FIS2005 - http://www.mdpi.org/fis2005/

Mathematical Foundation of Information Theory
A Set Theoretical Approach

Federico Flückiger
University of Lugano and
University of Applied Sciences of Southern Switzerland
Innovative Technologies Department, Co-Director eLab
Via Cantonale, Galleria 2, CH-6928 Manno / Switzerland
Tel: +41 58 66 66 588, Fax: +41 58 66 66 571
e-mail: federico.flueckiger@supsi.ch
http://www.elearninglab.org/, http://www.unisi.ch/ & http://www.dti.supsi.ch/
http://mypage.bluewin.ch/federico.flueckiger/

Since years we are seeking a unified approach putting together the different views of information into one concept without great success. There is one main reason for that: the absence of a clear and simple phenomenological foundation of the basic concepts together with a corresponding mathematical structure describing it. Both have been elaborated by the author of that abstract within a doctoral thesis done at university of Berne, Switzerland in 1995 [2]. The mathematical part of that work (chapter 4.4. in [2]) was never published or presented within a conference. That approach shows a mathematical structure based on set theory that defines information as an interdisciplinary concept showing the different facets of it such as
the structural-attributive information,
the functional-cybernetic information,
the information carrier,
the basic unit of information, and
the measure of information.

Using some easily equation transformation the second principle of thermodynamics (called first principle of information theory) is provable by the formula for the measure of information.

The Thing

The author has proposed a definition of the term information, based on knowledge from various scientific disciplines such as informatics, anthropology, psychology, situation theory, philosophy, etc., as part of a dissertation [2]. The definition is based on the finding that the world consists of things which individuals can recognize, describe, imagine, invent, etc. The term *thing* is introduced in greater detail in an another article by the author also published in this volume.

A thing is identified as such when it is recognized or postulated by an individual. Hence, its existence is not fundamentally independent of the individual. There is no reason why a thing cannot be merely an individual's invention, it need not be based in reality. Since by definition the things examined here are intimately related to the individual and, thus, are pragmatic from the outset, we differentiate here merely between syntactic and semantic relationships, in contrast to the term sign as understood in the field of semiotics. It can consist of other things or itself be a part of another thing. Hence, there are relations between things which equally must be seen as things because, from an ontological point of view, they have an equivalent structure. Those relationships pointing towards the thing are called syntactic, and those pointing away from the thing, semantic. Because the question of whether a relationship is syntactic or semantic is answered by the direction of the relationship, we have designated the relationships in the following discussion respectively as d_syntactic and d_semantic.

Basic Definitions

Firstly, I would like to say that such complex terms as information cannot be satisfactorily described by mathematical equations. But laying a mathematical foundation of initially phenomenologically determined terms helps in formulating the necessary engineering science aspects as a step towards technical realization. For this purpose, the following definitions, derived in part from the wording found in the textbooks by Halmos [3] and Wechler [7], are designed to form the mathematical foundations for the model of an interdisciplinary information theory. The theory is built on six auxiliary definitions, in which the set A of what are called apriori-things is developed. These include concrete as well as abstract objects and constructs which can be recognized or postulated by individuals.

We will first of all define a broad area of so-called a priori basic units from which the a priori things mentioned above are to be selected. The framework of these a priori basic units is begun with the definition of a set U_0 of a priori atoms that are axiomatic. According to the insights of chapter 4.3 in [2] this set has to be supplemented at once in two respects. First of all U_0 lacks reciprocal relations between the elements of U_0. Furthermore, finite subsets of U_0 have to be included as a priori basic units because only in this way can objects composed of other a priori basic units be considered as a priori basic units themselves. These supplements are taken into account in a set U_1 which is formed by the union of U_0 with the set of pairs of elements from U_0 and the finite power set of U_0. We have not reached our goal yet, though, because set U_1 does not contain all the required elements of a set of a priori basic units either. It still lacks the relations between the elements of U_1 as well as the finite subsets of U_1. Thus we have to construct, in an analogous way, a set U_2, which is again plagued by the same problems as the sets U_0 and U_1. In that way we construct a tower of sets U_n ($n \in \underline{N}$), where the set U_{n+1} always contains those elements that are still missing according to the previously outlined pattern in U_n.

