chmnddf 23/3/86

MASTER

CHANGES of MIND

Revision of "Interpretation" in Episodic Memory.

Antoine CORNUEJOLS

Computer Science Department

University of California, Los Angeles[1]

Abstract :

In this paper the distinction is made between ***Semantic Memory*** (the building components and links of any reasoning and knowledge structure) and ***Episodic Memory*** (the actual instantiation of a memory structure). Whereas most current research in Learning focus on constructing and modifying the Semantic memory, we concern ourselves here with the mechanisms by which can occur the ***modification and revision of an "Interpretation"***, or model, in the Episodic memory. This implies non-monotonic understanding and the judicious retraction and modification of erroneous inferences.

Since frame-based knowledge representation schemes have been tailored to embed inferencing, they are used here as a testing ground for our model, the main characteritics of which being that possible interpretations of a given set of inputs may be seen as local equilibrium states, and that re-interpretation may occur when the system is submitted to perturbations like new incoming informations, "day-dreaming" experiences (Mueller, 1985) and other methods. We propose a model based on domain-independent local mechanisms out of which emerges a global behaviour, closed to the ones exhibited by humans at a subconscious level or in their scientific endeavours. Several interesting considerations on the relations between local and global processes, the influence of the relative internal and external speeds of the system on its re-interpretation ability, and meta-learning, are equally brought up in this paper.

[1] **Current address : Antoine CORNUEJOLS, Computer Science Department, 3680 Boelter Hall, U.C.L.A., Los Angeles, CA 90024. USA.**

1- Motivation.

Systems that must interract with the real world face the challenge of having to make decisions based on an incomplete knowledge of their environment. Certain ***assumptions*** and ***inferences*** need therefore to be made in the course of the reasoning process, not only because the knowledge at hand is insufficient but also because the same set of "raw" data may be interpreted in more than one way. Consequently, the decisions taken may turn out to be maladjusted. In order to better perform in the future, the system must be able to revise its model of the world, that is, its knowledge base and particularly its set of assumptions, sometimes by removing or refining some of them. This is the essence of ***non-monotonic reasoning.*** Its necessity resides in the constraints imposed by the real world.

In the human realm of behaviours where all kinds of ambiguities, illusions, allusions, double-entendre, incongruities pervade the every day life, non-monotonic reasoning or re-interpreting, is at the core of many activities : getting a joke, re-interpreting scientific evidences, reading a fable or a censured opposition newspaper "between the lines" to reach its hidden meaning, and so on. All of these undertakings similarly involve the revision of one's ***Interpretation*** or model of the universe, that is the modification or corrections of ***expectations*** which are the bases of assumptions and inferences.

In the paradigm currently dominant in AI, expectations are, explicitly or implicitly (see [Brachman & Levesque, 1985]), embedded in ***frame-like based knowledge representation schemes.***[1] It has been found that this kind of structuring and packaging of the knowledge is conveniently expressive, and reasonably efficient computationally although recent studies [Brachman & Levesque, 1985] and [Brachman, 1985| balance somewhat that later assumption. Frames have therefore been widely adopted in AI from Natural Language Understanding to Vision.

This paper thus adopts this framework as a starting hypothesis and intends to propose a new methodology to deal with certain problems of non-monotonic reasoning.

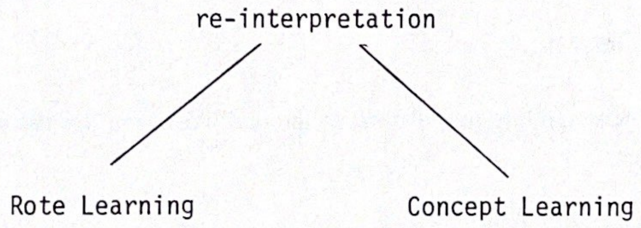
[1] Other paradigms exist. In the Connectionist view for instance, frames are somewhat replaced by bassins of attraction in the state space of the system.

Semantic and Episodic memories

The consideration of frame-based representation systems conducts to the distinction between what will be called from now on Semantic memory and Episodic memory.

Semantic Memory is the set of "virtual" or "potential" frames and links or relationships existing in the system prior to the task considered, which, once instantiated, will form the building blocks of the knowledge structures set up in ***Episodic Memory***. The later then contains an instantiation of whatever structure of frames and links has been found useful to interpret the perceived world.

Most research in Learning are concerned with Semantic Memory : how to acquire new ***concepts*** [Winston,75], [Lenat,76] or new rules [Smith et al.,85], how to refine them. This is considered to be the hard core of Learning. This paper on the other hand takes for granted what the semantic memory has in store, and is concerned with how changes of ***interpretation*** can occur in episodic memory. Our research is therefore centered on the spectrum between rote learning (pre-digested accumulation of data and procedures in classical data-bases) and semantic learning or learning of new concepts by discovery.



This doesn't mean that dynamic re-interpretation must be more easy than other learning tasks. The next section should prove it by outlining the main problems and difficulties.

Let us get at what we hint at by rapidly analyzing the following story understanding example [1].

(1)

[1] Story understanding has been chosen as an test-bed throughout this paper because this is a domain which has already been thoroughly investigated and which examplifies the frame-based approach. It is also simple to understand. But yet, the main ideas are valid in the other domains as well like visual perception for instance.

