

# Psychologic

*Reinhard Blutner, Berlin*

## **1 Introduction: Symbolic and connectionist approaches to cognition**

### **Symbolic paradigm**

(A) The basic units of cognition are (discrete) symbols handled by rule-based processes.

(B) Internal knowledge is represented by rules, principles, algorithms, and other symbol-like means.

(C) The computation performed by the system in transforming input representations to output representations is typically serial and digital in nature.

Problems:

- ☹ Scalability (as the domain grows larger, a system's performance degrades drastically)
- ☹ Robustness
- ☹ Flexibility
- ☹ Gradedness (graded factors determine discrete solutions)
- ☹ Self-organisation

## **Subsymbolic (connectionist) paradigm**

(A') The basic units of cognition are activations of neuronlike elements that interact to produce collectively emerging effects.

(B') Internal knowledge is represented by a matrix of real numbers (connection matrix).

(C') The computation performed by the system in transforming the input pattern of activity to the output pattern is massively parallel and continuously in nature.

## **The proper treatment of connectionism**

### **1. Eliminativist position**

Most concepts from symbolic theory are misguided or superfluous. This concerns, first at all, symbolically structured representations and rules. Such concepts may be eliminated by connectionism. This position represents the mainstream connectionist approach.

### **2. Implementationalist position**

The theses (A) and (B) are basically correct. Replace (C) by the following: The computation performed by the system can be implemented by connectionist aids.

This position is taken by Fodor & Pylyshyn. It aims to eliminate connectionism as a substantive cognitive paradigm.




### 3. Hybrid Systems

Link a current connectionist system with a current (physical) symbol system (exploiting the strengths of each)

### 4. Integrative connectionism

Unification of the symbolic and the connectionist paradigm. Symbolism as a high level description of the properties of neural nets.

Main thesis of this talk: Certain activities of connectionist networks can be interpreted as *nonmonotonic inferences*. In particular, there is a strict correspondence between Hopfield networks and weight-annotated Poole systems.

-  Nonmonotonic logic and algebraic semantics as descriptive and analytic tools for analyzing emerging properties of connectionist networks
-  Connectionist methods (randomised optimisation: simulated annealing) for performing nonmonotonic inferences
-  Certain logical systems are singled out by giving them a "deeper justification".

cf. Balkenius, C. & Gaerdenfors, P. (1991)

## 2 A concise introduction to neural networks

### 2.1 General description

A neural network  $N$  can be defined as a quadruple  $\langle S, F, W, G \rangle$ :

- $S$  Space of all possible states
- $W$  Set of possible configurations.  $w \in W$  describes for each pair  $i, j$  of "neurons" the connection  $w_{ij}$  between  $i$  and  $j$
- $F$  Set of activation functions. For a given configuration  $w \in W$  a function  $f_w \in F$  describes how the neuron activities spread through that network (fast dynamics)
- $G$  Set of learning functions (slow dynamics)

