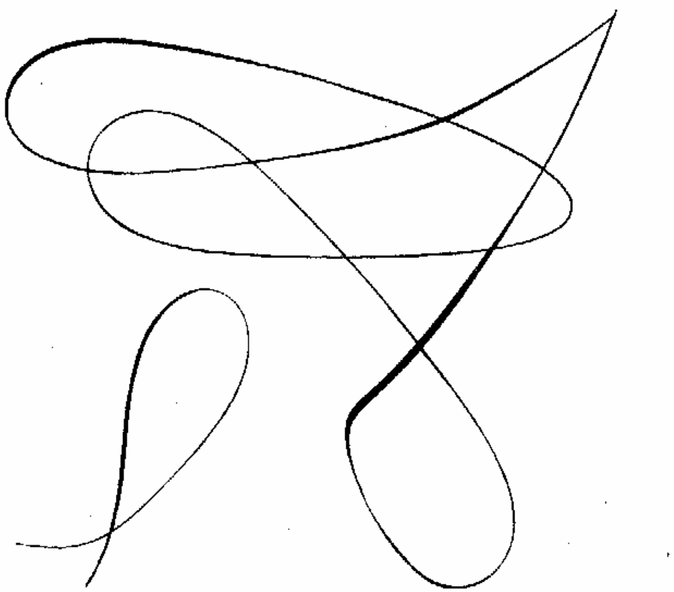
Neural Nets and Symbolic Reasoning

Integrating Connectionism and Symbolism: Nonmonotonic Reasoning and Hopfield Nets



Outline

- Introduction
- Neural networks as dynamical systems
- Information states as neural activation patterns
- Asymptotic updates and nonmonotonic inference
- Weight-annotated Poole systems and Hopfield networks
- The link to Optimality Theory

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

Intended results

- Better understanding of connectionist networks: Nonmonotonic logic and algebraic semantics as descriptive and analytic tools for analyzing their emerging properties

- New methods for performing nonmonotonic inferences: Connectionist methods (randomised optimisation: simulated annealing) can be adopted for realizing symbolic inferences

- Certain logical systems are singled out by giving them a "deeper justification".

- Understanding Optimality Theory: Which assumptions have a deeper foundation and which ones are pure stipulations?

2 Neural networks as dynamical systems

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 $w \in W$: the function $f_w \in F$ describes how the neuron activities spread through that network (fast dynamics)
- G: Set of learning functions (slow dynamics)

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

+1: maximal firing rate

0: resting

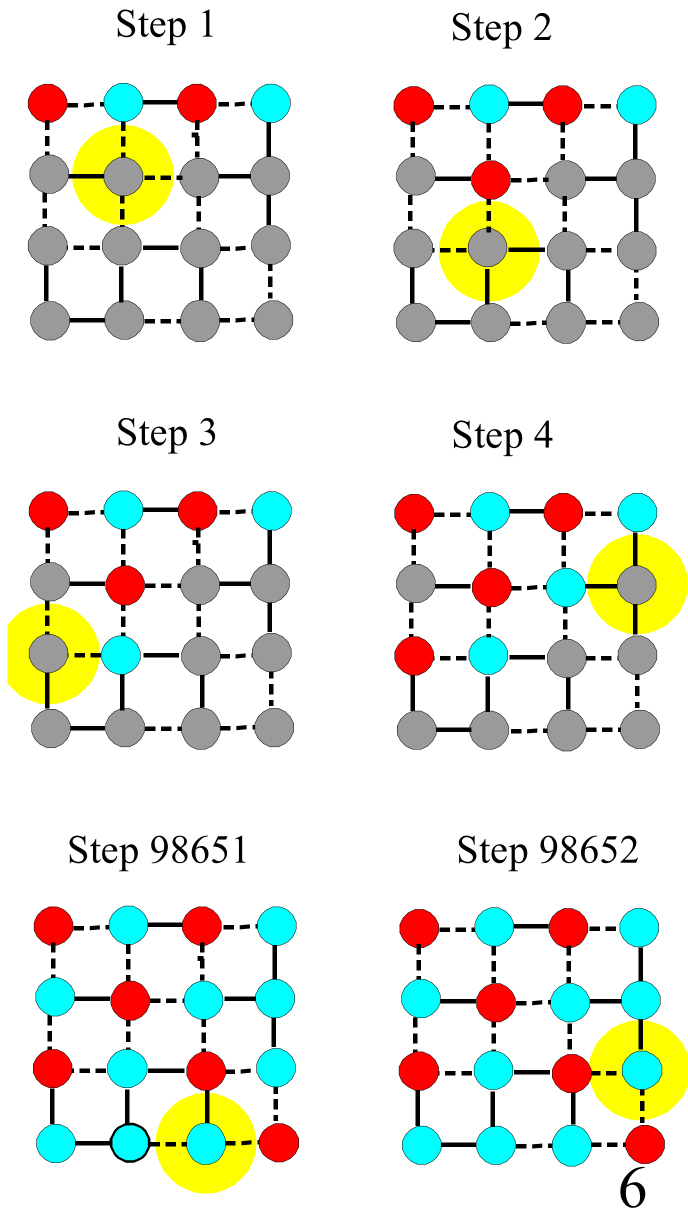
-1 : minimal firing rate)

