Nonmonotonic Inferences and Neural Networks

Reinhard Blutner (UvA)
Stanford, CA, June 30th, 2002
NASSLLI Workshop on Cognition

1 Introduction

- Puzzle: Gap between symbolic and subsymbolic (neuron-like) modes of processing
- Aim: Overcoming the gap by viewing symbolism as a high-level description of the properties of neural networks
- Method: standard methods of model-theoretic and algebraic semantics. Neural (Re)interpretation of information states as activation states of a neuronal network.
- Main thesis: Certain activities of connectionist networks can be interpreted as *nonmonotonic inferences*. In particular, there is a strict correspondence between certain network types and particular nonmonotonic inferential systems
- Optimality Theory as a general framework that integrates insights from symbolism and connectionism

Intended results

- e Better understanding of connectionist networks: Nonmonotonic logic and algebraic semantics as descriptive and analytic tools for analyzing their emerging properties
- e New methods for performing nonmonotonic inferences: Connectionist methods (randomised optimisation: simulated annealing) can be adopted for realizing symbolic inferences
- e Certain logical systems are singled out by giving them a "deeper justification".
- e Understanding Optimality Theory: Which assumptions have a deeper foundation and which ones are pure stipulations?

Overview

1 Introduction
2 Connectionism and symbolism
3 The idea of underspecification
4 A concise introduction to neural networks
5 Information states as neural activation patterns
6 Asymptotic updates and nonmonotonic inference
7 Relating connectionism and symbolism
8 Some remarks on Optimality Theory
9 Conclusions

2 Connectionism and symbolism

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 basic units of cognition are (discrete) symbols handled by rule-based processes. Internal knowledge is represented by rules, principles, algorithms, and other symbol-like means. 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. Integrative connectionism
Unification of the symbolic and the connectionist paradigm. Symbolism as a high level description of the properties of neural nets.

4. Hybrid Systems
Link a current connectionist system with a current (physical) symbol system (exploiting the strengths of each)

"Hybrid models, rather than being viewed simply as a short-term engineering solution, may be crucial to our gaining an understanding of the parameters and functions of biologically plausible cognitive models. From this understanding we might hope to see the development of a new, and potentially more correct, paradigm for the study of 'real', as opposed to artificial, intelligence." (James A. Hendler 1989)

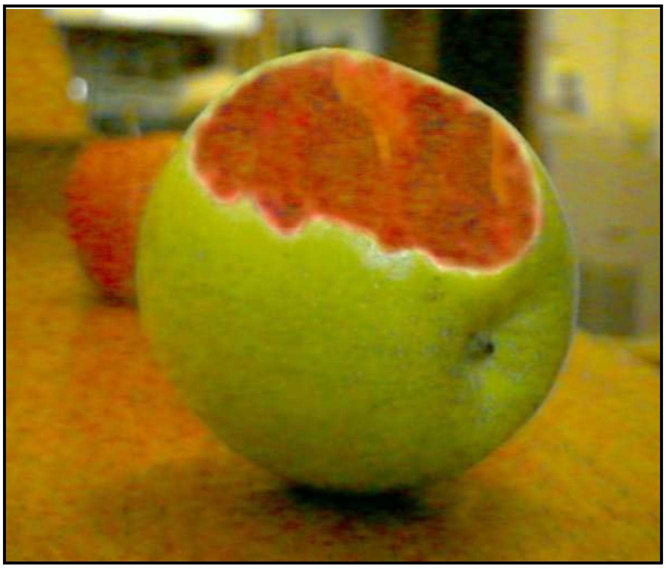
3 The idea of Underspecification

(1) *The tones sounded impure because the hem was torn.*
Linguistic meaning †Utterance meaning

(2) *In most democratic countries, most politicians can fool most of the people on almost every issue most of the time.*

(3)
(3)	a.	*a fast car*	[one that moves quickly]
	b.	*a fast typist*	[a person that performs the act of typing quickly]
	c.	*a fast book*	[one that can be read in a short time]
	d.	*a fast driver*	[one who drives quickly]

(4)	a.	*a red apple*	[red peel]
	b.	*a sweet apple*	[sweet pulp]
	c.	*a reddish grapefruit*	[reddish pulp]
	d.	*a white room/ a white house*	[inside/outside]



A red apple?