### Hopfield networks

Let the interval  $[-1, +1]$  be the working range of each neuron

**+1: maximal firing rate**

0: resting

**-1 : minimal firing rate**

$$S = [-1, 1]^n$$

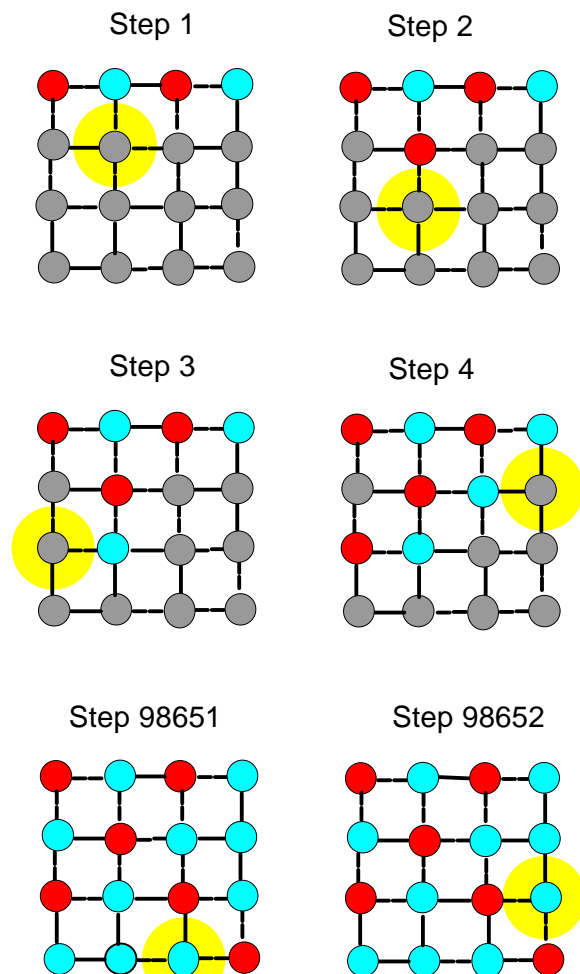
$$w_{ij} = w_{ji}, w_{ii} = 0$$

*Aynchronous Updating:*

$$s_i(t+1) = \Theta \left( \sum_j w_{ij} \times s_j(t) \right),$$

if  $i = \text{random}(1, n)$

$$s_i(t+1) = s_i(t), \text{ otherwise}$$



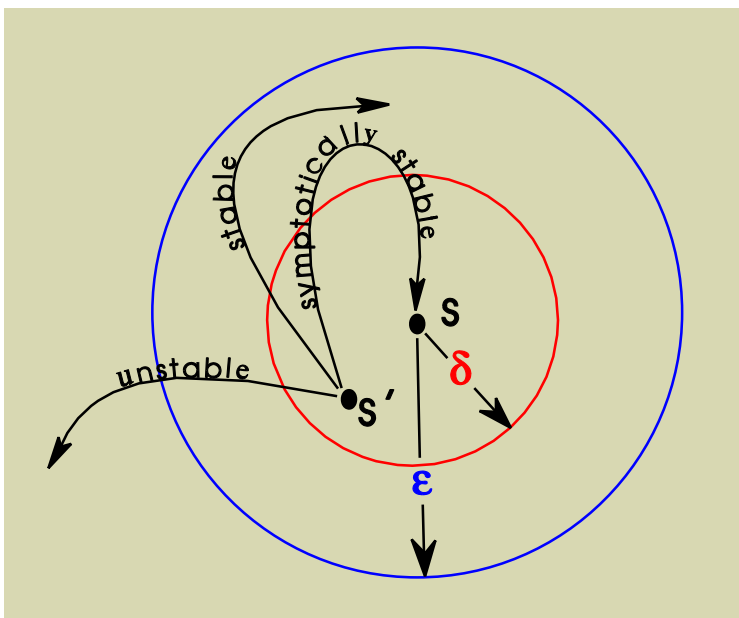
## 2.2 Hopfield networks as resonance systems

Let us consider Hopfield networks as *dynamical systems* (development of activation in time)

### Definition 2.1:

A state  $s \in S$  is called a resonance of a dynamic system  $[S, f]$  iff

1.  $f(s) = s$  (equilibrium)
2. For each  $\varepsilon > 0$  there exists a  $0 < \delta \leq \varepsilon$  such that for all  $n \geq 1$   
 $|f^n(s') - s| < \varepsilon$  whenever  $|s' - s| < \delta$  (stability)
3. For each  $\varepsilon > 0$  there exists a  $0 < \delta \leq \varepsilon$  such that  
 $\lim_{n \rightarrow \infty} f^n(s') = s$  whenever  $|s' - s| < \delta$  (asymptotic stabil.)



The existence of resonances is an emergent collective effect. Intuitively, resonances are the stable states of the network. They *attract* other states. When each state develops into a resonance, then the system produces a content-addressable memory. Such

memories have emergent collective properties (capacity, error correction, familiarity recognition.)

## Definition 2.2

A neural network  $[S, W, F]$  is called a resonance system iff  $\lim_{n \rightarrow \infty} (f^n(s))$  exists and is a resonance for each  $s \in S$  and  $f \in W$ .

### Theorem 2.1 (Cohen & Grossberg 1983):

Hopfield networks are resonance systems.

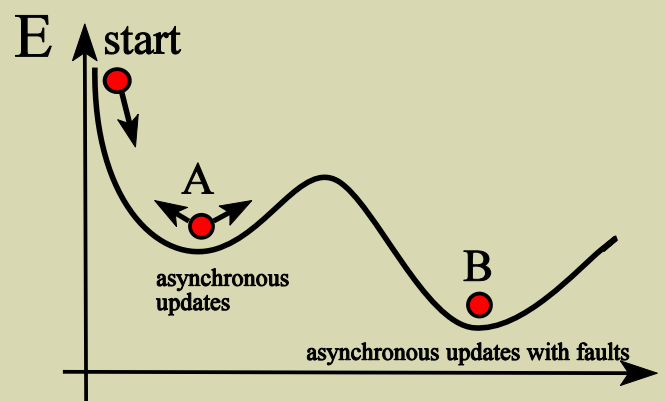
(The same holds for a large class of other systems: The McCulloch-Pitts model (1943), Cohen-Grossberg models (1983), Rumelhart's Interactive Activation model (1986), Smolensky's Harmony networks (1986), etc.)

### Theorem 2.2 (Hopfield 1982)

The function  $E(s) = -\sum_{i>j} w_{ij} \cdot s_i \cdot s_j$  is a Ljapunov-function of the system in the case of asynchronous updates. I.e., when the activation state of the network changes,  $E$  can either decrease or remain the same. The output states  $\lim_{n \rightarrow \infty} (f^n(s))$  can be characterized as *the local minima* of the Ljapunov-function.

### Theorem 2.3 (Hopfield 1982)

The output states  $\lim_{n \rightarrow \infty} (f^n(s))$  can be characterized as *the global minima* of the Ljapunov-function if certain stochastic update functions  $f$  are considered ("simulated annealing").



### 3 Information states in Hopfield networks

*Activation states can be partially ordered in accordance with their informational content*

<b>+1: maximal firing rate</b>	}	indicating maximal
<b>-1: minimal firing rate</b>		specification
<b>0: resting</b>		indicating underspecification

#### Definition 3.1

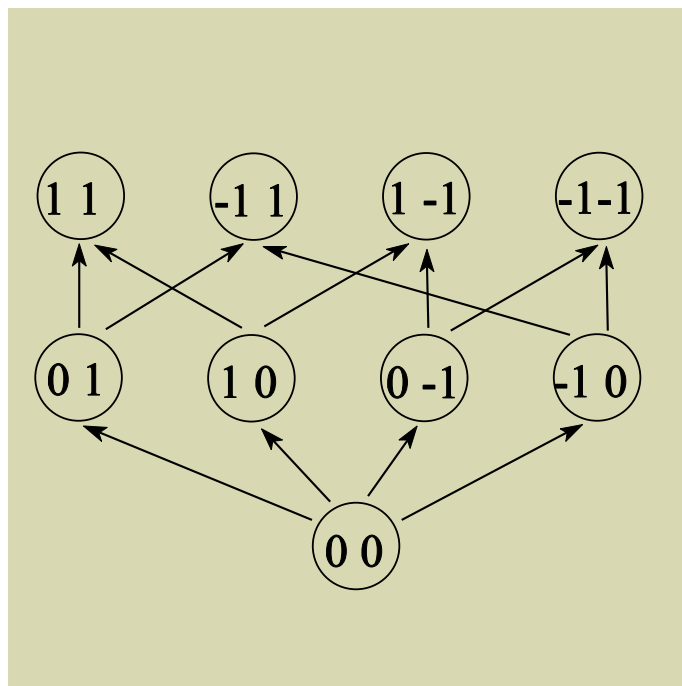
$\langle S, \geq \rangle$  is a poset of activation states iff

- (i)  $S = [-1, +1]^n$  (set of activation states)
- (ii)  $s, t \in S$ :  $s \geq t$  iff  $s_i \geq t_i \geq 0$  or  $s_i \leq t_i \leq 0$ , for all  $1 \leq i \leq n$ .

