Artificial neural networks – Simulations using a simple delta network

Josef Wiemeyer, Dept. of Sport Science, University of Technology Darmstadt

Abstract

The purpose of our simulations was to determine the impact of varying the degrees of freedom of a simple delta network consisting of two input neurons and one output neuron. Results of simulations clearly indicate that even in a very simple artificial neural network with eight degrees of freedom (i.e., network activation, learning rate, interference rate, forgetting rate, distribution schedule, number of learning trials, interference trials and retention trials) it is possible to produce any network performance by choosing network parameters in an appropriate way. The epistemic implications of these findings are discussed.

Introduction

Artificial Neural Networks (ANN) are becoming increasingly important in many areas. We can distinguish two kinds of application:

– Technological application, i.e., adopting ANN for the solution of practical problems

– Biological application, i.e., modelling and simulating neurophysiological information processing in man and animal

Whereas technological application does not pose any epistemic problems the question arises whether ANNs are the appropriate method for biological application. Part of this problem concerns the question which biological correspondence the several ANN parameters have and to what extent we can influence network performance by choosing the ANN parameters.

A simple delta ANN

The ANN used in our simulations consists of two input neurons and one output neuron (Schmidt, 1998). The input neurons can adopt the values 0 and 1 (i.e., feature existent/ not existent), whereas the output neuron can adopt values between 0 and 1 (i.e., degree of identification of learning stimulus S).

Synaptic weights (w_{11}, w_{21}) were changed by means of the generic sum function $\Delta w = x \cdot \sigma(nettoinput)$, with $\sigma(nettoinput) = (1+e^{-a \cdot nettoinput})^{-1}$ and $nettoinput = x_1 \cdot w_{11} + x_2 \cdot w_{12}$ (a – generalised activation, x_1, x_2 – activation state of input neurons 1 and 2; w_{11}, w_{12}

– synaptic weights of neural connection). This term can be interpreted as certainty of identifying the criterion stimulus S (1,1). A value of 1 means that S is surely identified.

For adapting the neural weights we applied the following formulae:

- Learning: $\Delta w_{11}=g\cdot(1-\sigma(w_{11}+w_{12})$, $\Delta w_{12}=\Delta w_{11}$ (*g* – rate of learning)
- Interference (pause): $\Delta w_{11}=-l\cdot\sigma(w_{11})$, $\Delta w_{12}=0$ (*l* – rate of interference)
- Forgetting: $\Delta w_{11}=-f\cdot\sigma(w_{11})$, $\Delta w_{12}= 0$ (*f* – rate of forgetting)

During learning trials synaptic weights w_{11} and w_{12} are changed, whereas during pauses and forgetting only synaptic weight w_{11} is changed.

Vector S (1,1) is the learning stimulus, whereas vector S_d (1,0) is the distractor we used during interference and forgetting trials.

Simulation procedure

The delta ANN has eight degrees of freedom: Learning trials, interference trials (pause), retention trials, distribution of learning and interference trials and parameters *a, g, l* and *f*.

In a first series of simulations we changed these parameters systematically (see table 1). First we isolated parameters *a, g, l* und *f*. Finally we examined the relationship of parameters *g/l, a/g* and *a/l*.

Table 1. Simulation parameters

Number of simulation	**Parameter varied**	**Initial value**	**Final value**	**Step**	**Remaining parameters**
1 - 4	*a, g, l, f*	0.1	5.2	0.3	1
5-7	*a/g, a/l, g/l*	0.1	5.2	0.3	1

In each simulation series consisting of 100 learning and 100 interference trials we varied the distribution of learning. We simulated six practice schedules:

- 1 block of 100 learning trials, followed by 100 interference trials (B100)
- 2 blocks of 50 learning trials and 50 interference trials (B50)
- 4 blocks of 25 learning trials and 25 interference trials (B25)
- 10 blocks of 10 learning trials and 10 interference trials (B10)
- 25 blocks of 4 learning trials and 4 interference trials (B4)
- 100 blocks of 1 learning trial and 1 interference trial (B1)

The number of retention trials was varied in three stages (0, 50 and 100 retention trials).