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

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

ASYNCHRONOUS UPDATING:

$$s_i(t+1) = \begin{cases} \theta\left(\Sigma_j w_{ij} \cdot s_j(t)\right), & \text{if } i = \text{rand}(1,n) \\ s_i(t), & \text{otherwise} \end{cases}$$

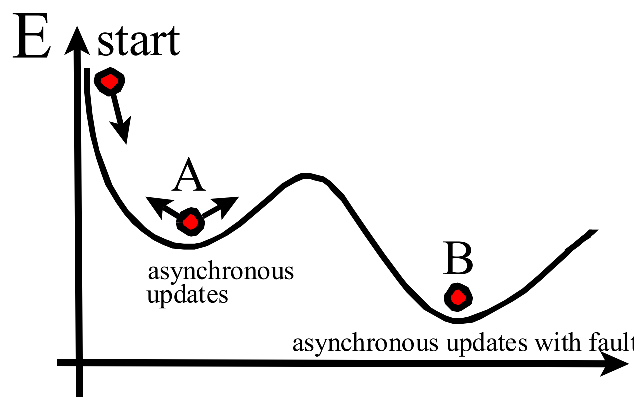


Theorem 1 (Cohen & Großberg 1983)

Hopfield networks are resonance systems (i.e. $\lim_{n\to\infty} f^n(s)$ exists and is a resonance for each $s \in S$ and $f \in F$)

Theorem 2 (Hopfield 1982)

$E(s) = -\frac{1}{2} \sum_{i,j} w_{ij} s_i s_j$ is a *Ljapunov-function* of the system in the case of asynchronous updates. The output states $\lim_{n\to\infty} f^n(s)$ can be characterized as *the local minima* of E



Theorem 3 (Hopfield 1982)

The output states $\lim_{n\to\infty} f^n(s)$ can be characterized as *the global minima* of E if certain stochastic update functions f are considered (faults!).

3 Information states as neural activation patterns

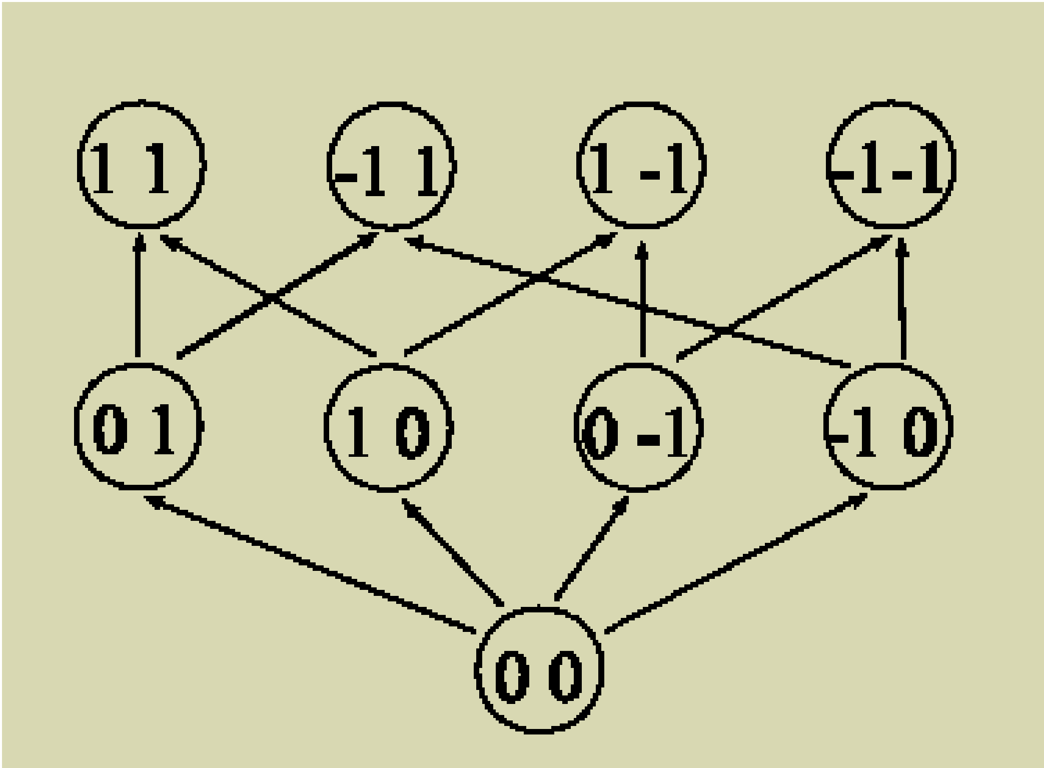
Activation states can be partially ordered in accordance with their informational content