What color is an apple?

Q_1 What color is its peel?

Q_2 What color is its pulp?

4 A concise introduction to neural networks

General description

A neural network N can be defined as a quadruple <S,F,W,G>:

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)

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

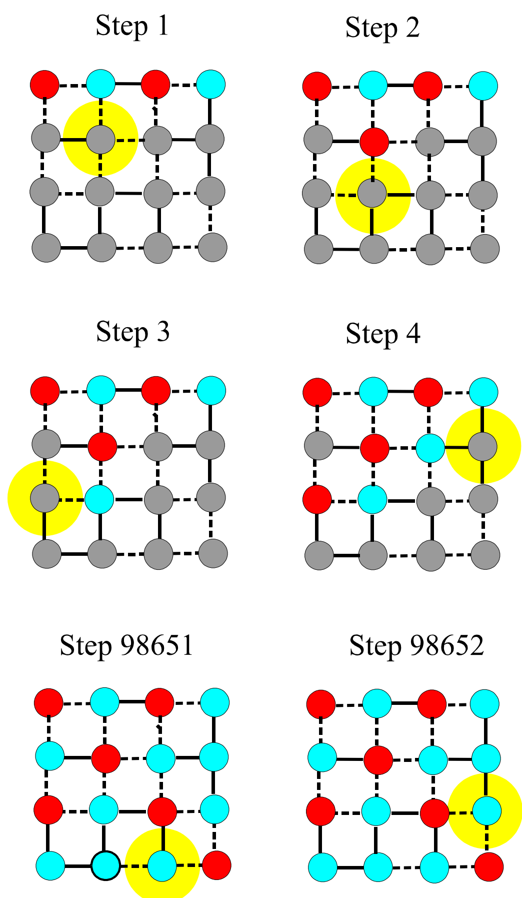
$w_{ij} = w_{ji}$, $w_{ii} = 0$

ASYNCHRONOUS UPDATING:

$s_i(t+1) = \sigma(\sum_j w_{ij} \cdot s_j(t))$, if $i = \text{random}(1,n)$

$s_i(t+1) = s_i(t)$, otherwise

(σ nonlinear function: threshold)



5 Information states as neural activation patterns

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

+1: maximal firing rate — indicating maximal

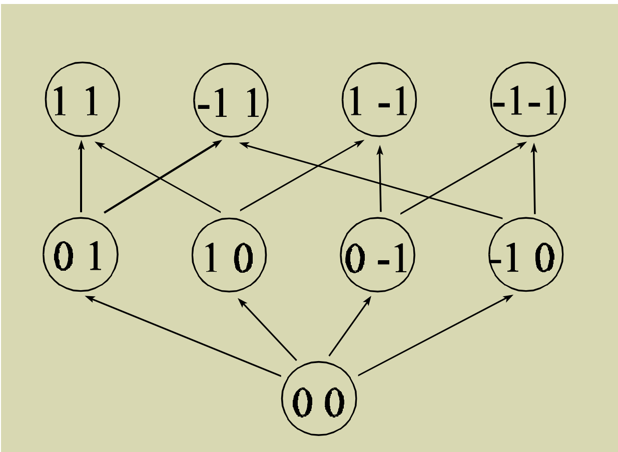
- 1: minimal firing rate A specification

0: resting — indicating underspecification

Poset of activation states:

$S = \{-1,0,+1\}^n$

$s\$t$ iff $s_i\$t_i\$0$ or $s_i\#t_i\#0$, for all $1\#i\#n$.