> **"He plunked down $5 at the window. She tried to give him $2.50 but he refused to take it. So when they got inside she bought him a large bag of pop-corns."**
>
> (from Collins et al.,1980)

In this context most people first assume that "she" is the cashier of a movie theater, and then, on encountering "... when they got inside, she bought him a large bag of pop-corns." re-attribute "she" to be the date of "he". Other people don't recognize at first a movie theater scene but instead a betting place for horse races for instance, and must modify their view accordingly in the course of reading the whole text.

But why and how do we do that? What makes an interpretation better than another ? How conflicts are detected? How are they solved? These are some of the basic questions that face the researcher in this area. Few approaches to this problems have been attempted yet, and none has given a satisfying definite solution. The next section shows why these are very hard problems, and what had been achieved so far.

2- Non-monotonic understanding : problems and related works.

2.1- Five basic questions.

We said that Understanding in real-time in the real world implies the use of faillible assumptions and inferences.

For instance, it is often the case that several interpretations are possible, and, furthermore, seemingly equally valid, for the same set of data. A famous example is **"I saw a man on the hill with a telescope"** which admits four different interpretations. Less trivial instances of the unavoidable ***ambiguousness*** of any message (because their meaning depends on both the emetteur and receiver models of the universe) are the SAM-II treaty which different interpretations from US and soviet sides make the joy and bread of many a politician, or sex jokes as :

> **"A house wife throws up her window and calls out to the iceman :**
>
> **- Have you the time ?**
>
> **And the iceman replies :**
>
> **- Yes, if I can find somebody to hold the horses."**

This "multi-interpretiveness" of one input is remarkably the mirror image of ***Analogy,*** where one looks for one interpretation or model that can satisfy several different sets of data. There may be there a link worth more considerations.

Another, though not completely different situation, where the inferencing process determines the final interpretation is when ***the analysis of the input can be pursued more or less*** and still yielding coherent pictures of the world, even though they may be different. An illustration is the chess game where the deepness of the causal reasoning may sometimes change dramatically the evaluation of the current position. Any activity involving chains of causal reasoning, like forecasting and diagnosis, is deemed to encounter the same problem. But other, more common, situations may be seen also as examples of this phenomenon. In :

> **"My son was born on the first of the month.**
> **_ Is that why you named him "Bill"?"**

the pun implies to perceive both senses of "bill". Numerous jokes are based on the same kind of tricks. But metaphors and fables also imply that several level of understanding be performed in order to be effective.

Thus, we have identified at least two fundamental sources of potential mistakes : ***ambiguity*** where one has to choose between several possible interpretations with insufficient informations to guide surely this choice, and *deepness of reasoning* where one has to guess where to stop. These problems are inherent to any communication process and there is no a priori way to escape them.

That's why we do not concern ourselves in this paper, except by side considerations, with the determination of what set of inferences and control strategies are the more appropriate ,(or with the discovery of so-called unambiguous languages). These features are embedded either explicitly or implicitly in the semantic component of the memory, but *whatever choice has been made on knowledge representation issues, there will always be the problem of recovering from erroneous inferences.* That is the concern of this paper. How to discover whenever it happends that an erroneous assumption has been made, and how to modify appropriately the episodic memory.

This requires that several questions be answered.

1. How to detect that something is unsatisfactory in the model of the world?

One sign would be a *conflict* showing up somewhere, that is two pieces of data contradicting each other. But its detection may imply a long chain of reasoning that is not likely to be routinely pursued in the memory. Also the system may be stuck with a coherent but otherwise very ***complicated interpretation*** of the world requiring a lot of additional assumptions to hold. (Think of the formidable Ptoleme model of the solar system with all its epicycles, compared to the simple and elegant model of Kepler). This situation is to be avoided because it is computationally inefficient and necessitates a lot of maintenance work. Here also the parsimony principle should be the guide and judge. The problem is to implement it.

2. What must be the respective status of input facts and the data that result from inferencing and reasoning?

The answer to that question is essential in order to be able to compare different pieces of knowledge. Should we trust all incoming data, and more generally should we introduce degrees of belief with all the methodological precautions and tools it implies?

Then of course, assuming that an unsatisfactory situation has been diagnosed, we must know :

3. How to proceed up to the initial cause of the mis-interpretation.

Because this is most likely the only place where a modification or a patch can durably and soundly fix the whole problem. Unfortunately, in most understanding systems a simple backtracking method is not practical, and more sophisticated reasons maintenance techniques must be employed.

Finally, these efforts would make sense only if we knew :

4. What changes must be made to modify the episodic memory toward a more appropriate state.

That in turn requires that we know what makes an interpretation more satisfying. But there is more. Because if the system is to be able at all to change its interpretation of the world then it must keep, at least partially, a trace of the raw facts. Thus a big issue is : ***what redundancy must be kept in memory to allow re-***

interpreting? [1]

A last question that desserves attention is, very naively :

5. How a system will function with inconsistent knowledge?

> There are two main reasons to ask that question. First of all, it may be quite unreasonable to do a lot of maintenance work each time an inconsistance arise when in fact the actual drop of performance of the system would be very slight. Second, it is very likely that the re-interpretation process when it takes place will entail some inconsistancies in episodic memory during some lapse of time however short they are. What would be the consequences for a system submitted to real-time constraints?