$s \geq t$  can be read as *s is at least as informative as t*, or *s is at least as specific as t*.

Poset of information states  
for  $n=2$ .

This poset doesn't form a  
lattice.

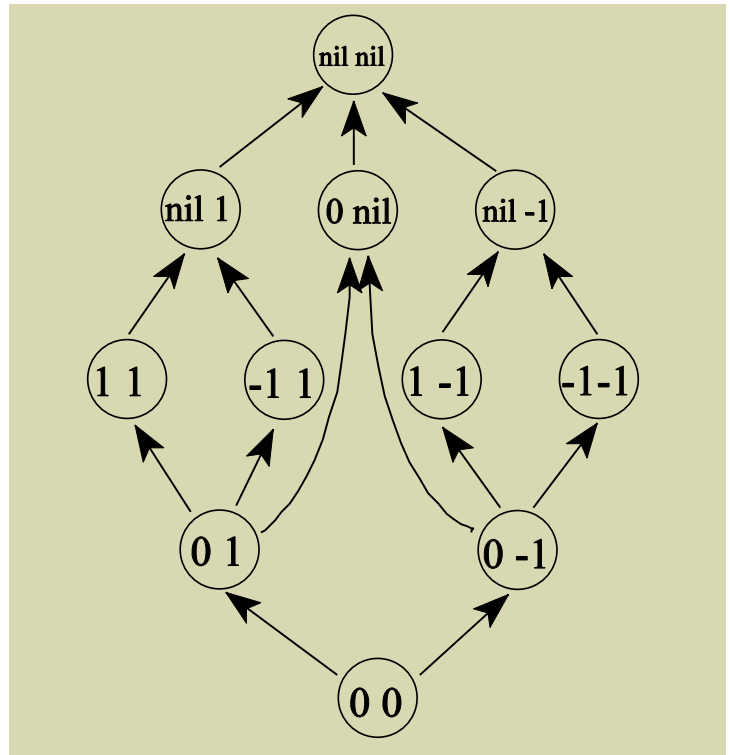


However, it can be extended to a lattice by introducing *impossible activation states*. Write "nil" for the impossible activation of an element.

### Definition 3.2

$\langle S \cup \perp, \geq \rangle$  is the extended poset of activation states iff

- (i)  $S = [-1, 1]^n$   
(the set of proper activation states)
- (ii)  $\perp = ([-1, 1] \cup \{\text{nil}\})^n - [-1, 1]^n$   
(the set of impossible activation states).
- (iii) for each  $s, t \in S$ :  $s \geq t$  iff  $s_i = \text{nil}$  or  $s_i \geq t_i \geq 0$  or  $s_i \leq t_i \leq 0$ , for all  $1 \leq i \leq n$ .



### Fact 3.1

The extended poset of activation states  $\langle S \cup \perp, \geq \rangle$  forms a DeMorgan lattice. The operation  $\sup\{s, t\} = s \circ t$  (CONJUNCTION) can be interpreted as the *simultaneous realization* of two activation states; the operation  $\inf\{s, t\} = s \oplus t$  (DISJUNCTION) can be interpreted as some kind of *generalization* of two instances of activation states; the COMPLEMENT  $s^*$  reflects a *lack* of information. The operations come out as follows:



$$(s \circ t)_i = \begin{cases} \max(s_i, t_i), & \text{if } s_i, t_i \geq 0 \\ \min(s_i, t_i), & \text{if } s_i, t_i \leq 0 \\ \text{nil}, & \text{elsewhere} \end{cases}$$

$$(s \oplus t)_i = \begin{cases} \min(s_i, t_i), & \text{if } s_i, t_i \geq 0 \\ \max(s_i, t_i), & \text{if } s_i, t_i \leq 0 \\ s_i, & \text{if } t_i = \text{nil} \\ t_i, & \text{if } s_i = \text{nil} \\ 0, & \text{elsewhere} \end{cases}$$

$$(s^*)_i = \begin{cases} 1-s_i, & \text{if } s_i > 0 \\ -1-s_i, & \text{if } s_i < 0 \\ \text{nil}, & \text{if } s_i = 0 \\ 0, & \text{if } s_i = \text{nil} \end{cases}$$

The fact that the extended poset of activation states forms a DeMorgan lattice gives the opportunity to interpret these states as propositional objects ("information states").

## 4 Asymptotic updates of information states

### 4.1 Asymptotic updates with clamping

In general, updating an information state  $s$  may result in a information state  $f \dots f(s)$  that doesn't include the information of  $s$ . However, for the following is important to interpret updating as specification. If we want  $s$  to be informationally included in the resulting update, we have to "clamp"  $s$  somehow in the network. A technical way to do that is as follows:

#### Definition 4.1

Let  $f$  be a (stochastic) update function. Define the following *update function with clamping* (cf. Balkenius & Gaerdenfors):

$$\underline{f}(s) = f(s) \circ s; \quad \underline{f}^{n+1}(s) = f(\underline{f}^n(s)) \circ s$$

#### Definition 4.2

Let  $\langle S, W, F \rangle$  be a resonance system with connection matrix  $w$  and an asynchronous (stochastic) update functions  $f$ . The asymptotic updates of  $s$  (with clamping) are defined as follows:

$$ASUP_w(s) = \{t: t = \lim_{n \rightarrow \infty} \underline{f}^n(s)\}$$

### 4.2 Energy-minimal specifications of activation states

From another perspective, we can consider specifications of  $s$  which minimize some cost function  $E$ .

