A Memory Model for Cognitive Agents

Guilherme Bittencourt

Departamento de Automação e Sistemas
Universidade Federal de Santa Catarina
88040-900 - Florianópolis - SC - Brazil
e-mail: gb@das.ufsc.br

Abstract. In this paper we present a memory model that, syntactically, consists of logical theories but whose semantic description includes, besides the usual truth value assignments, what we call emotional flavors, that correspond to the state of the agent’s body translated into cognitive terms. The combination between logical theories and emotional flavors allows the agent to memorize relevant theories that can be used for reasoning. These theories are represented in a specific format – prime implicants/implicates – which is enriched with annotations that explicitly store the internal relations among their literals. Based on this representation, the details of the proposed memory mechanism is described and some arguments to support its psychological plausibility are presented.

Keywords: Artificial Intelligence, Cognitive Science, Learning, Memory, Reasoning, Situated cognition, Knowledge representation, Logic.

1 Introduction

Roughly, cognitive models in Artificial Intelligence can be either based on language, the “sense-model-plan-act” paradigm [1] [2], or on action, the “behavior-based artificial creatures situated in the world” paradigm [3]. The language side uses symbolic knowledge representations, inference engines, planning (SAT) and reasoning about action algorithms and presents properties such as: compositionality, modularity, explicit semantic, etc. The action side uses reactive mechanisms that can be of different types: subsumption architecture, neural nets, fuzzy control, evolutionary computation, ant algorithms, reinforcement learning, and presents complementary properties such as: self-organization, embodiment, learning capability, etc. Although different combinations of these two paradigms have been tried, there is no consensus about how to smoothly integrate them.

In [4], we introduced an embodied model for higher-level cognition that support learning and reasoning mechanisms using logical expressions represent by their prime forms. This paper defined a generic model for a cognitive agent and proposed a computational architecture, coherent with this model, that is based on the following fundamental hypothesis:

– Cognition is an emergent property of a cyclic dynamic self-organizing process [5] [6] based on the interaction of a large number of functionally independent units of a few types [7].

Published in Angelo Loula & João Queiroz (Eds), Advances in Modeling Adaptive and Cognitive Systems. UEFS, 2010.
– Any model of the cognitive activity should be epistemologically [8] compatible with the Theory of Evolution. That applies not only to the “hardware” components of this activity but also to its “psychological” aspects [9].

– Learning and cognitive activity are closely related and, therefore, the cognitive modeling process should strongly depend on the cognitive agent’s particular history [10].

This paper maintains all the 1997 hypothesis and we describe in some detail the memory mechanism that support learning and reasoning in the model.

2 Preliminaries

Let \( P = \{ p_1, \ldots, p_n \} \) be a set of propositional symbols and \( \{ L_1, \ldots, L_{2n} \} \) the set of their associated literals, where \( L_i = p_j \) or \( L_i = \neg p_j \). A clause \( C \) is a disjunction [11] of literals: \( C = L_1 \lor \cdots \lor L_k \) and a dual clause, or term, is a conjunction of literals: \( D = L_1 \land \cdots \land L_k \). In the sequel, conjunctions and disjunctions of literals, clauses and terms are treated as sets.

Given a propositional logic language \( \mathcal{L}(P) \) and an ordinary formula \( \psi \in \mathcal{L}(P) \), there are algorithms for converting it into a conjunctive normal form (CNF) and into a disjunctive normal form (DNF) (e.g., [12], [13], [14]). The CNF is defined as a conjunction of clauses, \( \text{CNF}_\psi = C_1 \land \cdots \land C_m \), and the DNF as a disjunction of terms, \( \text{DNF}_\psi = D_1 \lor \cdots \lor D_w \), such that \( \psi \equiv \text{CNF}_\psi \equiv \text{DNF}_\psi \). Throughout the paper we will be dealing with one special type of CNF/DNF representation: prime implicates/implicates.

A clause \( C \) is an implicate [15–17] of a formula \( \psi \) iff \( \psi \models C \), and it is a prime implicate iff for all implicates \( C' \) of \( \psi \) such that \( C' \models C \), we have \( C \models C' \), or syntactically [18], for all literals \( L \in C \), \( \psi \not\models (C - \{ L \}) \). We define \( \text{PI}_\psi \) as the conjunction of all prime implicates of \( \psi \), clearly \( \psi \equiv \text{PI}_\psi \).

