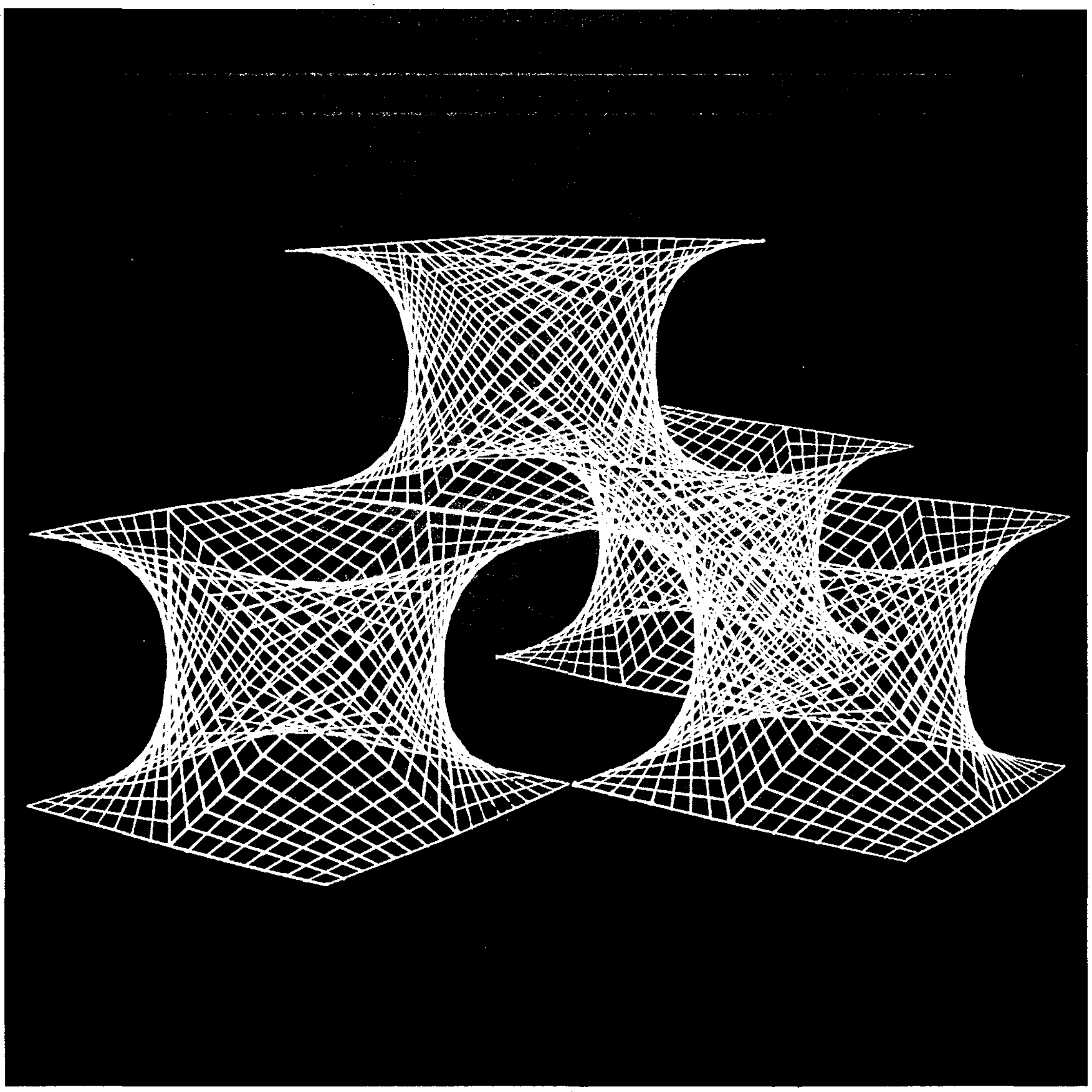
Symmetry: Culture and Science

ORDER / DISORDER
Proceedings, 4th Congress

The Quarterly of the
International Society for the
Interdisciplinary Study of Symmetry
(ISIS-Symmetry)

Volume 9, Numbers 2 - 4, 1998



Symmetry: Culture and Science
Vol. 9, Nos. 2-4, 383-392, 1998

THE LANGUAGE OF CELLS: A PARTITIONAL APPROACH TO CELL-SIGNALING

Pedro C. Marijuán – José Pastor – Morris Villarroel

Address. Department of Electronic and Communication Engineering, CPS, Universidad de Zaragoza, 50015 Zaragoza, Spain E-mail: marijuan@posta.unizar es