**Definition 4.3**

Let  $\langle S, \geq \rangle$  be a poset of activation states,  $E$  a real function on  $S$ .

The  $E$ -minimal specifications of  $s$  are defined as follows:

$$\min_E[s] = \{t: t \geq s \text{ and there is no } t' \geq s \text{ such that } E(t') < E(t)\}$$

**Fact 4.1** (Consequence of theorem 2.3):

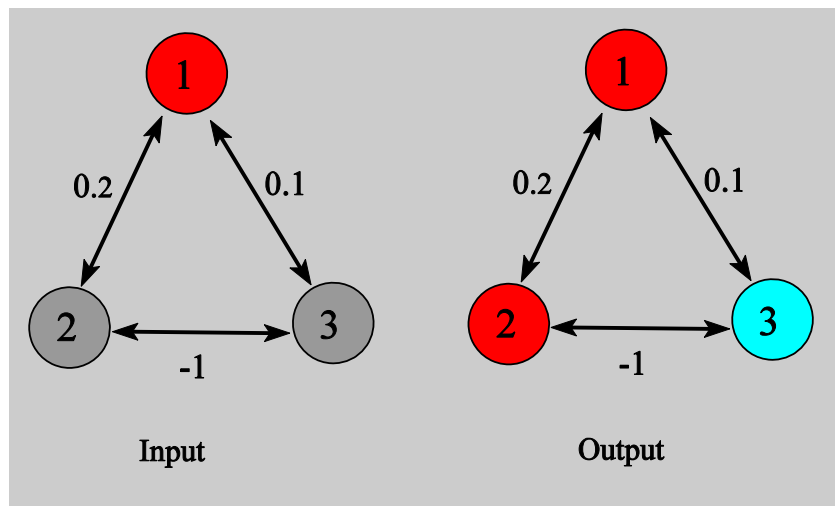
Consider Hopfield nets with asynchronous (stochastic) updates.

Let  $E(s) = -\sum_{i>j} w_{ij} \cdot s_i \cdot s_j$  be the Ljapunov-function of the system.

Then it holds:  $ASUP_w(s) = \min_E(s)$

**Example**

$$w = \begin{pmatrix} 0 & 0.2 & 0.1 \\ 0.2 & 0 & -1 \\ 0.1 & -1 & 0 \end{pmatrix}$$



	$E$
$\langle 1 \ 0 \ 0 \rangle \leq \langle 1 \ 0 \ 0 \rangle$	0
$\langle 1 \ 0 \ 1 \rangle$	-0.1
$\langle 1 \ 1 \ 0 \rangle$	-0.2
$\langle 1 \ 1 \ 1 \rangle$	0.7
$\langle 1 \ 1 \ -1 \rangle$	-1.1

$$ASUP_w(\langle 1 \ 0 \ 0 \rangle) = \min_E(s) = \langle 1 \ 1 \ -1 \rangle$$

### 4.3 Asymptotic updates and nonmonotonic inference

The propositional objects called information states are related by the partial ordering  $\geq$ . It is obvious that this relation can be interpreted as a strict entailment relation. In any case it satisfies the Tarskian restrictions for such an relation:

- $s \geq s$  (REFLEXIVITY)  
 if  $s \geq t$  and  $s \circ t \geq u$ , then  $s \geq u$  (CUT)  
 if  $s \geq u$ , then  $s \circ t \geq u$  (MONOTONICITY)

More interesting, Balkenius & Gaerdenfors (1991) have made clear that it is possible to define a nonmonotonic inference relation that reflects asymptotic updating of information states. Let  $\langle S, \geq \rangle$  be a poset of activation states,  $w$  the connection matrix and  $E$  the energy function. I consider two possibilities to define a nonmonotonic inferential relation (NIR)  $\sim$ :

#### Definition 4.4

- (A)  $s \sim_w t$  iff  $s' \geq t$  for each  $s' \in \text{ASUP}_w(s)$   
 (NIR based upon asymptotic updates)  
 (B)  $s \sim_E t$  iff  $s' \geq t$  for each  $s' \in \text{min}_E(s)$   
 (NIR based upon E-minimal specifications)

As an immediate consequence of fact 4.1 we find that both possibilities define the same relation, i.e.  $s \sim_w t$  iff  $s \sim_E t$ .

Furthermore, it is not difficult to prove the following facts:

### Facts 4.2

- |  |                          |
|--|--------------------------|
| (i) if $s \geq t$ , then $s \sim_E t$                              | SUPRACLASSICALITY        |
| (ii) $s \sim_E s$  | REFLEXIVITY              |
| (iii) if $s \sim_E t$ and $s \circ t \sim_E u$ , then $s \sim_E u$ | CUT                      |
| (iv) if $s \sim_E t$ and $s \sim_E u$ , then $s \circ t \sim_E u$  | CAUTIOUS<br>MONOTONICITY |

### Proof

I will only treat CUT and CAUTIOUS MONOTONICITY.

