symbols that denote them, the latter are used to bridge the relationship of the shared ontological background with the community languages. Thus, for example, verbs, substantives, adjectives and adverbs usually denote events, objects and attributes. Of course, any modern language such as English or French has more to it than only objects, attributes, relations, time, events and laws, but language is not the same as the SOB and this is not an enquiry into the domain of language.

Any new SOB element enters a process that is very similar to biological evolution. It is first used by a small minority; then it is adopted by other members that start using it; while this process is going on, minor variations may occur, changing the meaning or the usage of the element. Finally, those elements that are deemed useful or convenient by the community stay in the language and in the SOB while others may be discarded. There are countless instances of this process, going on in most communities. It is remarkable how these processes happen without any direction or organization, in a quite spontaneous way. This kind of process is what has been described for "memes", the powerful idea proposed by Richard Dawkins. (Dawkins, 1989).

The SOB is nowhere to be stored or officially defined. It is an abstract system, evolving with the community that shares and creates it. Dictionaries, glossaries and encyclopedias attempt to capture it, but fully capturing the SOB is by definition a never ending task. When members of two different communities come together and try to interact, a major obstacle that always emerges is the lack of intersection or commonality of their respective SOBs.

1.5. Descriptions

We may observe the crucial role played by descriptions and the act of describing. Business people need descriptions of the market state and of how sales are progressing; country leaders need descriptions of the economy; travelers need descriptions of weather and immigration policies; and so on. Descriptions perform a critical function for agents that strive to be effective in their decision making and planning. Also descriptions of past events are very important for education and learning. Every record in a database represents a description of an object or an event. Most media news are descriptions of events happening elsewhere.

It is important and interesting, though, to get a deeper grasp of what a description stands for. Descriptions are ordered triplets $<S, s, t>$, where S is any nonempty set, s is a subset of S and t is a time period. It may be interpreted as a selection of the subset s, from the larger set S, at time t. Subset s is a smaller group or set of possibilities from a larger one, regarding some situation in the world, at a specific time period. For example:

- A weather description selects, for a given time period, one atmospheric state from a range of feasible atmospheric states:

  <Set of all possible atmospheric states in a given place, set of most probable weather states for that place, date for which this weather prognostic is valid>

- A temperature report for a sick person, for a given time period, selects one number or quantity of degrees from a scale or range of feasible temperatures a human being may exhibit.

  <(33°C,42°C),(38,5°C,38,6°C), (Feb 23,2009, 16:43)>

- A description of a person's age that is stated as "She is in her sixties" chooses the subset of numbers between 60 and 69 from the usual range of ages that a person may achieve.

Also, there are a small number of different description types. The three examples above are of the type "system state descriptions", usually stating the value of attributes of an object in the world at a given time, or, in other words, partially describing the state of a system at a given time. On the other hand, "structural descriptions" select, for a given relation instance, the objects associated under that relation. For example, an instance of a Teacher-Student relation may be described by choosing who the teacher is and who the pupil is, from the set of all teachers and from the set of all students. Furthermore, "meta descriptions" will describe the systems or objects, attributes, types of events, relationships and laws, available in the SOB.

To give a complete account of what a description is we have to add the time dimension, for any description that selects the state of a system, does so for a given time period. Descriptions that are valid or true forever (see below about true and false descriptions) are simply descriptions valid for an infinite time period, covering all time relevant to the given community. Laws, for example, should be valid for ever.

It is important to note that there is no way of making a description without having a shared ontological background. In any form or in any language that the description will be expressed it will necessarily have to refer to the elements of that SOB. It is not possible to describe the state of system S if that system is not previously included in the background. The SOB provides the building blocks to build a description.

Descriptions as defined here may be consistent or inconsistent with the real state of affairs in the world. Usually, a consistent with reality description would be termed "true", whereas an inconsistent description would be labeled "false". Besides truthfulness, descriptions may be measured also by their precision. A description is more precise the smaller is the number of alternatives it chooses from the set of all possible alternatives. It is less precise to say "He had between four and six beers" than to say "He had five beers". In the first case the selected subset of alternatives has three members: 4, 5 and 6. In the second case the selected subset has only one member or element: five.