+1: maximal firing rate	indicating *maximal*
- 1: minimal firing rate	*specification*
0: resting	indicating *underspecification*

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

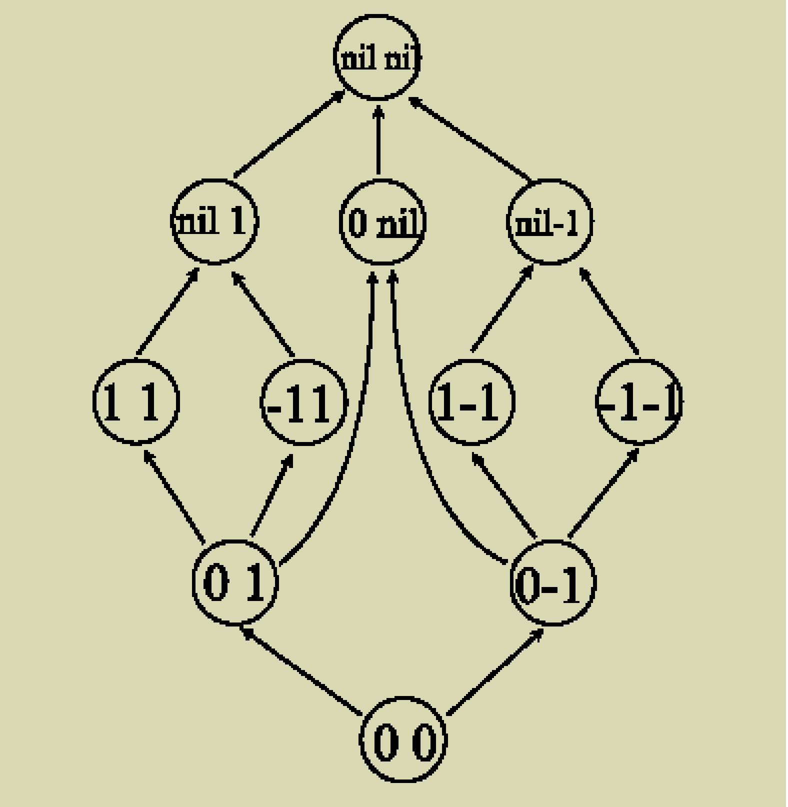
$s \geq t$ iff $s_i \geq t_i \geq 0$ or $s_i \leq t_i \leq 0$, for all $1 \leq i \leq n$ read: *s* is at least as specific as *t*

This poset doesn't form a lattice!



Extend it to a lattice by introducing impossible activation states:

$s \geq t$ iff $s_i = nil$ or $s_i \geq t_i \geq 0$ or $s_i \leq t_i \leq 0$, for all $1 \leq i \leq n$)



$s \circ t = \sup\{s,t\}$ (CONJUNCTION) *simultaneous realization* of two activation states

$s \oplus t = \inf\{s,t\}$ (DISJUNCTION) *generalization* of two instances of information states.

The COMPLEMENT s^* reflects a *lack* of information.