The set U of all 'a priori basic units' finally results from the union of all sets U_n constructed in this way. The formal structure of this set U is therefore as follows:

Def 1: $U_0 =$ set of all '**a priori atoms**'

$$U_1 = U_0 \cup U_0 \times U_0 \cup P_{fin}(U_0)$$
$$U_2 = U_1 \cup U_1 \times U_1 \cup P_{fin}(U_1)$$
:
:
$$U_{n+1} = U_n \cup U_n \times U_n \cup P_{fin}(U_n)$$

Thus we construct:

$$U = \bigcup_{n=0}^{\infty} U_n$$ the set of all '**a priori basic units**', and

$$U_R = U - U_0 - \bigcup_{n=0}^{\infty} P_{fin}(U_n)$$

the set of all '**directed relations between elements of U**' such that from $u \in U_R$ it follows that: $u = (b,c) = \{b,\{b,c\}\}$ with $b, c \in U$.

Thus the set U already includes everything that can be an a priori thing. For example, it is possible with Definition 1 to define the material parts of a piano as a priori atoms and to reciprocally relate them in such a way that the piano can be described as a structure. This structure is then itself an element of set U and thus an a priori basic unit even if it has not been defined as an a priori atom.

However, set U contains many more elements than are necessary for the description of a priori things. We require not just any subset of U_n ($n \in \underline{N}$), but only those subsets whose elements represent a coherent structured unit. It is therefore the aim of the following definitions to describe structures from set U, which are themselves composed of elements of this set, mathematically. Thus the next definition goes as follows:

Def 2: **Direct a priori relatedness**: Let the binary relation $E_\delta \subseteq U \times U$ be defined as:

$(a,b) \in E_\delta \Leftrightarrow (a,b) \in U_R$ or $(b,a) \in U_R$ for $a, b \in U$

We say that b is '**directly related a priori**' to a and vice versa if $(a,b) \in E_\delta$.

The analysis of a network of a priori basic units concerns not only the direct, but also the indirect a priori connections between elements of U. In order to capture this aspect in mathematical terms, we have to take one step further and describe the indirect relatedness of two a priori basic units which are linked via other a priori basic units:

Def 3: **Indirect a priori relatedness**: Let E_ι be the transitive closure of E_δ defined as:

$(a,b) \in E_\iota \Leftrightarrow \exists$ $P = \{p_1, ..., p_n$ ($n \in \underline{N}$), $p_i \in U$,

so that $(a,p_1), (p_1,p_2), ..., (p_{n-1},p_n), (p_n,b) \in E_\delta$, for $a, b \in U$

If: $(a,b) \in E_\iota$, then b is '**indirectly related a priori**' to a, via the elements of P in the order indicated. In that case P is the path from a to b and vice versa.

So far only pairs of a priori basic units have been examined as to their reciprocal direct or indirect a priori connections. However, this does not yet exhaustively describe a structure forming a network of a priori basic units connected in pairs. Thus we have to propose the following definition:

Def 4: The set $CO \subset U$ is called '**a priori coherent**' if for each pair $co_1, co_2 \in CO$ it is the case that: $(co_1,co_2) \in E_{\iota|CO} = E_\iota \cap CO \times CO$. Moreover, for the intervening path $P = \{p_1, ..., p_n\}$ ($n \in \underline{N}$) the rules are:

- $P \subset CO$
- $(co_1,p_1), (p_1,p_2), ..., (p_{n-1},p_n), (p_n,co_2) \in E_{\delta|CO} = E_\delta \cap CO \times CO$
- $(co_1,p_1), (p_1,p_2), ..., (p_{n-1},p_n), (p_n,co_2) \in CO$.

That is, each pair $co_1, co_2 \in CO$ must be indirectly related a priori, via the path, including the connections of relatedness, which lie in CO. We call any set CO which is a priori coherent an '**a priori structure**'.