1.6. Records

Records are material objects of an utmost importance to civilization. They play crucial roles in business, education, history, science, law and arts. They have had profound influence on humankind development and in advancing such important endeavors such as developing literacy, culture, democracy and the media. Electronics, the latest record handling technology, has brought with it one of the biggest revolutions in history, whose huge consequences are still to be wholly understood. However, little attention has been paid to them as subjects of study or research. Records are barely recognized as relevant objects at all and are often confused with other concepts, obscuring the analysis.

What is a record? All records are material objects, whose basic state may be altered by agents, changing it into one of predefined set of choices. For example, the basic state of a rectangular sheet of paper is to be blank or empty. The accepted choices to alter this basic state are pencil marks that are arranged in lines parallel to one of the paper borders. Each line is a linear sequence of characters and blank spaces, where the characters must be picked up from a small set, called an

alphabet. The resulting sequences must comply with the vocabulary and grammatical rules of a given language.

The SOB will include "writing" or record producing rules, and "reading" or interpretation rules for the adequately written records. Of course, "writing" and "reading" are used here in a more general sense than usual.

Records will include of course paper with ink or pencil marks, computer printouts, stones engraved by ancient symbols such as the cuneiform writing system, computer databases as well as printed or electronic books, newspapers and magazines. They will include also musical scores, electronic and optical symbol storage systems, and television and computer screens. All records have a material substrate and a range of states that represent the symbols.

Verbal utterances are records too. In this case the record substrate is air and the symbols are represented by air pressure waves; something similar happens with electromagnetic waves records. What happens here is that some records have longer live spans than others, some may last for years, and some may last only milliseconds. Some may be stored and transported; others cannot be stored, unless "translated" into a different type of record, such as a voice recording in a magnetic tape. Images and pictures, whether in X-rays, internet web pages, films, building blueprints, maps or other media are also included in the records concept, as well as flags, smoke signals, credit cards, forms, identification cards, RFID devices and so many other types of records. Bodily gestures use the body as record substrates to represent symbols.

Records are used to allow for coordination; to facilitate business and projects; to establish and honor commitments, rights and obligations; to educate and train; to amuse and entertain; to produce works of art; to convey authority; to establish laws and regulations; to keep accountabilities; to guide people around their cities and countries; to plan and structure activities, sometimes across continents; to rejoin widely separated family members and friends; to share pictures and music; to create new business; to replace the handling of big, cumbersome or dangerous objects, by handling symbols that represent them. Even animals use records to mark territories.

Records are so pervasive and fundamental that we sometimes fail to notice them and fail to distinguish records from their contents or referents. However no interaction between humans is possible without them. Talking, sending an email, taking a picture, sending a fax or a post card, drawing the blueprint of a gold mine or the map of an island are all record producing activities

Records may represent instructions, as in recipes book; they may contain poems or fiction narrative and they may contain declarations such as business contracts and national laws. For our discussion, however, one of the most important jobs records have is to represent descriptions and a very big proportion of them do that. Most forms and database records register one or more descriptions. Pictures and films are often used to describe situations, places or events. After a building is built, its blueprints cease to be normative, describing the architect goals and are kept, now to describe the building for future maintenance or enhancement.

In fact, whenever an agent produces a description it must do so by producing some sort of record or records. This is inevitable, since any expression of a description must be performed over a material substrate, using community shared symbols. The descriptions represented in a record may be denoted as "recorded descriptions". Note that a recorded description may be true or false, and may be more or less precise, in the same way that descriptions are.

2. Defining information, communication and knowledge

Building on the foundations laid out in the prior section, namely the world, the agents, the communities, the world texture, the SOB, the descriptions and the records, I may now present the proposed definitions for information, communication and knowledge.

