(Imprecise Topics about) Handling Imprecision in P Systems

Antonio Di Nola[1], **Gheorghe Păun**[2,3], **Mario J. Pérez-Jiménez**[2], **Francesc Rosselló**[4]

[1]Department of Mathematics and Computer Science
University of Salerno, 84081 Baronissi, Salerno, Italy
E-mail: adinola@unisa.it

[2]Institute of Mathematics of the Romanian Academy
PO Box 1-764, 014700 Bucureşti, Romania
E-mail: george.paun@imar.ro

[3]Research Group on Natural Computing
Department of Computer Science and Artificial Intelligence
University of Sevilla
Avda. Reina Mercedes s/n, 41012 Sevilla, Spain
E-mail: {gpaun, marper}@us.es

[4]Dept. Mathematics and Computer Science
Research Insitute of health Science (IUNICS)
University of the Balearic Islands, 07122, Palma de Mallorca, Spain
E-mail: cesc.rossello@uib.es

A mathematician from an office placed in the Sevilla central building of the university proudly says to one placed in his office from the southern Sevilla building of the university:

– Look, I am closer than you to Palma de Mallorca!

He is right, but a biologist placed in Palma de Mallorca smiles, because (s)he sees no difference...

1 (A Sort of) Introduction

The standard P systems are beautiful mathematical toys (maybe of interest for computer scientists, linguists, etc) which can easily get a smile from Palma-de-Mallorca-of-biologists. Working with multisets, hence precisely counting the objects from the compartments of a system (a cell), assuming a universal clock, which ticks uniformly for all compartments, and using the rules (reactions) in the maximal parallel manner, or, the other extreme case, in the sequential manner, are three features which make the life of mathematicians easier and nicer, but which are science fiction for today biologists. The specification *today* suggests that *tomorrow* can change

the things. After all, more than one biologist is convinced that cellular processes are exact and deterministic, and that it is our knowledge that is inexact and incomplete, and maybe someday this knowledge will improve. However, waiting for tomorrow is not always a good strategy (for sure, not for our employers...), so that a sort of obsession wanders the science in general and membrane computing in particular: to become more and more realistic! Many papers are motivated in these terms, most of them succeeding to move (at most? at least?) from the southern building of Sevilla university to the central one (we do not mention them in the bibliography which closes this note, but only those which succeed – or at least promise – bigger steps; for the general bibliography of the area, the reader can consult the known web address `http://psystems.disco.unimib.it`), and only part of them already greeting the biologist from the Valencia beaches...

The problem (difficulty) is not that numbers are no longer sure things, thus contradicting Galileo, but that *reality is imprecision* (and complexity, but this is a related but different topic). Even if God does not play dice (how could Einstein know it?! and, what did he mean with that sentence, actually?), for us, the humans, dices are indispensable. And not only dices (probabilities), but also many other sources/forms of uncertainty, in most cases originating in the partial knowledge we have about processes, phenomena, systems we want to understand. Probability, partial information, fuzzyness, randomness, stochasticity, ambiguity, noise (not to mention incompleteness, undecidability, intractability) are only a few of the main terms related to this issue.

Coping with uncertainty is not only a challenge coming from "reality", from practitioners – in our case, mainly from biology. Biology is invoked here (in general, in membrane computing) mainly because this is the field from where the membrane computing is inspired and where it promises to return in the near future tools and applications relevant for biologists. However, membrane computing started as a branch of (theoretical) computer science, with the aim of learning something useful, or at least intellectually interesting, for computer science from the study of the (structure and functioning of the) living cell. The initial goal had nothing related to any promise to biologists. And yet at this theoretical level the challenge to deal with various forms of uncertainty appears. Because mathematics has developed several tools (theories) for handling uncertainty —probability theory, with many branches, fuzzy set theory, rough set theory, approximate reasoning and approximate algorithmics, etc.— it is a natural task for the mathematician to bring such tools in membrane computing, with or without having in mind (and motivating the papers by) how realistic the models are, from a biological or non-biological point of view.

We want to add here a word of caution for those intending to be realistic. There are indeed many mathematical approaches to uncertainty, and it is mathematically correct to develop an "uncertain" formal computational model by mixing any previous "crisp" computational model with any one one of these approaches, but some of these models may not be sound from the point of view of real world uncertainty modelling.