In a second series of simulations we introduced a random factor. Within each iteration we multiplied Δw with a random factor. We performed three factor stages (0.1, 1 and 10). So Δw was modified by 10%, 100% or 1000%, respectively. We analysed four different practice schedules in order to examine the interaction of random factor and distribution:

– 1 block of 100 learning trials, followed by 100 interference trials (B100)
– 2 blocks of 50 learning trials and 50 interference trials (B50)
– 4 blocks of 25 learning trials and 25 interference trials (B25)
– 100 blocks of 1 learning trial, each block being followed by 1 interference trial (B1)

According to Schmidt (1998) we performed simulations with a=1, g=2, l=6 and f=1. For every parameter combination 20 simulations were performed. The order of parameter combinations was varied at random. The total number of simulations is 720 runs. In order to replicate our results we performed the whole simulation series twice.

In a third series of simulations we analysed the interaction of random factor and interference trials. We adopted 100 blocks of one learning trial and 0, 1, 2, 3, 4, 5 or 10 interference trial(s) respectively. Random factor was varied in three stages (1, 5, 10). Retention trials were also varied in three stages (0, 50 and 100 trials). We performed 20 simulations for every parameter combination (total number of runs: 1260). Again we chose simulation parameters according to Schmidt (1998).

In all the simulations we recorded degree of correct identification of stimulus S and degree of incorrect identification of stimulus S_d as dependent variables. According to the ANN algorithm these variables range from 0 (no identification) to 1 (clear identification). Because dependent measures in the second and third simulation series can be considered random variables we computed analyses of variance (SPSS for Windows, version 8.0).

Results and discussion

Impact of activation, learning, interference and forgetting

From figure 1 we can see that increasing activation leads to better network performance, except for the massed ANN (B1).