2.1.   Information

The proposed definition of information is motivated by the fact that many recorded descriptions may refer to the same state of affairs in the world:

a) "The Eiffel Tower is in Paris" written in a paper encyclopedia.
b) "La Torre Eiffel está en Paris", uttered by a Spanish speaking person.
c) "The Eiffel Tower location is Paris" written in a tourist web page.
d) One of the most important attractions in Paris is the Eiffel Tower.
e) "The Eiffel Tower is one of the projects of engineer Eiffel. It is in Paris", as written in a history of engineering book.
f) "From the Eiffel Tower, the Cheops pyramid and Big Ben, only the first one is in Paris", heard in a TV show.
g) Relation <Eiffel Tower, "is in", Paris, January 2005>, an abstract mathematical object, represented in writing.
h) "The Eiffel Tower is located in France's capital city", found in geography textbook.

Not all statements above are semantically equivalent, but they all have an element in common: the fact that the tower is in Paris in a certain time frame.

For me, it is very appealing and intuitively consistent to isolate that common essential factor and reify it as an entity by itself. It is obvious that this common factor must be independent of any material substrate, coding system, language and grammatical rules. It is also tempting to see that this common factor is independent of any observer. What is it, in general, this common factor of all recorded descriptions that refer to the same state of affairs? I suggest that the common factor is an abstract description, born from the very structure of the world and its dynamic development along the time dimension, as captured in the SOB. In other words, every state change and event in the world generates such an abstract description that uses the SOB elements to be structured. Equivalently, every time the value of an attribute for an object changes, a new abstract description is born.   These abstract descriptions are of course "stamped" with a time period, denoting the time period for which the description is true. If a rose is borne red at time t_0 and it changes to yellow at time t_1, then at least two descriptions will be born to describe both events and the situation in between. If a tree falls in the wood and no human observer is on sight, information is still generated.

These abstract descriptions are always true, since they are generated consistently from the world itself.

To these abstract descriptions I propose to apply the word "information", that is, every information is a partial description of the state of the world at a given time period, free of any material or linguistic representation.

This is in fact what many have in mind when they use the phrase "propositional content".   A proposition, after all, is a description, selecting a statement from a background of all alternative statements that could be posed. "a is G" means that a is not H, I, J, K, etc. And the word "content" is used to denote whatever is the essence of a materialized proposition or statement; that behind the scenes abstract essence.

This   definition   has   some   important consequences, namely:

- Information is not the same as a true recorded description. A true recorded description is a representation of that information.
- Information is dependent on the agents definition of the SOB, but a given information instance may exist independently of any agent or agents. In other words, information is observer independent, but is not community independent.
- Information is an abstract, intangible object. Agents cannot access information, only recorded descriptions of it.
- Information is by definition true, but recorded descriptions may be true or false.
- If all true recorded descriptions are destroyed or made inaccessible, the information represented by them still exists.
- Once a shared ontological background is defined for a given world a field of infinite informations is created and evolves and grows together with the world.

This information definition has some strong points to it:

- It is not dependent on unclearly defined or difficult concepts such as data, meaning or difference.
- It builds on common sense and easily grasped concepts.
- It is not dependent on any syntactic or coding factor, such as well formedness.
- It is easily amenable to be more precisely and rigorously defined as a mathematical system. For example, an information about the value of an attribute for a certain object at a certain time period is an element of the Cartesian product of the set of objects, the set of attributes, the set of values and the set of time periods. For example:
  (John, height in meters, 1.79, 24-11-2003)

Notice that the elements of the above vector are not names or symbols, they are the elements of the sets themselves, not denotations. Also, note that there are no syntactic issues involved. John is an element of the set of objects; "height in meters" is an element of the set of attributes; "1.79" is an element of the set of possible values of attributes; and "24-11-2003" is an element of the set of time periods.