For CUT, suppose all E-minimal specifications of  $s$  are specifications of  $t$  and all E-minimal specifications of  $s \circ t$  are specifications of  $u$ . Suppose any E-minimal specification  $s'$  of  $s$ .  $s'$  specifies both  $s$  as  $t$ , and, consequently, it specifies  $s \circ t$ . Since  $s \circ t \geq s$ , it results that  $s'$  is also a E-minimal specification of  $s \circ t$ . Consequently, it is a specification of  $u$ .

For CAUTIOUS MONOTONICITY, suppose all E-minimal specifications of  $s$  are specifications of  $t$  and  $u$ . We have to prove  $v \geq u$  for each E-minimal specification  $v$  of  $s \circ t$ . Assume any E-minimal specification  $v$  of  $s \circ t$ . Of course,  $v$  is a specification of  $s$ . We shall prove now that  $v$  is a E-minimal specification of  $s$ . If this were wrong, there would be a E-minimal specification  $v'$  of  $s$  such that  $E(v') < E(v)$ . But all E-minimal specifications of  $s$  are specifications of  $t$ , therefore  $v' \geq t$  and  $v' \geq s \circ t$ . This contradicts the E-minimality of  $v$  with respect to the specifications of  $s \circ t$ . Therefore  $v$  must be a E-minimal specification of  $s$ . Since all E-minimal specifications of  $s$  are specifications of  $u$ , one concludes that  $v \geq u$ . ■

Gabbay, Makinson, Gärdenfors, Kraus, Lehmann, Magidor, and others call such nonmonotonic consequence relations *cumulative*.  
CUMULATIVITY: If  $s \sim_E t$  and  $t \sim_E s$ , then  $s \sim_E u$  iff  $t \sim_E u$ .

## 5 Weight-annotated Poole systems

Knowledge base in

- (a) connectionist systems:
  - connection matrix
  - energy function
- (b) symbol systems
  - strong and weak (default-) rules

At least for Hopfield systems there is a strict relationship between connectionist and symbolic knowledge bases.

- ☺ Symbolic systems can be used to understand connectionist systems.
- ☺ Connectionist systems can be used to perform inferences.

### 5.1 Basic notions (cf. Poole 1988, 1994)

Let us consider the language  $L_{At}$  of propositional logic (referring to the alphabet  $At$  of atomic symbols)

#### Definition 5.1

A triple  $\langle At, \Delta, g \rangle$  is called a weight-annotated Poole system iff

- (i)  $At$  is a nonempty set (of atomic symbols)
- (ii)  $\Delta$  is a set of consistent sentences built on the basis of  $At$  (the possible hypotheses)
- (iii)  $g: \Delta \rightarrow [0,1]$  (the weight function)

**Definition 5.2**

Let  $T = \langle At, \Delta, g \rangle$  be a weight-annotated Poole system, and let  $\alpha$  be a consistent formula.

- (A) *A scenario of  $\alpha$  in  $T$*  is a subset  $\Delta'$  of  $\Delta$  such that  $\Delta' \cup \{\alpha\}$  is consistent.
- (B) *The weight of a scenario  $\Delta'$*  is
 
$$G(\Delta') = \sum_{\delta \in \Delta'} g(\delta) - \sum_{\delta \in (\Delta - \Delta')} g(\delta)$$
- (C) *A maximal scenario of  $\alpha$  in  $T$*  is a scenario the weight of which is not exceeded by any other scenario (of  $\alpha$  in  $T$ ).

**Definition 5.3**


Let  $T$  be a weight-annotated Poole system. Then the following cumulative consequence relation can be defined:

$\alpha \succ_{-T} \beta$  iff  $\beta$  is an ordinary consequence of each maximal scenario of  $\alpha$  in  $T$ .

## An elementary example

$$At = \{p_1, p_2, p_3\}$$

$$\Delta = \{p_1 \leftrightarrow_{0.2} p_2, p_1 \leftrightarrow_{0.1} p_3, p_2 \leftrightarrow_{1.0} \sim p_3\}$$

some (relevant) scenarios of $p_1$ :	G
$\{\}$	-1.3
$\{p_1 \leftrightarrow p_2\}$	-0.9
$\{p_1 \leftrightarrow p_2, p_1 \leftrightarrow p_3\}$	-0.7
$\{p_1 \leftrightarrow p_2, p_2 \leftrightarrow \sim p_3\}$	1.1 
$\{p_1 \leftrightarrow p_3, p_2 \leftrightarrow \sim p_3\}$	0.9

Consequently,  $p_1 \supset_{-T} p_2, p_1 \supset_{-T} \neg p_3$

## 5.2 The semantics of weight-annotated Poole systems

Let  $T = \langle At, \Delta, g \rangle$  be a weight-annotated Poole system, with  $At = \{p_1, \dots, p_n\}$ . Furthermore, let  $v$  denote a (total) interpretation function for the propositional language  $L_{At}$ :

$$v: At \mapsto \{-1, 1\}.$$

The usual clauses apply for the evaluation of the formulas of  $L_{At}$  relative to  $v$ :

$$[\alpha \wedge \beta]_v = \min([\alpha]_v, [\beta]_v)$$

$$[\alpha \vee \beta]_v = \max([\alpha]_v, [\beta]_v)$$

$$[\sim \alpha]_v = -[\alpha]_v.$$



The following defines a function which indicates how strong a given interpretation  $v$  conflicts with the space of hypotheses  $\Delta$ :

**Definition 5.4**

$$\mathcal{E}(v) = -\sum_{\delta \in \Delta} g(\delta) \cdot \llbracket \delta \rrbracket_v \quad (\text{the energy of the interpretation})$$

Next, the notions of *model* and *preferred model* can be defined as follows:

**Definition 5.5**

- (A) An interpretation  $v$  is called a *model* of  $\alpha$  just in case  $\llbracket \alpha \rrbracket_v = 1$ .
- (B) An interpretation  $v$  is called a *preferred model* of  $\alpha$  just in case it is a model of  $\alpha$  with minimal energy (w.r.t. the other models of  $\alpha$ ).