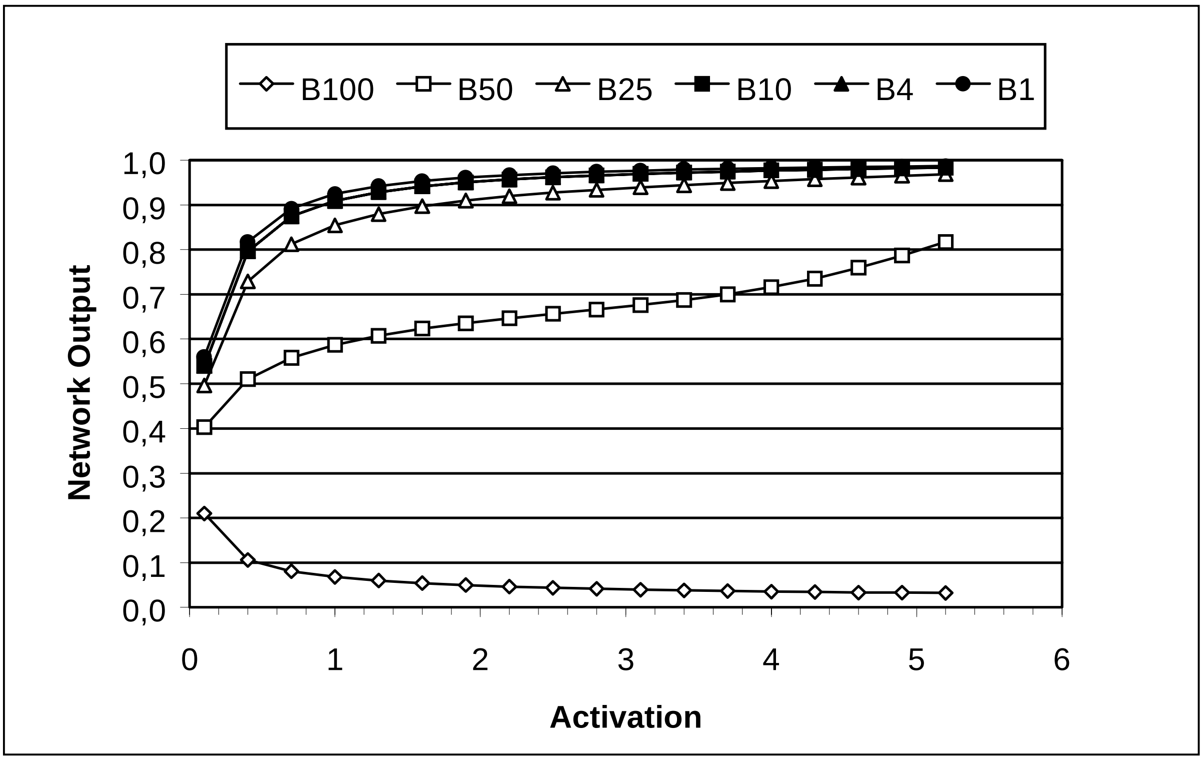

Figure 1. Simulation of a delta ANN (variation of activation), 100 retention trials, network output is correct stimulus identification.

Figure 2 illustrates that increasing learning rate always increases performance regardless of training schedules. However, best performance is achieved by distributed learning (i.e., network B1).

In figure 3 we can see that increasing interference decreases massed learning (network B100) and increases distributed learning (particularly network B1).

Increasing forgetting has a negative impact on network performance, regardless of distribution schedule. Again we find particular advantages for distributed learning. Network B1 forgets much less as compared to network B50. Because of the lowest performance level of all networks network B100 does not forget as much as the other networks.

