Knowledge and Information Systems (2001) 3: 96–119

Knowledge and Information Systems

An Artificial Network Simulating Cause-to-Effect Reasoning: Cancellation Interactions and Numerical Studies

L. B. Romdhane, B. Ayeb and S. Wang

Department of Mathematics and Computer Science, University of Sherbrooke, Sherbrooke, P.Q., Canada

Abstract. During the last few decades, a variety of models have been proposed to address causal reasoning (known also as abduction); most of these dealt with a probabilistic or a logical framework. Recently, a few models have been proposed within a neural framework. The investigation of neural approaches is mainly motivated by the computational burden of the causal reasoning task and by the satisfactory results given by neural networks in solving hard problems in general. A particular class of causal reasoning that raises several difficulties is the cancellation class. From an abstract point of view, cancellation occurs when two causes (hypotheses) cancel each other's explanation capabilities with respect to a given effect (observation). The present work is twofold. First, we extend an existing neural model to handle cancellation interactions. Second, we test the model on a large database and propose objective criteria to quantitatively evaluate the scenarios (explanations) produced. Simulation results show good performance and stability of the model.

Keywords: Abduction; Causal interactions; Causes; Competition; Diagnosis; Effects

1. Problem Statement and Preliminaries

1.1. Introduction

Causal reasoning (also known as abduction; Ayeb et al, 1998) is one of the reasoning forms which is used virtually everywhere – e.g. legal reasoning, planning, natural language processing. The intuition behind causal reasoning could be stated as follows. Given an observable effect (say e), a possible cause (say c), and

Received 17 Nov 1999
Revised 12 May 2000
Accepted 16 Jun 2000

a causal interaction/association between c and e, causal reasoning consists in hypothesizing that c occurred. The main task of causal reasoning is to compute, from the possible causes, a 'composite cause' that covers the entire set of observed effects. Dependent on the interactions between causes and effects, several classes of causal reasoning problems could be introduced. These include the simplest class, called independent class, to more complex ones such as incompatibility or cancellation classes. A complete illustrative example is given in the next section, where different classes are smoothly introduced and illustrated.

This paper focuses on the cancellation class. Typically, cancellation occurs when one cause may suggest that the value of an effect increases, while another cause may suggest that this same value decreases. As may be expected, this sort of interaction between causes raises several difficulties and challenges. The major aims of this paper are to construct a neural model that addresses these interactions, illustrates the cancellation class and provides a sensitive analysis of neural model convergence. The rest of this paper is organized as follows. After introducing preliminary material, Section 2 provides the neural model as well as illustrative examples. Section 3 includes related work, while Section 4 is devoted to benchmarking numerical studies. Concluding remarks and further research are in Section 5.

1.2. Complex Causal Reasoning: The Greek/Roman Warfare Example

Although our introductory example is simple, it motivates our discussion. Suppose that we are observing an ancient ruler's advisor analyzing a Greek/Roman warfare situation. Figure 1(a) describes part of the strategic knowledge used by the advisor; (b) formalizes it as a set of causal relationships.

Relations $R1$–$R7$ represent causal relationships between causes (for example, *Tyrant*, *Moderate*) and effects (for example, *UnriskySituation*, *TroopDeployment*). The signs positive (+) and negative (−) model the *nature* of the causality between causes and effects. The sign '+' denotes a positive influence, whereas '−' denotes a negative influence. For example, $R2$ tells us that *Negotiator* covers or explains *UnriskySituation*, whereas $R5$ tells us that *Tyrant* denies or cancels the presence of *UnriskySituation*. Consequently, the causes *Tyrant* and *Negotiator* cancel each other's explanation capabilities with respect to the same effect *UnriskySituation*. The same remark applies for *Tyrant* and *Moderate*. This means that when *UnriskySituation* is observed, *Tyrant* will try to extinguish *UnriskySituation*; that is, cancel its presence. *Negotiator* and *Moderate* will do the opposite and support its presence. Relations $R8$ and $R9$ represent incompatibility relationships between elementary causes. For instance, $R8$ tells us that *Moderate* and *Tyrant* are mutually exclusive and cannot coexist. In our example, $w_1, w_2, \cdots$, and w_7 are all real numbers over the interval $[0, 1]$. They measure the strengths of causality between causes and effects. Similarly, incompatibility causal relationships could also be weighted rather than being either/or (compatible/incompatible).

Now suppose that, in the current Greek/Roman, we observe that the situation is far from unrisky and it seems that troops are already deployed. The atmosphere remains unstable and an agreement is rather unlikely. These observed effects are said to form a *manifestation*. Our objective is to determine what kind of adversary we are dealing with; that is, to compute a set of causes that cover the observed effects (see Josephson et al, 1987; Fischer et al, 1991; Goel et al, 1995 for a concise review). Such causes are said to form a *scenario*

There is little risk when dealing with an adversary who is a moderate or a negotiator. An agreement remains possible provided that the adversary is both a moderate and a negotiator. Whenever we are involved with a tyrant, a tense atmosphere reigns. Furthermore, the situation tends to be very risky. If the adversary is a tyrant and a warrior, then he will probably order the deployment of his troops. It is also possible that a negotiator will deploy his troops in order to divert attention.

(a)

The set of causal relationships $\mathscr{R}$ is comprised of:

$(R1)$ $Moderate \overset{+}{\longmapsto} UnriskySituation < w_1 >$
$(R2)$ $Negotiator \overset{+}{\longmapsto} UnriskySituation < w_2 >$
$(R3)$ $Moderate \wedge Negotiator \overset{+}{\longmapsto} PossibleAgreement < w_3 >$
$(R4)$ $Negotiator \overset{+}{\longmapsto} TroopDeployment < w_4 >$
$(R5)$ $Tyrant \overset{-}{\longmapsto} UnriskySituation < w_5 >$
$(R6)$ $Tyrant \overset{+}{\longmapsto} TenseAtmosphere < w_6 >$
$(R7)$ $Warrior \wedge Tyrant \overset{+}{\longmapsto} TroopDeployment < w_7 >$

$(R8)$ $Moderate \oplus Tyrant$
$(R9)$ $Tyrant \oplus Negotiator$

(b)

Fig. 1. The Greek/Roman warfare example. (a) strategic knowledge; (b) causal relationships.

(explanation in Goel et al, 1988; Peng and Reggi, 1989; Thagard, 1989; Ayeb and Wang, 1995; Ayeb et al, 1998). Back to our example. Let us first specify our manifestation as three sets: (1) $\{TroopDeployment\}$, the set of effects observed to be present; (2) $\{UnriskySituation\}$, the set of effects observed to be absent; and (3) $\{TenseAtmosphere,\ PossibleAgreement\}$, the set of effects considered to be unknown; that is, neither present nor absent. Indeed, an effect is unknown if there is no piece of 'information' confirming either its presence or its absence. In a practical situation, this means that the 'value' of such an effect cannot be accessed (measured) for various reasons.[1] The scenario with respect to our manifestation is $SCEN_1 = \{Tyrant, Warrior\}$, since $SCEN_1$ together with the relations in $\mathscr{R}$ cover *TroopDeployment*, the present effect in our manifestation. Moreover, $SCEN_1$ does not cover the absent effect *UnriskySituation*. Finally, the causes included in $SCEN_1$ preserve the consistency of $SCEN_1 \cup \mathscr{R}$ since they do not create incompatibilities with respect to relations $R8$ and $R9$.

Although simple, the above example has the characteristics of a complex causal problem. It is a *monotonic* causal problem since the causal relationships involve an additive interaction between causes to cover some effects – see relation $R3$. It is also an *incompatibility* causal problem since the construction of scenarios has to prevent the coexistence of incompatible causes – relations $R8$ and $R9$. It is a *cancellation* causal problem since there are causes that cancel each other's explanation capabilities with respect to a given effect – relations $(R1), (R_2)$ and (R_5). Considering that our manifestation specifies not only present effects but also absent ones and that we require a scenario that avoids covering absent effects, our example is an *open* causal problem. In the next section, we give formal definitions of the distinct classes of causal problems.

1.3. Preliminary Material

From the Greek/Roman warfare example, one can say that a causal problem is mainly characterized by a set of possible effects, a set of possible causes and a set

[1] Possible reasons include time and safety.

of causal relationships $\mathscr{R}$. First, let us mention that computation of a scenario for a causal problem raises many difficulties. First, many conflicting criteria must be taken into consideration simultaneously (for example, the inferred scenario should be minimal *and* coherent). Second, solving causal problems is NP-Hard (Bylander et al, 1991). For a causal problem consisting of N possible causes, a best solution must be chosen among 2^N possibilities. This ultimately explains the investigation of neural-based techniques, which have shown robustness and efficiency in solving many hard problems (see, for example, Peterson and Anderson, 1988).

Causal reasoning can be formalized in several ways (see, for example, Peng and Reggia, 1989; Thagard, 1989; Bylander et al, 1991; Ayeb et al, 1993; Ayeb, 1994; Goel et al, 1995). For our purposes, we have borrowed the framework of the neural model developed in Ayeb and Wang (1995) and Ayeb et al (1998), and adapted it to formalize the cancellation class of causal reasoning.

Definition 1.1. A *causal problem* **CP** is a 3-tuple $\langle \mathscr{C}, \mathscr{E}, \mathscr{R} \rangle$, where $\mathscr{C}$ and $\mathscr{E}$ are two disjoint finite sets of constants, and $\mathscr{R}$ is a finite set of causal relationships between $\mathscr{C}$ and $\mathscr{E}$.

In fact, $\mathscr{C}$ is the set of possible causes, $\mathscr{E}$ the set of possible effects, and $\mathscr{R}$ denotes a mapping from $\mathscr{C}$ to $\mathscr{E}$. For convenience, let us first introduce the following notation to express the concept of a causal relationship.

Notation 1.1. Consider $C \subset \mathscr{C}$, a subset of elementary causes and $e \in \mathscr{E}$, a single observable effect. We write: $\bigwedge_{c \in C} \overset{w}{\longmapsto} e$ to express a causal relationship from C to e with a given weight w; where w is a real number in $[0, 1]$.