For any weight-annotated Poole system  $T = \langle At, \Delta, g \rangle$ , the following definition associates a scenario  $sc(\Delta, v)$  with each model  $v$ :

**Definition 5.6**

$$sc(\Delta, v) =_{\text{def}} \{ \delta \in \Delta : \llbracket \delta \rrbracket_v = 1 \}$$

**Fact 5.1**

With regard to the weight-annotated Poole system  $T = \langle At, \Delta, g \rangle$  and any model  $v$ :

$$G(\text{sc}(\Delta, v)) = - \mathcal{E}(v)$$

**Proof**

$$\begin{aligned} G(\text{sc}(\Delta, v)) &= \sum_{\delta \in \Delta \ \& \ [\delta]_v = 1} g(\delta) - \sum_{\delta \in \Delta \ \& \ [\delta]_v = -1} g(\delta) \\ &= \sum_{\delta \in \Delta} g(\delta) \cdot [\delta]_v = - \mathcal{E}(v) \quad \blacksquare \end{aligned}$$

**Fact 5.2**

Let  $T = \langle At, \Delta, g \rangle$  be a weight-annotated Poole system,  $\alpha$  a consistent formula,  $\Delta'$  a maximal scenario of  $\alpha$  in  $T$ , and  $v$  a model of  $\{\alpha\} \cup \Delta'$ . Then  $\text{sc}(\Delta, v) = \Delta'$ .

**Proof**

In order to show that  $\text{sc}(\Delta, v) \subseteq \Delta'$ , let's take any  $\delta \in \text{sc}(\Delta, v)$ . In case that  $\delta \notin \Delta'$ , the set  $\{\delta\} \cup \Delta'$  would be a scenario of  $\alpha$  in  $T$ . Because of  $G(\{\delta\} \cup \Delta') = G(\Delta') + 2G(\delta)$ , the set  $\Delta'$  would be not a *maximal* scenario. However, this conflicts with the premises. Consequently, we have shown that  $\delta \in \Delta'$ .

In order to show that  $\text{sc}(\Delta, v) \supseteq \Delta'$ , assume any  $\delta \in \Delta'$ . It follows  $[\delta]_v = 1$  and  $\delta \in \text{sc}(\Delta, v)$ .  $\blacksquare$

The following notion is the semantic counterpart to the syntactic consequence relation  $\alpha \supset_{-T} \beta$ :

**Definition 5.7**

$\alpha \supset_{=T} \beta$  iff each preferent model of  $\alpha$  is a model of  $\beta$ .

There is the following soundness and completeness result:

### Theorem 5.1

For all formulas  $\alpha$  and  $\beta$  of  $L_{At}$ :  $\alpha \succ_{-T} \beta$  iff  $\alpha \succ_{=T} \beta$ .

### Proof

It is sufficient to show that the following clauses are equivalent:

- (A) There is a maximal scenario  $\Delta'$  of  $\alpha$  in  $T$  such that  $\{\alpha\} \cup \Delta' \cup \{\neg\beta\}$  is consistent.
- (B) There is a preferent model of  $\alpha$  such that  $\llbracket \beta \rrbracket_v = -1$ .

(A)  $\Rightarrow$  (B): Let's assume that  $\Delta'$  is a maximal scenario of  $\alpha$  in  $T$  and  $v$  is a model of  $\{\alpha\} \cup \Delta' \cup \{\neg\beta\}$ . Show that  $v$  is a preferent model of  $\alpha$  in  $T$ ; i.e., show that for any model  $v'$  of  $\alpha$  in  $T$ ,  $\mathcal{E}(v') \geq \mathcal{E}(v)$ . From fact 5.1 it follows that  $\mathcal{E}(v') = -G(\text{sc}(\Delta, v'))$ , and the facts 5.1 & 5.2 necessitate  $\mathcal{E}(v) = -G(\Delta')$ . Since  $\text{sc}(\Delta, v')$  is a scenario of  $\alpha$  in  $T$  and  $\Delta'$  is a maximal scenario, it follows that  $\mathcal{E}(v') \geq \mathcal{E}(v)$ .

(B)  $\Rightarrow$  (A): Assume a preferent model  $v$  of  $\alpha$  and assume  $\llbracket \beta \rrbracket_v = -1$ . Obviously, the set  $\text{sc}(\Delta, v) \cup \{\alpha\} \cup \{\neg\beta\}$  is consistent ( $v$  is a model of it). We have to show now that the scenario  $\text{sc}(\Delta, v)$  is a maximal scenario of  $\alpha$  in  $T$ . Otherwise there would exist a maximal scenario  $\Delta'$  with  $G(\Delta') > G(\text{sc}(\Delta, v))$ . Because we have  $G(\Delta') = -\mathcal{E}(v')$  for any model  $v'$  of  $\{\alpha\} \cup \Delta'$  and  $G(\text{sc}(\Delta, v)) = -\mathcal{E}(v)$  [facts 5.1 & 5.2], this would contradict the assumption that  $v$  is a preferent model of  $\alpha$  in  $T$ . ■

## 6 Integrating Poole systems and Hopfield nets

Bringing about the correspondence between connectionist and symbolic knowledge bases, we have first to look for a symbolic representation of information states.

### 6.1 Symbolic representation of information states

We consider neural networks with  $n$  elements and we take the elementary language  $L_{A_t}$  (with  $A_t = \{p_1, \dots, p_n\}$ ) in order to speak about the activation states of the net. The symbol  $p_i$  are intended as corresponding to the node  $i$  of the network.