- It helps to close the debate about information being true or not. (Floridi, 2005).
- It provides a clear definition of the semantics of messages and records.
- It provides a powerful distinction between records and what those records represent. By isolating the information factor, it opens up the possibility of developing a science, a technology and a management discipline for records and databases, (partially existing, but not recognized as such). Furthermore, it opens the possibility of creating an information curriculum for the knowledge worker, to have at least the same importance as the mathematics and natural sciences curricula in the information society schools.

2.2. Messages

The proposed definition for messages is quite simple, given the context already build: messages are records representing information transported from one place to other place.

2.3. Knowledge

Confronting the definition of knowledge will not be as easy as the definition for messages. Knowledge is commonly conceived as information in the head, brain or mind of people. In philosophy the definition "justified true belief" is common. Knowledge is accepted also as a property of organizations and, as a consequence, the discipline of knowledge management has started to develop.

Some authors think of knowledge as the capacity of an agent to act effectively in one or more domains of its live and world. For a noted example, Maturana and Varela deny the conventional distinction between knowledge and action. Instead they say: 'All doing is knowing, and all knowing is doing.' (Maturana & Varela, 1992).

"Effectively" means that the potential behavior and its outcomes will be better for the agent or for its community, will better promote its survival and well being and will also be more efficient, i. e. require less time and resources to be performed, as compared to non effective behaviors. I like this stance for the following reasons:

- Current science development still does not allow us to look into the heads, brains or minds of people and see whether they have knowledge. However, we may test their knowledge by observing their behavior.
- The expression "justified true belief" is too ambiguous. It does not lend itself to measurement or comparison of knowledge. It does not state clearly who the believer is, who is declaring the truthfulness of the belief and how is it that the justification has to take place, how do we know that there is enough of it and who is responsible for building the justification case. One consequence of these unclear points is that we are not sure whether the "justified true

belief" is a subjective property of a person or is a judgment of an observer.

- The "justified true belief" is also difficult to bring into the practical world of "knowledge management".

Defining knowledge as capacity to act effectively has as its main drawback its potential dimension. We as observers may verify that an agent exhibits effective behavior. However, if no such behavior is exhibited, we cannot affirm that the knowledge is not there.

I believe that knowledge is better understood if we separate or distinguish between some knowledge types.

2.3.1. Memorized information knowledge

The easiest type of knowledge is the capacity to answer questions with appropriate or right answers, without using records. If I ask a person a number of questions about or in a given domain and I get a high proportion of right answers, I am justified to declare that the person "has knowledge" about that domain. Rightly answering questions is an example of effective behavior, and I need no theory about the inner workings of that person's brain to come to the conclusion that he or she knows. Moreover, the more different (non redundant) right answers the agent is able to give, the more knowledge it "has". We may use expressions such as "this person has a good memory", "this person knows a lot", and "this person has a lot of information".

This type of knowledge behavior also includes the producing of (recorded) descriptions without waiting for a question to be posed. That kind of behavior will happen when a person is observing a certain situation whose resolution may benefit of him or her contributing some information that other participants may not have or may have overlooked.

We will say that an agent that has the capability of producing a true recorded description "has" the corresponding information. The agent may "lose" or "forget" that information, temporarily or permanently. This does not deny or contradict the possibility

of an agent of producing a false recorded description.

2.3.2. Indirectly memorized information knowledge

A second type of knowledge will be the capacity to answer questions, but now with the help of records and other agents. This is the case when a person has no answer but knows where and how to look for it. This is again a type of effective behavior. We could say now "this person knows how to find information in the Internet" or "She knows who to ask for". This is also a kind of effective behavior.

We may generalize the above described two types of behavior into a more general class of information providing behaviors, such as teaching and reporting.

2.3.3. Thinking

A third class or type of effective behavior is when an agent infers or deduces from information that it already "has", new information. This may happen by thinking, drawing conclusions or applying logical rules. If a person looks at the rings in a hacked tree he or she may come to a conclusion, that is, to build an information about the tree's age. In the same way, any observation of the world or of recorded descriptions may trigger a thinking process and cause the generation of new descriptions or of new information.