2.2- Previous works.

2.2.1- The BORIS understander system. [Dyer,83]

Michael Dyer's work on Language Understanding described in [Dyer,1983] doesn't deal with the issue of the retraction of inferences and non-monotonic reasoning per se. But it is one of the most integrated and complete system yet devised using a frame-based methodology in this domain, and it is taken here both as an example of how far one can go without non-monotonic reasoning capabilities and as an ideal starting point for future improvements.

Dyer's system, BORIS, shows how far disembiguation can be carried through the use of "multi-level" understanding, that is the bearing of all types of knowledge available, from scriptic activities to human motivations (goals) and affects, in the process of understanding. This is a foremost example of the simultaneous building and use of an highly interconnected network of frames in the Episodic Memory. Nonetheless, as we have seen, not all errors can be prevented that way, and BORIS provides yet another recourse to recover from them in an early stage of understanding, the *Working memory*. In it are contained the bindings created between word senses while reading a sentence word by word. In this store word senses can be tentatively binded and then retracted if a discrepancy is discovered with the Episodic memory. The necessary redundancy to retrieve alternative word senses results from the fact that all

[1] As an illustration, in story (1), once the begining has been interpreted as a betting place scene, how do we go back to the other possible interpretation as a movie theater place?

word senses are kept until the whole sentence has been parsed. For instance, in :

"... the judge would award the case to George once the judge learned that Ann had been cheating on HIM"

"him" is first tentatively bound to **"the judge"** before this presupposition is checked against the Episodic Memory which in this case shows a violation of the interpretation of the story made so far. **"Him"** is therefore re-assigned to the other possible choice : **George.**

Unfortunately, not all ambiguities can be resolved in the scope of a single sentence, as in the case where the knowledge needed for the revision of the interpretation will only be available in sentences ahead.

Therefore, several attempts have been made to attack the global problem.

2.2.2- The ARTHUR system. [Granger,80]

Granger was among the firsts who concern themselves specifically with the recognition and correction of erroneous initial inferences about context. His system, ARTHUR, was able to modify its first assumption about the context of a short narrative like : **"Mary picked up a magazine. She swatted a fly."** where it would switch from the inferred goal of "reading the news" to the one of "destroying a fly".

In doing so ARTHUR applies a Parsimony principle : *The best context inference is the one which accounts for the most actions of a story character.* Here it could explain the action of Mary by a combination of both of the goals stated above, but it is more parsimonious to explain it on the ground of a single goal : "destroying a fly".

Granger draws up a taxonomy of erroneous inferences related to the kind of schema used in understanding a story like scripts, plans and goals, and hints at the possibility of making a distinction between strong and weak inferences.

The problem with ARTHUR is that it is contrieved to correct erroneous inferences only in specific situations implying the recognition of a simple obvious redundancy in the interpretation structure involving only scripts, plans and goals, in fairly simple contexts. It therefore apparently lacks generality and power.

2.2.3- The RESUND system, and non-monotonic dependencies. [O'Rorke,83]

One of the most serious investigation in the domain of non-monotonic understanding has been led by Paul O'Rorke with the design of RESUND. With it O'Rorke tried, by attaching a justification to each inference drawned (not unlike the XPLAIN system [Swartout]), to make use of the apparatus developped for the so-called Truth-maintenance systems and for the dependency-directed backtracking [Doyle,79], [De Kleer,79], [Charniak,80].

In the process of studying the particular inferences used in Natural Language Processing systems, O'Rorke came up with a first set including Elaboration, Intentionality, Identification and Criteriality rules. (*Elaboration* : "John purchased tickets" → John paid someone, .../ *Intentionality* : looks for goals/ *Identification* : the different parts of a narrative are assumed to be co-referential/ *Criteriality* : the occurence of a sub-event evokes a whole event (scripts for instance)). These rules in turn can give birth to mostly two types of problems : *identity conflicts* and *schematic constraint violations*. When these happend _note that their detection is immediate_ a dependency-directed backtracking process determines the underlying incompatible assumptions by looking back along dependencies and rules out one of these assumptions, based on the application of *preference policies* which represent different criteria for gauging the relative strength of assumptions.

This model is more general and powerful than the Granger's approach, and can be readily extended. I see however three major weaknesses in it. First, it can only treat "local" inconsistencies which are relatively easy to point out, but it would be helpless in discovering and curing a global unsatisfying knowledge structure like the one of Ptoleme's view of the world. Second it requires on top of the mechanism for frame-based understanding an external agent with global knowledge to judge and compare the strength of the various assumptions. This expert's knowledge and computational ability are likely to have to grow exponentially as grows the number of different types of inferences, justifications and preference policies. Third, as Johan de Kleer himself has pointed out recently [De Kleer,], Truth maintenance mechanisms are intrinsically incapable of working with multiple contradictory choices at once, and are futhermore very inefficient in both time and space. Thus they appear to be poor candidates to underly the non-monotonic reasoning capabilities of real-world understanding systems.