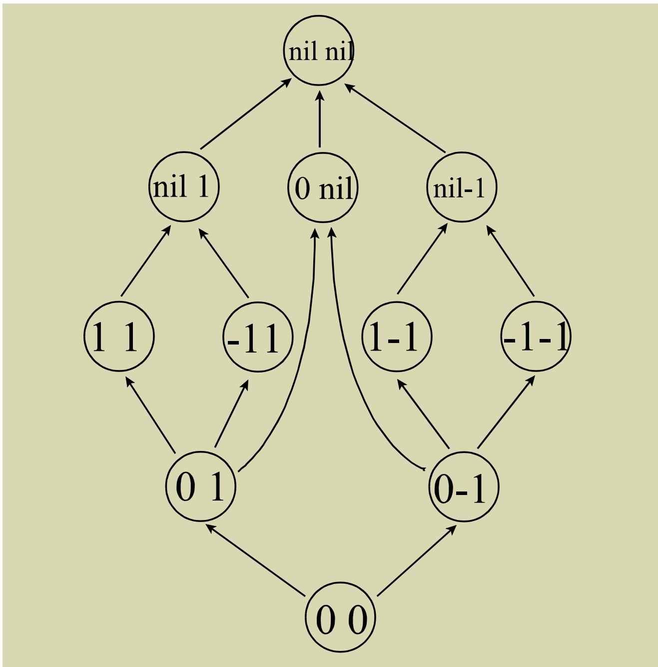
This poset doesn't form a lattice

Extended poset of activation states

$S = \{-1,0,+1,nil\}^n$

nil = "impossible activation"

$s\$t$ iff s_i = nil or $s_i\$t_i\$0$ or $s_i\#t_i\#0$, for all $1\#i\#n$.



DeMorgan lattice

CONJUNCTION : simultaneous realization of two states

DISJUNCTION r : some kind of generalization.

This fact enables us to interpret activation states as propositional objects (*information states*).

6 Asymptotic updates and nonmonotonic inference

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

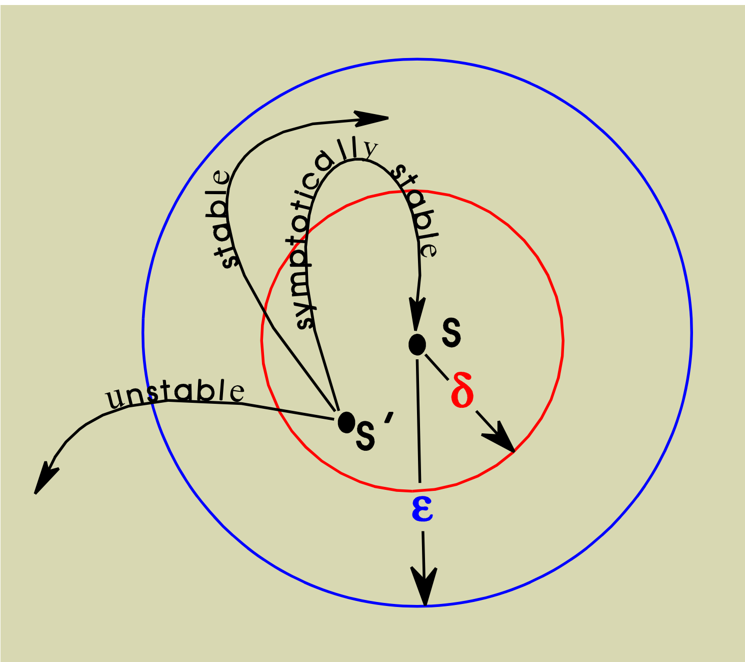
Definition 1

A state sOS is called a resonance of a dynamic system [S, f] iff

1. f(s) = s (equilibrium)

2. For each g>0 there exists a 0<*#g such that for all n$ 1
   |f^n (sN)-s| < g whenever |sN-s| < * (stability)

3. For each g>0 there exists a 0<*#g such that
   $\lim_{n64}$ (sN) = s whenever |sN-s| < * (asymptotic stabil.)



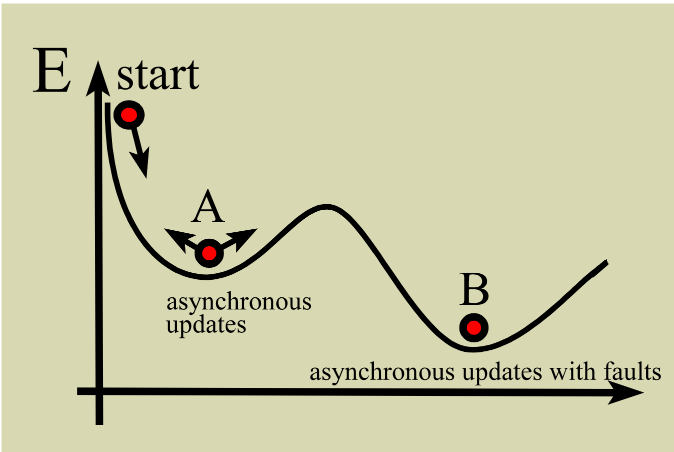
The existence of resonances is an emergent collective effect. Intuitively, resonances are the stable states of the network. They *attract* other states. When each state develops into a resonance, then the system produces a content-addressable memory. Such memories have emergent collective properties (capacity, error correction, familiarity recognition.)