A term \( D \) is an implicant of a formula \( \psi \) iff \( D \models \psi \), and it is a prime implicant iff for all implicants \( D' \) of \( \psi \) such that \( D \models D' \), we have \( D' \models D \), or syntactically, for all literals \( L \in D \), \( (D - \{ L \}) \not\models \psi \). We define \( \text{PI}_\psi \) as the disjunction of all prime implicants of \( \psi \), again \( \psi \equiv \text{PI}_\psi \). This normal form is also known as Blake canonical form.

In propositional logic, implicates and implicants are dual notions, in particular, an algorithm that calculates one of them can also be used to calculate the other [14]. Alternatively, prime implicates and implicants can be defined as special cases of CNF (or DNF) formulas, exactly those that consist of the smallest sets of clauses (or terms) that represent the theory and are closed for inference, without any subsumed clause (or term). In general, (prime) implicates and (prime) implicants are allowed to contain a literal and its negation. In this case, they will be, respectively, tautologic or contradictory. In the sequel, when we refer to (prime) implicates and (prime) implicants we assume that they do not contain such pairs of tautologic or contradictory literals.
3 Framework

Before we begin the model description itself, it is necessary to define the abstract framework in which the model is based. In the sequel, italic expressions should be understood as technically defined terms in the model and “quoted” expressions are intended to evoke a certain “meaning” whose precise definition is out of the scope of the model.

Consider a cognitive agent immersed in an unknown environment represented as a set of primitive propositional symbols \( P = \{ p_1, \ldots, p_n \} \). In a first approximation, the only property of each propositional symbol is its truth value. The states of the environment are defined as the set of all possible truth assignments to this set of propositional symbols \( A_P \). We also suppose that, as time goes by, the environment drifts along the possible states through flips of the primitive propositional symbols truth values.

From the agent point of view, the environment, including the agent itself, appears as an unbound temporal sequence of assignments \( \ldots, \epsilon_{i-1}, \epsilon_i, \epsilon_{i+1}, \ldots \) where \( \epsilon_i \in A_P, \epsilon_i : P \to \{ \text{true}, \text{false} \} \) are semantic functions that map propositional symbols into truth values. The primitive propositional symbols can be of two kinds: controllable and uncontrollable. Roughly, uncontrollable symbols correspond to perceptions and controllable ones to actions. Perceptions may include internal perceptions, i.e., internal properties of the agent that are “felt”, such as proprioceptive information [21], and actions may include orders to the agent body. Both controllable and uncontrollable symbols are “neutral”, in the sense that, a priori, the agent is indifferent to which semantic values (true or false) they assume.

Primitive propositional symbols can be combined into theories. Each theory is simply a well formed formula of propositional logic and its semantics is derived from the truth values of its components as usual. Because the proposed cognitive agent is a computational one, the particular adopted syntactical representation can have an important impact on the computational properties of the cognitive mechanism, such as efficiency, modularity and reuse potential. Theories can be used as components of other theories, i.e., it is possible to associate an abstract propositional symbol with a theory and to use it as an ordinary proposition symbol in another proposition. Let \( Q = \{ q_1, q_2, \ldots \} \) be the set of all abstract propositional symbols. Theories in which abstract propositional symbols occur are called abstract theories.

In order to “embody” the agent, we assume that the state of the agent’s “body” can be perceived through emotional flavors. An emotional flavor has two complementary aspects: from the cognitive point of view it can be true or false and from the agent’s body point of view it can be, in a first approximation, either “good” or “bad”, in the sense that the agent has the motivation that good emotional flavors be true and bad ones false. From the cognitive point of

\[1\] This property could be generalized to allow fuzzy [19] or multiple values [20]. The adopted logic could also be first-order or even a higher-order logic, instead of propositional logic.
view, any motivation is directly or indirectly derived from the desire to control the truth values of emotional flavors and therefore, beyond their truth values, theories only have “meaning” with respect to the emotional flavors to which they are associated.