Abstract: *Cell signaling is the fastest growing subject in biochemistry yet no general mathematical principles have been found to decipher a "language" of cells using the discrete positional approach (Shannonian Information Theory does not appear to provide a workable framework). Instead, cellular information may be compositional; messages are exchanged as the presence or absence of symbols and few meaningful positional relationships are involved. Here we introduce a new informational grammar using uni- and multi-dimensional partitions that may help us to better understand signal processing in the eukaryotic cell.*

1. INTRODUCTION: POSITIONAL VERSUS COMPOSITIONAL INFORMATION

The notion of a "language" of cells does not seem consistent with the standard views of Information Theory applied to biology. Although Shannon (1949) distinguished between discrete, continuous, and mixed information sources, the standard application (and possibly overextension) of his ideas to cell biology have been heavily influenced by the sequential structure of DNA and RNA and, traditionally, only the discrete-positional case has been considered (e.g., Gatlin 1972, Schneider 1995). As a consequence, the lack of distinction between "positional" and "compositional" forms of information and the subsequent neglect of the latter have implied an analytical dead-end concerning the possibilities of elucidating formal mechanisms of cellular languages.

The assumed preconditions for information transmission, and particularly for any workable language, refer to sequences of messages containing combinations of symbols which are deciphered or transmitted always following a positional order (only broken for crytographic purposes – Pastor and Sarasa 1998). Shannon's formula appears to be the natural way of measuring the average combinatory content of these positional messages and of establishing their relative index of surprise in order to design appropriate channels, codes, etc. Subsequently, a workable language can be created by following a set of grammatical (Markovian) rules to connect successive positional messages comprised within the dictionary scope of the language.

However, one can point to a number of instances in natural and social communication where symbols are used in a rather different way. Instead of a "positional" context (which also generally implies the assumptions of sequence, stability and hierarchy – see Marijuán and Villarroel 1998) symbols may be used in a "compositional" way. In this alternative context, messages are exchanged as presences or absences of symbols which have been accumulated upon predetermined sets of objects. No meaningful positional relationships are assumed among the objects within the set or among the symbols accumulated on these objects. For example, several glasses on a tray may contain a variable number of different symbolic items (ice cubes, soda, vermouth, olives, cherries). We may consider the set of glasses on the tray as the message, each glass being an individual object that accumulates several symbols which make it distinguishable. Then two subjects could communicate by exchanging trays with a variable number of glasses and contents (Marijuán and Pastor 1998). That messages can be reliably distinguished and transmitted by the "concurrent processing of discrete states of media", has already been postulated by Karl Javorszky (1995). A whole body of partitional calculus (or granularity algebra) has been envisioned by this author (Javorszky 1995b, Steidl and Javorszky 1996). Interestingly, partitional reasoning has also been applied to problems in pattern recognition (Frigui and Krishnapuram 1997), logic (Mosterín 1987, Modica and Rustichini 1994) and even economics (Caianiello 1985).

Biological examples of compositional information exchanges may be found in the communicational use of colors, odors and tastes. We may also consider pheromones in social insects and, anecdotally, the etiquette "language of flowers", and perhaps even musical compositions and the formative frequencies of vowels and consonants of our own spoken languages. The "language of cells" we shall discuss here may be one of the most interesting instances of communication by means of such compositional tools; and it has been the forerunner of all further means of biological communication. Marshall McLuhan's famous dictum "the medium is the message" and the particular disdain this author showed about Shannon's information theory (McLuhan 1962) are worth recalling when considering this fundamental distinction between positional and compositional forms of information exchange.