This type of behavior may become a higher, more complicated class of behavior when the thinking generates questions; those questions are answered through some sort of investigation; that investigation may involve actions by other agents and perusal of records, structuring a complex dance or web of information looking actions. A good example would be a police investigation, searching for the author of a crime.

2.3.4. Non informational knowledge

A last type of knowledge would be that capacity of effective behavior that is not linked neither to information nor to recorded descriptions. This is more easily illustrated by artistic or sport effective behavior. The difference between this latter type of behavior

and the ones formerly discussed is the same as the difference between the expert tennis journalist and the ATP number one tennis player. Neither could exhibit effective behavior in the role of the other, but they are very skilled and competent in their own action domains. However, the artist or the player may be able to produce partial information that may help others to become better artists or better players. They may "articulate" their knowledge.

In summary, I prefer to define knowledge as the capacity of acting effectively, according to some standard of effectiveness. This way of defining knowledge makes it a more objective and observable phenomena, and in some cases a measurable one. It also emphasizes the close relation between information, descriptions and knowledge, while keeping those concepts clearly apart.

The definition is also consistent with the organizational knowledge concept, if we look at an organization as an agent. Organizations do exhibit behavior and there are a number of standards that help to measure or evaluate its effectiveness, including those defined by the organization itself.

3. Some observations about other approaches to information from this work's viewpoint.

This last section presents observations about other approaches to define information, communication and knowledge from the standpoint of the herein proposed definitions.

3.1. Shannon's communication theory

It is amazing how this very important development has been and continues to be mistaken for a general information theory. Currently known as the "The mathematical theory of information", everybody knows it is only a communications theory, dealing with the design of efficient codes. (Shannon, 1948).

In fact, Shannon's seminal work tells us how to optimize the size of records, so to minimize the cost of communications, as channel capacity was a scarce resource at his time. He shows how the design of a code may yield an average message length less

than the maximum one ($\log_2 n$) when the messages to be send have non uniform frequencies. The more diverse the frequencies are, the less uniform they are, the shorter the average message may be achieved. He also showed what the best design case was and showed mathematically that when the messages have equal estimated frequencies, the average message length will be maximized. Thus frequency uniformity will entail maximum cost.

As Shannon himself clarified, his theory deals only with the codification of records and their transmission over channels. It says nothing about the SOB, about language, descriptions, meanings or anything else. His result, stating the average number of bits needed to code for a given set of potential messages and their relative frequencies has been mistaken to be a definition of the quantity of information. This is one of the best known cases where an undefined concept has been object of measurement. To me, it sounds quite lacking of seriousness, a bad generalization from a perfectly simple fact. Also, the application of the word "entropy" to the phenomena researched by Shannon has only helped to confuse things and has led people to develop esoteric theories about the universe made of information or about the information being the essential material out of which everything else is made.

Shannon's theory could have been named "A theory of records, their coding and their transportation".

3.2. Information quantity

As mentioned above, Shannon's theory has been misused to account for something that has been named "information quantity". This term has led to at least two types of confusions.

3.2.1. Information quantity "in" a message

First, we may look at the concept of the quantity of information a message carries. It has been said that this quantity is a function of the probability distribution of the set of possible character strings to be carried by the message. This function is, unfortunately, called the entropy of the messages source,

usually designated with the letter H. This is not what Shannon states. Instead, his theory tells us only how to codify the messages so that their average length will be minimal, if and only if the expected frequency distribution actually occurs after the codification is put to use. If the actual frequency distribution turns out to be quite different from the expected one, the chosen codification may result in more record space and transmission time and cost. Notice that we have identified here two different probability distributions: the one used to design the codification system for the source (HS), and the actual distribution observed after the design (HA).

But also the amount of information carried by a message has been said to be a function of the message's receiving agent state and expectations. In this approach the relevant frequency or probability distribution is the one subjectively held by the receptor end. Let us designate this third distribution by HR. There is no assurance that HR is similar to HA or to HS. This simply shows that the "quantity of information carried by a message is", at least, an ambiguous notion.