With this definition, structured a priori objects can be formally described as entities: A structured a priori object CO can mathematically be viewed as a finite a priori coherent subset of U. Base on this insight, we can now construct the set of a priori things A. In analogy to the set U of all a priori basic units, A is based on the set $A_0 = U_0$ of all a priori atoms. Following the pattern of Definition 1 we construct, on the basis of A_0, a tower of sets A_{n+1} ($n \in \underline{N}$) of which each is composed of the union of set A_n with the set of pairs of elements of A_n and the set of all finite a priori coherent subsets of A_n. The set A of all a priori things then results from the union of all sets A_n. The formal definition of this set must be as follows:

Def 5: Let $P_{co}(X) = \{CO \in P_{fin}(X) \mid CO \text{ is a priori coherent}\}$, so that for the set A of all a priori things:

$A_0 = U_0 =$ set of all '**a priori atoms**'

:

:

$$A_{n+1} = A_n \cup A_n \times A_n \cup P_{co}(A_n)$$

Thus we construct:

$$A = \bigcup_{n=0}^{\infty} A_n$$ set of all '**a priori things**', and

$$A_R = A - A_0 - \bigcup_{n=0}^{\infty} P_{co}(A_n)$$

set of all '**directed a priori relations between elements of A**' such that it follows from $a \in A_R$: $a = (b,c) = \{b,\{b,c\}\}$ with $b, c \in A$.

Now the set A includes all elements that are eligible as a priori things: Apart from the a priori atoms, A also comprehends all directed a priori relations as well as all possible a priori structures. In [2] (chapter 3.1) we found that every non-atomic thing and thus every non-atomic a priori thing has an internal structure that is determined by so-called d_syntactic relations.

Moreover, every thing has a meaning space which is delimited by d_semantic relations. This state of affairs can be represented in the following definition:

Def 6: Let $c = (a,b) \in A_R$ with a, b, $\in$ A, then we call c a **'d_semantic relation a priori'** with reference to a and a **'d_syntactic relation a priori'** with reference to b.

Furthermore let $DSEA(a) = \{c \in A_R \mid \exists\, b \in A \text{ with } c = (a,b)\}$ be the **'d_semantic closure a priori'** of $a \in A$ and $DSYA(b) = \{c \in A_R \mid \exists\, a \in A \text{ with } c = (a,b)\}$ the **'d_syntactic closure a priori'** of $b \in A$.

Definition 6 represents the thesis proposed in [2] whereby a directed relation, depending on whether its beginning or its end is the centre of interest, can either be interpreted as d_semantic or as d_syntactic[1]. Thus we now have all we need to define the thing as the basis of a mathematically constituted definition of information.

Information as a Structured Object

The thing as a structured object, relevant to the information theory, is then defined in definition 7 in the following manner as 4-tuple:

Def 7: Let $I \subset A$ be a set which we can interpret as the set of all organic individuals capable of grasping, i.e. of constructing as such, the a priori things postulated in Definition 5. Further, let $\Re^+_0$ be the set of positive real numbers including 0 and $\underline{N}_0$ the set of all natural numbers including 0. The **'set TH of all things'** is then defined as:

$$TH = \{th \mid th \in A \times I \times \Re^+_0 \times \underline{N}_0\}$$

A thing $th \in TH$ is thus a 4-tuple (a, i, t, s), with:

a = the a priori thing corresponding to th,
i = the individual who constructs th (i.e. recognizes or postulates),
t = the time when th is constructed, with the origin $t_0 = 0$ set arbitrarily to the origin of the universe, the so-called Big Bang, and
s = the selection counter with an initial value of 0, which indicates how often the thing th has been used to generate a message (cf. further down).

Moreover for the set $TH_R \subset TH$ of all relations:

$$TH_R = \{th \mid th \in A_R \times I \times \Re^+_0 \times \underline{N}_0\} \text{ holds.}$$

With definition 7 we can derive the relevant set of all things that can be constructed (i.e. recognized or postulated) by an individual:

[1] Note that in mathematical logic the use of the terms syntax and semantics is different from the use in the present study.