Now, we are ready to define the independent, monotonic and cancellation classes of causal problems.

Definition 1.2. Let $\mathbf{CP}\langle \mathscr{C}, \mathscr{E}, \mathscr{R} \rangle$ be a causal problem. Then:

- **CP** belongs to the *independent* class if each causal relationship can be expressed as $c \overset{w}{\longmapsto} e$. In other words, each causal relationship involves exactly *one* elementary cause and *one* observable effect.
- **CP** belongs to the *monotonic* class if there is at least one causal relationship which must be expressed as $\bigwedge_{c \in C} \overset{w}{\longmapsto} e$. In this case, we have an additive interaction among some elementary causes to explain an observable effect.
- **CP** belongs to the *cancellation* class if there is at least one causal relationship which must be expressed as $c \overset{w,-}{\longmapsto} e$, where we use the sign $-$ to denote a 'negative' influence of cause c on effect e.

According to Definition 1.2, in the monotonic class additive interactions occur between causes to cover the same effect. However, the crux of additivity could be stated as follows: each cause (or subset of them) covers the considered effect only *partially*, but together they cover it *fully*. Cancellation interactions arise when two causes c_i and c_k (or more) cancel each other's explanation capabilities with respect to a given effect e_j. For example, c_i might imply that the value of e_j should increase, whereas c_k might imply that the value of e_j should decrease. Now, we need to introduce the concept of a manifestation for a causal problem as follows.

Definition 1.3. Let $\mathbf{CP}\langle \mathscr{C}, \mathscr{E}, \mathscr{R} \rangle$ be a causal problem. A *manifestation* MM for **CP** is a 3-tuple $\langle E_P, E_A, E_U \rangle$, where $E_P \subseteq \mathscr{E}$ denotes the present effects, $E_A \subseteq \mathscr{E}$ denotes the absent effects and $E_U \subset \mathscr{E}$ denotes the unknown effects. Naturally, the

same effect cannot be present, absent and/or unknown in the same manifestation MM. Consequently, we have $(E_P \cap E_A) = (E_P \cap E_U) = (E_A \cap E_U) = \emptyset$.

A manifestation $MM\langle E_P, E_A, E_U\rangle$ is said to be *multiple* if $E_A \neq \emptyset$; otherwise it is said to be a *single* manifestation. Now we will introduce a convention and two definitions related to the open and incompatibility classes of causal problems.

Convention 1.1. We make use of the 'distinguished' symbol $\perp$ to denote an effect that can never be observed.

Definition 1.4. Let $\mathbf{CP}\langle \mathscr{C}, \mathscr{E}, \mathscr{R}\rangle$ be a causal problem.

- **CP** belongs to the *open* class if, for every manifestation, say $MM\langle E_P, E_A, E_U\rangle$, we have $E_A \neq \emptyset$. In other words, in an open causal problem there are effects that are known to be absent.
- **CP** belongs to the *incompatibility* class if there is at least one causal relationship which is expressed as $\bigwedge_{c\in C} \overset{w}{\longmapsto} \perp$. Hence, we have a set of causes which cannot coexist.

Before defining the notion of a scenario for a causal problem, we need the following notation.

Notation 1.2. Let c denote an elementary cause and assume that c causes $\{e_1, e_2, \cdots, e_k\}$; that is, $c \overset{w_i}{\longmapsto} e_i\,(i = 1, \ldots, k)$. We define the following 'function': $\textsc{Solve}(c) \overset{def}{=} \{e_1, e_2, \cdots, e_k\}$. If C denotes a subset of elementary causes, then $\textsc{Solve}(C) \overset{def}{=} \bigcup_{c\in C} \textsc{Solve}(c)$.

Notation 1.3. Let c denote an elementary cause and assume that c cancels $\{e_1, e_2, \cdots, e_k\}$; that is, $c \overset{w_i^-}{\longmapsto} e_i\,(i = 1, \ldots, k)$. We define by $\textsc{Cancel}(c) \overset{def}{=} \{e_1, e_2, \cdots, e_k\}$. Similarly, given a subset of elementary causes C, $\textsc{Cancel}(C) \overset{def}{=} \bigcup_{c\in C} \textsc{Cancel}(c)$.

Definition 1.5. Let $\mathbf{CP}\langle \mathscr{C}, \mathscr{E}, \mathscr{R}\rangle$ be a causal problem. Suppose that $MM\langle E_P, E_A, E_U\rangle$ is a given manifestation for **CP**. A *scenario* for **CP** with respect to MM is a subset of elementary causes $C \subseteq \mathscr{C}$, such that: (1) $E_P \subseteq \textsc{Solve}(C)$; (2) C is minimal; (3) $E_P \cap \textsc{Cancel}(C) = \emptyset$; (4) $E_A \cap \textsc{Solve}(C)$ is minimal; and (5) $\perp \notin \textsc{Solve}(C)$.

In Definition 1.5, the minimality criterion is framed in terms of set cardinality. This convention applies throughout this paper for any finite set. Note the potential conflict among criteria 1 to 5, which renders the inference process very hard to tackle, particularly in the presence of incompatibility and cancellation interactions (Bylander et al, 1991). However, to address this conflict, a global belief function could be used to rank the computed scenarios.[2] In addition, it should be noted that $\perp \notin E_P$, since $\perp$ denotes an effect that is *never* observed. However, if the causal problem at hand belongs to the incompatibility class, then necessarily we have $\perp \in E_A$.

All types of causal relationships in our formalization are weighted with real

[2] In neural networks such a function is also called an energy/target function.

numbers over $[0, 1]$.[3] Weighted causal interactions have been used in many well-known models (e.g., Pearl, 1987; Peng and Reggia, 1989). The semantic equivalent for these weights depends on the framework used and the approach adopted (logical, probabilistic or neural). In Pearl (1987), the weight of a causal interaction is interpreted as a 'conditional probability'; whereas in Peng and Reggia (1989) it is interpreted as the 'frequency' of the causality. For our purposes, we interpret these weights as the 'strengths' of the causal interactions. More precisely, we treat them as being 'temporary beliefs'. We use the term 'temporary' to indicate that these weights are *adapted* during the competition process. The causes included in a computed scenario are those that gain greater 'beliefs'.

There are two stages in the work described here. First, we extend the neural model for causal reasoning proposed first in Ayeb and Wang (1995), then in Ayeb et al (1998) to handle the cancellation type of causality. Indeed, the model presented in Ayeb et al (1998) is able to handle the open, incompatibility and monotonic classes of causal reasoning, but not the cancellation class. Second, we test the model in Ayeb et al (1998) on a real-world medical problem using a large battery of cases. We propose objective criteria to evaluate the computed scenarios. Simulation results show good performance of the algorithm. The remainder of this paper is organized as follows. Section 2 outlines our extended model and provides an illustration of the cancellation class of causal reasoning. Section 3 reviews related research. In Section 4, we present a numerical evaluation of the model. The final section provides some concluding remarks and suggests avenues for future research.

2. Neural Model: Description and Extension

Before going into the details of the algorithm, let us introduce the methodologies used in the mechanization of causal reasoning within a neural framework. There are two main methods. The first, which is more conventional, consists of two steps: (1) propose a target or energy function; (2) design the neural architecture to optimize (maximize or minimize) the proposed target function. The second method, which is less conventional, also consists of two steps: (1) design the architecture of the network to address the causal problem at hand; (2) analyze the dynamics of the network and the properties of the generated scenarios with respect to the chosen competition mechanisms. The main advantage of the conventional method is that it makes the target function available at an earlier stage. However, if not properly designed with respect to the adopted criteria, such a target function may impose severe constraints on the network's topology as well as on the optimization techniques to be used. On the other hand, the second method allows more flexibility in the design of the neural architecture with respect to the causal problem at hand. In particular, the problem of using high-order neural networks is avoided (Goel et al, 1998). However, it makes analysis of the network dynamics more difficult due to the lack of a 'global criterion', since the mechanics of the competition process are designed 'locally'. The model presented in Ayeb et al (1998) adopts the second method. Since the model presented here is a variant of the one in Ayeb et al (1998) extended to the cancellation class of causal reasoning, naturally we adopt the same methodology, described below.

[3] Our restriction to the interval $[0, 1]$ is not necessary and is used for conventional reasons only, since any value over $[0, \infty]$ could be easily adjusted to $[0, 1]$.

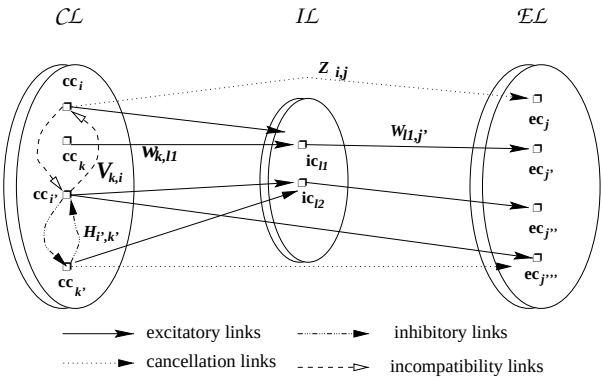


Fig. 2. The three-layered neural architecture for a causal problem.