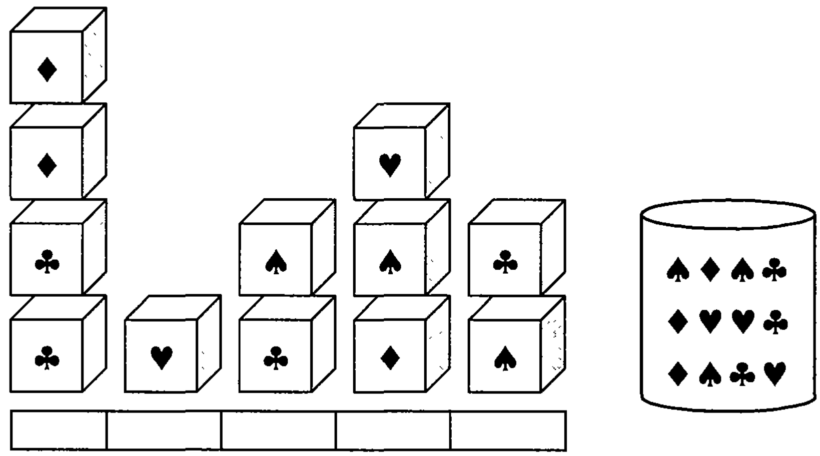
2. ANALYZING A COMPOSITIONAL MESSAGE

2.1 Unidimensional Partitions

The theory of compositional messages is formally based on the partitional-additive properties of natural numbers. Following this theory, messages are distinguished and analyzed by measuring the relative frequency of partitions in the overall structure of the message. Mathematically speaking, partitions are a very straight forward concept, i.e., the additive decompositions of natural numbers. For instance, the set { (5), (4,1), (3,2), (3,1,1), (2,2,1), (2,1,1,1), (1,1,1,1,1) } represents all the unidimensional partitions of the number five. By adhering to this mathematical treatment, one can use the well-known partitional properties of numbers to discuss the most probable logical states of a compositional message encrusted upon the elements of the set N.

Compositional messages have to be mapped onto a different kind of state space, where what counts (what generates dynamics) is the absence or presence of specific symbols on a set of N elements. When receiving a compositional message, the presence of the different symbols on each element of the set has to be counted and grouped in homogeneous classes of overlapping or non-overlapping nature. For instance, the message of Figure 1 generates the following partitions of 5: hearts (3, 2), spades (3, 2), clubs (2, 2, 1), diamonds (3, 1, 1).

Each class is defined by the presence of a specific symbol, and this symbol effectively creates a partition of the set of N elements. After the classes are defined by single symbols, the more complex coincidences of combinations of symbols (class overlaps) among the elements can also be considered. It can be easily proved that, in the first case of linear of unidimensional partitions for single symbols, all the possible countings of symbolic presences among the N elements of the set lead to the whole set of partitions of N, called $E(N)$. The successive consideration of two, three, four symbols, etc. can then be considered multidimensional partitions and their mutual coincidences would generate families of unidimensional second, third order partitions, etc.



Message set:	the five slots
Elements:	each one of the slots
Symbol:	the distinctions of each cube
Object:	one slot with the symbols on it
Sign:	the group of symbols on the element

Figure 1: A Compositional Message - An indefinite number of symbols (taken here from a "jar") are placed upon the discrete elements that make up the compositional message No positional relationships of order are involved.

It is worth noting that, whereas Shannonian entropy increases with the total number of symbols, partitional entropy reaches a limit. Thus, not only do partitions convey the abstract "form" of messages but they establish boundaries on the state space of possible messages using three important logical principles that characterize this approach:

The Principle of Parsimony precludes the addition of a symbol *different* from those already present if that symbol does not introduce further distinctions. It follows that a maximum of N-1 *different* symbols may accumulate on a single element,

e.g., the message [AXYZH, A, 0, 0] = [AB, A, 0, 0]

The Principle of Economy precludes the addition of a symbol the *same* as those present if this symbol does not introduce further distinctions (redundant symbols on all objects will not be perceived by the receiver). It follows that, a maximum of N-1 *equal* symbols may accumulate on a single element,

e.g., the message [AAA, AA, 0, 0] = [AA, A, 0, 0]

The Principle of Symmetry precludes the distinctions derived from the mutual exchange among symbols. So, commutative relations apply among the set of symbols,

e.g., the message [AB, AA, B, 0] = [AB, BB, A, 0]

2.2 Kmax- the most probable partitional state

After the above principles have been applied, the set of partitions $E(N)$ can be immediately transformed into a probability body (for the unidimensional case). The probability of any state of the set to exist as described by a specific partition is given by the relative frequency of this partition among all partitions. For instance, on $E(5)$ the probability is 1/7 - for states (5), (2,1,1,1) and (1,1,1,1,1), 2/7 - for states with either 2 or 3 summands each, 15/20 - for any summand to be an odd number, etc.