2.2.4- Dynamic memory. [Schank,82]

In his book, "Dynamic memory", Schank advocates a general need for a self-organizing dynamic memory which would adapt by itself its structure and content depending on the inputs it receives from the world (in that case written narratives). This theory actually concerns only the Episodic part of the memory, but even in that sub-domain of the whole problem of memory and learning, the proposed model falls short of a complete solution. For instance it doesn't deal at all with potential conflicts and inconsistencies, but merely suggests schemes for adapting the indexing system of the memory, such that further ramifications are developped when needed along with the acquisition of new knowledge in certain fields, whereas other parts of the indexing structure can be eliminated if their usefulness has not been confirmed. The whole system described in the book actually looks like a rediscovery of principles studied 15 years ago for Pattern Recognition and Automatic Classification, but without their theoretical soundness and justification. It is in any case insufficient for the general purpose stated in this research.

3- A new approach: stability and perturbations.

3.1 The guiding principles.

This section stresses the main methodological guidelines of the research reported here.

Let us remind the problem. We want a system which would be able to detect a discrepancy or an unsatisfactory state whenever it occurs in episodic memory, and would appropriately modify the knowledge structure in episodic memory to reach a more satisfactory interpretation of the world.

This statement of the problem points out that there would be no point changing an interpretation in memory if no, more satisfactory, one is available. That means in my opinion that the first question is not how to detect a discrepancy in memory, which would still leave open the problem of unsatisfactory states appreciation, but rather how to detect that a better interpretation is possible. Therefore the problem is turned over : we want a system that changes its model of the world whenever a better model is possible. ***The detection of contradictions or awkward structures becomes then a side-effect of the main process of re-interpreting.***

The second point is that ***we want that re-interpretation process to occur spontaneously.*** By that I mean that it must be the product of an internal essential necessity of the knowledge representation scheme, and not the result of the decision of some external omniscient and omnipotent "mega-demon". To give an image, complex molecules or crystals in chemistry reconfigure automatically themselves when the necessity arises (collisions with other molecules or new stresses). They don't wait for the chemist to measure their energy and give them instructions to reconfigure.

More concretely, that implies that the *acting forces* of re-interpretation will be found ***at the local level*** of knowledge representation, i.e. in the frame-based paradigm, at the level of the links between different pieces of data, and of the data themselves.

This confront us with the problem of realizing a given global function or behaviour out of local processes.

Furthermore we still have to define clearly what is the global function we want to implement, that is ***what makes a satisfactory interpretation***, on what grounds to compare two models of the world. E.g. what makes us laugh at :

> **Blanche : Gracie! Where did you get all these lovely flowers?**
> **Gracie : Well, George told me Bettie was in the hospital and that I should go visit her and take her flowers. So, when she wasn't looking, I did!**

why do we think that Margie had a false understanding?

We must begin by examining that question.

3.2 The value of an interpretation : its fitness function.

Two properties seems to be required of any good model of the world. One is to ***give "right" answers*** when asked, that's a functional requirement : a mute or dumb data-base is of little use. The second is to obey the *parsimony principle* advocated in science : don't multiply unnecessary hypotheses.

A third desirable property should also in my opinion be added to these requirements. Indeed, it is a distinctive feature of any good scientific theory that it leads to "good" questions for future research. Likewise, a good model of the world in episodic memory should also be one ***ready to ask "intelligent" questions***.

Basically a frame-based memory is made up of instantiated frames tied together by links. A *frame* is simply a package of pieces of knowledge, organized in such a way as to represent a concept or a class of objects. Certain attributes are part of the frame, other are to be find in other frames. The attributes are therefore represented as *slots* which value is found by following a link to a frame. Sometimes, a slot takes a value "by default", meaning that a connection to a frame or data is provided apriori and can be used in the absence of any specific information concerning the value of that slot. Additionally, a slot is of a specified type, for instance HUMAN-BEING, with possibly procedural restrictions like HUMAN-BEING different from the one in slot such and such. As would be the case for instance in the frame MOVIE-THEATER, where the HUMAN-BEING acting as casher is different from the ones who are custommers. Also, procedural knowledge may be attached as to tell what to do when the slot is filled, or what to do to fill it.

For instance an airline pilot frame could look like :

AIRLINE PILOT	
Airline :	TWA (instantiated)
Sex :	MALE (default-value)
Trade-union :	(procedure to find it)

Associated frames :	TRANSPORT CONTRACT

The **links** are thus simply pointers between frames, or slots and frames. Their destination is either given explicitly, as TWA for the airline slot above, or implicitly, as with the procedure attached to the trade-union slot.

In a story understanding system, frames includes MOPs, scenes, goals, plans [Schank, 1982], TAUs [Dyer, 1983], and so on, and are linked together with various types of connections like Intentional-links, Enablement-links, and so on. When a question is asked to the system then, as : **"Why did Paul write to Richard?"**, the proper frame (here LETTER) is selected and the related link traversed, here an Intentional-link, to retrieve the answer which may be embedded in a frame or a complex structure of frames, yielding perhaps : **"Because he wanted a lawer to represent him in a divorce case."**

It is therefore apparent that in order to be able to answer questions, a knowledge structure must possess, in addition to the adequate frames, a rich interconnections network between them.