2.1. Building the Three-Layered Neural Architecture

Given a causal problem $\mathbf{CP}\langle \mathscr{C}, \mathscr{E}, \mathscr{R} \rangle$, the corresponding general architecture (shown in Fig. 2) is built following the distinct steps summarized in Fig. 3. The architecture of the network is comprised of three layers: a cause layer $\mathscr{CL}$, an intermediate layer $\mathscr{IL}$ and an effect layer $\mathscr{EL}$. The cells in $\mathscr{CL}$ (resp. $\mathscr{EL}$) model the finite set of causes $\mathscr{C}$ (resp. the finite set of effects $\mathscr{E}$). Connections in the network are created using the set of causal interactions $\mathscr{R}$ – see Fig. 3. An inhibitory connection is created between cause cells if they cover at least one common observable effect. In this way, cause cells will inhibit each other for exclusivity in covering the shared effect. Regarding incompatibility interactions, they are modeled by using lateral inhibitory connections between each pair of incompatible causes. Hence, two incompatible causes will try to extinguish each other as long as both are still active. Cancellation and independent interactions are modeled by creating forward connections between the corresponding cause cells (in the $\mathscr{CL}$ layer) and effect cells (in the $\mathscr{EL}$ layer). An additive interaction is modeled by creating an intermediate cell in the $\mathscr{IL}$ layer and then connecting it to the corresponding cause cells and effect cell. An intermediate cell thus plays the role of a 'macro-cause' representing the subset of causes involved in the same additive interaction.

Before going further, we need the following notation. Subscripts i and k will be used to index all quantities related to causes, and subscript j to index all quantities related to effects. Subscript l will be used to index all quantities related to intermediate cells. a_j denotes the activity of effect cell ec_j, x_i the activity of cause cell cc_i and m_l the activity of an intermediate (hidden) cell. $\Psi = (\psi_1, \cdots, \psi_p)$ is a finite set of real numbers over $[0, 1]$. ψ_j $(j = 1, \cdots, p)$ is the degree of certainty of effect e_j in the manifestation $MM = \langle E_P, E_A, E_U \rangle$. Naturally, we have $\psi_j = 0$ $\forall e_j \in E_U$, since E_U is the set of unknown effects.

Now, we are ready to introduce the initialization process for the network's connections and inputs, as follows. Connections in the network are initialized using the set of causal strengths in $\mathscr{R}$; whereas the inputs to the network are initialized using Ψ. The whole initialization process is summarized in Fig. 4. In the initialization of hidden-to-output connections, we use θ_i, a real number in $[0, 1]$ to indicate cc_i's contribution to coverage of ec_j by the additive interaction being

BUILD(NET_{CP})

I **Create the three layers of the NET_{CP}: input ($\mathscr{EL}$), intermediate ($\mathscr{IL}$) and output ($\mathscr{CL}$).**

I.1) Layer $\mathscr{CL}$ includes a cause cell (cc) for each elementary cause in $\mathscr{C}$.

I.2) Layer $\mathscr{IL}$ includes an intermediate cell (ic) for each 'additive' causal relationship in $\mathscr{R}$; for example, $c_i \wedge \ldots \wedge c_k \overset{w}{\longmapsto} e_j$.

I.3) Layer $\mathscr{EL}$ includes an effect cell (ec) for each observable effect in $\mathscr{E}$.

II **Create forward connections: input-to-hidden and hidden-to-output.**

II.1) There is a link $W_{i,j}$ (called an excitatory connection) between each cell cc_i of $\mathscr{CL}$ and ec_j of $\mathscr{EL}$ representing independent causal interactions in $\mathscr{R}$; that is, $c_i \overset{w}{\longmapsto} e_j$.

II.2) There is a link $Z_{i,j}$ (called a cancellation connection) between each cell cc_i of $\mathscr{CL}$ and ec_j of $\mathscr{EL}$ representing cancellation causal relationships in $\mathscr{R}$; that is, $c_i \overset{w^-}{\longmapsto} e_j$. Hence, although cancellation interactions occur between causes with respect to a given effect, cancellation links are created *only* between the 'cancelling' cause and the 'cancelled' effect.

II.3) For 'additive' relationships such as $c_i \wedge \ldots \wedge c_k \overset{w}{\longmapsto} e_j$, we create:
(a) $W_{i,l}, \cdots, W_{k,l}$ excitatory connections between cause cells $cc_i, \cdots, cc_k$ and an intermediate cell ic_l;
(b) an excitatory connection $W_{l,j}$ between ic_l and ec_j cells.

III **Create lateral connections: incompatibility and inhibitory.**

III.1) There is a link $H_{i,k}$ (called an inhibitory connection) between cc_i and cc_k cells if there is at least one observable effect e covered by c_i and c_k; that is, $\{c_i \overset{w_1}{\longmapsto} e; c_k \overset{w_2}{\longmapsto} e\} \in \mathscr{R}$.

III.2) There is a link $V_{i,k}$ (called an incompatibility connection) between cc_i and cc_k cells if causes c_i and c_k are incompatible; that is, $\{c_i \wedge c_k \overset{w}{\longmapsto} \perp\} \in \mathscr{R}$.

Fig. 3. Steps in designing the neural architecture for a causal problem.

NETWORK INITIALIZATION

⊡ **Network Inputs.**

$\forall ec_j \in \mathscr{EL}$ such that $e_j \in E_U, a_j = 0$

$\forall ec_j$ such that $e_j \in \{E_P \cup E_A\}, a_j = \psi_j$

⊡ **Connections corresponding to independent causal relationships.**

- `Excitatory:` $\forall W_{i,j}$ such that $\exists \; c_i \overset{w}{\longmapsto} e_j$, Let $W_{i,j} = w$

- `Cancellation:` $\forall Z_{i,j}$ such that $\exists \; c_i \overset{w^-}{\longmapsto} e_j$, Let $Z_{i,j} = w^-$

⊡ **Incompatibility connections.**

$\forall cc_i, cc_k$, such that $\exists \quad c_i \wedge c_k \overset{w}{\longmapsto} \perp$, Let $V_{i,k} = w$

⊡ **Connections corresponding to 'additive' causal relationships.**

- `Input-to-Hidden Connections:` $\forall W_{l,j}$ such that $\exists c_i \wedge \cdots \wedge c_k \overset{w}{\longmapsto} e_j$, Let $W_{l,j} = w$

- `Hidden-to-Output Connections:` $\forall W_{i,l}$ such that $\exists W_{l,j}$, Let $W_{i,l} = \theta_i \; W_{l,j}$

⊡ **Inhibitory Connections.**

$\forall cc_i, cc_k$; Let:

- $E_{i,k} = \{e_j$ such that $W_{i,j} \; W_{k,j} \neq 0\}$
- $H_{i,k} = CARD(E_{i,k})/CARD(\mathscr{E})$

Fig. 4. Initialization of network's inputs and connections.

considered. The parameter θ_i could be either user-defined or uniformly distributed by the algorithm. In other words, if the user has a piece of 'knowledge' regarding the contribution of cc_i to the additive interaction (for example; weak, medium, strong), the information can be used to set θ_i. Otherwise, θ_i is set uniformly by the algorithm to $1/k$, with k being the number of causes in the additive interaction. The inhibitory connections between cause cells are initialized using $E_{i,k}$, the finite set of effects common to causes c_k and c_i. The term $CARD$ denotes set cardinality.

2.2. The Algorithm

Our algorithm, called SOLVE, is run in five main steps and is depicted in Fig. 5. Steps I to IV are computational steps that engage the network in the competition process, whereas step V is a termination test. SOLVE is an extension of the algorithm UNIFY developed in Ayeb and Wang (1998), and designed to mechanize the cancellation class of causal reasoning. Implementation of cancellation interactions in UNIFY alters its mechanics. In particular, network inputs are no longer fixed but subject to modification. Moreover, the halting stage and the normalization of forward connections are modified. Here, we describe each step of SOLVE with particular emphasis on the 'cancellation competition'. SOLVE accepts as input NET_{CP}, the network for $\mathbf{CP}\langle \mathscr{C}, \mathscr{E}, \mathscr{R} \rangle$ created using the method shown in Fig. 3 and initialized as in Fig. 4.

Before going into the details of SOLVE, let us focus on elementary cause cell activities. They are updated using a shunting competitive mechanism introduced by Grossberg (1988) and adapted from Ayeb and Wang (1998):

$$\frac{dx_i}{dt} = -\tau x_i + \Sigma_{j,e_j \in E_P}\ W_{i,j}\ a_j - \Sigma_{j,e_j \in E_A}\ W_{i,j}\ a_j + \Sigma_l W_{i,l} m_l$$
$$-\Sigma_{j,e_j \in E_P}\ Z_{i,j}\ a_j - x_i\ \left(\Sigma_{k \neq i}\ H_{i,k}\ f(x_k) + \Sigma_{p \neq i}\ V_{i,p}\ f(x_p)\right) \quad (1)$$

where τ is a small real number modeling the decay rate of x_i. Let us group the significant terms in formula (1) by defining the following quantities.

$$EX_i^+ = \Sigma_{j,e_j \in E_P}\ W_{i,j} a_j \quad (2)$$

defines the excitatory (positive) input to cc_i from present effects. cc_i also receives a positive input from intermediate cells by:

$$AX_i = \Sigma_l W_{i,l} m_l \quad (3)$$
$$EX_i^- = \Sigma_{j,e_j \in E_A}\ W_{i,j} a_j + \Sigma_{j,e_j \in E_P}\ Z_{i,j} a_j \quad (4)$$

models the inhibitory input to cc_i from absent and present effects. Indeed, x_i, the activity of cc_i, is penalized by absent effects since our concern is to infer a set of causes that *avoids* coverage of absent effects. Moreover, x_i is weakened by present effects with which it has a cancellation type of causality since a cancellation link models a 'negative' influence between causes and effects. However, it should be noted that in the case of a cancellation interaction, an effect e_j which is *absent* or *unknown* has no influence on the activity of the involved cause cell. This is natural since cancellation occurs only in the *presence* of e_j.