$E(N)$ is obtained by Ramanuyan:

$$E(N) = 1\frac{1}{\pi\sqrt{2}}\sum_{k=1}^{\infty}\sqrt{k}\,A_k(n)\frac{d}{dn}\frac{\sin h\left\{\frac{\pi}{k}\sqrt{\frac{2}{3}}\sqrt{n-\frac{1}{24}}\right\}}{\sqrt{n-\frac{1}{24}}}$$

(see Javorszky 1995)

The partition with the highest relative frequency is called the Kmax. In this most probable partitional state, the set shows Kmax distinct summands with respect to a one-describing dimension. In the case of $E(5)$, there is a Kmax shared both by 2 { (4,1) (3,2) } and 3 { (3,1,1) (2,2,1) }.

Heuristically, it appears that a compositional message can be univocally described by its corresponding "trace" of unidimensional partitions (Steidl and Javorszky 1996), if a few additional statistical measures that act as a sort of context or shared background in the communicational process have been previously established: most probable message length, ratio of symbols/elements, structural depth, shallowness, etc. Then the possible use of partitions of further combinations of symbols becomes redundant – and its inclusion would notably complicate the mathematical description of the message. Only the unidimensional-multidimensional problem may be pertinent.

The Kmax of every property or symbolic presence may be used as the origin or natural cannon to which the respective deviations of successive messages can refer. Kmax is the main message. Therefore, the information in a complex compositional message is represented by a comparatively small collection of distinguishing maxima instead of an overwhelming collection of symbols (and their permutations and combinations) that make up the message. The Kmax of a message would represent a natural property to which successive messages can refer. This simplifies the description of a specific message in the context of a continuous communication process and may speed up the process while diminishing errors.

2.3 Multidimensional partitions

Karl Javorszky (1995) has argued that an efficient massively parallel communication procedure – using multidimensional partitions – can be built around minimized partitional traces of the above Kmax. It seems to work particularly well with data sets of moderate size, which are preferably prestructured and come in a quasi continuous stream, so that the number of possible symbols is always kept rather finite. Although symbols might come from an infinite multitude, there should be a relatively small collection of distinguishing items employed at the communicational session, and their group relations should not generate a cardinality overstatement symbols/elements above a certain limit.

To the extent that Javorszky's estimates are correct, the overall capacity of a multidimensional compositional channel making use of discrete states of media can be generically expressed as:

$$T(N) = E(N) \; exp \; ln \; E(N),$$

where $T(N)$ is the number of different logical states which can be distinguished by means of collections of symbols put on the elements of the set N. Only non-redundant states are counted (see the principle of economy), because redundant symbol groups can always be substituted by single symbols, coalescing into a unique logical state. $E(N)$ is the already mentioned number of unidimensional partitions of the set N.

It is also interesting to compare $T(N)$ and the strictly positional use of the same elements of the set N in a combinatory way. According to the positional approach, a total of $N!$ different messages or logical states can arise using the same elements.

Surprisingly, $T(N)$ yields a larger number of logical states than $N!$ for values of N in between 31 and 95, with a maximum around 63-64. However, for N=12, the number of combinations $N!$ reaches a maximum with respect to $T(N)$. Apparently, several parameters of the genetic code would correspond with such max./min. extremes that characterize the compositional-positional interrelationship (see Javorszky 1995, for a detailed expression of all these formulae and calculations). Even a cursory analysis of the multidimensional partitions for N=3 and N=4 shows the emergence of an intricate "geometrical" (compositional) realm where symmetry patterns can be replicated by means of partitional operations (Villarroel, Pastor and Marijuán, in prep.).

2.4 The Emergence of Power Laws

Numerical partitions are characterised by exponential growth, and, heuristically, compositional messages have elements which are, by their very nature, contingent. The set seems to follow power laws and thus be vulnerable to small changes (Bak 1996). It is worth noting that a very simple way to obtain a power law is by the superposition of the whole partitional summands of a given number. We have graphed numerical partitions up to 20 (see Figure 2) and, except for some interference in the frontiers of the numeric interval, every summand's relative presence cleanly depends on a power law (Marijuán and Villarroel 1998). We have yet to solve for the general expression of the exponent of the power law which may describe partitional growth.

The power law theme leads towards a physical paradigm that apparently shares basic formal properties with the above compositional dynamics: self-organized criticality. The generalization built upon the well-known sand-pile paradigm that seems to apply to numerous natural phenomena (geologic, chemical, physical ones), leaving the characteristic signature of "power laws" in the involved structures and processes (Bak 1996), could also apply to the critical exchange of compositional messages that biosocial informational entities are collectively orchestrating by means of their communicational activities superimposed upon the structural ones... The fact is that power laws are omnipresent in cellular, organismic, economic, and social realms too (Scarrott 1996; Bak 1996; Marijuán 1998). Inescapably, this biologically-inspired train of thought on compositional messages has to be linked not only with self-organized criticality but also with the engineering-inspired views of G. C. Scarrott on "recurrence" (and related power laws) postulated as one of the basic tenets of natural information systems.