To satisfy this motivation, the cognitive agent should be able to learn the relevant relations between specific emotional flavors and theories built up from primitive propositional symbols. Once a theory is known to be true whenever a given “good” emotional flavor is true, this theory can be associated with an abstract propositional symbol that becomes the cognitive counterpart of the emotional flavor. Using this theory, the agent can “rationally” act on the truth values of the theory’s controllable symbols in such a way that the theory truth status is preserved when the values of the uncontrollable symbols change. This relation between an emotional flavor and an abstract propositional symbol (and its learned theory) is called a thought. Now, we turn to the concrete representation of a thought.

4 Quantum Representation

Given a formula \( \psi \), represented by a conjunctive normal form \( CNF_\psi \) and by a disjunctive normal form \( DNF_\psi \), we introduce the concept of a conjunctive quantum, defined as a pair \((L, F_c)\), where \( L \) is a literal that occurs in \( \psi \) and \( F_c \subseteq CNF_\psi \) is its set of conjunctive coordinates that contains the subset of clauses in \( CNF_\psi \) to which literal \( L \) belongs. A quantum is noted \( L F_c \). Dually, we define a disjunctive quantum as a pair \((L, F_d)\), where \( L \) is a literal that occurs in \( \psi \) and \( F_d \subseteq DNF_\psi \) is its set of disjunctive coordinates that contains the subset of terms in \( DNF_\psi \) to which literal \( L \) belongs. The rationale behind the choice of the name quantum is to emphasize that we are not interested in an isolated literal, but that our minimal unit of interest is the literal and its situation with respect to the theory in which it occurs.

**Example 1.** Consider the theory \( \psi \) given by the following CNF:

\[
0 : (p_0 \lor \neg p_1 \lor p_2) \land \\
1 : (p_1 \lor \neg p_2 \lor p_3) \land \\
2 : (p_0 \lor p_1 \lor \neg p_2) \land \\
3 : (\neg p_1 \lor p_2 \lor p_3) \land \\
4 : (p_0 \lor p_1 \lor p_2)
\]

The literals that occur in \( \psi \) can be represented by the following set of conjunctive quanta:

\[
\{p_0^{(0,2,4)}, \neg p_0^{\{\}}, p_1^{(1,2,4)}, \neg p_1^{\{0,3\}}, p_2^{\{0,3,4\}}, \neg p_2^{\{1,2\}}, p_3^{\{1,3\}}, \neg p_3^{\{\}}\}
\]

\( ^2 \) To simplify the notation, the sets of conjunctive coordinates contain clause numbers instead of the clauses themselves. These numbers do not mean anything with respect to the representation, they should be seen as simple pointers to the real clauses.
The quantum representation can be used to characterize the $PI_\psi$ or the $IP_\psi$ of a formula $\psi$, given that we have available, respectively, either a $DNF_\psi$ or a $CNF_\psi$ that represents the formula.

Let $L_1 \wedge \cdots \wedge L_k$ be a term. It can be represented by a set of conjunctive quanta, $D = L_1^{F_1} \wedge \cdots \wedge L_k^{F_k}$. $D$ will represent an implicant of $\psi$, if $\cup_{i=1}^{k} F_i^{c} = CNF_\psi$ and $L_i \not\equiv \neg L_j$, $i, j \in \{1, \ldots, k\}$, for some $CNF_\psi$ representing $\psi$. In words, $D$ contains at least one literal that belongs to each clause in $CNF_\psi$, spanning a path through $CNF_\psi$, and no pair of contradictory literals. To represent a prime implicant, a quantum set $D$ have to satisfy a non redundancy condition, i.e., each of its literals should represent alone at least one clause in $CNF_\psi$. To define this condition, we introduce the notion of exclusive coordinates. Given a quantum set $D$ and a quantum $L_i^{F_i} \in D$, the exclusive conjunctive coordinates of $L_i^{F_i}$ in $D$ are defined by $\hat{F}_i = F_i^c - \cup_{j=1,j \neq i}^{k} F_j^c$, i.e., the clauses in the set $F_i^c$, to which no other literal represented in $D$ quanta belongs. Using this notion, the non redundancy condition, that guaranties that $D$ represents a prime implicant, can be written as: $\forall i \in \{1, \ldots, k\}, \hat{F}_i \neq \emptyset$.