In addition to the forward input (positive/negative), a cause cell cc_i also has

ALGORITHM SOLVE(NET_{CP})

BEGIN SOLVE

□ **STEP I: UPDATING CAUSE CELL ACTIVITIES**

i.1) `Elementary cause cells;`
For each $cc_i \in \mathscr{CL}$ Do:
$x_i = (EX_i^+ - \xi_i EX_i^- + AX_i)/(\tau + IN_i + IC_i)$

i.2) `Composite cause cells;`
For each $ic_l \in \mathscr{IL}$ Do:
$m_l = min(\{x_i \text{ such that } W_{i,l} \neq 0\})$;

□ **STEP II: UPDATING ALL CONNECTION WEIGHTS**

ii.1) `Excitatory links between the` $\mathscr{CL}$ `and` $\mathscr{CL}$ `layers:;`
For each $W_{i,j} > 0$ Do:
a) Compute: $T_{i,j} = W_{i,j}/(1 - \alpha\, x_i\, a_j)$
b) Update:
$W_{i,j} = T_{i,j}/\sqrt{\sum_i T_{i,j}^2 + \sum_i (Z_{i,j})^2}$

ii.2) `'Additive' excitatory links:;`
For each $W_{l,j} > 0, W_{i,l} > 0$ Do:
a) Compute: $T_{l,j} = W_{l,j}/(1 - \alpha\, m_l\, a_j)$
b) Update: $W_{l,j} = T_{l,j}/\sqrt{\sum_l (T_{l,j})^2)}$
c) Redistribute: $W_{i,l} = \theta_i\, W_{l,j}$

ii.3) `Incompatibility links:;`
For each $V_{i,k}$ such that $(x_i > \epsilon) \wedge (x_k > \epsilon)$ Do:
$V_{i,k} = (V_{i,k} - \beta\, \Delta_{x_i}\, \Delta_{x_k})\, \delta_{i,k}$

ii.4) `Inhibitory Links;`
For each $H_{i,k}$ such that $(W_{i,j_0}\, W_{k,j_0} \neq 0)$ Do:
$H_{i,k} = H_{i,k} - \alpha\, \Delta_{x_i}\, \Delta_{x_k}$

ii.5) `Cancellation links;`
For each $Z_{i,j}$ such that $Z_{i,j} > 0$ Do:
$Z_{i,j} = Z_{i,j} + \eta x_i a_j - \mu x_i Z_{i,j}$

□ **STEP III: UPDATING INPUT CELL ACTIVITIES**

Let $C_{tmp} = \{c_i / x_i > 0\}$.
For each ec_j such that $e_j \in$ CANCEL(C_{tmp}) Do:
iii.1) Compute:
- $F_j^+ = \sum_i W_{i,j} x_i + \sum_l W_{l,j} m_l$
- $F_j^- = \sum_i Z_{i,j} x_i$

iii.2) Update: $a_j = f(F_j^+/(F_j^+ + F_j^-))$

□ **STEP IV: NETWORK PRUNING**

iv.1) For each $W_{i,j} < \epsilon$ Set $W_{ij} = 0$
iv.2) For each $Z_{i,j} < \epsilon$ Set $Z_{i,j} = 0$

□ **STEP V: TERMINATION TESTING**

Let $C_{tmp} = \{c_i / x_i > 0\}$.
If $\big((\text{CANCEL}(C) \neq \emptyset)$ OR (*LimitIter* is reached) $(\exists c_i \in C_{tmp}$ such that $\neg$EXCLUSIVE(cc_i)))
Then **Goto** STEP I;
Else **(Return** C_{tmp}**)**.

END SOLVE

Fig. 5. Our SOLVE algorithm.

a *lateral* input modeled by the two quantities IC_i and IN_i:

$$IC_i = \Sigma_{p \neq i} V_{i,p} f(x_p) \tag{5}$$

$$IN_i = \Sigma_{k \neq i} H_{i,k} f(x_k) \tag{6}$$

where $f(\cdot)$ is a sigmoid-type function.

Now, let us rewrite formula (1) as:

$$\frac{dx_i}{dt} = -\tau x_i + ((EX_i^+ - \xi_i EX_i^-) + AX_i) - x_i(IN_i + IC_i) \tag{7}$$

where ξ_i is defined as:

$$\xi_i = \frac{1}{1 + \left(\frac{EX_i^-}{EX_i^+}\right)} \tag{8}$$

The parameter ξ_i is a control factor introduced in (7) to balance the influence of the 'inhibitory signal' EX_i^- with respect to the 'excitatory signal' EX_i^+ on x_i (Ayeb and Wang, 1998).

The competition process consists in solving the system of non-linear equations (7) in its equilibrium state; that is, when $(dx_i/dt) = 0 \forall i$. The activities of cause cells are computed according to:

$$x_i = \frac{(EX_i^+ - \xi EX_i^-) + AX_i}{\tau + IC_i + IN_i}, \quad \forall cc_i \in \mathscr{CL} \tag{9}$$

From formula (9), we can observe that x_i is strengthened by EX_i^+ and AX_i. Naturally, AX_i is null when the causal problem at hand belongs to the independent class; that is, when there are no additive interactions between causes to cover some effects. The negative inputs EX_i^-, IC_i and IN_i weaken x_i. However, through the use of the control factor ξ_i in formula (9), we ensure that a cause cell never vanishes as long as it is still receiving some positive (excitatory) input.

The initialization of cause cell activities should be derived using formula (9). Let us assume that cause cells enter an equilibrium state before the competition process begins. This means that the 'negative' lateral feedback to a cause cell from other competing cells is null. Hence, we have $IC_i = IN_i = 0 \ \forall cc_i$. Each cause cell is initialized as follows:

$$x_i^{init} = \frac{1}{\tau}\left(EX_i^+ - \xi EX_i^-\right) \quad \forall cc_i \in \mathscr{CL} \tag{10}$$

Naturally, hidden cells have null initial activities since the competition has not started yet. Thus, the initial input AX_i to cc_i is null and is not included in the above equation. Now, we are ready to describe each step of our algorithm SOLVE.

Step I updates the activities of elementary and composite cause cells. As discussed earlier in this section, cause cells (in the $\mathscr{CL}$ layer) are updated using formula (9). Intermediate (hidden) cells (in the $\mathscr{CL}$ layer) are updated using the following formula:

$$m_l = min(\{x_i \text{ such that } W_{i,l} \neq 0\}), \forall ic_l \in \mathscr{IL} \tag{11}$$

where *min* is the minimum operator. Indeed, an intermediate cell acts as an information transmitter cause cells and effect cells. Consequently, the amount of transmitted 'information' is set to the minimum conveyed by the involved causes.

Step II updates all connection strengths in the network. In II.1, excitatory connections between cause and effect cells are updated as follows:

$$T_{i,j} = \frac{W_{i,j}}{1 - \alpha\, x_i\, a_j}, \forall W_{i,j} > 0 \tag{12}$$

$$W_{i,j} = \frac{T_{i,j}}{\sqrt{\sum_i (T_{i,j})^2 + \sum_i (Z_{i,j})^2}}, \forall W_{i,j} > 0 \tag{13}$$

Formulas (12) and (13) involve a redistribution of weight strengths $W_{i,j}$ for the belief that e_j is covered by cc_i. They reflect the current levels of the network's belief in the plausibility of each cause. Clearly, causes that gain greater belief are normally those which resist the competition more effectively. The normalization in (13) is important to the competition process. In fact, it keeps excitatory weights within the interval [0, 1] and thereby prevents their exponential growth, which could lead the network to a non-stable state/non-valid solution. Moreover, (13) ensures that an exclusive cause cell always gains the maximal belief (= 1), no matter how strong the other competing cells are. It guarantees that an effect is either cancelled or confirmed by the competing causes (depending on the cancellation and excitatory strengths). In fact, excitatory and cancellation weights

are in competition to cancel or cover the corresponding effect. This competition is considerably enhanced by limiting its resources, and the normalization process is one frequently used technique (see, for example (Amari and Arbib, 1982; Rumelhart and Zipser, 1985).

In II.2, input-to-hidden connections are updated according to:

$$T_{l,j} = \frac{W_{l,j}}{1 - \alpha\, m_l\, a_j}; \forall W_{l,j} > 0 \tag{14}$$

$$W_{l,j} = \frac{T_{l,j}}{\sqrt{\sum_l T_{l,j}^2}}; \quad \forall W_{l,j} > 0 \tag{15}$$

Then, we update the hidden-to-output connections by:

$$W_{i,l} = \theta_i W_{l,j}; \ \forall W_{i,l} \tag{16}$$

Indeed, formula (16) redistributes the beliefs (weights) between elementary cause cells and an effect cell that are involved in the same additive interaction. This clearly demonstrates that an intermediate cell is simply a mediator between causes and effects.

Before commenting on the update scheme for incompatibility connections, we need to formalize the meaning of an exclusive cause. A cause c_i is *temporarily* exclusive if there is an observable effect $e_j \in E_P$ such that c_i is the only *remaining* cause covering e_j. Exclusiveness is thus related to the temporary belief of the network and is formalized by the predicate EXCLUSIVE(c_i) as in Ayeb and Wang (1998).

Exclusive(c_i) holds iff $\exists ec_{j_0} \in \mathscr{E}\mathscr{L}$ such that $W_{i,j_0} > 0 \wedge W_{k,j_0} = 0, \forall\, k \neq i$

In II.3, incompatibility connections are updated using the following formulas:

$$V_{i,k} = (V_{i,k} - \beta\ \Delta_{x_i}\Delta_{x_k})\delta_{i,k} \tag{17}$$

where Δ_{x_i} (resp. Δ_{x_k}) stands for the change in activity of cause cell cc_i (resp. cc_k) from one iteration to the next, and $\delta_{i,k}$ is a Kronecker-like symbol defined by:

$$\begin{aligned} \delta_{i,k} &= 1 \quad \text{if } \neg\text{EXCLUSIVE}(cc_i) \wedge ((x_k > x_i) \vee \text{EXCLUSIVE}(cc_k)) \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{18}$$

Formula (17) coupled with formula (18) guarantee an asymmetry in the update of incompatibility connections. This results in increased inhibition of the activity of non-essential causes as well as decreased inhibition of the activity of highly plausible causes.