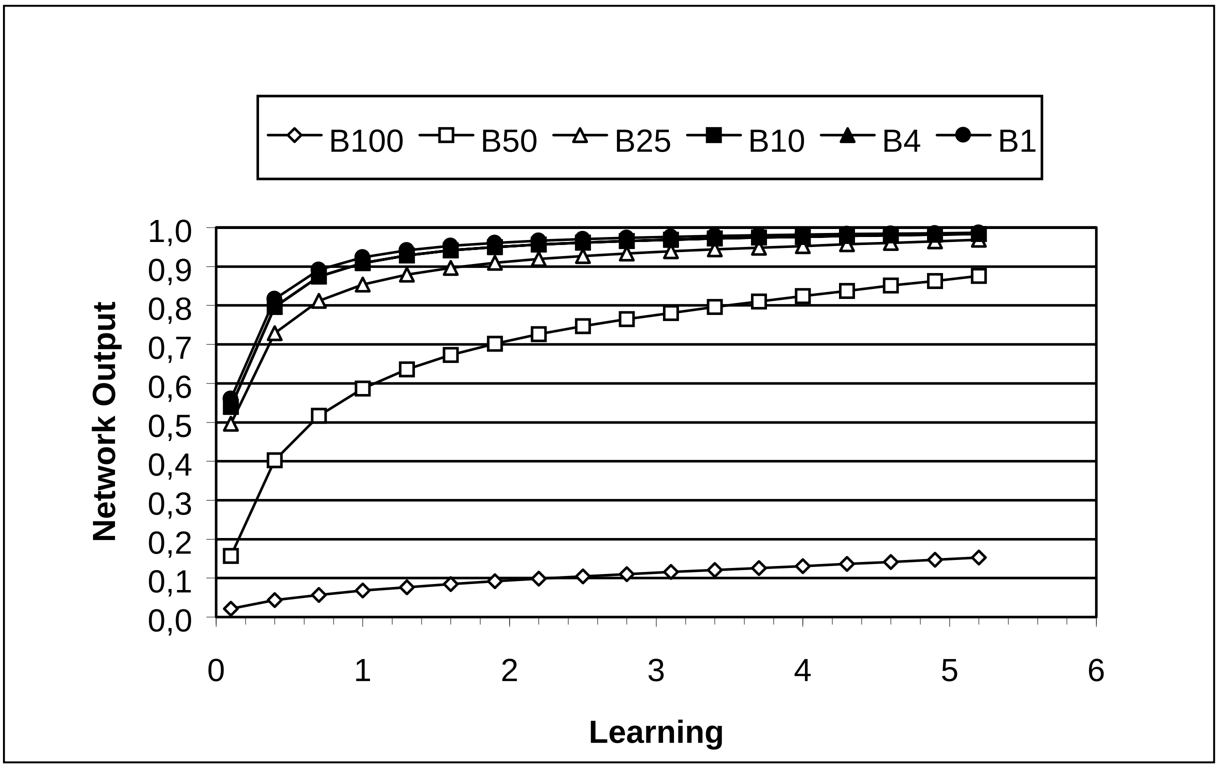

Figure 2. Simulation of a delta ANN (variation of learning), 100 retention trials, network output is correct stimulus identification.

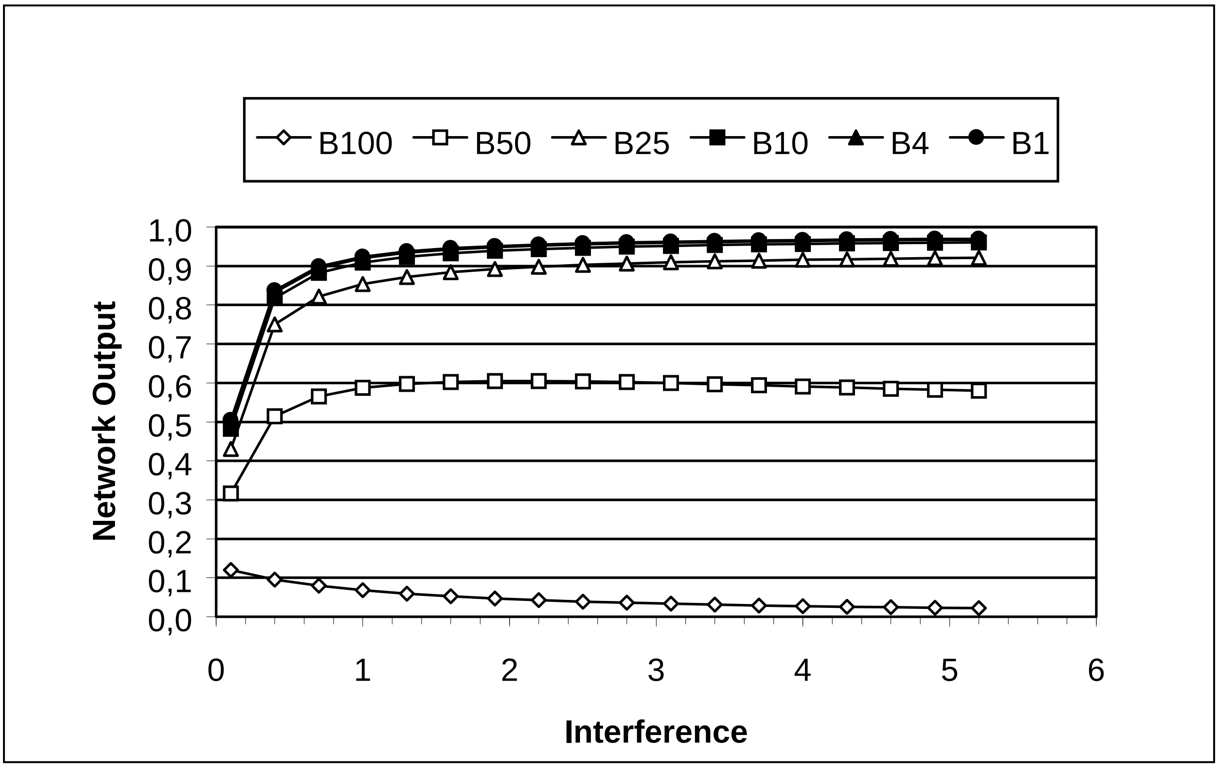

Figure 3. Simulation of a delta ANN (variation of interference), 100 retention trials, network output is correct stimulus identification.