The parsimony principle on its side dictates that questions should be answered by involving the least possible number of assumptions. In particular arbitrar and otherwise unnecessary assumptions should be chased away. One characteristic of such assumptions is (always ?) to be weakly interconnected with the rest of the knowledge.

Thus a very alluring characterization of a good structure in episodic memory would be ***one which presents the maximum ratio of links per frame.*** And from now on, when we speak of a good interpretation, this will be indeed the objective we have in mind.[1]

Of course this definition needs also further refinements. They will come later with the study of a concrete scheme to achieve this global goal. From now on also, and in order to be coherent with other researchs on massively parallel processing [Packard, 1985], we will call this objective the **fitness function**. (Because this is the function that memory structures must strive to perform if they are to survive among competitors.)

In this view, intelligent questions are the ones that try to select the right frame or macro-frame (like macro-molecule) among a set of candidates claiming a high potential for connections with the existing structure.

We will try now to show what processes at the local level could, once working in parallel, realize the fitness function, and possibly tend at the same time to provide intelligent queries and guesses as well.

3.3- The basic model.

3.3.1- Three characteristics.

The previous section has put large constraints on the type of model we are looking for. We now want to add three new requirements that should further reduce the space of possibilities.

The first one is in fact a re-statement at the local level of the fitness function defined in 3.2. ***If each pending link of each frame, considered as an autonomous*** entity ***tries to establish a connection with the most connected candidate frame, and if each isolated frame is then discarded, then it ensues as a consequence that the whole memory will present the maximum***

[1] But we must not forget that this intuitive characterization might turn out to be incorrect. At this stage only experimentation can tell. So even if we are going to take it as the goal to be achieved in the rest of this paper, should further data seem to invalidate our approach, this would be a potential cause for failure, and would therefore need to be re-examined ~~at~~ in the light of the new findings.

ratio of links per frame which is our objective.

The second necessary feature of a re-interpretation model is ***perturbation.*** That is apparently random breaks of connections and modifications of data even during periods when there is no input to the system. This is a necessity for at least three reasons : first, to avoid that the system remains locked in *"local minima"* i.e. in a certain interpretation which although seemingly in accordance with the data from the world is less parcimonious than another one, second, to overcome monotonicity, if nothing is broken or changed then there is not much place for non-monotonicity, and third, to allow the system to *re-interpret* its data even *in the absence of new inputs*.

This feature is a crucial one, it is also exacting a toll. It means that the memory will be "dissipative", that is, will tend to forget certain things over the time and will need a more or less continuous flow of informations from the outer world.

Finally the last desirable characteristic is that the system commits itself to ***only one interpretation at a time.*** There are obvious functional reasons, we don't want a system giving ambiguous or cryptic answers! It corresponds also to what is observed in humans. In the following famous picture, one can see alternatively the silhouette of a young "elegante" or the face of an old woman, but never both at once. That means that we won't have several attributes at the same time for one slot, and that the time change from one interpretation to another must be short compared to the rate of incoming informations, and choices to be made.

3.3.2- Proposition for a solution.

The following model is offered as a possible scheme for a solution. The previous sections intended to constrain the space of choices for the determination of the local processes which form the basis of the model. Unfortunatly, there is no methodology available as now to determine the set of necessary and sufficient conditions to be verified by local processes in order to achieve a global performance. We can only try to define some necessary or at least useful conditions, and rely on our intuition and on guess and test work for the rest. That means in particular that, even in the case of a successful implementation, there is no guarantee that it should work on the whole set of possible situations, and still less that it is an optimal solution or is close to it.

The mechanism described here focuses on links. A link, which connects a given slot of a frame to another piece of data : another frame or a simple data (like a name), is characterized by a ***belief coefficient k*** ($0<k<1$) which corresponds to the confidence we put in the established link. For instance in analyzing **"John wrote a letter to Mary"**, we can be reasonably confident that the slot *sender* of the frame LETTER is filled with the role **John.** The belief coefficient of this link should accordingly be high, as opposed to the situation : **"John decided to write a letter to Mary"**, where it is not so sure that John actually did it, or even that any letter to **Mary** was eventually sent. The determination and computation of these belief coefficients are domain-dependent and will not be studied further here.

For our purpose the k coefficient determines the ***stability of the connection.*** That is we consider that a connection with a higher k will be more stable than with a lower one.

That means that at any time a connection may "decide" to "uncommit" itself and look for a candidate, and that the frequency of these "infidelities" will be determined by the attached coefficient k.

A candidate is any slot or frame which specifications correspond to the ones looked for by the free connection. For instance, let's suppose that the link relating **John** to ***sender*** in the previous example uncommits itself. Then any candidate should be for instance of the class "HUMAN-BEING" or "INSTITUTION", but not a GOAL or a SCENARIO.