Intuitively, the expressions of the language  $L_{A_t}$  may provide *a symbolic means* to speak about activation states. Following usual practice of *algebraic semantics*, we can grasp this idea by interpreting the non-logical symbols of the language in terms of activation states. Some logical symbols of the language may be (re)interpreted as certain operations on the algebra of information states:

#### Definition 6.1

Let  $\langle S_{U\perp}, \geq \rangle$  be the extended poset of activation states for a neural network with  $n$  elements.

(A) The triple  $\langle S_{U\perp}, \geq, \uparrow \downarrow \rangle$  is called *a Hopfield model* (for  $L_{A_t}$ ) iff  $\uparrow \downarrow$  is a function assigning some element of  $S_{U\perp}$  to each

atomic symbol and obtaining the following conditions:

$$\uparrow \alpha \wedge \beta \downarrow = \uparrow \alpha \downarrow \circ \uparrow \beta \downarrow, \uparrow \sim \alpha \downarrow = -\uparrow \alpha \downarrow.$$

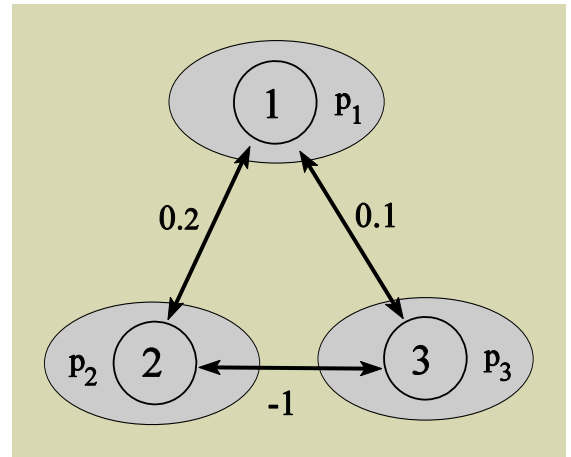
(B) A Hopfield model is called *local* (for  $L_{At}$ ) iff it realizes the following assignments:

$$\uparrow p_1 \downarrow = \langle 1 \ 0 \ \dots \ 0 \rangle$$

$$\uparrow p_2 \downarrow = \langle 0 \ 1 \ \dots \ 0 \rangle$$

...

$$\uparrow p_n \downarrow = \langle 0 \ 0 \ \dots \ 1 \rangle$$



### Definition 6.2

An information state  $s$  is said to be *represented* by a formula  $\alpha$  of  $L_{At}$  (relative to a Hopfield model  $M$ ) iff  $\uparrow \alpha \downarrow = s$ .

In our **example**, the following formulae *represent* proper activation states:

$p_1$  represents  $\langle 1 \ 0 \ 0 \rangle$

$p_2$  represents  $\langle 0 \ 1 \ 0 \rangle$

$p_3$  represents  $\langle 0 \ 0 \ 1 \rangle$

$p_1 \wedge p_2$  represents  $\langle 1 \ 1 \ 0 \rangle$

$\sim p_1$  represents  $\langle -1 \ 0 \ 0 \rangle$

$p_1 \wedge p_2 \wedge \sim p_3$  represents  $\langle 1 \ 1 \ -1 \rangle$

Note, that in contrast to the complement  $s^*$  (reflecting a *lack* of information), the *inner negation*  $-s$  realizes a conversion from positive into negative information, and *vice versa*.

A state  $s \in S$  is said to be *symbolic* (relative to  $M$ ) iff it can be represented by some formula  $\alpha$  in  $L_{At}$ . With regard to a local model each state is symbolic, and it can be represented by a conjunction of literals (atoms or their inner negation).

## 6.2 Translating Hopfield networks into weight-annotated Poole systems

Local Hopfield models give the opportunity to relate connectionist and symbolic knowledge bases in a way that allows to represent nonmonotonic inferential relation (NIRs) based upon asymptotic updates by inferences in weight-annotated Poole systems. The crucial point is the translation of connection matrixes  $w$  into associated Poole-system  $T_w$ :

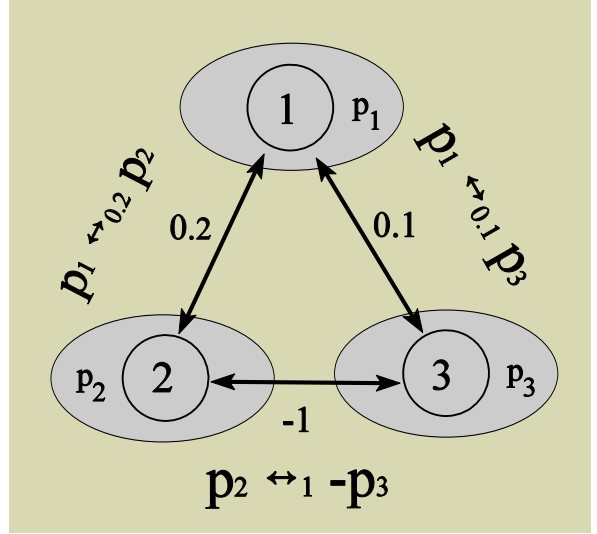
### Definition 6.3

Consider a Hopfield system ( $n$  neurons) with connection matrix  $w$ , and let  $At = \{p_1, \dots, p_n\}$  be a set of atomic symbols. Take the following formulae of  $L_{At}$ :

$$\alpha_{ij} = (p_i \leftrightarrow \text{sign}(w_{ij}) p_j), \text{ for } 1 \leq i < j \leq n.$$

For each connection matrix  $w$  the associated Poole system is defined as  $T_w = \langle At, \Delta_w, g_w \rangle$  where the following clauses apply:

- (i)  $\Delta_w = \{ \alpha_{ij} : 1 \leq i < j \leq n \}$
- (ii)  $g_w(\alpha_{ij}) = |w_{ij}|$



In Section 4 updating information states came out as a kind of specification. For these systems it is simply to show that each (partial) information state tends toward completion. I.e., in case the class of asymptotic updates of  $s$ ,  $ASUP_w(s)$ , contains any partial information state, then it contains one of its total specifications. Taking some additional condition (no isolated nodes) it can be shown that  $ASUP_w(s)$  contains only total information states. In this case, each information state is completed asymptotically. In the following I consider only this case.