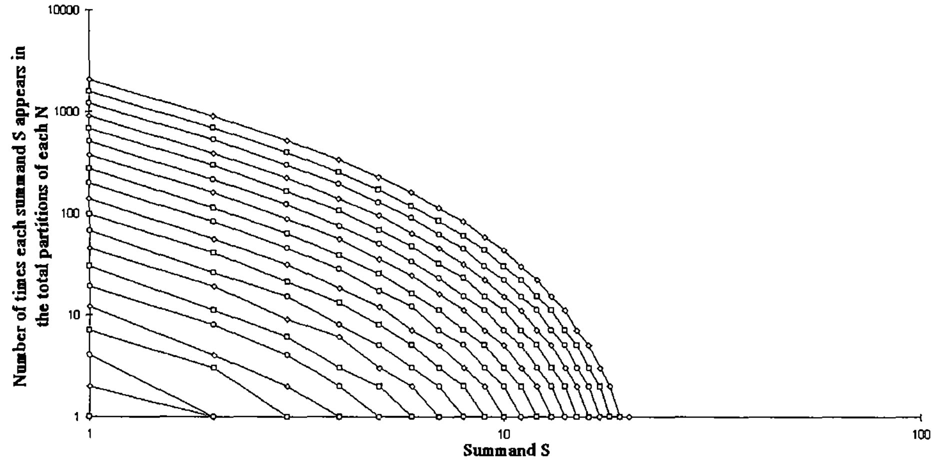


Figure 2: Power laws in partitions. Each curve stands for a particular value of N starting from 1 until 20. The ordinate represents the number of times each summand appears in the total partitions of each N. The x-axis represents the summand S

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
S																				
1	1																			
2	2	1																		
3	4	1	1																	
4	7	3	1	1																
5	12	4	2	1	1															
6	19	8	4	2	1	1														
7	30	11	6	3	2	1	1													
8	45	19	9	6	3	2	1	1												
9	67	26	15	8	5	3	2	1	1											
10	97	41	51	13	8	5	3	2	1	1										
11	139	56	31	18	12	7	5	3	2	1	1									
12	195	83	45	28	17	12	7	5	3	2	1	1								
13	272	112	63	38	25	16	11	7	5	3	2	1	1							
14	373	160	87	55	35	24	16	11	7	5	3	2	1	1						
15	508	213	122	74	50	33	23	15	11	7	5	3	2	1	1					
16	684	295	164	105	68	47	32	23	15	11	7	5	3	2	1	1				
17	915	389	222	139	94	63	45	31	22	15	11	7	5	3	2	1	1			
18	1212	526	298	190	126	89	61	44	30	22	15	11	7	5	3	2	1	1		
19	1597	686	395	250	170	117	84	59	43	30	22	15	11	7	5	3	2	1	1	
20	2087	911	519	336	226	159	112	82	58	43	30	22	15	11	7	5	3	2	1	1

3. A PARTITIONAL APPROACH TO CELLULAR COMMUNICATION

How can cells reliably communicate without any consideration about positional order in the "letters" of the chemical "words" they exchange? The experimental evidence is that every organismic cell, and every tissue, has sculpted its own coding and decoding apparatus, the Cellular Signaling System, basicly devoted to the analysis of communicational concentrations found in the extracellular-intracellular milieu, i.e., compositional messages. It implies the combined workings of thousands of receptors, hundreds of related protein kinases and phosphatases, and less than ten second messengers, all of them interconnected in order to make sense of the incoming stream of diluted messages. A very complex array of internal states (control of functionality, growth, cell-cycle stages, migration, apoptosis...) is regularly communicated among cells, tissues, and organs by means of such a peculiar molecular-processing apparatus.

In the Shannonian sense, one could conceive of a superimposed state-space built out from the whole variable concentrations that participate in the communication games, so that message patterns would map onto cellular states or onto molecular actuators leading to such cellular states (an automata table, or a grammar could be built). But there appear troubling evidences. The molecular adaptation of receptors (for instance by methyl or phosphate groups), the abundance of sigmoid curves and saturating effects, the vertical organization of signaling pathway components in "transducisomes", and the generalized cross talking among such pathways imply quasi-unsurmountable barriers for handling a regular information-thermodynamic state space. The tools used to describe physical states may not necessarily be meaningful for the description of a communicational space in the cellular milieu (Marijuán and Villarroel 1998).