In II.4, inhibitory connections are updated according to:

$$H_{i,k} = H_{i,k} - \alpha\ \Delta_{x_i}\ \Delta_{x_k}, \forall H_{i,k} \text{ such that } W_{i,j_0}\ W_{k,j_0} \neq 0 \tag{19}$$

Remember that cc_i and cc_k have inhibitory connections if they explain common effects. According to formula (19), inhibitory weights vary in the opposite direction from cause cell activities. Hence, a cause cell with a weak activation is strongly inhibited by other competing cells and vice versa.

In II.5, cancellation connections are updated according to a modified version of Hebb's rule (Brown et al, 1990):

$$Z_{i,j} = Z_{i,j} + \eta x_i a_j - \mu x_i Z_{i,j} \quad \forall Z_{i,j} > 0 \tag{20}$$

where η and μ denote a small learning rate and a 'forgetting' factor, respectively.

An increase in a_j, the activity of ec_j, will lead to an increase in the inhibition strength against cc_i. Hence, the activity of cause cell cc_i will decrease accordingly – see formulas (4) and (9). An important feature is that the evolution of cancellation weights is proportional to the activities of both cause and effect cells. The forgetting term in (20) is a commonly used technique in self-organizing networks to prevent a catastrophic amplification of weights (Kohonen, 1988).

Step III updates the inputs of the network. This update process involves only those input cells corresponding to a present effect that have cancellation links with some causes. Indeed, unlike the model discussed in Ayeb and Wang (1998), the inputs of our network are not fixed but are modified as follows:

$$a_j = f(F_j) \tag{21}$$

$$F_j = \frac{F_j^+}{F_j^+ + F_j^-} \tag{22}$$

$$F_j^+ = \sum_i W_{i,j} x_i + \sum_l W_{l,j} m_l \tag{23}$$

$$F_j^- = \sum_i Z_{i,j} x_i \tag{24}$$

where $f(\cdot)$ is a sigmoid-type function. The quantity F_j is the *feedback* to the input cell ec_j from elementary and composite cause cells. Because ec_j can have two types of causal interactions with causes, two types of feedbacks are modeled: an inhibitory feedback from causes that have cancellation links with ec_j quantified by F_j^-; and an excitatory feedback F_j^+ from causes that have excitatory links with ec_j. From formulas (21) and (22), we can say that a_j is weakened by F_j^- and strengthened by F_j^+. Depending on F_j^+ and F_j^-, a_j may increase or decrease.

From the update scheme for cancellation weights in formula (20) and the update of input cells in (21), we can state the following: if $F_j^- \ll F_j^+$, then a_j is a strictly positive value. This would result in an increase in $z_{i,j}$ $\forall i$. Consequently, inhibition against cause cells that are involved in a cancellation relationship with ec_j will increase, ultimately eliminating the activity of such cells. Likewise, if an effect cell does not receive enough excitatory feedback F_j^+, then its activity vanishes since $a_j \approx 0$ when $F_j^+ \ll F_j^-$. In both cases, a cancellation type of causality leads either to the 'death' of the involved cause or to rejection of the involved effect from the set of present effects. Obviously, this depends on weights of the excitatory and cancellation links; that is, on both feedbacks, F_j^+ and F_j^-.

Although cancellation appears to be a *between-cause-and-effect* competition, it is actually a *between-cause-and-cause* competition by means of effects. More precisely, two causes try to cancel each other's explanation capabilities via the activation of the relevant effect. Indeed, a closer look at formula (21) reveals that competition depends on the two feedback signals F_j^+ and F_j^-. And by definition, these two quantities involve *only* cause cells. Depending on the strengths of these two signals, an effect may be either confirmed or 'cancelled' (that is, its activation level is weakened and reduced to 0). Finally, it should be noted that the activity of input cells not involved in a cancellation interaction remains constant throughout the competition. In fact, when $F_j^- = 0$ we have $a_j = f(1)$ which is a constant. Ultimately, this explains why in equation (21) we restrict ourselves to effect cells that have cancellation interactions with the competing causes.

$(R1)$ $Moderate \overset{+}{\longmapsto} UnriskySituation < w_1 = 0.83 >$
$(R2)$ $Negotiator \overset{+}{\longmapsto} UnriskySituation < w_2 = 0.90 >$
$(R3)$ $Moderate \wedge Negotiator \overset{+}{\longmapsto} PossibleAgreement < w_3 = 0.78 >$
$(R4)$ $Negotiator \overset{+}{\longmapsto} TroopDeployment < w_4 = 0.75 >$
$(R5)$ $Tyrant \overset{-}{\longmapsto} UnriskySituation < w_5 = 0.80 >$
$(R6)$ $Tyrant \overset{+}{\longmapsto} TenseAtmosphere < W_6 = 0.80 >$
$(R7)$ $Warrior \wedge Tyrant \overset{+}{\longmapsto} TroopDeployment < w_7 = 0.86 >$
$(R8)$ $Moderate \wedge Tyrant \longmapsto \perp < w_8 = 0.75 >$
$(R9)$ $Tyrant \wedge Negotiator \longmapsto \perp < w_9 = 0.70 >$

Fig. 6. The Greek/Roman warfare example: the set of causal interactions.

Step IV is merely a pruning phase in which all non-significant connections (excitatory and cancellation) are set to zero and thus no longer considered in the competition process. Although such a phase is not crucial to the competition process, it may accelerate convergence of the network to a stable state.

Step V is the termination test. The reasoning process is stopped when each of the elementary causes sustaining the competition is: (1) exclusive; *and* (2) has already cancelled all present effects with which it has cancellation interactions. Ultimately, the goal of SOLVE is to eliminate all non-essential, non-consistent causes and to retain only exclusive, plausible ones. However, SOLVE could also be stopped when a maximal number of allowed iterations (*LimitIter*) is reached. This is useful because finding a solution is not always guaranteed, since for the incompatibility, monotonic and cancellation classes of causal reasoning there is no proof that a solution necessarily exists (see Bylander et al, 1991, for further details).

2.3. Illustration

The main purpose of this section is to illustrate the capabilities of SOLVE in handling cancellation interactions. For a description of other aspects of the competition on other classes, we refer the reader to the original model in Ayeb and Wang (1998).

2.3.1. The Greek/Roman Warfare Example Revisited

As a first example, we consider the Greek/Roman warfare problem introduced in Section 1.2 and depicted in Fig. 6. Remember that the cause *Tyrant* has a cancellation causal relationship with the effect *UnriskySituation*. On the contrary, the causes *Moderate* and *Negotiator* both have excitatory causal relationships with the same effect *UnriskySituation*.

To illustrate the mechanics of the competition, we consider a manifestation $MM\langle E_P, E_A, E_A\rangle$; where the set of *present* effects is $E_P = \{UnriskySituation, TroopDeployment, TenseAtmosphere\}$, the set of *unknown* effects is $E_U = \{PossibleAgreement\}$, and the set of *absent* effects is $E_A = \emptyset$. We assign all present effects the highest belief; that is, $\psi = 1$. This means that the activations of the effect cells corresponding to *UnriskySituation*, *TroopDeployment* and *TenseAtmosphere* are initialized to 1; whereas the remaining one (*PossibleAgreement*) is initialized to 0. The simulation results are reported in Fig. 7. Figure 1(a) traces the

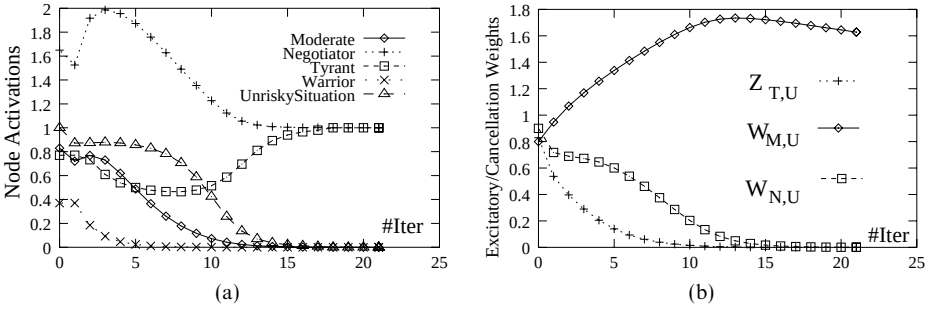


Fig. 7. Simulation results for the Greek/Roman warfare problem. (a) Final scenario = {Negotiator, Tyrant}; (b) Excitatory/cancellation weights.

evolution of cause cell activities. We note that at the beginning of the competition the activity of *Tyrant* is strongly diminished by the causes *Moderate* and *Negotiator*, and by the present effect *UnriskySituation*. In fact, *Moderate* and *Negotiator* support the presence of *UnriskySituation* and thereby increase its activation – see formulas (22) and (24). Raising the activity of *UnriskySituation* decreases that of *Tyrant* – see formulas (4) and (10). Conversely, the fact that effect *TenseAtmosphere* is present strengthens the activity of *Tyrant*. As the competition continues, *Tyrant* lowers the activity of both *Moderate* and *Negotiator*, whose support for the present effect *UnriskySituation* is reduced during the competition. In fact, we can see in Fig. 7(b) that the excitatory links between *Moderate* and *UnriskySituation* ($W_{M,U}$) and between *Negotiator* and *UnriskySituation* ($W_{N,U}$) drop to towards 0 at the end of the competition. Since the activation of *UnriskySituation* is reduced, its cancellation link to *Tyrant* ($Z_{T,U}$) is likewise weakened – see Fig. 7(b).

At the end of the competition, *Tyrant* extinguishes *UnriskySituation* and excludes it from E_P, the initial set of present effects. In other words, *Tyrant* cancelled the explanation capability of both *Moderate* and *Negotiator* with respect to *UnriskySituation*. However, despite the exclusion of *UnriskySituation*, *Negotiator* survives the competition since it takes exclusivity in covering the present effect *TroopDeployment*. From this simulation, we remark that the 'cancellation' competition is a complex process since the death of an effect cell may lead to the death of 'non-essential' cause cells covering this effect (*Moderate* in our case); whereas other causes (*Negotiator* in our case) may survive and be included in the final scenario.