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

(2) $(s \oplus t)_i = \begin{cases} \min(s_i, t_i), & \text{if } s_i,t_i \geq 0 \\ \max(s_i, t_i), & \text{if } s_i,t_i \leq 0 \\ s_i, & \text{if } t_i = \text{nil} \\ t_i, & \text{if } s_i = \text{nil} \\ 0, & \text{elsewhere'} \end{cases}$

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

4 Asymptotic updates and nonmonotonic inference

- In general, updating an information state *s* may result in an information state f...f(*s*) that doesn't include the information of *s*
- In what follows it 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
- Balkenius & Gärdenfors (1991). Let f designate the original update function (1) and $\underline{f}$ the *clamped* one:

$$\underline{f}(s) = f(s) \circ s;$$

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

Asymptotic updates and minimal specifications

Definition 1 (asymptotic updates)

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

Definition 2 (E-minimal specifications of s)

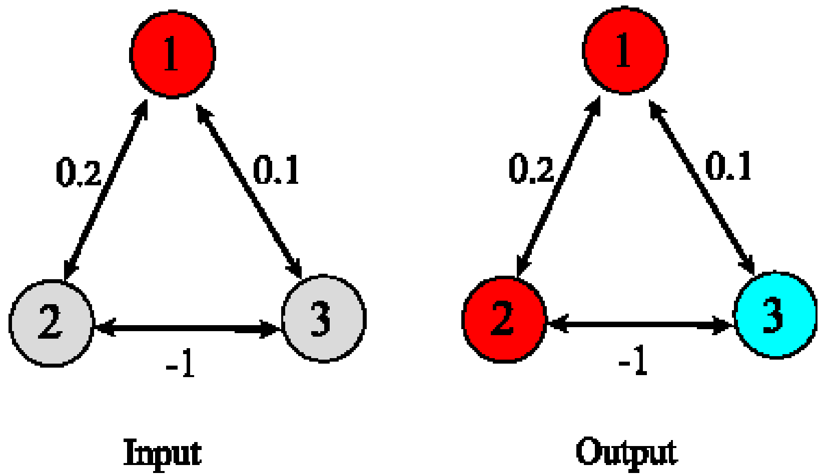
$min_E(s) =_{def} \{t: t \geq s \text{ and there is no } t' \geq s \text{ such that } E(t') < E(t)\}$

Consequence of Theorem 3

$ASUP_w(s) = min_E(s)$, where $E(s) = -\sum_{i<j} w_{ij}\, s_i\, s_j$ (energy function)

(Remark: $E(s) = -\frac{1}{2} \sum_{i,j} w_{ij}\, s_i\, s_j = -\sum_{i<j} w_{ij}\, s_i\, s_j$ :::symmetry!)

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



$E(s) = -0.2s_1s_2 - 0.1s_1s_3 + s_2s_3$

			E
<1 0 0>	$\le$	<1 0 0>	0
		<1 0 1>	-0.1
		<1 1 0>	-0.2
		<1 1 1>	0.7
		<1 1-1>	-1.1 ☜

$\mathrm{ASUP}_w(\langle 1\ 0\ 0\rangle) = \min_E(s) = \langle 1\ 1\ {-1}\rangle$

Definition 3 (Nonmonotonic consequence relation)

$s \mid\!\sim_w t$ iff $s' \geq t$ for each $s' \in ASUP_w(s)$

In our example $<1\ 0\ 0> \mid\!\sim_w <1\ 1\ {-1}>$

$<1\ 0\ 0> \mid\!\sim_w <0\ 1\ 0>$

Theorem 4 (cumulative consequence relation)

Let $\mid\!\sim_w$ be a relation between information states as defined in Definition 3, then

(i) if $s \geq t$, then $s \mid\!\sim_w t$ (SUPRACLASSICALITY)

(ii) $s \mid\!\sim_w s$ (REFLEXIVITY)

(iii) if $s \mid\!\sim_w t$ and $s \circ t \mid\!\sim_w u$, then $s \mid\!\sim_w u$ (CUT)

(iv) if $s \mid\!\sim_w t$ and $s \mid\!\sim_w u$, then $s \circ t \mid\!\sim_w u$. (CAUTIOUS MONOTONICITY)

5 Weight-annotated Poole systems and Hopfield networks

Consider the knowledge base in

Connectionist Systems	Symbol Systems
- 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.