Instead, I suggest that we should go back a few steps and look at what is it that the message is carrying. Remember that a message is indeed a record on the move, traveling from an origin to a destination. We are focusing here on messages that represent information, true partial descriptions of the world. We could reasonably say that the message carries the information it represents. The first quantity to be defined here should be the number of true descriptions carried by the message. This is not as naïve as it first may sound. After all, this is the way that most people think when pondering the size of a certain database. Most IT professionals never estimate the probability distribution of the set of possible records in a database, and librarians do not engage usually in entropy analysis.

A second way to apply some numerical quantity function to a message could be exemplified as follows. Let's assume that the message represents only one information about the state a system S was at time T. Let's also assume that the system S may be in n different states and that all these states

are equiprobable. In this simple case, the absolute quantity of information about the system S carried by the message could be arbitrarily defined as $\log_2 n$, the average number of questions necessary to identify the given state. If the states are not equiprobable, the average information carried by a set of messages describing S could be defined as the corresponding H function.

This latter, Shannon compatible way of defining a message's quantity of information, may seem a more elegant or more scientific way to go about it, but we must agree that it is seldom applied in the current information pragmatics and is not seen as useful, perhaps for the enormous availability of cheap and efficient space for record storage and band width for message sending.

It is interesting to note that a message will usually entail many informations besides the core information represented by its codified content. For example:

- The information that the events of preparing, sending and receiving the message have taken place.
- Information about the validity or authority of the message that could be conveyed by the type and format of the material substrate utilized and the nature of any certifying marks made on it, such as signatures, rubber stamp impressions, and preprinted logos.
- Information about the sender.

This collateral or secondary information carried by a message somewhat contradicts the observation that receiving the same message for a second time does not convey more information than the first time it was received.

3.2.2. Entropy

A second type of confusion is associated to the use of the word "entropy" in connection with information. A glass of water with an ice cube in it has less entropy than the same glass of water after the ice cube has melted down. This is true by the physical definition of entropy. How is this associated with information? The glass of water, before the

ice melting is more organized or ordered, has less randomness. This means that in this kind of state the probability distribution of the water molecule positions is less uniform, has more variations. The probability distribution for the water molecules positions in the ice cube is very different than the same quantity for the molecules not in the ice cube, whereas, after the ice is melted, every molecule could be anywhere in the glass, with approximately the same probability. This is similar to the order in an adolescent's room. In a room lacking order, a pair of socks could be anywhere (high entropy) whereas in a tightly ordered room (low entropy) it could only be found in a specific drawer. According to Shannon, the description of a lesser entropy, more organized state requires fewer bits in average to be expressed, while the description of the melted ice glass will require close to the maximum number of bits. And that is all that Shannon says. He does not allow us to say that there is more or less information "in" the glass. This latter connection maybe psychologically attractive to sustain and explore, but it does not help to clarify things, on the contrary it helps to confuse them.

It is interesting to note that it is accepted, on the grounds explained before, that an observer of a system with less entropy will have in general less uncertainty about the system's state. What is seldom noticed is that the reduced uncertainty does not flow from the fact that there is less entropy in the system, but from the fact that the observer has been told about that reduced level of entropy. If the observer does not know that the glass of water contains an ice cube, the reduced entropy does not help her or him. In other words, to benefit from the more organized way the system is in, I have to pay the cost of learning about that organization first.

3.3.   Paradoxes and abstract objects

In Russel's theory of descriptions (Russel, 1905) he analyzed a by now well known description, namely:

"The present King of France is bald."

By the definitions proposed here this is not a true description, since it is not consistent with the world state of affairs, as France has at present no king. However, the following one is a true description:

"Don Quixote is thin."

The reason to assert that it is true is that while Don Quixote has never existed physically he does exist in our SOB as an abstract object.