2.3.2. The Chest-Clinic Problem

In this second set of experiments, we consider a medical problem composed of 10 symptoms (effects in our model), 4 diseases (causes in our model), and 17 causal associations. $\mathscr{E} = \{cough, dyspnea, expectoration, fever, inflamed\_throat, headache, lost\_voice, nasal\_discharge, nose\_pain, thoracic\_pain\}$ are the possible symptoms; $\mathscr{C} = \{laryngitis, pneumonia, sinusitis, tonsillitis\}$ are the possible diseases; and the set of causal interactions $\mathscr{R}$ is summarized in Fig. 8. We note that the disease *pneumonia* has a cancellation interaction with the symptom *cough* (φ_6); whereas the disease *laryngitis* has a 'positive' (excitatory) interaction (φ_3) with the same symptom.

We considered a manifestation in which the set of present effects is E_P =

$$\mathscr{R}\begin{cases}
(\varphi_1) & laryngitis \overset{+}{\longmapsto} inflamed_throat < w = 0.8 > \\
(\varphi_2) & laryngitis \overset{+}{\longmapsto} expectoration < w = 0.5 > \\
(\varphi_3) & laryngitis \overset{+}{\longmapsto} cough < w = 0.5 > \\
(\varphi_4) & laryngitis \overset{+}{\longmapsto} lost_voice < w = 0.45 > \\
(\varphi_5) & pneumonia \overset{+}{\longmapsto} fever < w = 0.25 > \\
(\varphi_6) & pneumonia \overset{-}{\longmapsto} cough < w = 0.8 > \\
(\varphi_7) & pneumonia \overset{+}{\longmapsto} headache < w = 0.5 > \\
(\varphi_8) & pneumonia \overset{+}{\longmapsto} dyspnea < w = 1.0 > \\
(\varphi_9) & pneumonia \overset{+}{\longmapsto} expectoration < w = 0.5 > \\
(\varphi_{10}) & pneumonia \overset{+}{\longmapsto} thoracic_pain < w = 1.0 > \\
(\varphi_{11}) & sinusitis \overset{+}{\longmapsto} headache < w = 0.5 > \\
(\varphi_{12}) & sinusitis \overset{+}{\longmapsto} noise_pain < w = 1.0 > \\
(\varphi_{13}) & sinusitis \overset{+}{\longmapsto} nasal_discharge < w = 0.5 > \\
(\varphi_{14}) & tonsillitis \overset{+}{\longmapsto} inflamed_throat < w = 0.25 > \\
(\varphi_{15}) & tonsillitis \overset{+}{\longmapsto} fever < w = 0.25 > \\
(\varphi_{16}) & sinusitis \wedge pneumonia \longmapsto \perp < w = 1.0 > \\
(\varphi_{17}) & laryngitis \wedge tonsillitis \longmapsto \perp < w = 0.85 >
\end{cases}$$

Fig. 8. The set of causal interactions for the chest-clinic problem.

$\{cough,\ expectoration,\ inflamed\_throat,\ lost\_voice,\ nasal\_discharge,\ nose\_pain\}$, the set of absent effects is $E_A = \{dyspnea,\ headache\}$, and the set of unknown effects is $E_U = \{fever,\ thoracic\_pain\}$. Present and absent effects were assigned the highest belief; that is, 1. Remember that unknown effects are automatically assigned a null degree of certainty by SOLVE. The simulation results are summarized in Fig. 9, where the activation of cause cells and that of the 'cancelled' effect, *cough*, are plotted over the course of the competition. The scenario computed by SOLVE is $\mathscr{S} = \{laryngitis,\ sinusitis\}$ since these are the two diseases (causes) that win the competition. As the competition begins *pneumonia* increases its cancellation to *cough*, whose activation is slightly diminished. Conversely, *laryngitis* tries to support the presence of *cough* and, thanks mainly to its high activation level, succeeds in extinguishing *pneumonia*. In fact, *laryngitis* is highly activated, since it covers most of the present symptoms in E_P. We should remark that, unlike the Greek/Roman warfare example (Section 2.3.1), in this case the 'cancelled' symptom (*cough*) is not excluded from the set of present effects. We should note also that $\mathscr{S}$ is a *best* scenario since: (1) it covers all present effects in E_P; (2) it does not include incompatible causes; (3) it avoids coverage of absent effects in E_A; and (4) it is minimal since removing either *laryngitis* or *sinusitus* from $\mathscr{S}$ will lead to only *partial* coverage of E_P.

3. Related Research

The main aim in this section is to compare the capabilities of other proposed models, with particular emphasis on the implementation of cancellation interactions. Existing models for causal reasoning can be categorized into three main classes on the basis of the framework adopted: logical, probabilistic, and neural. Naturally, there are several models that cannot be strictly categorized in a single class; for example a model may simultaneously use a probabilistic and a neural framework. These models could be named 'hybrid models' for causal reasoning. In this section, we will give a brief overview of neural-based models.

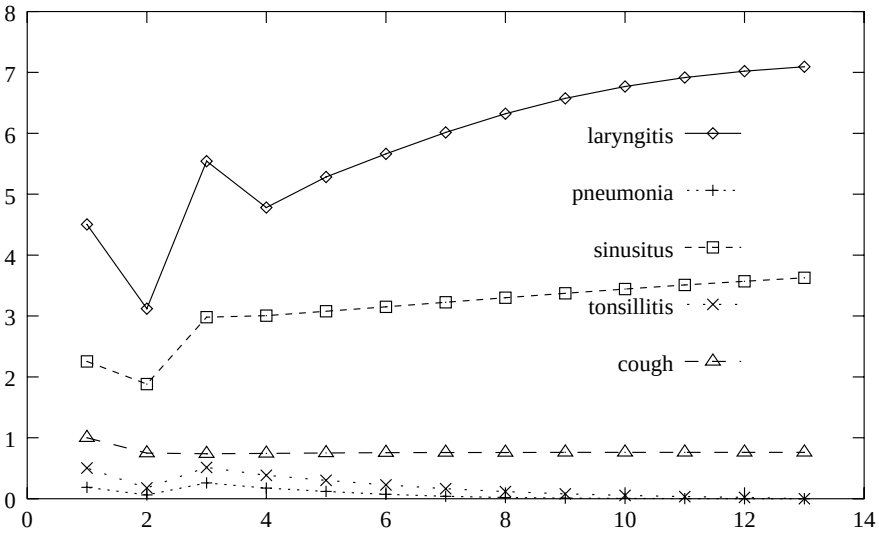


Fig. 9. Simulation results for the chest-clinic problem.

The model of Thagard (1989) aims to mechanize a range of problems formulated in a coherence theory. Thagard's algorithm, called ECHO, has been successfully applied to a variety of complex problems (see Eliasmith and Thagard (1997) and Thagard (1989) for simulation results). Despite the fact that both ECHO and SOLVE use excitatory/inhibitory connections and are based on competitive mechanisms between causes, there are fundamental differences between the two models. Indeed, the explanations (scenarios) in ECHO are constructed within a theory of explanatory coherence. Briefly, this theory is based on a set of seven axioms, called principles in Thagard (1989). In SOLVE scenarios are characterized by conventional criteria such as set inclusion and minimal cardinality – see Definition 1.5. In the neural architecture of ECHO, cancellation interactions could be mechanized as follows: for each effect e, two units are created, the first modeling e and the second, its negation $\bar{e}$. Thereafter, causes having 'positive' interactions with e are linked to the e cell, and causes having cancellation interactions with it are linked to the $\bar{e}$ cell. Finally, an inhibitory link is created between the e and $\bar{e}$ cells. Thus, acceptance of e (resp. $\bar{e}$) will encourage the rejection of the 'cancelling' causes (resp. 'supporting' causes). Due to the inhibitory connection between the e and $\bar{e}$ cells, only one of them will sustain the competition and be included in the final explanation. Obviously, the acceptance of $\bar{e}$ should be interpreted as the cancellation of e.

Another neural model, proposed by Peng and Reggia (1989), is based on a 'likelihood measure' criterion L developed in Peng and Reggia (1987). The update rules for node activations (manifestations and disorders) are derived from L; and inference of the best explanation is based on the maximization of L (Peng and Reggia 1989). Unfortunately, there is no simple way to implement cancellation interactions in Peng and Reggia's model. In fact, in order to derive L (Peng and Reggia 1987), the causal strengths in Peng and Reggia's model are interpreted as 'positive' influences. However, cancellation interactions have to be interpreted as 'negative' (inhibitory) influences. Although in a probabilistic

framework a 'negative influence' could be modeled by using *low* conditional probabilities, there is no simple way to do this in Peng and Reggia's (1987) proposal without altering the likelihood measure L and thereby the mechanics of the whole model.

In a recent model Goel et al (1995) proposed a new characterization of the abduction task whereby it is perceived as 'inference to a confident explanation' rather than 'inference to the best explanation'. A confident explanation is characterized by the conventional criteria of parsimonious coverage, consistency and plausibility. Briefly, the set of hypotheses is decomposed into *essential* and *non-essential* ones. A hypothesis is said to be essential if it is the *only* one covering at least one observable effect; otherwise it is non-essential. The set of essential hypotheses is then decomposed into *confident* and *non-confident* ones. Next, a concurrent mechanism for synthesizing confident explanations is proposed. In their characterization, Goel et al (1995) mention cancellation and incompatibility interactions, among others.

Regarding cancellation interactions, the authors do not give a method/algorithm to handle them. However, they argued that this new characterization of the abduction task can help in handling incompatibility interactions efficiently, as follows: during the concurrent mechanism, non-essential hypotheses incompatible with the essential hypotheses are discarded from the explanation being constructed (Goel et al, 1995). Unfortunately, the authors have not developed a concise theoretical analysis of this method but instead illustrate its usefulness with a simple example. All of these caveats prevent us from making a fair theoretical comparison between our model and theirs. In particular, we cannot tell with certainty whether the model in Goel et al (1995) is able to mechanize cancellation interactions.