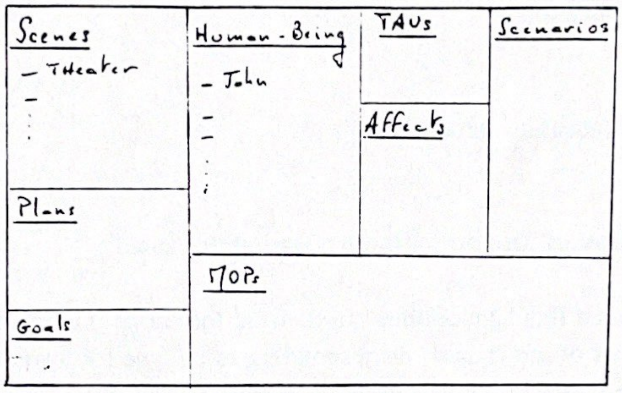
Let's further precise what we mean by ***disengaged link.*** By that we don't mean that the connection breaks right away, but rather that it behaves like someone looking for a new job, by looking through the adds without quitting its current employment until a new place is found. Furthermore, we admit that the new place may be the old one. This prevents from the lack of answers the system could give that would otherwise occur if the uncommitted link was actually free, "pending in the void".

Now how does function the "adds service" ?

First, each frame or piece of data is characterized by an ***"Influence" weight*** Which is a function of both the certainty attached to it when it was first instantiated in Episodic Memory i_0 and its ***connectivity*** with other pieces of knowledge in Episodic Memory : $I=f\,(I_0,d^0$ *of connectivity of* $i)$. For instance $I_i=I_{0_i}+\sum_{s,j} k_{i_s j}\,.\,I_j+\sum_{s,j} k_{j,i}\,.\,I_j$ where i_s stands for a particular slot s of the frame i and j for the frame linked to i_s. Thus the summation is over all the links from i, plus the links going to it.

The degree of connectivity is directly related to the number of links established with other pieces of data.

Second, we have a "Classified advertisements" board organized around classes of frames and data such as:



"Classified advertisement" board

And each frame or data sends to its corresponding part of the board signals at randomly distributed times t with a distribution depending upon their influence I, i.e. an "influent" frame will send more signals on average to the board than a less influent one.

Now when a connection disengages itself, like the *sender* slot of LETTER, it begins to look at the board for a signal coming from an adequate candidate, here in the "Human-being" or "Institution" part of the board, and as soon as it reads such a signal it unconnects itself, this time for good, from the previous site, and connects to the selected candidate (we suppose that the signal is accompanied by the name of the candidate), with a new strength k, with the calculation of which we don't concern ourselves here.

This is the basis of the mechanism proposed here. It is of course arbitrary. It may also need some additional hypotheses, such as a decreasing "activity" of the frames with time, to avoid jamming the board.

3.3.3- Mathematical analysis of the model.

□ Meantime of commitment of a link l.

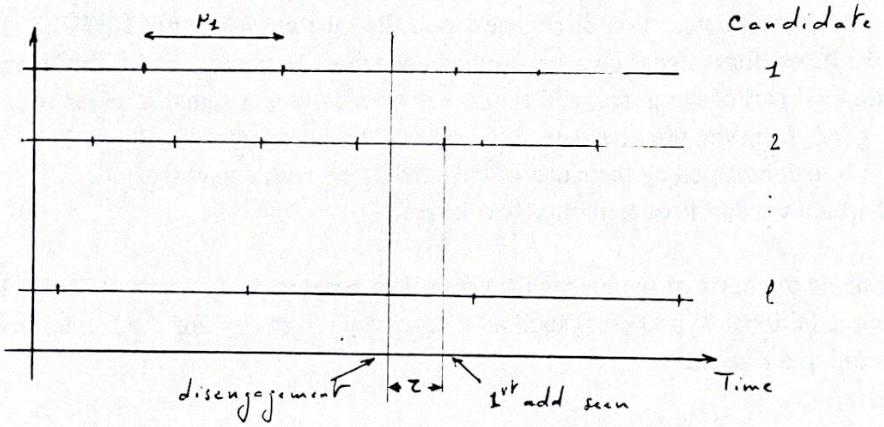
It depends on the strength of the connection : $t_c = f(\lambda \frac{k_l}{1-k_l})$, where λ is a universal constant chosen for the system, and f is a monotonically increasing function to be chosen accordingly to the domain of application.

If $k_l = 0$, then $t_c = 0$
$k_l = 1$, then $t_c = \infty$, (absolute certainty)

□ Meantime of "uncommitment" (waiting for an ad).

Here we suppose that a link "uncommits" itself from the frame it is connected to and then begins to watch the part of the "board" corresponding to its *type* for a signal from any candidate including the frame it is still connected to. This ensures that a link is always connected, and changes its destination only when it has found a new appropriate candidate.
All frames send ads : an impulse with their name, to the "board" with a distribution $\phi(t)$ related to their own *Influence*, with a mean frequency : $f = \alpha I_i$, α being a constant of the system.