Dually, a clause $L_1 \vee \cdots \vee L_k$ represented by a set of disjunctive quanta, $C = L_1^{F_1} \vee \cdots \vee L_k^{F_k}$, with no pair of tautological literals allowed, is an implicate of $\psi$, if $\cup_{i=1}^{k} F_i^{d} = DNF_\psi$, for some $DNF_\psi$ representing $\psi$. Again $C$ represents a prime implicate if it satisfies the non redundancy condition, expressed by $\forall i \in \{1, \ldots, k\}, \hat{F}_i \neq \emptyset$, where $\hat{F}_i = F_i^d - \cup_{j=1,j \neq i}^{k} F_j^d$ is the set of exclusive disjunctive coordinates of $L_i^{F_i}$ in $C$.

Example 2. Consider the theory $\psi$ introduced in example 1. The set:

$$D = \{p_0^{[0,2,4]}, p_1^{[1,2,4]}, p_2^{[0,3,4]}\}$$

represents an implicant of $\psi$, because the union of the conjunctive coordinates associated with its quanta is equal to the set of clauses in $CNF_\psi$:

$$\{0, 2, 4\} \cup \{1, 2, 4\} \cup \{0, 3, 4\} = \{0, 1, 2, 3, 4\}$$

The exclusive conjunctive coordinates of the quanta in $D$ are given by:

$$D = \{p_0^{[1]}, p_1^{[1]}, p_2^{[1]}\}$$

The fact that $p_0$ has empty exclusive coordinates indicate that $D$ is not a prime implicant. Actually, the prime implicants of theory $\psi$, with respect to the $CNF$ of example 1, are the following terms, where the exclusive conjunctive coordinates are in boldface:

$$
\begin{align*}
0 : p_0^{[0,2,4]} & \vee p_3^{[1,3]} \\
1 : p_0^{[0,2,4]} & \vee \neg p_1^{[0,3]} \vee \neg p_2^{[1,2]} \\
2 : p_1^{[1,2,4]} & \vee p_2^{[0,3,4]}
\end{align*}
$$

3 In an abuse of notation, we use the conjunction symbol $\wedge$ to separate the members of the set, instead of the more usual comma, and do not use the usual delimiters $\{\}$. The aim is to simplify the understanding of the set of quanta as a representation of a term.
Given a theory $\psi$, it is possible to determine the sets of conjunctive and disjunctive quanta that, respectively, represent $IP_\psi$ with respect to $PI_\psi$ and $PI_\psi$ with respect to $IP_\psi$. This particular quantum representation of a theory is special because it is unique, in a sense to be clarified soon. We call $PIP$ the set of all such pairs of prime representations:

$$PIP = \{(PI_\psi, IP_\psi) \mid \psi \in \mathcal{L}(X)\}$$

**Example 3.** Consider the theory $\psi$ introduced in example 1. Below we present the PIP pair associated with this theory:

$$\begin{pmatrix}
0 : p_1^{[2]} \lor \neg p_2^{[0]} \lor p_3^{[1]} \\
1 : \neg p_1^{[0]} \lor p_2^{[2]} \lor p_3^{[1]} \\
2 : p_0^{[0,1]} \lor p_2^{[2]} \\
3 : p_0^{[0,1]} \lor p_1^{[2]}
\end{pmatrix}$$

Note that the coordinates of the quanta in the terms in the new $IP_\psi$ now refer to $PI_\psi$ and not to the arbitrary CNF representation introduced in example 1 and therefore they are completely different. The order of the terms has also been changed, highlighting that the numbers that appear in the coordinates of the terms are just pointers to clauses, without any special meaning with respect to the representation.

## 5 Prime Form Manipulation

The main hypothesis underlying the proposed memory model consists in restricting the syntactical form of the theories that participate in thoughts in such a way that these theories are always represented using prime normal forms. In particular, these prime normal forms are represented using the quantum representation, introduced above. This choice is based on several properties of these normal forms that make them suitable for knowledge representation.

Propositional theories can be separated into different classes, called *families*, according to an equivalence relation that consider equivalent any two theories that differ only with respect to renaming of their propositional symbols and to the group of permutations and complementations. The prime representations in $PIP$ are unique in the sense that, each pair represents one of these families of structurally identical theories.

The problem of determining the number of families for a given number of propositional symbols has been studied in the more general context of Boolean functions [22] for a long time.