4. Benchmarking Numerical Studies

We consider a medical problem[4] drawn from neurology and composed of 26 diseases (causes), 56 symptoms (effects) and 384 causal interactions. This problem was first used as a testbed in Wald et al (1989), where the reader will find a detailed description of the specific symptoms, diseases and causal associations.

Running SOLVE requires setting of the following free parameters: $\alpha, \beta, \epsilon, \tau, \eta$ and μ. There are no prescribed rules for this; their values are set heuristically, using their role in the algorithm as a guide. For the stopping criterion ϵ, values between 0.01 and 0.1 are recommended. For the learning rates α, β, η and μ, values as small as 0.1 should be used to allow a smooth adaptation of weights. The decay term τ also should be chosen very small, say between 0.01 and 0.1, to prevent the 'sudden' death of a cause cell. In all our simulations, we have used the following parameters: $\epsilon = 0.01, \tau = 0.1, \alpha = \beta = 0.01, \mu = 0.01$ and $\eta = 0.05$. We stress that changing these parameters by small amounts does not affect the quality of the computed scenarios; however, it may result in slower convergence of the algorithm. Now, we need the following notation. We use $CARD(\cdot)$ to denote the set cardinality, $\overline{\mu}(\cdot)$ the *mean*, and $\sigma(\cdot)$ the *standard deviation*.

[4] This causal network was kindly provided to us by Prof. James Reggia.

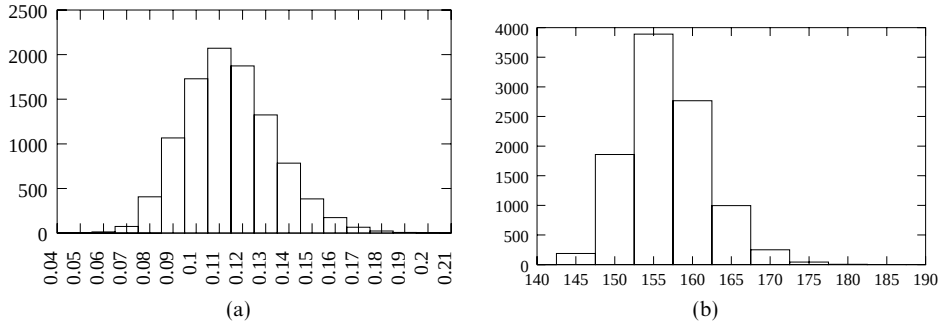


Fig. 10. Simulation results for single manifestations. (a) The $Error_{present}$ criterion; (b) iterations to terminate.

Table 1. First set of experiments – $\overline{\mu}$ and σ of simulation variables.

$\overline{\mu}(Error_{present}) = 0.109$	$\overline{\mu}(Iterations) = 153.7$
$\sigma(Error_{present}) = 0.019$	$\sigma(IIerations) = 5.1$

4.1. Single Manifestations

In this first set of experiments, we consider single manifestations in which there are only present (E_P) and unknown (E_U) effects, with no absent ones (E_A). A possible effect $e_j (j = 1 \cdots 56)$ is randomly assigned the state present or unknown. Moreover, the degree of certainty of a present effect is generated randomly over $]0, 1]$. In this way, we generated 10,000 random single manifestations. We used the following criteria to evaluate the performance of SOLVE: (1) the number of iterations required for SOLVE to terminate; and (2) a measure of the coverage of the present effects by the winning causes, defined by:

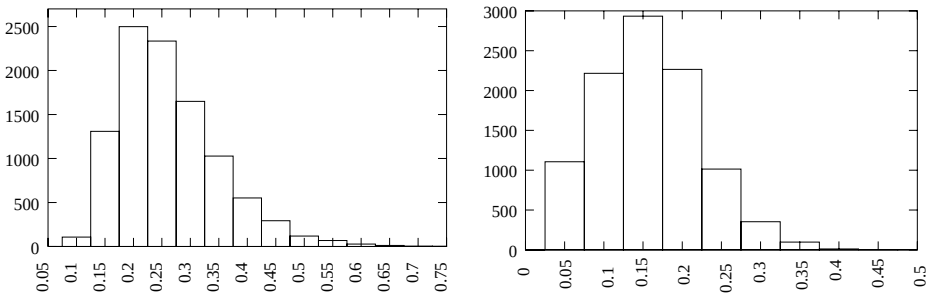
$$Error_{present} = \frac{\sum_{ej \in E_P} (a_j - S_j)^2}{CARD(E_P)} \tag{25}$$

where S_j is given by:

$$S_j = \sum_{cc_i} W_{i,j} f(x_i) \tag{26}$$

Intuitively, $Error_{present}$ quantifies the *average cover error* of the computed scenario with respect to present effects. Typically, $Error_{present} \approx 0$ if all observed effects are covered by the computed scenario.

The simulation results are summarized in Fig. 10 and Table 1. Figure 10 should be read from left to right as follows: in 74 cases, $Error_{present}$ takes values over $[0.06, 0.07]$; and so on. Figure 10 should be read from left to right as follows: in 188 cases, SOLVE required between 140 and 145 iterations to compute a scenario; and so on. To analyze the results for $Error_{present}$ and the number of iterations, we considered the mean and the standard deviation of these two variables – see Table 1. It will be observed that on the average, SOLVE has an error of 0.109 in covering present effects, which could be considered as low. In addition, we have $\sigma(E_{present}) = 0.019$. Consequently, one can conclude that in the majority of



(a) $Error_{present}$ (b) $Error_{absent}$

Fig. 11. Evaluation with respect to $Error_{present}$ and $Error_{absent}$ criteria.

Table 2. Multiple manifestations – μ and σ of simulation variables.

$\overline{\mu}(Error_{present}) = 0.237$	$\overline{\mu}(Error_{absent}) = 0.132$	$\overline{\mu}(Iterations) = 118.4$	$\overline{\mu}(R_{A/P}) = 0.341$
$\sigma(Error_{present}) = 0.088$	$\sigma(Error_{absent}) = 0.067$	$\sigma(Iterations) = 7.2$	$\sigma(R_{A/P}) = 0.105$

cases SOLVE succeeds in covering the present effects with an error of about 0.11. For the number of iterations, we can conclude from Table 1 that SOLVE took approximately 153 iterations to compute a scenario in the majority of cases.

4.2. Multiple Manifestations

In this second set of experiments, we consider multiple manifestations; that is, those composed of present (E_P), unknown (E_U), and absent (E_A) effects. An effect is randomly assigned the state present, absent or unknown. Moreover, the degree of certainty of an absent or a present effect is generated randomly over $]0, 1]$. In this way, we generated 10,000 random manifestations. We used three criteria to evaluate the performance of the algorithm: (1) the number of iterations for SOLVE to terminate; (2) $Error_{present}$, defined by formula 25; and (3) $Error_{absent}$, the coverage of absent effects by the winning causes, defined by:

$$Error_{absent} = \frac{\sum_{ej \in E_A} a_j S_j}{CARD(E_A)} \tag{27}$$

where S_j is by (26). Intuitively, $Error_{absent}$ models how well SOLVE avoids coverage of absent effects. In particular, $Error_{absent}$ is null if none of the surviving causes covers any of the absent effects.

The simulation results are summarized in Figs 11, 12 and 13 and Table 2. For $Error_{absent}$, we can deduce that SOLVE effectively avoids coverage of absent effects, given the small values for its mean and standard deviation presented in Table 2.

Now, let us consider the $Error_{present}$ criterion plotted in Fig. 11. Although its mean and standard deviation values remain small (see Table 2), they are higher than those obtained in the first set of simulations (see Table 1). This is mainly due to the following. Remember that, unlike the first set of simulations (Section 4.1), here we consider the absent effects. An absent effect inhibits the set of cause cells to which it is connected. Hence, a large number of absent effects would lead to strong inhibition of cause cells, whose activation levels will be greatly diminished.

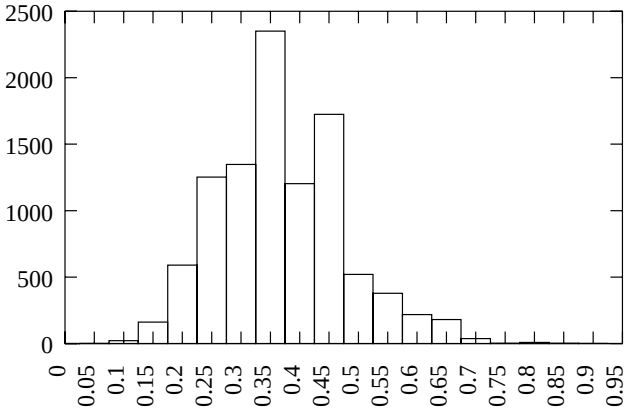


Fig. 12. The ratio $R_{A/P}$.

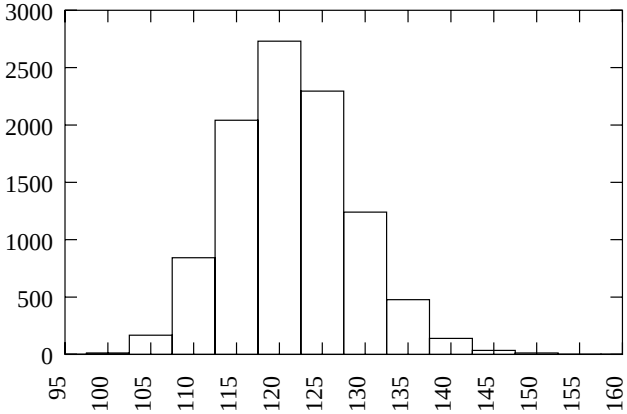


Fig. 13. Iterations to terminate.