Let's say that l candidates exist. We have the following situation :



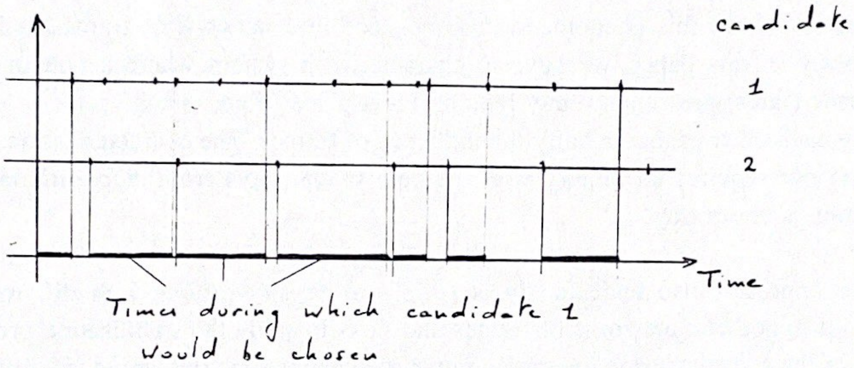
If there was only one candidate, the mean waiting time would be : $\frac{1/f_i}{2} = \frac{\mu_i}{2}$, where μ_i stands for the meantime between two signals from frame i.

For two candidates i and j : $\tau = \frac{\frac{1}{f_i + f_j}}{2} = \frac{\mu_i \mu_j}{2(\mu_i + \mu_j)}$.

For l candidates : $\tau = \frac{\prod_i \mu_i}{2\sum_i \prod_{i \neq j} \mu_j}$

□ Probability that the i^{th} candidate be chosen after a link l has disengaged.

If there are two candidates, we would have the following situation :



and the probabilty of the candidate i to be chosen is related to the mean waiting time by the relation :

$$P_i = \frac{1}{1 + \mu_l / \mu_k}$$

If several candidates are competing with i then :

$$P_i = \frac{1}{1 + \frac{\textit{meantime before signal of } i}{\textit{meantime before any other signal}}}$$

$$P_i = \frac{1}{1 + \frac{\sum_{k \neq i} \prod_{m \neq k} \mu_m}{\prod_{m \neq i} \mu_m}}$$

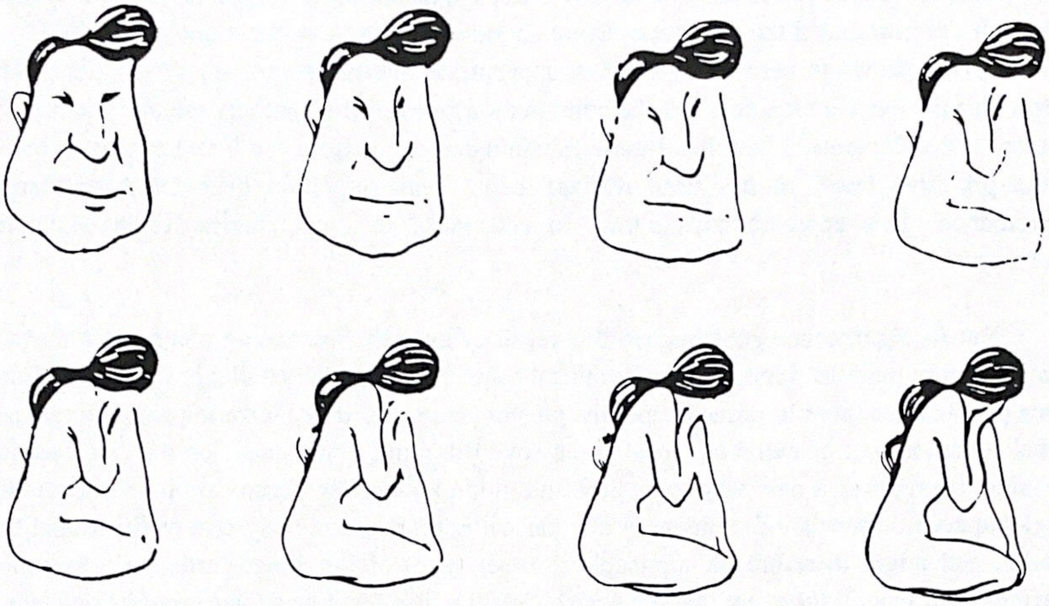
Now that we have obtained the equations governing the behaviour of individual links, the difficulties begin with the study of the global behaviour of the system. It is indeed not a trivial task to put in equations the passage from one state to another of the system.

Ideally, we would like to determine the ***stability of the system***, and in particular, we would expect that among the set of all possible states, only a small number of "small" sub-sets are attractive and stable. They would correspond to our idea of an interpretation, that is of a stable overall structure of knowledge with only small fluctuations quickly resorbed from place to place.

In order to exhibit this phenomenon taking place in a network of frames and links as defined previously in this paper, we have to characterize a system where a potentially large number of agents (links) are concurrently looking for a place to go, changing their destination and influencing each other as they modify the influence of frames. The characterization of such a collective behaviour requires an inquiry with the help of the tools from ***stochastic modelling***. This investigation is under way.

Another approach also under study is simply to do *simulations* with different initial configurations of hypothetic networks of frames and links to study the stabilisation process and the sensitivity of the stable states to externally imposed perturbations (incoming informations).

One phenomena which should show up, is an ***hysteresis*** in the interpretations transitions. That is that there should be an asymmetry between the transitions from one state i to state j and the inverse transition. More accurately the switching from one state to the other shouldn't happend at the same place, as is apparent in the following sequence of drawings where one is invited to switch from a face interpretation to a woman figure interpretation and vice-versa if one goes back along the sequence. In the first case the switching may typically occur around the 6^{th} drawing, in the second, around the 4^{th} may be, pointing out an hysteresis phenomena. The mechanism described in this paper is expected to give rise to the same behavior.



