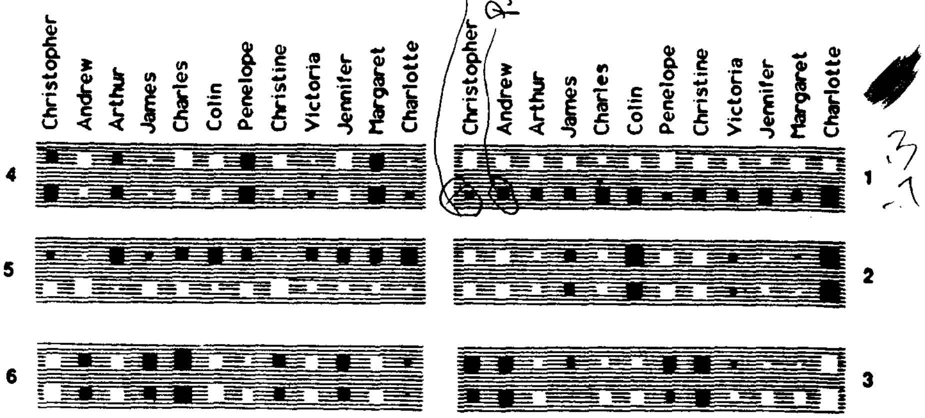

Figure 8.4
The weights from the 24 input units that represent people to the 6 units in the second layer that learn distributed represenions of people. White rectangles, excitatory weights; black rectangles, inhibitory weights; area of the rectangle encodes the magnitude of the weight. The weights from the 12 English people are in the top row of each unit. Unit 1 is primarily concerned with the distinction between English and Italian and most of the other units ignore this distinction. This means that the represention of an English person is very similar to the representation of their Italian equivalent. The network is making use of the isomorphism between the two family trees to allow it to share structure and it will therefore tend to generalize sensibly from one tree to the other. Unit 2 encodes which generation a person belongs to, and unit 6 encodes which branch of the family they come from. The features captured by the hidden units are not at all explicit in the input and output encodings, since these use a separate unit for each person. Because the hidden features capture the underlying structure of the task domain, the network generalizes correctly to the four triples on which it was not trained. We trained the network for 1500 sweeps, using $\varepsilon = 0.005$ and $\alpha = 0.5$ for the first 20 sweeps and $\varepsilon = 0.01$ and $\alpha = 0.9$ for the remaining sweeps. To make it easier to interpret the weights we introduced 'weight-decay' by decrementing every weight by 0.2% after each weight change. After prolonged learning, the decay was balanced by $\partial E/\partial w$, so the final magnitude of each weight indicates its usefulness in reducing the error. To prevent the network needing large weights to drive the outputs to 1 or 0, the error was considered to be zero if output units that should be on had activities above 0.8 and output units that should be off had activities below 0.2.

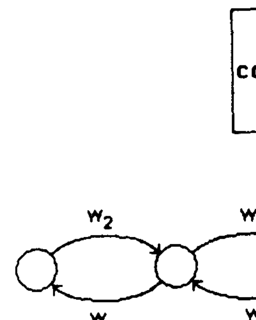

Figure 8.5
A synchronous iterative r
step in the recurrent net
ered nets can be mapped
performing this mapping
layers during the forwar
and (6)). So in an iterativ
ond, for a layered net to
layers must have the san
in each set of correspon
portional to this average
directly to iterative nets
sequential structures[3].

mines the relative c
the weight change.
To break symme
learning procedure
sonal communicatio
One simple task t
output units is the c
levels of a one-dim
centre point, it is es
an individual inpu
symmetry or non-sy
evidence from the i
that intermediate u