Unfortunately, our informations are subject to the same recursive paradoxes as discovered by Russell himself and others. For example:

"This description is false"

3.4.   The standard definition of information

The Standard Definition of Information (SDI) is defined in (Floridi, 2005) as:

1. (Information) consists of n data (d), for n = 1, 2, 3, …;
2. The data are well-formed (*wfd*);
3. The *wfd* are meaningful (*mwfd* = *d*);
4. The *d* are truthful.

The self attributed characteristic of "Standard" in the name of this definition is apparently referring to "commonly used" rather than defined by a standards setting board.

My strongest concerns with this definition are that the fundamental concept of data is mostly left obscure or undefined, and that information as an abstract (semantic) concept is made dependent on syntactic characteristics (well formedness).

Data seems to be associated with difference, and difference would be a difference in the state of a physical system. If this physical system is the world, the difference seems to refer to a state of affairs in the world or to an actual event. In this case, data in SDI could be assimilated to the concept of information presented in this work, although it is clearly not the same idea. Information, as defined here, is not the actual world phenomenon neither the record that could represent its description. If the referred to physical system is, on the contrary, not the world but a record, as when ink marks

produce a difference on a blank paper sheet, then again, it seems to confuse the record with the information, a confusion that is the main goal of this paper to dissolve. This is the same concern that the well formedness condition raises. Syntaxis is a record property, not an information property.

4. Further developments

Assuming that the goal of clarifying some confusion in the information, communication and knowledge fields has been achieved and hopefully having demystified the concepts of quantity of information and information entropy, there is now room to go for newer and may be more important challenges. Any philosophy of the information, communication and knowledge fields should concern itself with issues as follows:

- What to teach children about information to allow them to be more productive actors in the information society and by doing so raising their life quality expectations? We teach children detailed facts about chemistry and physics, but they do not know how to organize and retrieve information records.
- How is education going to change or how should it change given the information superabundance. Most education, until now has been limited to provide information to students, but they now can get far more information from the Internet. What is it that we should be teaching them in this context and how?
- How is it that people get and store information (representations) by observing the world and reading records? How are light waves and other sensory stimulus

converted into high level conceptual conclusions?
- In the anonymous Internet world: What is a person identity? How are we going to deal with personal identity, with information authenticity and with individual privacy?
- Knowledge management
- In the globalized world: How are we going to deal with intermixing cultures that do not share ontological backgrounds? How does this influence violence and tolerance? How are we going to move from discussion into dialogue?

How is scientific knowledge to be dealt with given the exponential growth of knowledge producing? How are we going to deal with the challenges posed by an ever deepening specialization and its entailing isolation into smaller and smaller pockets of knowledge? How is scientific truth to be preserved?

5. Conclusions

It has been said that a unified theory of information is impossible to achieve, given the multiple domains where information is used. I agree that a "theory" may be hard to build, but I clearly think than common definitions and distinctions such as those presented here are achievable and beneficial, to ground on them further developments. What may and does differ between disciplines and knowledge domains is their shared ontological backgrounds, but, once those are provided as contexts, descriptions, records, communication and knowledge will not need discipline specific definitions.

References

Berger, P. L. and Luckmann, T. (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*, Garden City, NY: Anchor Books.

Chen, P. (1976). The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems (TODS),1(1)*, 1975, Framingham, MA , 9 – 36.

Dawkins, R. (1989). Memes: the new replicators. in *The Selfish Gene*. Oxford University Press.

Floridi, L. (2005). Is Information Meaningful Data?. *Philosophy and Phenomenological Research. 70(2)*, 351-370.

Maturana, H. and Varela, F. (1992). The tree of knowledge. Shambhala.

Russell, B.(1905). On Denoting*, Mind, 14*, 479-493.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*, 379–423 & 623–656.

About the Author

Roberto Gejman
Roberto Gejman holds a B. Sc and M. SC. in Computer Science from the Technion, Israel; he is Adjunct Assistant Professor in Information Systems Management at the Catholic University of Chile; he is a partner of the Chile Member Firm of Deloitte, engaging in information systems consultancy; participates in the "What is information?" BITrum project.