To have a feeling about the relation between the number of theories and that of families, consider a set $P$ containing $n$ propositional symbols, there are $2^n$
different truth value assignments for these \( n \) symbols. Any theory build up with symbols in \( P \) will be true in some subset of these \( 2^n \) assignments, therefore the number of non trivial theories is given by\(^5\):

\[
2 \sum_{i=1}^{2^{n-1}} C_i^n
\]

<table>
<thead>
<tr>
<th>( n )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>( \sum )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 all</td>
<td>4</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>pip</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>3 all</td>
<td>8</td>
<td>28</td>
<td>56</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>162</td>
</tr>
<tr>
<td>pip</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>4 all</td>
<td>16</td>
<td>120</td>
<td>560</td>
<td>1820</td>
<td>4368</td>
<td>8008</td>
<td>11440</td>
<td>12870</td>
<td>39202</td>
</tr>
<tr>
<td>pip</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>19</td>
<td>27</td>
<td>50</td>
<td>56</td>
<td>58</td>
<td>221</td>
</tr>
<tr>
<td>5 all</td>
<td>32</td>
<td>496</td>
<td>4960</td>
<td>35960</td>
<td>201376</td>
<td>906192</td>
<td>3365856</td>
<td>10518300</td>
<td>2448023842</td>
</tr>
<tr>
<td>pip</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>47</td>
<td>131</td>
<td>472</td>
<td>1326</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
</tbody>
</table>

Table 1. Number of Theories.

Table 1 shows the number of theories and the number of \( PIP \) pairs obtained for \( 2 \leq n \leq 5 \), with the last line incomplete. It should be noted that, although both numbers, of general theories and of \( PIP \) pairs, grow very fast, the latter grows much slowly than the former. Each line in the table represents the number of theories and the number of \( PIP \) pairs, or families, obtained for a given number \( n \) of propositional symbols. The theories in the same columns have the same number of models. Each of these families has a certain population of theories not uniformly distributed. For instance, theories with \( n = 3 \) have the population profiles shown in table 2.

Given both prime representations of a given formula \( \psi \), the prime representations of its negation can be obtained directly\(^6\) because the following equalities are valid:

\[
PI_{\neg \psi} = \overline{PI}_\psi \quad IP_{\neg \psi} = \overline{PI}_\psi
\]

\(^5\) The theories with more than \( \frac{2^n}{2} \) models are equivalent to the negations of the theories with less than \( \frac{2^n}{2} = 2^{n-1} \) models.

\(^6\) We note \( \overline{\neg \psi} \) the formula \( A \) with the truth values of all its literals flipped.
As stated in Section 3, one of the main characteristics of the proposed cognitive model is that propositional symbols can be combined into theories, represented by PIP pairs and associated with abstract propositional symbols that can be used in other theories. This property presupposes mechanisms to revert this decomposition process to obtain a PIP representation that uses only primitive propositional symbols. The relationship between emotional flavors and such an expanded theory may be clearer than with the compact theory where the emotional flavors are entangled by the abstraction process. The next subsections treat this problem.

5.1 Expanding a Theory

This mechanism can be used to obtain an expanded version of a theory when its abstract propositional symbols are replaced by their definitions. Consider a theory, expressed by a PIP pair – \((\Pi IP)\), that contains an abstract propositional symbol \(q\) among its propositional symbols. This abstract propositional symbol is also associated with a PIP pair: \((\Pi q, IP q)\). The following algorithms allow to incorporate the definition \((\Pi q, IP q)\) into the PIP pair of the original theory.

\[
\text{Expand}((\Pi IP), (\Pi q, IP q))
\]

1. **For all** clauses \(C \in PI\) do
   (a) if \(q \in C\) then replace \(C\) by \(\text{Combine}(C - \{q\}, \Pi q)\)
   (b) if \(\neg q \in C\) then replace \(C\) by \(\text{Combine}(C - \{q\}, IP q)\)

2. **For all** terms \(D \in IP\) do
   (a) if \(q \in D\) then replace \(D\) by \(\text{Combine}(D - \{q\}, IP q)\)
   (b) if \(\neg q \in D\) then replace \(D\) by \(\text{Combine}(D - \{q\}, IP q)\)

3. **Return** \((\Pi IP)\)