   $\lceil p_1 \rfloor = \langle 1\ 0\ \ldots\ 0 \rangle$

   $\lceil p_2 \rfloor = \langle 0\ 1\ \ldots\ 0 \rangle$

   ...

   $\lceil p_n \rfloor = \langle 0\ 0\ \ldots\ 1 \rangle$

   Local Representation

2. Assigning combinations

   $\lceil \alpha \wedge \beta \rfloor = \lceil \alpha \rfloor \circ \lceil \beta \rfloor$, $\lceil \neg\beta \rfloor = -\lceil \beta \rfloor$

3. Translating Hopfield networks into weight-annotated Poole systems: Translate the connections w_{ij} into weight-annotated defaults $p_i \leftrightarrow \mathrm{sign}(w_{ij})\ p_j$ plus weight $|w_{ij}|$, for $1 \leq i < j \leq n$.

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

(i) Δ is a set of consistent sentences built on the basis of At (hypotheses)

(ii) $g: \Delta \Rightarrow [0,1]$ (the weight function)

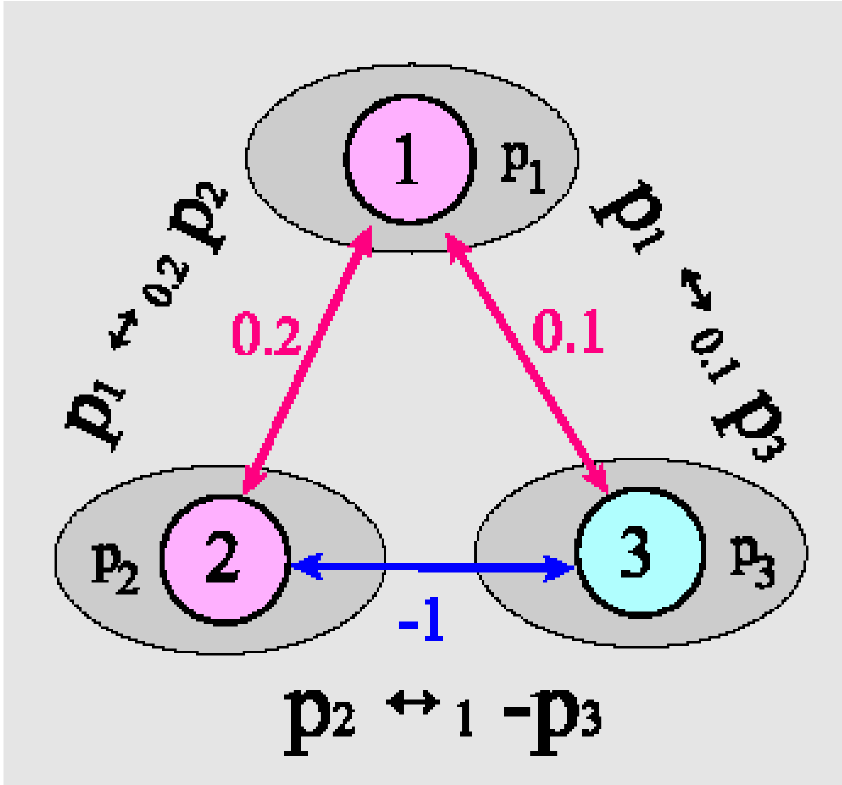
A scenario of a formula α in T is a subset Δ' of Δ such that $\Delta' \cup \{\alpha\}$ is consistent

The weight of scenario Δ' is $G(\Delta') = \sum_{\delta \in \Delta'} g(\delta) - \sum_{\delta \in (\Delta - \Delta')} g(\delta)$

Inference: $\alpha \sim>_T \beta$ iff for each maximal scenario Δ' of α in T: β is an ordinary consequence of $\Delta' \cup \{\alpha\}$ (nonmonotonic inference as entailment in maximal scenarios)

Theorem 4: Let α and β be formulas that are conjunctions of literals. Assume further that the Poole system T is *associated* with the connection matrix *w*. Then