As a matter of fact, each total information state  $t$  corresponds to a total propositional interpretation function  $v/t$  where  $v/t(p_i) = t_i$ . Now the following facts are simply to prove:

### Facts 6.1

- (i)  $\llbracket p_i \rrbracket_{v/t} = t_i$
- (ii)  $\llbracket \sim \alpha \rrbracket_{v/t} = -\llbracket \alpha \rrbracket_{v/t}$

$$(iii) \llbracket \alpha \leftrightarrow \beta \rrbracket_{v/t} = \llbracket \alpha \rrbracket_{v/t} \cdot \llbracket \beta \rrbracket_{v/t}$$

(iv)  $t \geq \alpha$  iff  $\llbracket \alpha \rrbracket_{v/t} = 1$ , in case that  $\alpha$  is a conjunction of literals

$$(v) \mathcal{E}(v/t) = E(t) \quad (\text{i.e. } \sum_{\alpha \in \Delta} g(\alpha) \cdot \llbracket \alpha \rrbracket_{v/t} = \sum_{i>j} w_{ij} \cdot t_i \cdot t_j)$$

here  $E$  is the energy function of a Hopfield network with the connection matrix  $w$  and  $\mathcal{E}$  is the energy function of the weight-annotated Poole-system  $T_w$ .

The following theorem states that NIRs based upon asymptotic updates can be represented by inferences in weight-annotated Poole systems.

### Theorem 6.1

Assume that the formulae  $\alpha$  and  $\beta$  are conjunctions of literals. Assume further that the Poole system  $T$  is associated to the connection matrix  $w$ . Then

$$\llbracket \alpha \rrbracket \vdash_{w} \llbracket \beta \rrbracket \quad \text{iff } \alpha \supset_{=T} \beta \quad (\text{iff } \alpha \supset_{-T} \beta)$$

### Proof

Exercise (use theorem 5.1, facts 6.1, and the fact that with regard to a local Hopfield model each state is symbolic and can be represented by a conjunction of literals)



## 7 Conclusions

- ☺ Weight-annotated Poole systems can be used to understand connectionist systems. Nonmonotonic inferences ( $\alpha \triangleright_{-T} \beta$ ) as an analytic tool to understand emerging properties of connectionist networks.
- ☺ Weight-annotated Poole systems are singled out by giving them a "deeper justification".
- ☺ Connectionist systems can be used to perform non-monotonic inferences. Efficiency?

## Appendix: A simple example from phonology

Consider the following fragment of the English vocal system:

-back	+back	
/i/	/u/	+high
/e/	/o/	-high/-low
/æ/	/ɔ/	+low
	/a/	

The phonological features may be represented as by the atomic symbols BACK, LOW, HIGH, ROUND. The generic knowledge of the phonological agent concerning this fragment

may be represented as a Hopfield network using exponential weights with basis  $0 < \epsilon \leq 0.5$ . Furthermore, make use of the following **Strong Constraints**:

LOW  $\rightarrow \sim$  HIGH;      ROUND  $\rightarrow$  BACK

VOC		/a/	/i/	/o/	/u/	/ɔ/	/e/	/æ/
BACK	$\epsilon^1$	+	-	+	+	+	-	-
LOW	$\epsilon^2$	+	-	-	-	+	-	+
HIGH	$-\epsilon^4$	-	+	-	+	-	-	-
ROUND		-	-	+	+	+	-	-

### Assigned Poole-system

VOC  $\leftrightarrow_{\epsilon^1}$  BACK;      BACK  $\leftrightarrow_{\epsilon^2}$  LOW  
 LOW  $\leftrightarrow_{\epsilon^4}$   $\sim$  ROUND;      BACK  $\leftrightarrow_{\epsilon^3}$   $\sim$  HIGH

(These default rules are in strict correspondence to Keane's markedness conventions)

## References:

- Balkenius, C. & Gärdenfors, P. (1991): "Nonmonotonic inferences in neural networks". In J.A. Allen, R. Fikes, & E. Sandewall (Eds.), *Principles of knowledge representation and reasoning*. San Mateo, CA: Morgan Kaufmann.
- Derthick, M. (1990): "Mundane reasoning by settling on a plausible model". *Artificial Intelligence* 46, 107-157
- Hopfield, J.J. (1982): "Neural networks and physical systems with emergent collective computational abilities". *Proceedings of the National Academy of Sciences* 79, 2554-2558.
- Poole, D. (1988): "A logical framework for default reasoning". *Artificial Intelligence*, 36, 27-47.
- Poole, D. (1996): "Who chooses the assumptions?". In P. O'Rorke (Ed.), *Abductive Reasoning*. Cambridge: MIT Press.
- Smolensky, P. (1988): "On the proper treatment of connectionism". *Behavioral and Brain Sciences* 11, 1-23.
- Smolensky, P. (1996): "Computational, dynamical, and statistical perspectives on the processing and learning problems in neural network theory". In P. Smolensky, M.C. Mozer, & D.E. Rumelhart (Eds.), *Mathematical Perspectives on Neural Networks*. Mahwah, NJ: Lawrence Erlbaum Publishers. 1-13.
- Steedman; M. (1997): Connectionist sentence processing in perspective. Unpublished ms.