As the procedure \text{Combine} is dual with respect to either PI or IP, we adopt the following notation that allows us to present it just once. Let \(P = \{B_1, \ldots, B_k\}\); and \(B \in \{C, D\}\) and \(P \in \{\Pi, IP\}\), respectively.

\[
\text{Combine}(B, P)
\]

1. \(P \leftarrow \{\text{Simplify}(B_1 \cup B), \ldots, \text{Simplify}(B_k \cup B)\}\)
2. \( P \leftarrow \text{Simplify}(P) \)
3. \textbf{Return } \( P \)

The procedure \textbf{Simplify} is intended to solve minor syntactical details (e.g., repeated or contradictory literals in clauses/terms, update coordinates, etc). To fix these details is facilitated by the used of coordinates.

### 5.2 Decomposing a Theory

Expanding a theory, as described in section 5.1, is a \textit{direct problem}, in the sense that knowing the PIP representation of the theory to be expanded and of the PIP representation of all the abstract propositional symbols that occur in it and, recursively, of all PIP representations of the abstract propositional symbols until some primitive propositional symbol is reached. This process has just one solution.

On the other hand, decomposing a theory into a (hopefully) simpler theory, is an \textit{inverse problem}. This means that we must “choose” one or more of its propositional symbols, and all the tree below it, to be represented by a new abstract propositional symbols in such a way that when all the abstractions are expanded using the algorithm of Section 5.1, we obtain the original theory. This problem has infinite solutions, because of the completness of several sets of logical operators (e.g., \( \{\vee, \wedge, \neg\} \) or \( \{\vee, \neg\} \) or \( \{\rightarrow, \neg\} \) and therefore of several sets of PIP pairs with no more than two variables (see table 3).

Our goal is to decompose the theories in such a way that the resulting theory uses a minimum number of PIP pairs, and with a limited number of propositional symbols (typically 3 or 4). These decompositions are in principle arbitrary but, on the one hand, in an agent society these definitions, to be useful, should be consensual to allow the emergence of a communication language in which these definitions are shared, allowing learning by being told (remember, they are simply names, if their “meaning” is not shared they are useless...). On the other hand, these definitions have an important impact on the efficiency of the reasoning methods because they directly affect the size of the final PIP representation of the compressed theory. We assume that these decompositions and the underlying abstract theories are constructed by some kind of \textit{evolutionary computation} algorithm.

As usual in evolutionary computation, sensible decompositions of a given \textit{PIP} pair can be found by trial and error, but some heuristics can help to improve the efficiency of the process, for instance helping to choose the initial population. We present below some heuristics for finding sensible decompositions of a given \textit{PIP} pair; other heuristics are of course possible. The first heuristic is to search for repeated patterns of propositional symbols that appear in the pair’s \textit{IP}. Each such pattern is always a good candidate for an abstract propositional symbol with a single term in its \textit{IP}.

A second heuristic takes into account the fact that there are controllable and uncontrollable symbols and analyzes repeated configurations of symbols of the same type that appear in the pair’s \textit{IP}. If the number of configurations is
less than would be possible, given the involved propositional symbols, then to decompose these configurations into abstract propositional symbols may simplify the original proposition.

A third heuristic to abstract propositions can be obtained by exploring the structure of the prime implicants of a proposition. Let $IP_\psi = \{D_1, \ldots, D_k\}$ be the set of prime implicants of the proposition $\psi$. We define the relation $Resol(D_i, D_j, D_{ij})$ as the set of n-tuples of terms such that $D_{ij}$ is the resolvent of $D_i$ and $D_j$. A good candidate for a new abstract propositional symbol would be a proposition with the following set of prime implicants: $IP_{p_{new}} = [D_i - (D_i \cap D_j), D_j - (D_i \cap D_j)]$. If $D_i \cap D_j$ contains more than one literal, it can also be defined as a new abstract propositional symbol: $p_{\cap} = [D_i \cap D_j]$. These definitions would reduce the original set of terms $\{D_i, D_j, D_{ij}\}$ to a single term given by: $(p_{\cap}, p_{new})$.