$\lceil\alpha\rfloor \mid\!\sim_w \lceil\beta\rfloor$ iff $\alpha \sim>_T \beta$



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

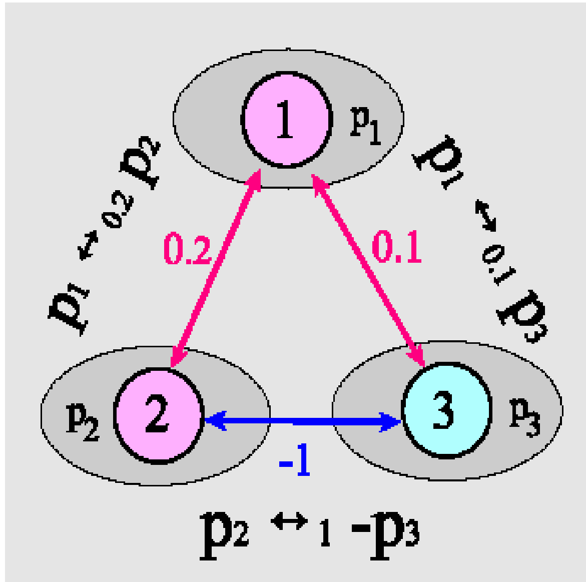
$\Delta = \{p_1 \leftrightarrow_{0.2} p_2,\ p_1 \leftrightarrow_{0.1} p_3,$
$p_2 \leftrightarrow_{1.0} \neg p_3\}$

Example

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

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

$\qquad p_2 \leftrightarrow_{1.0} \neg p_3\}$



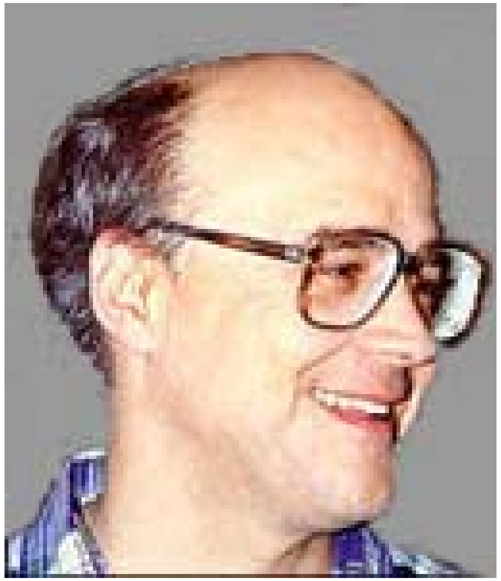
some (relevant) scenarios of p_1:	G
$\{\}$	-1.3
$\{p_1 \leftrightarrow p_2\}$	-0.9
$\{p_1 \leftrightarrow p_2,\ p_1 \leftrightarrow p_3\}$	-0.7
$\{p_1 \leftrightarrow p_2,\ p_2 \leftrightarrow \neg p_3\}$	1.1
$\{p_1 \leftrightarrow p_3,\ p_2 \leftrightarrow \neg p_3\}$	0.9

Consequently: $p_1 \sim>_T p_2,\ p_1 \sim>_T \neg p_3$

corresponding to: $<1\ 0\ 0>\ |\!\sim_w\ <1\ 1\ {-1}> \geq\ <0\ 1\ 0> \circ <0\ 0\ {-1}>$

- Symbolic systems can be used to understand connectionist systems
- Connectionist systems can be used to perform inferences
- As with weight-annotated Poole systems, Hopfield systems can be interpreted as looking for an optimal satisfaction of a system of conflicting constraints.

6 The link to Optimality Theory

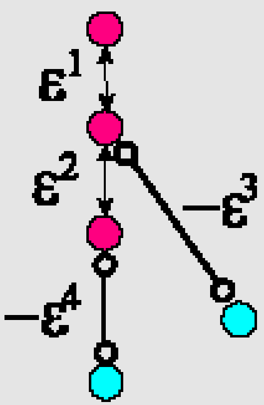


Example from phonology

–back	+back	
/i/	/u/	+high
/e/	/o/	–high/–low
/æ/	/]/ /a/	+low

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 $0 < \varepsilon \leq 0.5$.

Strong Constraints: LOW → ¬HIGH; ROUND → BACK



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

Assigned Poole-system

VOC $\leftrightarrow\varepsilon^1$ BACK; BACK $\leftrightarrow\varepsilon^2$ LOW

LOW $\leftrightarrow\varepsilon^4$ ¬ROUND; BACK $\leftrightarrow\varepsilon^3$ ¬HIGH

Keane's markedness conventions

- 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.
- The grammar doesn't count!