Def 8: Let $i_0 \in I$, then the **'set $TH(i_0)$ of all things that can be constructed by an individual i_0'** is defined as:

$$TH(i_0) = TH \cap A \times \{i_0\} \times \Re^+_0 \times \underline{N}_0$$

or in other terms:

$$TH(i_0) = \{th \in TH \mid th = (a, i_0, t, s) \text{ for suitable } a \in A, t \in \Re^+_0, s \in \underline{N}_0\}$$

Furthermore the set $TH_R(i_0) \subset TH(i_0)$ of all relations that can be constructed by i_0 is defined as:

$$TH_R(i_0) = TH \cap A_R \times \{i_0\} \times \Re^+_0 \times \underline{N}_0$$

Because the set TH(i) contains all things that can be constructed by the individual i, TH(i) in a way represents the knowledge of i. The set TH(i) can basically contain one or more things $th \in TH(i)$ for each a priori thing $a \in A$. Thus the existence of structures in TH(i) on the pattern of the a priori structures in Definition 4 becomes conceivable. Yet since TH(i) is more complex than A, Definitions 2-4 and Definition 6 have to be adapted to the new situation. First of all the direct relatedness between two things has to be redefined:

Def 2*: **Direct relatedness**: Let $b, c \in TH(i)$ with $b = (a_b, i, t_b, s_b)$ and $c = (a_c, i, t_c, s_c)$. Let the binary relation $E_\Delta(i) \subseteq TH(i) \times TH(i)$ with $i \in I$ be defined as:

$(b,c) \in E_\Delta(i) \Leftrightarrow ((a_b,a_c), i, t, s) \in TH_B(i)$ or
$((a_c,a_b), i, t, s) \in TH_B(i)$

We say that c is '**directly related'** to b and vice versa if $(b,c) \in E_\Delta(i)$.

In a similar way, we adapt the definition of indirect relatedness:

Def 3*: **Indirect relatedness**: Let $E_I(i)$ be the transitive closure of $E_\Delta(i)$ defined as:

$(b,c) \in E_I(i) \Leftrightarrow \exists$ $P = \{p_1, ..., p_n\}$ $(n \in \underline{N})$, $p_j \in TH(i)$ with $i \in I$, so that $(b,p_1), (p_1,p_2), ..., (p_{n-1},p_n), (p_n,c) \in E_\Delta(i)$ for $b, c \in TH(i)$

If: $(b,c) \in E_I(i)$, then c is '**indirectly related**' to b via the elements of P in the order indicated and vice versa. In that case P is the path from b to c or from c to b respectively.

Thus the definition of an information structure follows almost naturally :

Def 4*: The set $CO \subset U$ is called **'coherent'**, if for every pair $co_1, co_2 \in CO$ the following holds: $(co_1,co_2) \in E_I(i)|_{CO} = E_I(i) \cap CO \times CO$. Moreover for the intervening path $P = \{p_1, ..., p_n\}$ $(n \in \underline{N})$ it is true that:

- $P \subset CO$
- $(co_1,p_1), (p_1,p_2), ..., (p_{n-1},p_n), (p_n,co_2) \in E_\Delta(i)|_{CO} = E_\Delta(i) \cap CO \times CO$
- $(co_1,p_1), (p_1,p_2), ..., (p_{n-1},p_n), (p_n,co_2) \in CO$.

Thus every pair co_1, $co_2 \in CO$ has to be indirectly related and the path including the relations of relatedness has to lie in CO. We call every coherent set CO an '**information structure**'.

The knowledge of i represented by TH(i) is thus composed of structured things, the so-called information structures. We are now left with the adaptation of Definition 6 which describes the internal structure and the meaning space of a thing:

Def 6*: Let $th = (a_{th}, i, t_{th}, s_{th}) \in TH(i)$ with $i \in \mathbf{I}$, then

- $DSE(th) = \{e = (a_e, i, t_e, s_e) \in TH_B(i) \mid a_e \in DSEA\,(a_{th})\}$ is the '**d_semantic closure**' of th.
- $DSY(th) = \{c = (a_c, i, t_c, s_c) \in TH_B(i) \mid a_c \in DSYA\,(a_{th})\}$ is the '**d_syntactic closure**' of th.

It is self-evident that every $e \in DSE(th)$ is always a d_semantic relation pointing away from th towards another thing and every $c \in DSY(th)$ is a d_syntactic relation pointing from another thing towards th.