A neural network [S,W,F] is called a resonance system iff
$\lim_{n64}$ (f^n (s)) exists and is a resonance for each sOS and fOW.

Fact 1 (Cohen & Großberg 1983):
Hopfield networks are resonance systems. (The same holds for a large class of other systems including the McCulloch-Pitts model (1943), and Smolensky's Harmony networks (1986))

Fact 2 (Hopfield 1982)
The function $E(s) = -G_{i>j}\, w_{ij}\, s_i\, s_j$ 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 $\lim_{n64} (f^n (s))$ can be characterized as *the local minima* of the Ljapunov-function.

Fact 3 (Hopfield 1982)
The output states $\lim_{n64} (f^n (s))$ can be characterized as *the global minima* of the Ljapunov-function if certain stochastic update functions f are considered ("simulated annealing").



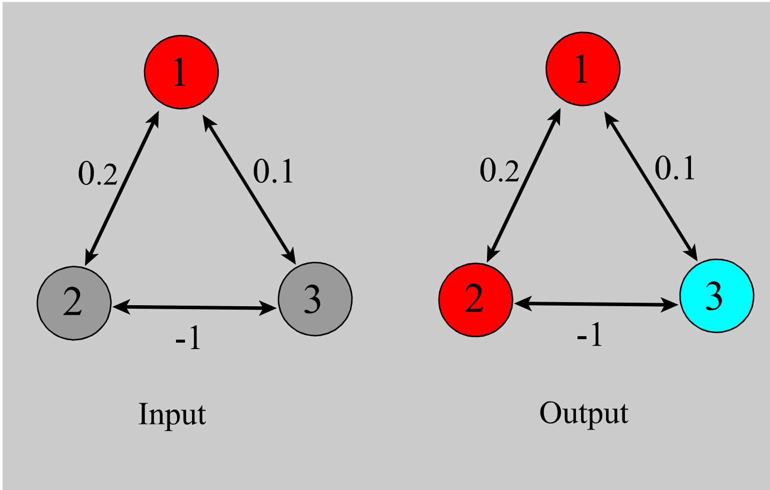
Definition 2 (asymptotic updates)
$ASUP_w(s) =_{def} \{t: t = \lim_{n64} f^n(s)\}$ [f asynchronous updates with clamping]

Definition 3 (E-minimal specifications of s)
$min_E(s) =_{def} \{t: t\$s$ and there is no $t N\$s$ such that $E(tN)<E(t)\}$

Consequence of fact 3
$ASUP_w(s) =_{def} min_E(s)$, where $E(s) = !\,G_{i>j}\, w_{ij}\, s_i\, s_j$
(energy function)

Example

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



			E	
<1 0 0>	#	<1 0 0>	0	
		<1 0 1>	-0.1	
		<1 1 0>	-0.2	
		<1 1 1>	0.7	
		<1 1-1>	-1.1	7

$ASUP_w$(<1 0 0>) = min_E(s) = <1 1-1>

Definition 4 (Nonmonotonic inference relation)
s “- $_w$ t iff s' $ t for each s' O $ASUP_w$(s)

In our example <1 0 0> “- $_w$ <1 1-1>
<1 0 0> “- $_w$ <0 1 0>

Fact 4

(i)	if s $ t, then s “- $_E$ t	SUPRACLASSICALITY
(ii)	s “- $_E$ s	REFLEXIVITY
(iii)	if s “- $_E$ t and s B t “- $_E$ u, then s “- $_E$ u	CUT
(iv)	if s “- $_E$ t and s “- $_E$ u, then s B t “- $_E$ u	CAUTIOUS MONOTON.