6 Memory Structure

The memory contains thoughts. A thought $\tau$ is defined as a relation between an abstract propositional symbol (associated with a logical theory represented by a PIP pair) and an emotional flavor. This relation consists of three elements:

- A generic pair $\pi_\tau \in PIP$ with variable propositional symbols $V(\pi_\tau) = \{x_1, \ldots, x_k\}$.
- A set of propositional symbols $P_\tau = \{p_1, \ldots, p_k\}$ associated with an emotional flavor and with its cognitive counterpart, the abstract propositional symbol $q_\tau$.
- A mapping $\mu_\tau : P_\tau \rightarrow V(\pi_\tau)$ that associates with each $p_i$ a $x_i$ that occur in the PIP pair.

It should be noted that: (i) every thought is an operational recipe to control the truth value of an emotional flavor; (ii) each emotional flavor can be associated with different PIP pairs (different ways to control its value) and thus can participate in different thoughts; (iii) the PIP pairs are independent of the thought contents and can also participate in different thoughts.

Modeling the emotional flavors is out of the scope of the proposed model; to keep it simple each emotional flavor was just given a truth value and indirectly, a degree of activity that makes a thought that is evoked by an emotional flavor to become one of which the agent is aware. Nevertheless, emotional flavors are what the model is all about: the thoughts that are (best) learned are those that were (successfully, or dramatically unsuccessfully) associated with (very) active

\footnote{These three elements are analogous to the three “subjects” in the semiosis definition: “(...) an action, or influence, which is, or involves, a cooperation of three subjects, such as a sign, its object, and its interpretant; this tri-relative influence not being in any way resolvable into actions between pairs.” [23]}

\footnote{We explicitly avoid representing the emotional flavor in the formalism, because only its syntactical counterpart actually belongs to the proposed cognitive model.}
emotional flavors. On the other hand, each learned thought refines its associated emotional flavors, providing an extra way of controlling it and simultaneously imposing conditions on its satisfaction.

6.1 Memory Indexing

These thoughts are “indexed” according to two different dimensions: a semantic one, that associates thoughts that share the same emotional flavor, and a syntactical one, that associates thoughts that share the same PIP pair and at least one of its decompositions.

Although the memory contains thoughts, it does not need to contain the logical structure of the theories associated with the thoughts, only their “indexes”, as long as a “table” of known PIP pairs is available. The fact that the structure of the thoughts’ theories, the PIP pairs, can be shared by several thoughts reduces significantly the storage requirements, especially if abstract propositional symbols are defined in such a way that preferentially low dimension PIP pairs are used. Besides reducing the storage requirements, using low dimension PIP pairs increases the probability that different thoughts share the same PIP pair.

6.2 Memory Functioning

The “daylight” activity of the memory is to provide relevant thoughts to be used to control the active emotional flavors.

![Fig. 1. Reactive Level](image)

**Memory Optimization** The “dreaming” activity consists of organizing the structure of memorized thoughts in such a way that the “remembering” mechanism works effectively. The goal of the organizing mechanism is to find (or better, to evolve) sensible decompositions of theories that facilitate the storage, inference and communication of thoughts. The search for these decompositions is
What is it like to be BAT

Perception/Action

Memory of Thoughts

Emotion

Genetically determined by the emotions

Genetically determined by the environment

To be or not to be

name y

name z

name x

Thought

Cognition

Perception/Action

Emotions

Fig. 2. Instintive level

Fig. 3. Cognitive Level
done by some evolutionary algorithm. The detailed description of this algorithm is also out of the scope of the paper (see [24]).

A central element in any evolutionary algorithm is the fitness function that, in our case should take into account the number of propositional symbols that occur in each abstracted proposition. The smaller this number, the better is the decomposition policy. The idea would be to choose decomposition policies that simultaneously reduce the number of propositional symbols in the original proposition and that introduce the minimum number of new propositional symbols.

Up to this point, we have discussed the operations necessary to abstract and expand thoughts. These operations are used to optimize the structure of the memory contents. Once these best definitions are found for several different thoughts, the semantic “meaning” of these thoughts that belong to different emotional flavors domains but share analogous structural decompositions are related.

The goal of this optimization is to find and reuse common “properties” that are shared by theories. These common properties may lead to sensible metaphors [25]. We claim that metaphors are especially important in human intelligence because they allow the production of thoughts that entangle originally unrelated emotional flavors and therefore can lead to the transference of expertise between totally different domains, that is so characteristic of human intelligence.