Definition 6* makes it possible to mathematically describe the thing in the way outlined in [2]: All things $th \in TH(i)$ with $i \in \mathbf{I}$ thus have a d_syntactic and a d_semantic closure. Moreover the reflections on pragmatics in [2] (chapter 3.1) also have their mathematical correspondence in the series of definitions proposed here. The fact that according to Definition 7 every thing $th \in TH(i)$ contains, apart from the reference to the a priori thing a_{th}, references to a constructing individual i and a construction time t_{th}, might - together with the selection counter s_{th} which in a way expresses a 'tendency of the individual i to make th known' - be taken to point to the pragmatic aspect of a thing.

With Definitions 1-8 we have created a general framework to describe the structural-attributive view of information. The new findings now necessitate the following two notes:

Note 1: Every thing $x \in TH$ which is an information structure will be considered as an **'information carrier'** if and only if $DSE(x) \neq \emptyset$.

Note 2: If a thing $x \in TH$ is an information carrier according to Note 1, then every relation $c \in DSE(x)$ is an **'information element'**.

Thus a thing th is an information carrier if and only if a d_semantic relation is pointing away from th. From this the fact that the d_semantic relation is an information element follows trivially. With other words: structural-attributive information can be called a non-trivial information structure whose d_semantic closure is not empty. In this way we can formalize both the structure of the external world and the structure of individual knowledge. The cardinality of such an information structure CO, i.e. the number of elements in CO, is expressed according to [3] as $CARD(CO) = |CO|$.

Information as process

Functional-cybernetic information theories describe information as a process IP which over time modifies information structures in general and knowledge in particular. A structure Y, for example, which has been transmitted at a time t_1 has a greater cardinality at the time $t_2 > t_1$ than at the time $t_0 < t_1$. This process IP is to be elucidated on the basis of the model of a universal communication system of C.E. Shannon [6] ; it is to be understood as a process during which a message is produced by an information source and transmitted to a destination which then integrates the structure of this message in its own structure. All the components of this model - message, information source, channel and destination - are considered as reciprocally independent information structures according to Definition 4* in [2] constructed by an individual $i \in I$ in its role as an external observer of a situation.

Of course, the successful completion of such an information process requires that the destination have a certain internal structure, since the destination must always have certain structural similarities to the message. It would, for example, be extremely difficult to inform an English-speaking individual by means of a message in Chinese. In other words, the question arises how the qualities of any two information structures can be compared so that differences and similarities become visible:

Def 9: Let $X, Y \in TH(i)$ with $i \in I$ be two information structures, and let further $x = (a_x,i,t_x,s_x) \in X$ and $y = (a_y,i,t_y,s_y) \in Y$, then:

- The two things x and y are called '**equivalent**', if $a_x = a_y$, that is if x and y refer to the same a priori thing. We note this fact with $x \sim y$.
- The set $SC_{X,Y} \subset X$ is called the '**structural community**' of X relative to Y, if it is true that: $SC_{X,Y} = \{x \in X \mid \exists\, y \in Y \text{ with } x \sim y\}$.
- Finally the set $SD_{X,Y} = X - SC_{X,Y}$ is called the '**structural difference**' of X relative to Y.

Of course, the statements of Definition 9 apply to all information structures in TH. Now, if the two information structures X and Y represent intelligent beings, this definition can be interpreted as follows: The set $SC_{X,Y}$ stands for the a priori knowledge of X with respect to Y and $SD_{X,Y}$ for the difference in knowledge of X relative to Y. On the other hand, $SC_{Y,X}$ designates the a priori knowledge of Y relative to X and $SD_{Y,X}$ the difference in knowledge of Y relative to X.

In the course of the information process, the message M plays a central part. It transmits a structural part S of an information source X to a destination Y in coded form. Such a message need not correspond directly to S, but there must be an unambiguous translation rule from S into M and vice versa which must be known both to the information source and the destination. Further, the message M must not differ structurally from S, since M cannot in any case contain more than the information source X knows about the fact S to which M refers. Moreover each instances of the generation of a message M increments the selection counter s_S of information structure S by the value of 1. Finally, M as a product of information structure X is always linked to S by a directed relation $r \in TH_R(i)$. Thus we get the following definition of a message:

Def 10: Let $M, S \in TH(i)$ with $i \in \mathbf{I}$, where both M and S are information structures. We call M a '**message**' of S, if it is true that:

- There is a relation $r \in TH_R(i)$ with $r \in DSE(S) \wedge r \in DSY(M)$.
- $SD_{M,S} = \varnothing$.
- The selection counter s_S of information structure S is incremented by the value 1.