7 Relating connectionism and symbolism

Consider the knowledge base in

<u>Connectionist Systems</u>	<u>Symbol Systems</u>
• connection matrix	• strong and weak (default-) rules
• energy function	

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

1. Assigning activation states to each atomic symbol of an elementary language L_{At} , e.g.

   ¼p_1¿ = <1 0 ... 0>
   ¼p_2¿ = <0 1 ... 0>
   ...
   ¼p_n¿ = <0 0 ... 1>

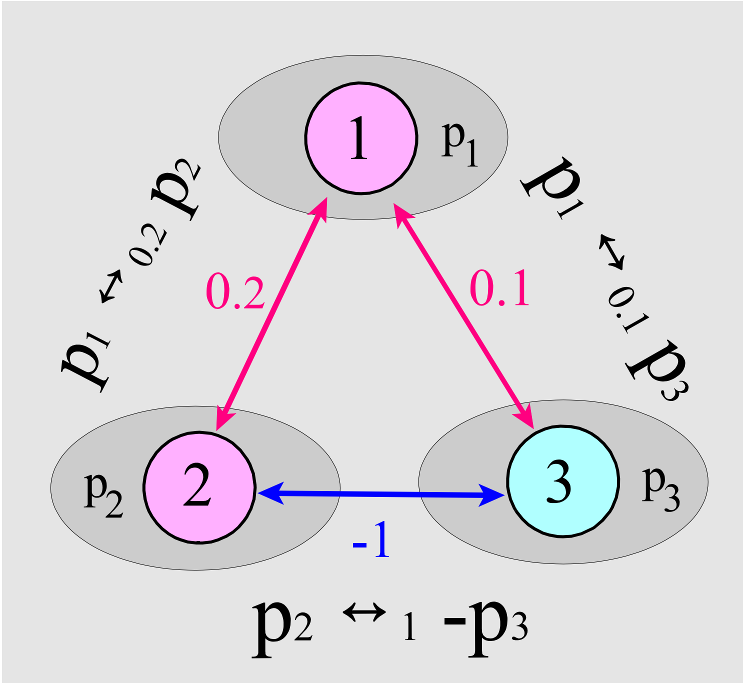
   Local Representation

2. Assigning combinations:
   ¼"v$¿ = ¼"¿B¼$¿, ¼-"¿ = -¼"¿

3. Translating Hopfield networks into weight-annotated Poole systems: Translate the connections w_{ij} into weight-annotated defauls p_i : sign(w_{ij}) p_j with weight |w_{ij}|, for 1#i<j#n

Fact 5
Assume that the formulae " and $ are conjunctions of literals. Assume further that the Poole system T is associated to the connection matrix w. Then: ¼"¿ “-$_w$ ¼$¿ iff " Å)$_T$ $

Example



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

) = {p_1 : $_{0.2}$ p_2,
p_1 : $_{0.1}$ p_3,
p_2 : $_{1.0}$ - p_3}

- *A scenario of " in T* is a subset)N of) such that)NC{"} is consistent.
- *The weight of a scenario*)N is G()N) = G$_{*0)N}$ g(*) - G$_{*0()-)N)}$ g(*)
- *Nonmonotonic inference* as entailments in *maximal* scenarios

some (relevant) scenarios of p_1:	G
{}	-1.3
{p_1 : p_2}	-0.9
{p_1 : p_2, p_1 : p_3}	-0.7
{p_1 : p_2, p_2 : $\neg p_3$}	1.1 7
{p_1 : p_3, p_2 : $\neg p_3$}	0.9

Consequently: p_1 Å)$_T$ p_2, p_1 Å)$_T$ $\neg p_3$

corresponding to: <1 0 0> "-$_w$ <1 1-1> $ <0 1 0> B<0 0 -1>

(Symbolic systems can be used to understand connectionist systems.

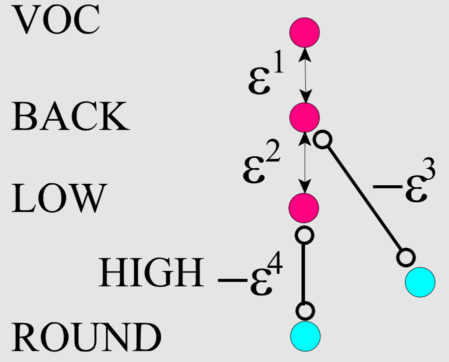
(Connectionist systems can be used to perform inferences.