Instead of the classical analysis of DNA sequences, it seems that the natural target to explore the possibilities of the partitional approach should be the "mysterious" processing operations performed by the cellular signaling system. In this sense, the system of receptors, membrane-bound enzyme and protein complexes, second messengers, and the dedicated kinase and phosphatase chains, could be understood as an abstract partitional processing-system capable of extracting the relative information differences within the stream of incoming compositional messages and physically transport these differences down to final effectors at the nucleus, cytoplasm, or membrane. That's the basic hypothesis that we are presently trying to explore.

If (and what a big "if") cells would make use of formal tools of logico-partitional nature in their management by means of the cellular signaling system of the compositional messages they receive, then the notion of a genuine cellular language, with specific dialects for every organismic tissue, could be seriously argued. And perhaps more interesting than that, quite a few other bizarre aspects of the signaling system could receive some more formal (and simpler) treatment: the cross-talk between signaling pathways, the checkpoints relating signaling operations with cell-cycle stages, the chaotic fluctuations of second messengers, and even the widespread formation of aggregates and complexes (transducisomes) among signaling components.

The studies by Caianiello (1985) on the partitional dynamics inherent in monetary systems and the suggestion by one of us (Marijuán 1998) about the "currency" role played by the set of second messengers in the internal measurement of cellular function might finally be stepping stones pointing out in the same direction: the foundations of information processing in nature and society.

REFERENCES

Bak, P. (1996) *How Nature Works· The Science of Self-organized Criticality*, New York: Copernicus.

Caianiello, E R. (1985) A Thermodynamic Approach to Self-Organizing Systems, in Yates, F E, Garfinkel, A. , Walter, D.O., Yates, G.B., eds. *Self-Organizing Systems, theEmergence of Order*, New York. Plenum Press.

Frigui, H. and Krishnapuram, R (1997) Clustering by competitive agglomeration, *Pattern Recognition* 30, pp 1109-1119.

Gatlin, L L. (1972) *Information Theory and the Living System*, Columbia: New York.

Javorszky, K. (1995) *Granularity Algebra*, Vienna: Mackinger-Verlag

Javorszky, K (1995b) The Logic of Self-Sustaning Sampling Systems, *IPCAT95 Proceedings*, pp. 386-398, University of Liverpool, UK.

Marijuán, P. (1996) Gloom in the Society of Enzymes: on the Nature of Biological Information, *Biosystems* 38, pp 163-171

Marijuán, P. (1996b) Information and Symmetry in the Biological and Social Realm: New Avenues of Inquiry, *Symmetry Culture and Science* 7(3), pp. 281-294.

Marijuán, P and Pastor, J. (1998) *The Language of Cells Foundation of Information Science*, http //fis.iguw.tuwein ac.at/igw/fis/fis htm.

Marijuán, P and Villarroel, M (1998) On Information Theory Stumbling Blocks. Some Biological Considerations about the Concepts of "Sequence", "Stability" and "Hierarchy", *Cybernetics*, in press.

Marijuán, P C. (1998) Information and the Unfolding of Social Life· Molecular Biological Resonances Reaching up to the Economy, *Biosystems* 46, pp. 145-151.

McLuhan, M (1962) *The Gutenberg Galaxy*, Toronto· University of Toronto Press

Modica, S. and Rustichini, A (1994) Awareness and Partitional Information Structures, *Theory and Decision* 37(1), pp 107-124.

Mosterín, J (1987) *Conceptos y teorías en la ciencia*, Madrid: Alianza Editorial.

Pastor, J. and Sarasa, M A. (1998) *Criptografía Digital Fundamentos y aplicaciones*, Zaragoza, Spain Prensas Universitarias de Zaragoza.

Scarrott, G (1996) *The purpose and nature of information*, (comm. workshop: Information, Instrument for the Survival of Society, UK Institution of Electrical Engineers, London, October 1996).

Schneider, T.D (1995) *New Approaches in Mathematical Biology Information Theory and Molecular Machines. Physics and the Origin and Evolution of Life*, Trieste. Italy.

Shannon, C E (1949) *A Mathematical Theory of Communication*, Chicago. University of Illinois Press

Steidl, R. and Javorszky, K (1996) *Message transmission by means of counting states of sets*. Unpublished manuscript.