Further let $MA(S) = \{M \in TH(i) \mid M \text{ is a message of } S\}$ be the set of all messages of S and further let $MB(S) = \{r \in TH_R(i) \mid r \in DSE(S) \wedge r \in DSY(M)\ \forall\ M \in MA(S)\} \subseteq DSE(S)$ be the set of all relations of S to its messages.

Definition 10 only stipulates formal conditions for a message, not conditions that concern its content. Since moreover Definition 10 only posits that the structural difference of M relative to S must be empty and not vice versa, any element of S can basically be generated as a message. This agrees very well with the observation that a living being can at all times transmit anything it knows in any combination as a message to its environment.

An information structure $M \in TH(i)$ is a possible message of another information structure $S \in TH(i)$ if there is a relational element $r \in DSE(S)$ such that at the same time $r \in DSY(M)$. This leads one to assume that all information structures $T \in TH(i)$ that are referred to by an element r' from DSE(S) can also be considered as possible messages of S. This would mean that for each of these T it is true that: $SD_{T,S} = \varnothing$, which is generally not likely. Nevertheless there will be an information structure $T' \in T$ for every such T that satisfies that criterion. If $T' \supset \varnothing$, then T' stands for a possible message. But if $T' = \varnothing$, for example because $SC_{T,S} = \varnothing$, then no non-empty message can be generated on the basis of $r' \in DSE(S) \cap DSY(T')$. This need not disturb us. It simply marks a limit case. We can now formulate the following theorem:

Theorem 1: Let $S \in TH(i)$ with $i \in \mathbf{I}$ be an information structure and let further be $r \in DSE(S)$ and $r \in DSY(T)$ for a suitable $T \in TH(i)$, then there is an information structure $T' \in T$ with $SD_{T',S} = \varnothing$.

Proof: It follows from Definition 9 that

$\exists\ T' \in SC_{T,S}$ with $SD_{T',S} = \varnothing$ QED

An information process whereby a destination $Y \in TH(i)$ is informed by an information source $X \in TH(i)$ can now be viewed as a process that reduces the structural difference $SD_{X,Y}$ of the destination relative to the information source and at the same time integrates the new elements in the knowledge structure of Y in such a way that Y is again an information structure. Therefore the next definition is:

Def 11: Let $X(t_1)$, $Y(t_1) \in TH(i)$ with $i \in I$ be two information structures at a time t_1 with $SD_{X,Y}(t_1) \neq \emptyset$, then we can say that:

- An element $x \in X(t_1)$ at the time $t_2 > t_1$ is called '**integrated**' in $Y(t_2)$ if: $Y(t_2)$ is an information structure and $x \in SD_{X,Y}(t_1)$ and $x \notin SD_{X,Y}(t_2)$. We note this as $Y(t_2) = Y(t_1) \hookleftarrow \{x\}$.
- The set $X(t_1)$ at the time $t_2 > t_1$ is called '**partially integrated**' in $Y(t_2)$ if: $Y(t_2)$ is an information structure and $SD_{X,Y}(t_1) \supset SD_{X,Y}(t_2) \neq \emptyset$.
  We note this as $Y(t_2) = Y(t_1) \hookleftarrow (X(t_1) - SD_{X,Y}(t_2))$.
- The set $X(t_1)$ at the time $t_2 > t_1$ is called '**completely integrated**' in $Y(t_2)$ if: $Y(t_2)$ is an information structure and $SD_{X,Y}(t_2) = \emptyset$. We note this as $Y(t_2) = Y(t_1) \hookleftarrow X(t_1)$.

The complete integration of a set $X(t_1)$ in the set $Y(t_2)$ with $t_2 > t_1$ can thus be interpreted as meaning that all elements of $X(t_1)$, that have no equivalent in $Y(t_1)$ yet, will acquire one by a complete integration. Thus the information process can be represented as follows:

Def 12: Let $X(t_1)$, $Y(t_1) \in TH(i)$ with $i \in I$ be two information structures at the time t_1 and let $M \in TH(i)$ be a message of $X(t_1)$. The '**information process**' $IP(M,Y(t_1))$ which transmits the message M from $X(t_1)$ to $Y(t_1)$ at the time t_1 has the following effect at the time $t_2 > t_1$:

- $Y(t_2) = Y(t_1) \hookleftarrow M$.
- $\exists\, r \in TH_R(i)$ with $r \in DSE(Y(t_2)) \wedge r \in DSY(M)$ as well as $r \notin DSE(Y(t_1))$.