To give an idea of how much storage such decompositions can spare, using only theories with two symbols, it is possible to represent all three symbol theories with at most three level decomposition trees. The same is true for 4 symbol theories: using only 3 symbol theories, those with 2 or less propositional symbol it is possible to represent all 221 4 symbol theories with at most three level decomposition trees.

<table>
<thead>
<tr>
<th>PIP pair</th>
<th>$T(i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: affirmation/negation</td>
<td>$x_0$</td>
</tr>
<tr>
<td>2: conjunction/disjunction</td>
<td>$x_0 \land x_1$</td>
</tr>
<tr>
<td>3: exclusive (or)</td>
<td>$(x_0 \lor x_1)\land(\neg x_0 \land \neg x_1) \lor \neg x_0 \land x_1$</td>
</tr>
</tbody>
</table>

Table 3. Table of PIP pairs with 2 propositional symbols.

Assume that $m = 2$, then the table of known PIP pairs contains only three entries (see table 3). The PIP pair of example 3, call it $q_1$, can be represented as: $q_1 = T(3)$, with $x_0 = p_0, x_1 = q_2$ and $q_2 = T(2)$, with $x_0 = p_1, x_1 = p_2$.

Consider again the PIP pair of example 3. A possible thought associated with it could be related to the emotional flavor “hunger” in which the abstract theories would have the following semantics: $p_0$ - to forage, $q_2$ - to hunt, $p_1$ -
that cognition (and the immunological system, and life itself) can be modeled as Autopoiesis theory, in an over simplified way, proposes the model, for instance Wittgenstein’s work [27] and the work on metaphors by inspiration sources also influenced the intended epistemological interpretation of Autopoiesis theory [26] and on the plausibility. These arguments are mainly inspired by Maturana and Varela’s that modeling memory using prime form representation has some psychological plausibility. In this section, we would like to present some epistemological arguments to argue to see the possible metaphors connecting both thoughts.

# Psychological Plausibility

In this section, we would like to present some epistemological arguments to argue that modeling memory using prime form representation has some psychological plausibility. These arguments are mainly inspired by Maturana and Varela’s Autopoiesis theory [26] and on the systemic approach in general [5], but other inspiration sources also influenced the intended epistemological interpretation of the model, for instance Wittgenstein’s work [27] and the work on metaphors by Lakoff and Johnson [25].

**Autopoiesis theory** Autopoiesis theory, in an over simplified way, proposes that cognition (and the immunological system, and life itself) can be modeled as...
A structurally determined process, i.e., a mechanism that can change its structure, during interaction with the environment, without destroying its basic organization, where this organization is exactly what defines the mechanism as a structurally determined process of a specific type. An autopoietic cognitive model must do something that modifies itself, but without losing its identity. A PIP pair already has a well determined organization, but to fulfill the autopoiesis requirements, it should incorporate, besides its static “storage model” nature, a dynamical aspect. A possible elementary action underlying the proposed model would be the continuous shift of point of view that the representation supports: from IP, the world situations, into PI, the rules to be respected, and conversely. This activity is incorporated into the behaviors (learning, reasoning, memorizing) of an agent that develops a specific instance of a PIP representation as a way to better handle a complex environment.

**Metaphors We Live by** In the proposed model, all meaning is based on emotional flavors. The truth values of theories associated with thoughts are just operational criteria that help to control emotional symbols in a given environment situation. Because of that, the meaning of any abstract thought is necessarily established through a metaphor with a “concrete” thought with a similar structural mechanism. In this sense the proposed model defines a formal model of the metaphorical mechanism proposed by Lakoff and Johnson [25].

**Proust’s Madeleine Effect** In his novel “In Search of Lost Time”, Marcel Proust tells how eating a cookie, as an example of involuntary memory, can take you back into your past through recalling, in just a moment, a whole period of your life. The proposed prime based model, because of its fractal nature, adapts itself very well to this kind of effect. Suppose that the initial stimulus brings back a thought, if its related sub thoughts are activated in a breath first way, we can simulate the Proust’s madeleine effect.

8 Conclusion

The paper presented through (very) simple examples a possible memory model that has both a logical definition and an embodied nature. The use of prime form representation to represent thoughts and the two indexing dimension that this representation endorse is also described. Some arguments in support of its psychological plausibility are also presented.

References