For instance, a fuzzy set $F : X \to [0, 1]$ (or with range any frame) assigns to each element of the (crisp) set X "the value with which this element belongs to the set F," but it can also be understood, dually, as assigning to each element of X "its value of having property F" or even "its value of approximating the element F." In this way, fuzzy set theory appears as suitable to work with mixtures, non-crisp properties, and degrees of approximation, but not to work with other types of uncertainties. Thus, for instance, our lack of knowledge of the place where a molecule is at a given moment shouldn't be modelled by means of a raw fuzzy set, but using the probability values of all possible places. Other approaches, like possibility theory of belief

theory, can be used here, but then they model more than simply our "lack of knowledge." And although a probability distribution is a fuzzy set, it is not sound to use general fuzzy set theory methods, like aggregation techniques or distances between fuzzy sets, to handle probabilities. On its turn, rough set theory approximates crisp sets "from below" and "from above," and it can be used in a natural way to model approximations set-theoretically, but not numerically. And so on...

This long (and, admittedly, imprecise) discussion is intended to stress the fact that the topics/suggestions mentioned below are not necessarily meant to "bring P systems closer to biology" (sometimes, one even writes "back to biology"), although such a goal is implicit and it would be a nice "by-product" of the possible results obtained in the study of these topics/suggestions. The questions formulated below are just natural from a mathematical point of view (although in their formulation we will use biological motivation/metaphora).

Then, an important point we want to make is the fact that this note is explicitly meant to foster discussions, researches, collaborations during the Brainstorming Workshop on Uncertainty in P Systems, Palma de Mallorca, November 2004. This is not a research paper, is only a positional paper, a provocation to the participants in the meeting. The choice of issues is subjective, their ordering has no significance (of importance), the list is not meant to be exhaustive, the classification below is approximate. And, of course, many formulations are imprecise enough; already formulating in a rigorous manner such topics would be a matter of investigation.

Finally, a warning/precaution related to the bibliography: we have mentioned many titles at the end, all we know in this moment in membrane computing area related to the topic of this discussion, but we will cite very few of them in the text, although many of them are directly related to the issues we discuss. Also, we do not mention any book or paper related to the mathematical approaches to uncertainty/imprecision, for instance, about fuzzy or rough set theories; there are huge bibliographies (on the web) about these topics and the reader can easily find such an information.

2 The Identification of Objects

Let us start "from inside", from the objects swimming and evolving in the compartments of a membrane structure.

Are we (always) entitled to say that the object a is in region i? Actually, what means "the object" a? If we "see" a molecule x, how much are we sure that it is of type a and not of type b? What about considering estimations for x to be of one of the types from a given list, for instance, expressed in the form of probabilities? We can then discuss about objects with probability .8 to be a and .2 to be b. This can be also a matter of possibilities, leading to the use of possibility values and theory instead of probability ones, of beliefs, closer to fuzzy set theory, or of similarities, which lead to classification in terms of rough sets theory. We can go further (making the life of the mathematician still harder), working with objects x identified by statements of the form "x is an object from the set $\{a, b, c\}$ with probability .66 and from the set $\{d, e\}$ with probability .34." The two sets are here disjoint, but in a general case they can be not. What can we do with additional statements of the form "with probability 1, x is not from the set $\{a, f\}$?"

We have here also another sensitive issue, related to the fact that in the compartments of a P system we deal with *copies* of objects. What is a copy? (S. Marcus posed somewhere a much more dramatic question: do there exist copies?) If we cannot precisely identify an

object, then we cannot precisely say when one object is a copy of the other (hence that they are indistinguishable, for instance, for the evolution rules). Can we work with *similar* objects, instead of *identical* objects? Equality is an equivalence relation, similarity is only a tolerance relation (it is not transitive), leading to tolerance classes which are not necessarily disjoint. This raises problems when handling the objects, as they are not "crispy" classified. Should we return to numerical estimations of the similarity, which could be used for instance to model inexact, mutated or simply modified copies, or should we remain in a qualitative framework (e.g., working with a similarity relation)?

And finally, do always these questions matter? For instance, in molecular biology, methylated DNA can act in some reactions as usual, non-methylated DNA, but it may lead to errors (with a certain probability!) in other reactions, yielding inexact results, and it cannot be involved at all in other reactions that need definitely non-methylated DNA molecules. Should we impose consequently that some distinctions, or similarities, matter as far as some rewriting rules are concerned, but not for the remaining ones?