We designate by $SG(IP(M,Y(t_1))) = Y(t_2) - Y(t_1)$ the '**real structural growth**' of $Y(t_2)$ due to the information process $IP(M,Y(t_1))$.

Definition 12 captures two important features of information: First of all it shows that a message which is integrated by the information process of a destination can be reproduced as a message by this destination. Secondly, Definition 12 makes clear that an information process is only non-trivial, i.e. can only modify a structure, if: M, $SG(IP(M,Y(t_1))) \neq \emptyset$. We can now derive the following theorem:

Theorem 2: Let $X(t_1)$, $Y(t_1) \in TH(i)$ with $i \in I$ be two information structures at the time t_1 and let $M \in TH(i)$ be a message produced by $X(t_1)$. After every information process $IP(M,Y(t_1))$ which transmits the message M from $X(t_1)$ to $Y(t_1)$ at the time t_1:

$$CARD(DSE(Y(t_2))) \;\geq\; CARD(DSE(Y(t_1))) \qquad \text{with } t_2 > t_1$$

Thus the information process $IP(M,Y(t_1))$ entails that the cardinality of the d_semantic closure of information structure $Y(t)$ and thus the capacity to form messages increases or remains the same.

Proof: According to Definition 12 it follows that:

$$DSE(Y(t_2)) \supseteq DSE(Y(t_1)) \cup \{r\} \qquad (\text{with } r \in DSE(Y(t_2)) \wedge r \in DSY(M))$$

From this the proposition of Theorem 2 follows trivially!

Thus Theorem 2 means that with each information process the cardinality of the destination structure increases or remains the same. For this reason the knowledge of the destination system always increases or at least remains constant in this kind of information process, but never decreases.

A measure for information

In conclusion to the formal definition of the concept of information the question arises what is the formal measure of information. To be more precise, we would like to know what is the measure for the amount of information of an information carrier. Since according to Note 2 an information structure S is called an information carrier if and only if DSE(S) $\neq \emptyset$, i.e. if its d_semantic closure is not empty, and because according to Theorem 1 it is precisely the elements of DSE(S) that refer to the possible messages of S, this measure must logically make a quantitative statement about the d_semantic closure of S. Such a measure must increase in line with the cardinality of DSE(S), and it must be capable of taking into account the elements of DSE(S) weighted according to the probability that an element will be part of a message. These requirements are fulfilled by Shannon's formula for the amount of information H (cf. [6] as well as below in Def 14).

One problem that still awaits its solution is the definition of the above-mentioned selection probability for the elements of the d_semantic closure of an information structure S. The basis for a solution is already there in the structure of these elements, for the selection counter s_r for every relation r $\in$ DSE(S) contains an empirically produced value that expresses the selection frequency of r as part of a message. The selection probability q(r) relative to the other elements of DSE(S) can now be calculated by dividing s_r by the sum of all s_c with c $\in$ DSE(S) , as can be seen from the following definition:

Def 13: Let S $\in$ TH(i) with i $\in$ I be an information structure and let further r = (a_r, i, t_r, s_r) $\in$ DSE(S). Let the function q: DSE(S) $\rightarrow$ [0,1] be defined as follows:

$$q(r) = \frac{s_r}{\sum_{c \in DSE(S)} s_c}$$

We call q(r) the '**selection probability**' of the relation r $\in$ DSE(S) relative to the other relations c $\in$ DSE(S).

Thus we have a basis for the definition of the measure for the amount of information:

Def 14: Let S $\in$ TH(i) with i $\in$ I be an information structure and let further r = (a_r, i, t_r, s_r) $\in$ DSE(S), then:

$$H(S) = -K \sum_{r \in DSE(S)} q(r) \log q(r)$$

H(S) is the '**amount of information**' or the **entropy** of the thing S.