Consequently, the S_j values in formula (26) will decrease and the $Error_{present}$ will increase accordingly – see formula (25). To verify this behavior, we define the ratio $R_{A/P}$ as follows:

$$R_{A/P} = \frac{CARD(E_A)}{CARD(E_P)} \tag{28}$$

$R_{A/P}$ measures the ratio of absent effects to present ones; thus, high values of $R_{A/P}$ indicate strong inhibition of cause cells. We observe in Table 2 that $\overline{\mu}(R_{A/P}) = 0.34$ and $\sigma(R_{A/P}) = 0.105$, which means the number of absent effects is approximately one-third the number of present ones in the majority of the manifestations considered. In particular, we can see in Fig. 12 that $R_{A/P}$ is greater than 0.3 in 6626 cases. Doubtless, this explains the relatively high values taken by $Error_{present}$ compared to those obtained in Section 4.1. Despite this strong inhibition of cause cells and the resulting hardness of the competition process, the number of iterations required by SOLVE remains acceptable – see Fig. 13 and Table 2.

To summarize, the performance of SOLVE is good with respect to the proposed criteria. In fact, SOLVE: (1) ensures good coverage of the present effects; (2) effectively avoids coverage of absent effects; and (3) computes scenarios within reasonable time scales.[5]

5. Concluding Remarks and Future Research

The work presented in this paper is twofold. First, we extended the neural model for causal reasoning presented in Ayeb and Wang (1998) to handle the cancellation class of causality. The extended model, called SOLVE, handles a variety of complex causal interactions in the independent, open, incompatibility, monotonic and cancellation classes. Second, we tested SOLVE on a medical problem using a large battery of cases and proposed 'objective' criteria, namely $Error_{present}$ and $Error_{absent}$, to evaluate the model's performance. Intuitively, $Error_{present}$ quantifies the 'average coverage' of present effects. On the other hand, $Error_{absent}$ measures how efficiently SOLVE avoids covering absent effects. Simulation results reveal that SOLVE performs well with respect to these criteria, for both multiple and single manifestations.

Our mechanization of the cancellation class is based on a competitive mechanism between causes. The activation of an observed effect is based on two feedback signals: a 'positive' signal from causes having excitatory interactions and a 'negative' (inhibitory) signal from causes having cancellation interactions. In this way, 'excitatory' causes try to support the observed effect by strengthening its activity. On the other hand, 'cancelling' causes try to reject the same effect by weakening its activity. Thus it could be said that in SOLVE the cancellation competition between causes is implemented by means of effects. We stress the fact that the cancellation competition between causes leads either to acceptance or rejection of the relevant effect. This is an important feature of SOLVE, since we do not restrict ourselves to the single type of cancellation defined in Bylander et al (1991), whereby an effect *must* always be covered. Our modeling of the cancellation class is attractive and may have many practical implications; for example, SOLVE could correct our manifestations in the case of inconsistency[6] in our observations.

Naturally, as with any model, there are shortcomings in the current work. An important issue is related to the experimental standpoint. SOLVE must be compared numerically with other neural-based models. Unfortunately, it is difficult (if not impossible) to define a global criterion that could be used for evaluation. In fact, there are fundamental differences between SOLVE and existing neural-based proposals in terms of neural architecture and mechanisms. Obviously, one could define criteria such as $Error_{present}$ and $Error_{absent}$. However, such a methodology would lack objectivity and be highly arbitrary.

Another important issue is related to the theoretical standpoint. First, we should recall that causal reasoning is NP-Hard. A major question arises at this stage: *Could we circumvent the intractability of causal problems by developing suitable representations or reasoning methods*? As pointed out by many researchers, abduction problems are hard no matter what representations or tools are chosen

[5] In extreme cases, SOLVE took up to 20 seconds of CPU time on a Sun machine to compute a final scenario.

[6] Consistency is taken with respect to an interpretation level.

(see, for example, Bylander et al, 1991; Goel et al, 1995). If there are no tractable algorithms for causal problems, then we have no choice but to solve them heuristically (unless we are willing to wait a long time). Neural-based models fall under this category (see, for example Goel et al, 1988; Peng and Reggia, 1989; Thagard, 1989; Goel et al, 1995; Ayeb and Wang 1998). Unfortunately, in neural-based models (including our model SOLVE) a solution is not guaranteed to be optimal. For example, one can build counter-examples in which SOLVE computes solutions that do not cover some present effects or that include incompatible causes.

From the methodological standpoint, a 'belief' or energy function should be built to characterize the distinct classes of causal reasoning. Such an energy/belief function would provide a means of analyzing the dynamics of SOLVE, and it could be used as a confidence measure to rank the computed scenarios. Moreover, that would make it possible to implement many criteria (for example, minimality of the final scenario) as 'soft' constraints.[7] Naturally, the main advantage would be to considerably reduce the number of connections in the network, and thereby simplify its architecture. These issues are currently under investigation.

Acknowledgements. The authors wish to thank Prof. James Reggia, who kindly provided the medical causal problem used in our benchmarks. Mr L. B. Romdhane is supported by a CIDA (ACDI)/MUET (Montreal) fellowship. Thanks also to NSERC Federal Grants (OGP0121468) awarded to RiFa Lab to support this research.

References

Amari S, Arbib, M (1982) *Competition and cooperation in neural nets.* Springer-Verlag, New York.

Ayeb B (1994) Towards systematic construction of diagnostic systems for large industrial plants: methods, languages and tools. *IEEE Transactions on Knowledge and Data Engineering* 6: 698–712.

Ayeb B, Wang S (1995) Abduction-based diagnosis: a competition-based neural model simulating abductive reasoning. *Journal of Parallel and Distributed Computing*, 24: 202–212.

Ayeb B, Marquis, P, Rusinowitch, M (1993) Preferring diagnoses by abduction. *IEEE Transactions on SMC*, 23: 792–808.

Ayeb B, Ge J, Wang S (1998) A unified model for abduction based reasoning. *IEEE Transactions on SMC*, 28: 408–425.

Brown T, Kairiss E, Keenan, C (1990) Hebbian synapses: biophysical mechanisms and algorithms. *Annual Review of Neuroscience*, 13: 475–511.

Bylander T, Allemang D, Tanner MC, Josephon JR (1991) The computational complexity of abduction. *Artificial Intelligence*, 49: 25–60.

Eliasmith C, Thagard P (1997) Waves, particles, and explanatory coherence. *British Journal for the Philosophy of Science*, 48: 1–19.

Fischer O, Goel A, Svirbely J, Smith J (1991) The role of essential explanations in abduction. *Artificial Intelligence in Medicine*, 3: 181–191.

Goel A, Ramanujam J, Sadayappan, P (1988) Towards a 'neural' architecture of abductive reasoning. In *Proceeding of Second International Conference on Neural Networks*, pp. I-681–I-688.

Goel A, Josephson J, Fischer O, Sadayappan, P (1995) Practical abduction: characterization, decomposition and distribution. *Journal of Experimental & Theoretical Artificial Intelligence*, 7: 429–450.

Grossberg S (1988) Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, 1(1): 17–60.

Josephson JR, Chandrasekaran B, Smith JW, Tanner, MC (1987) A mechanism for forming composite explanatory hypotheses. *IEEE Transcations on SMC*, 17: 445–454.

Kohonen T (1988) Statistical pattern recognition with neural networks: benchmarking studies. In *Proceedings of the International Conference on Neural Networks*, pp. 61–68; San Diego, CA, USA.

[7] By soft constraints, we mean here the use of 'cost' terms in the energy function.

Pearl J (1987) Distributed revision of composed beliefs. *Artificial Intelligence*, 33: 173–215.
Peng Y, Reggia, JA (1987) A probabilistic causal model for diagnostic problem solving (two parts). *IEEE Transactions on SMC*, 17: 146–162/395–406.
Peng Y, Reggia, JA (1989) A connectionist model for diagnostic problem solving. *IEEE Transactions on SMC*, 19: 285–289.
Peterson C, Anderson JR (1988) Neural networks and NP-complete optimization problems: a performance study on the graph bisection problem. *Complex Systems*, 2: 59–89.
Poole, D (1993) Logical programming, abduction and probabilities: a top-down algorithm for computing prior and posterior probabilities. *New Generation Computing*, 11(3-4): 377–400.
Rumelhart D, Zipser D (1985) Feature discovery by competitive learning. *Cognitive Science*, 1985; 9: 75–112.
Thagard P (1989) Explanatory coherence. *Behavioral and brain sciences*, 12: 435–502.
Wald J, Farach M, Tagamets M, Reggia JA (1989) Generating plausible diagnostic hypotheses with self-processing causal networks. *Journal of Experimental & Theoretical Artificial Intelligence*, 2: 91–112.

Author Biographies

Bechir Ayeb is a full professor in Computer Science with the Faculty of Sciences, University of Sherbrooke, P.Q., Canada. In 1990, Dr Ayeb founded RiFa Research Laboratory, focusing on diagnosis, causal reasoning, and recently on diagnosis with respect to algorithmic and distributed systems. He heads RiFa lab, where different aspects of diagnosis are explored. These include theoretical, methodological, and experimental issues. Dr Ayeb has authored/co-authored more than forty research papers on these issues during the last few years. Dr Ayeb is member of IEEE Computer Society.

Lotfi Ben Romdhane was born in Hammam-Sousse, Tunisia. He earned the engineering degree from the Ecole Nationale des Sciences de l'Informatique, Tunisia, in 1994, and the Ph.D. degree at RiFa Laboratory, Sherbrooke University, Canada, in 1999 – both in computer science. His research interests include neural networks, hypothetical reasoning and fuzzy clustering. Dr Lotfi Ben Romdhane was awarded the CIDA (ACDI) Doctoral fellowship from 1995 to 1999.

Shengrui Wang is a full professor at the Department of Mathematics and Computer Science of the University of Sherbrooke. He received his Ph.D. in computer science from Institut National Polytechnique de Grenoble, France, in 1989. His research interests include neural networks, image processing and understanding, content-based image retrieval, automatic target detection, remote sensing and expert systems.

Correspondence and offprint requests to: B. Ayeb, University of Sherbrooke, RiFa Laboratory, Faculty of Sciences, Dept. of Math and Computer Science, Sherbrooke, P.Q., Canada J1K 2R1. Email: ayeb@dmi.userb.ca