3 The Place of Objects

After identifying the objects, we have to place them in the compartments of a membrane structure, maybe also taking into account the environment, and this is again a rich source of imprecision. Even if we try in a laboratory to introduce ourselves a molecule in a compartment of a cell, we cannot be always sure that the result is the one we expect; still more difficult is to fish for a molecule in a given compartment. Furthermore, many chemicals from the cell are macromolecules, maybe long chains of atoms, which can be placed with an end in a compartment and the other end in another compartment (like the many proteins embedded in the membranes). We have also to mention that the molecules move continuously across membranes.

In short, it makes sense to assign probabilities (estimations) to the fact that a given object is placed in a given compartment, and this is directly related to the next point.

4 Describing the Multisets

Which objects are at a given time in a given compartment, and how many copies of each? From the probabilities assigned to the presence of objects in a given compartment we can have a probabilistic estimation of the contents of that compartment. Maybe more natural is to have fuzzy set estimations or rough set approximations of multisets from compartments (at least, because the general study of fuzzy and rough multisets is well developed and can provide tools to use in our area).

Rough set theory looks particularly attractive, because the basic issue of this theory is to approximate a set by an upper and a lower approximation, the latter one *surely* included in the set, the former one *surely* including the set, in between having a border, as larger as higher the imprecision is. By enlarging the available information, the two approximations can converge to the real set, thus making smaller and smaller the border. How this attractive idea can be used in a P system? For instance, which rules should be used for evolving the objects from the border? Should they participate in the same (cooperative) rules with objects from the lower approximation and/or from outside the upper approximation?

On its turn, fuzzy set theory can be used to represent statements like "there are around seven copies of reactive a in that membrane," by defining, for instance, the content of a membrane as

a mapping sending each reactive a to a fuzzy natural number, i.e., a mapping $n_a : \mathbb{N} \to [0,1]$ that has some suitable properties.

Other, not so well-know approaches already introduced to handle approximately defined quantities, or specially tailored *a posteriori*, can be used to define generalized multisets that model other kinds of uncertain contents of membranes. For instance, techniques imported from interval calculus can be used to manipulate multisets sending a reactive to, say, "somewhere between 5 and 9."

Of course, we can try to learn something (or to get challenges) from the biologist or from the bio-chemist. On the one hand, the former one will push us towards linguistic logic, because (s)he currently works with statements using such terms as "many molecules", "sufficiently large population", "high enough pressure," and so on. Fuzzy set theory can be used in this connection, since statements like "there are many copies of reactive a in that membrane" correspond to define the multisets that describe the membranes contents as mappings that send each reactive to a linguistic variable, and these multisets can be manipulated using techniques imported from fuzzy control theory.

On the other hand, bio-chemists work with probabilities, reaction rates, concentrations, gradients, stoichiometric constants, etc.. All these mean real (well, rational) numbers, associated both with the contents of membranes and with the reactions taking place in/on the membranes. Multisets with real multiplicities associated with objects have been already considered, e.g., in [9, 15, 16]. In particular, the first cited paper started a systematic study of P systems with non-discrete multisets, but the topic is far from being exhausted.

It should be mentioned that in most biological applications of P systems reported so far, the rules have associated probabilities/reaction rates, sometimes dynamically computed, in accordance with the current population of objects, using standard techniques from biochemistry (e.g., stoichiometric constants), which, interestingly enough, brings continuous mathematics aspects in the functioning of P systems, which, in the basic version, are essentially discrete machineries.

5 The Level of Rules

All the previous sources of uncertainty and ideas about ways to capture/handle them in P systems at the level of objects and multisets have a direct connection with the way the evolution rules are defined and applied. How precise is a rule defined? Are the multisets from its left and right hand members precise or not? If not, what this means? What about the relation between the left and the right hand multisets, whatever their definition is? Furthermore, in the case of multiset-rewriting rules, where targets are associated with the objects newly introduced by a rule, we can question the precision of these targets, and associate with each object all targets, with probabilities assigned to them: for instance, something like $(a : here.5, out.2, in.3)$. In this way, probabilities are assigned to the presence of an object in a given compartment, hence even if we start from a precisely known system, after a while the place of objects will be only probabilistically known.

Besides, are the reactions the same at different moments, or their result also depends on parameters other than the contents of the compartment where they are used?