Thus it is possible to calculate the amount of information for each information structure S, regardless of the part it plays in an information process. According to Definition 14, the amount of information of S calculated in this way is primarily a measure for the cardinality of the d_semantic closure of S, which could be understood as a measure for the 'semantic content' of an information structure. Thus with his formula for the calculation of the amount of information, Shannon unknowingly created the basis for a measure for the 'semantics' of an information structure. According to that finding we may now formulate the following conclusion entitled the **'Law of Information Theory'**:

Theorem 3: Let $X(t_1)$, $Y(t_1) \in TH(i)$ with $i \in I$ be two information structures at the time t_1 and let $M \in TH(i)$ be a message of $X(t_1)$. Let further $IP(M,Y(t_1))$ be an information process which transmits the message M from $X(t_1)$ to $Y(t_2)$ at the time t_1 with $t_2 > t_1$, so that an $r' \in TH(i)$ results, with
$r' = (a_{r'}, i, t_{r'}, s_{r'}) \in DSE(Y(t_2)) \wedge r' \in DSY(M)$. To such an information process the following formula applies:

$$\Delta H \geq 0$$

Proof:

$$\Delta H = H(Y(t_2)) - H(Y(t_1))$$

$$\text{according to } Def\ 13,14: \Delta H = -K\left(\sum_{r \in DSE(Y(t2))} q_r \log q_r - \sum_{b \in DSE(Y(t1))} q_r \log q_r\right)$$

$$\text{according to } Def\ 12,14: \Delta H \geq -K\left(\sum_{r \in DSE(Y(t1))} q_r \log q_r + q_{r'} \log q_{r'} - \sum_{r \in DSE(Y(t1))} q_r \log q_r\right)$$

$$\Rightarrow \Delta H \geq K * q_{r'} \log q_{r'}$$

$$\Rightarrow \Delta H \geq 0$$

$$QED.$$

Thus the non-trivial information processes according to Definition 12 show an affinity to irreversible physical processes and accordingly the trivial information processes an affinity to reversible physical processes.

Conclusions

We have established that it is possible to bring together the most disparate approaches of existing information theory under one umbrella and to formalize them with a single mathematical model. It turned out that applying Shannon's term of entropy to derive a mass of information, in contradiction to the differing opinions of many information theorists, is highly suitable for all information theories based on the term thing. The entropy proved itself as a mass of the cardinality of the r_semantic shell around the thing. Thus, it represents, so to speak, an estimate of the semantic contents of a thing as a unit of information. The larger its contents are, the greater in turn is the probability of the individual interacting with its environment, which indicates a close correspondence with the thermodynamic term of entropy. For these reasons we postulate the application of the Second Law of Thermodynamics as the fundamental law of information theory.

Bibliography

[1] Barwise, Jon, Perry, John, 'Situations and Attitudes', Cambridge, Mass. and London, England: MIT Press, 1983.

[2] Flückiger, Federico, „Contributions Towards a Unified Concept of Information“, Doctoral thesis at Faculty of Science of university of Berne, Switzerland, 1995. (http://mypage.bluewin.ch/federico.flueckiger/Download/Ctrb2uci.pdf)

[3] Halmos, Paul R., 'Naive Mengenlehre', moderne Mathematik in elementarer Darstellung, aus dem Amerikanischen übersetzt von Manfred Armbrust und Fritz Ostermann, 4. Auflage, Vandenhoeck & Ruprecht, Göttingen, 1976.

[4] Morris, Charles W., 'Signs, Language and Behavior', Englewood Cliffs, New York, 1946.

[5] Quine, Willard van Orman, 'Word and Object', Cambridge, Mass.: MIT Press, 1960.

[6] Shannon, Claude E., Weaver, Warren, 'The Mathematical Theory of Communication', the University of Illinois Press, Urbana, Chicago, London, Fourth Printing of the Paperback Edition, 1969, Original Edition, 1948.

[7] Wechler, Wolfgang, 'Universal Algebra for Computer Scientists', ETACS Monographs on Theoretical Computer Science, J.-Springer-Verlag, Berlin, Heidelberg, 1992.

[8] Wiener, Norbert, 'Cybernetics, or Control and Communication in the Animal and the Machine', 2nd revised and enlarged edition, New York and London: MIT Press and Wiley, New York, London, 1961.