Exponential weights and strict constraint ranking

! back	+back	
/i/	/u/	+high
/e/	/o/	! high/ ! low
/æ/	/]/	+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 < g # 0.5.

Strong Constraints: LOW 6 ¬HIGH; ROUND 6 BACK



	/a/	/i/	/o/	/u/	/ɔ/	/e/	/æ/
BACK	+	−	+	+	+	−	−
LOW	+	−	−	−	+	−	+
HIGH	−	+	−	+	−	−	−
ROUND	−	−	+	+	+	−	−

Assigned Poole-system

VOC :g1 BACK; BACK :g2 LOW
LOW :g4 ¬ROUND; BACK :g3 ¬HIGH

Keane's markedness conventions

Smolensky's speculation

Exponential weights correspond to an *automatic* processing mode

8 Some remarks on Optimality Theory

- As with weight-annotated Poole systems, OT looks for an optimal satisfaction of a system of conflicting constraints.
- The exponential weights of the constraints realize a *strict ranking* of the constraints
- Violations of many lower ranked constraints count less than one violation of a higher ranked constraint.

Find the optimal vowels (satisfying the strong constraints)

Find the optimal high vowels (”)

	Back	Low	High	Round	Voc ; Back	Back ; Low	Back ; ¬High	Low ; ¬Round
	+	+	!	+				*
L	+	+	!	!				
	!	!	+	!	*			*
L	+	!	+	+		*	*	
	+	!	+	!		*	*	*
	+	!	!	+		*		
	+	!	!	!		*		*
	!	+	!	!	*	*	*	
	!	!	!	!	*		*	*

The candidates can be seen as information (activation) states. The Harmony (or NegEnergy; H = -E) can be recognized immediately from the violations of the (strictly ranked) constraints.

Principles of OT and their relation to connectionism

Optimality: The correct output representation is the one that maximizes Harmony. (H = - E, Ljapunov-function)

Containment: Competition for optimality is between outputs that include the given input. (Clamping the input units restricts the optimization in a network to those patterns including the input.)

Parallelism: Harmony measures the degree of simultaneous satisfaction of constraints.

Conflict. Constraints conflict: it is typically impossible to simultaneously satisfy them all. (Positive and negative connecti-ons typically put conflicting pressures on a unit's activity.)

Domination: Constraint conflict is resolved via a notion of differential strength: stronger constraints prevail over weaker ones in cases of conflict.

Minimal violability: Correct outputs typically violate some constraints, but do so only to the minimal degree needed to satisfy stronger constraints.

Strictness of domination: Each constraint is stronger than all weaker constraints combined. (Corresponds to a strong restriction on the numerical constraint strengths, and makes it possible to determine optimality without numerical computation.)

Universality: The constraints are the same in all human grammars. (Corresponds to a strong restriction on the content of the constraints, unclear how to explain)

9 Conclusions

- Certain activities of connectionist networks can be interpreted as *nonmonotonic inferences*. In particular, there is a strict correspondence between certain network types and particular nonmonotonic inferential systems.
- The relation between nonmonotonic inferences and neural computation must be of the type that holds between higher level and lower level systems of analysis in the physical sciences.
  (For example, statistical mechanics explains significant parts of thermodynamics from the hypothesis that matter is composed of molecules, but the concepts of thermodynamic theory, like temperature" and "entropy," involve no reference whatever to molecules.)
- So far we have considered only local representations where symbols correspond to single nodes in the network. As a consequence, we are confronted with a very pure symbol system only that fails to express
  - Constituent structure
  - Variable binding
  - Quantification
- There are several possibilities to overcome these shortcomings:
  - Using distributed representations for realizing constituent structures (e.g. Smolensky's tensor product representations)
  - Realizing dynamic binding by using temporal synchrony. Representation of rules with variables; variable binding (Shastri & Ajjanagadde 1993; based on ideas of Feldman 1982 and von der Malsburg 1986)