When a rule can/should be used? This has to do with the probability for a reaction to take place, which, in turn, depends on the concentration of reactants, but also on reaction conditions (temperature, pH, available energy, etc). We do not repeat the previous discussion about reaction rates, stoichiometry, etc., but these terms are highly relevant here. An important issue

concerns the relationship between non-determinism and various forms of uncertainty; intuitively, by assigning probabilities (rates of reaction) to rules, we diminish the non-determinism in using the rules, as the probability induces some priority relation among rules. This topic deserves a more detailed examination, for instance, in relation with the resolution of computationally hard problems by means of P systems.

Finally: if the multisets of objects from compartments are fuzzy or rough sets, what about applying to them sets (or multisets) of rules which are also fuzzy or rough?

6 The Clock and the Parallelism

These two issues are related. Without an external clock it is difficult (but probably not impossible) to define the functioning and the result of a P system, but assuming that all compartments have the same clock and that all rules/reactions last the same amount of time is far from... Palma de Mallorca. Getting rid of the internal clock is a great topic and the first results started to appear, see, e.g., [7, 8, 22]. "No clock" does not necessarily mean "no internal time" (for instance, no known duration of rules), but the internal time can be different from a compartment to another one, with rules of different durations, maybe expressed in non-integer numbers, maybe simply unknown. The collaboration between (cooperative) rules can be achieved through the objects they produce, the rules can also be synchronized by signals or promoters/inhibitors, with a great flexibility in what concerns the moment where the necessary objects become available.

The question of time is directly connected to that of parallelism. Maximal parallelism is powerful (e.g., because it can provide information about the whole multiset from a compartment), the sequential use of rules is easy to handle, but the truth is somewhere *in media res*. How to deal with "partial parallelism," what this means and how can it be estimated?

7 Other (Related) Issues

Of course, there are many other things to discuss in this framework. We have said nothing explicit about the environment, which also can be described in imprecise terms. Then, we mentioned above only multiset-rewriting rules, but the same issues can be formulated for symport and antiport rules, which move multisets (of which type?) of objects (how precisely known?) from a region to another one (how precisely defined?) and that can even degenerate (with some probability?) with each movement.

In the previous sections we have talked about imprecision in general, without mentioning possible *degrees of imprecision*. Given a system, can we evaluate its degree of imprecision? Given two systems, can we say that one of them is better identified than the other one? (Of course, such a comparison asks for an external observer —a crisp one, maybe.) In what terms, using which measures? Maybe entropy, maybe other criteria. When a system contains "too much" imprecision, so that it makes no sense to further work with it, because the information we get is irrelevant (non trustful)? How the degree of imprecision can be decreased, what is the information we can get and which we have to look for in order to improve the knowledge about a system? Working with degrees of approximation and the asymptotic convergence of approximations to the set one looks for, are standard issues in rough set theory; it remains to implement them also in P systems.

The previous topic is related to using approximation in an operational manner, for instance, in terms of probabilistic or randomized algorithms. If we cannot solve a problem with reasonable

resources (polynomial space and time), then let us try to have an approximate solution, with a well estimated degree of accuracy, or an optimal solution which is however found only with some precisely estimated chance, but using reduced resources (larger the resources, bigger the probabilities to get a good solution or to get a solution at all). In membrane computing the challenge is clear: "solving" hard problems in polynomial time, but not using an exponential workspace (even created in the natural way provided by membrane division or string replication), but a polynomial workspace, by paying in the accuracy of the certainty of the solution. The question waits to be systematically approached.

Finally, there are many other sources of imprecision. Non-determinism and confluence (in the strong sense, all configurations lead to a unique configuration, or in the weak sense, all configurations lead to configurations from a given class, having a specified property) are standard properties of P systems. How are they related to imprecision? What means in our context ambiguity or synonymy, to mention only two sources of imprecision from linguistics?

We hope that the reader will take the turn and address these questions (at least by reformulating them in mathematical terms) or related ones, thus moving the area closer to reality/biology (and the reader her-himself to Palma...).

References

[1] I.I. Ardelean, M. Cavaliere, Modelling biological processes by using a probabilistic P system software, *Natural Computing*, 2, 2 (2003), 173–197.

[2] I. Ardelean, M. Cavaliere, Playing with a probabilistic P system simulator: Mathematical and biological problems, *Brainstorming Week on Membrane Computing*, Tarragona, February 2003, TR 26/03, URV, 2003, 37–45.