In this context an interesting conjecture could also be checked, namely that ***the interpretations reached by the system subjected to a certain environment would differ sensibly depending on the relative speed of the internal processes compared with the rate of incoming stimuli.*** I have the strong intuition that one would observe much more radical re-interpretations in the case of a "slow" understanding system, at the price obviously of a certain "laziness" and most likely also of some forgettings since one retains an information only to the extent that it can be related to an existing structure of knowledge.[1]

Finally, as was noted when we first introduced the fitness function, this rule and its translation into the local level could turn out to be ill-adapted to the real-world, that is, to lead to re-interpretations that are discomfirmed by the reality perceived by the system. It would then seem logical ***to apply the fitness function to itself !*** Indeed the fitness function is in some way also an interpretation or model of the world. Therefore it should fall under the same juridiction as any other interpretation, provided it is stated in the same terms. Here however lies the heart of the difficulty, and nothing has yet been done in this research to solve this problem. But this is a fascinating idea altogether and certainly worth more considerations in the future.

[1] Here we assume the existence of a short-term memory where items sink into oblivion after a short period of time whether or not they have been related to the existing Episodic memory in the meantime.

Conclusions

This research centers on how to find better explanations of the perceived universe by a system. It has introduced the concept of using a kind of meta-rule or meta-constraint : the fitness function, that allows to zero in on the few appropriate interpretations among all the possible ones under the current situation and the combination rules of the symbols making the primitive representation elements. When this fitness function can result from the interaction of processes at the primitive level, as has been realized here, re-interpretation becomes a spontaneous phenomenon. This approach thus allows to get rid of the cumbersome Truth Maintenace mechanisms.

But its significance goes beyond this result or even the interesting conjectures that arise from it, and is twofold. First, on the "technical" side, this research challenges Artificial Intelligence people once more to probe deeper the phenomenon of self-organization and emergence of global behaviour out of numerous local processes. But more importantly, on the "methodological" side, it proposes a new way to control and guide knowledge organization and acquisition, by global considerations independent of any particular representation system or functional taxonomies, and might therefore be applicable to other types of Learning, particularly to concept Learning, and that all the more that the control function itself can be subjected to its own action, giving rise to an interesting and hopefully convergent bootstrapping process.

This is at least the hope and object of current research.

Aknowledgements.
This research was sponsored by a grant from AEROJET N^o 442510-57180-2.

REFERENCES

- Ronald J. BRACHMAN : **"I lied about the trees"**, The AI magazine, Fall 85, pp.80-93.

- Ronald J. BRACHMAN, Hector J. LEVESQUE : **"A Fundamental Tradeoff in Knowledge Representation and Reasoning"**, in Readings in Knowledge Representation, Morgan & Kaufmann, 1985, pp.41-70

- E. CHARNIAK, C. RIESBECK, D. McDERMOTT : **"Data Dependencies"** in Artificial Intelligence Programming, Lawrence Erlbaum Associates, 1980,pp.193-226.

- Allan COLLINS, John S. BROWN, Kathy M. LARKIN : **"Inference in text understanding"** in Theoretical Issues in Reading Comprehension, Spiro, Bruce, Brower eds., LEA, 1980. pp.385-407.

- Jon DOYLE : **"A Truth Maintenance System"**, Artificial Intelligence 12, (1979), 231-272.

- Michael G. DYER : **In depth-Understanding**, MIT Press, 1983.

- Richard H. GRANGER : **"When expectation fails : Toward a self-correcting inference system."** Proceedings of the first National Conference on Artificial Intelligence, Stanford, Cal., 1980.

- Johan de KLEER, Jon DOYLE, G. L. STEELE, G. J. SUSSMAN : **"Explicit Control of Reasoning"** in Artificial Intelligence : An MIT perspective, vol.1, P. H. Winston and r. H. Brown (ed.), MIT Press, 1979,pp.33-92.

- Johan de KLEER : **"Choices without backtracking."**, Proceedings of the NCAI-84.

- Douglas B. LENAT : **"AM : An Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search"**, Ph.D. dissertation, Stanford University, 1976.

- Erik T. MUELLER : **"Toward a computational theory of human day-dreaming."**, Proceedings of the seventh Annual Conference of the Cognitive Science Society, Irvine, Cal. 1985.pp.120-128.

- Paul O'RORKE : **"Reasons for beliefs in understanding : applications of non-monotonic dependencies to story processing."**, Proceedings of the NCAI-83, pp.306-309.

- R. SCHANK : **Dynamic Memory**, Cambridge University Press, 1982.

- R.G. SMITH, H.A. WINSTON, T.M. MITCHELL, B.G. BUCHANAN : **"Representation and use of explicit justifications for knowledge base refinement"**, Proceedings of the IJCAI-85, pp.673-680.

- Patrick H. WINSTON : **"Learning Structural Descriptions from Examples"**, The Psychology of Computer Vision, Winston P.H. (ed.), Mc Graw Hill, New-York, ch5, 1975.