[3] D. Besozzi, C. Zandron, Dynamical probabilistic P systems, DNA10 (poster?).

[4] M. Buzzi, *Calcolo con membrane. P sistemi probabilistici*, Master Thesis, Univ. of Como, 2003.

[5] J. Casasnovas, F. Rosselló, Scalar and fuzzy cardinalities of crisp and fuzzy multisets, submitted, 2003.

[6] J. Casasnovas, J. Miró, M. Moyà, F. Rosselló, An approach to membrane computing under inexactitude, *Intern. J. Foundations of Computer Sci.*, to appear.

[7] M. Cavaliere, Towards asynchronous P systems, *Pre-proceedings of Fifth Workshop in Membrane Computing, WMC5*, Milano, Italy, 2004, 161–173.

[8] M. Cavaliere, D. Sburlan, Time-independent P systems, *Membrane Computing. International Workshop WMC5, Milano, Italy, 2004*, LNCS ??, Springer, 2005.

[9] A. Cordón-Franco, F. Sancho-Caparrini, Non-discrete P systems, *Pre-proceedings of Fifth Workshop in Membrane Computing, WMC5*, Milano, Italy, 2004, 205–207.

[10] R. Freund, Asynchronous P systems, *Pre-proceedings of Fifth Workshop in Membrane Computing, WMC5*, Milano, Italy, 2004, 12–28.

[11] V. Manca, On the dynamics of P systems, *Pre-proceedings of Fifth Workshop in Membrane Computing, WMC5*, Milano, Italy, 2004, 29–43.

[12] S. Marcus, Tolerance multisets, *Multiset Processing. Mathematical, Computer Science and Molecular Computing Points of View*, LNCS 2235, Springer, 2001, 217–223.

[13] S. Miyamoto, Fuzzy multisets and their generalizations, *Multiset Processing. Mathematical, Computer Science and Molecular Computing Points of View*, LNCS 2235, Springer, 2001, 225–236.

[14] M. Mutyam, Probabilistic rewriting P systems, *Int. J. Found. Computer Sci.*, 14, 1 (2003), 157–166.

[15] T.Y. Nishida, Multiset and K-subset transforming systems, *Pre-proc. Workshop on Multiset Processing*, Curtea de Argeş, Romania, TR 140, CDMTCS, Univ. Auckland, 2000, 193–202, and *Multiset Processing. Mathematical, Computer Science and Molecular Computing Points of View*, LNCS 2235, Springer, 2001, 255–266.

[16] T.Y. Nishida, Simulation of photosynthesis by a K-subset transforming system with membranes, *Pre-proc. Workshop on Membrane Computing*, Curtea de Argeş, 2001, *Technical Report* 17/01 of RGML, URV, Tarragona, 223–228, and *Fundamenta Informaticae*, 49, 1-3 (2002), 249–259.

[17] A. Obtulowicz, Probabilistic P systems, *Pre-proceedings of Workshop on Membrane Computing*, Curtea de Argeş, Romania, August 2002, MolCoNet Publication No 1, 2002, 331–332, and LNCS 2597, Springer, 2003, 377–387.

[18] A. Obtulowicz, Mathematical models of uncertainty with a regard to membrane systems, *Brainstorming Week on Membrane Computing*, Tarragona, February 2003, TR 26/03, URV, 2003, 241–246, and *Natural Computing*, 2, 3 (2003), 251–263.

[19] A. Obtulowicz, General multi-fuzzy sets and fuzzy membrane systems, *Pre-proceedings of Fifth Workshop in Membrane Computing, WMC5*, Milano, Italy, 2004, 316–326.

[20] A. Obtulowicz, Gh. Păun, (In search of) Probabilistic P systems, *BioSystems*, 70, 2 (2003), 107–121.

[21] F. Sancho-Caparrini, A note on complexity measures for probabilistic P systems, *Proceedings of the Second Brainstorming Week on Membrane Computing, Sevilla, February 2004*, Technical Report 01/04 of Research Group on Natural Computing, Sevilla University, Spain, 2004 443–448, and *JUCS*, 10, 5 (2004), 559–539.

[22] D. Sburlan, Clock-free P systems, *Pre-proceedings of Fifth Workshop in Membrane Computing, WMC5*, Milano, Italy, 2004, 372–383.

[23] A. Syropoulos, Fuzzifying P systems, submitted, 2004.