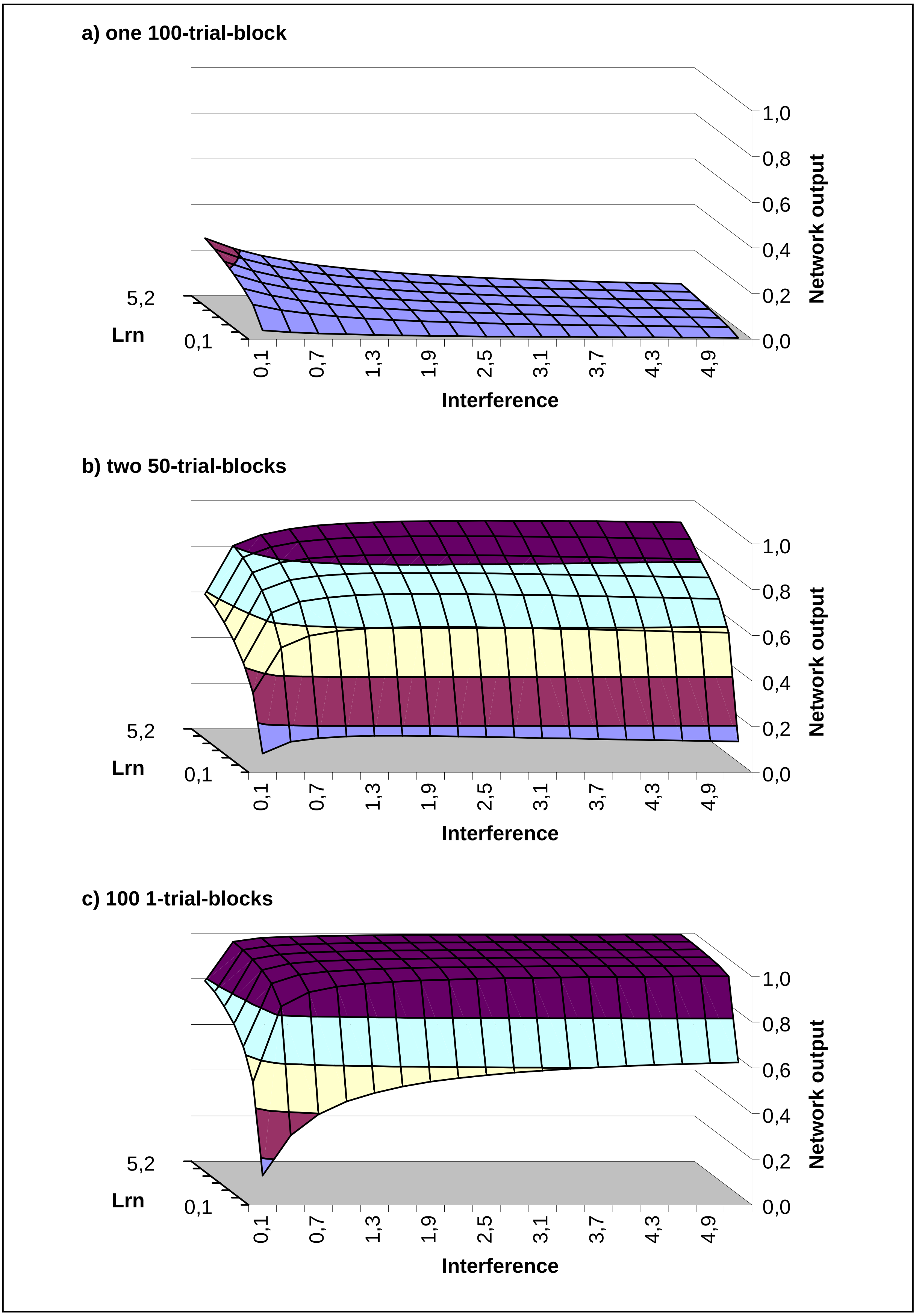

Figure 4. Influence of learning (Lrn) and interference on network performance

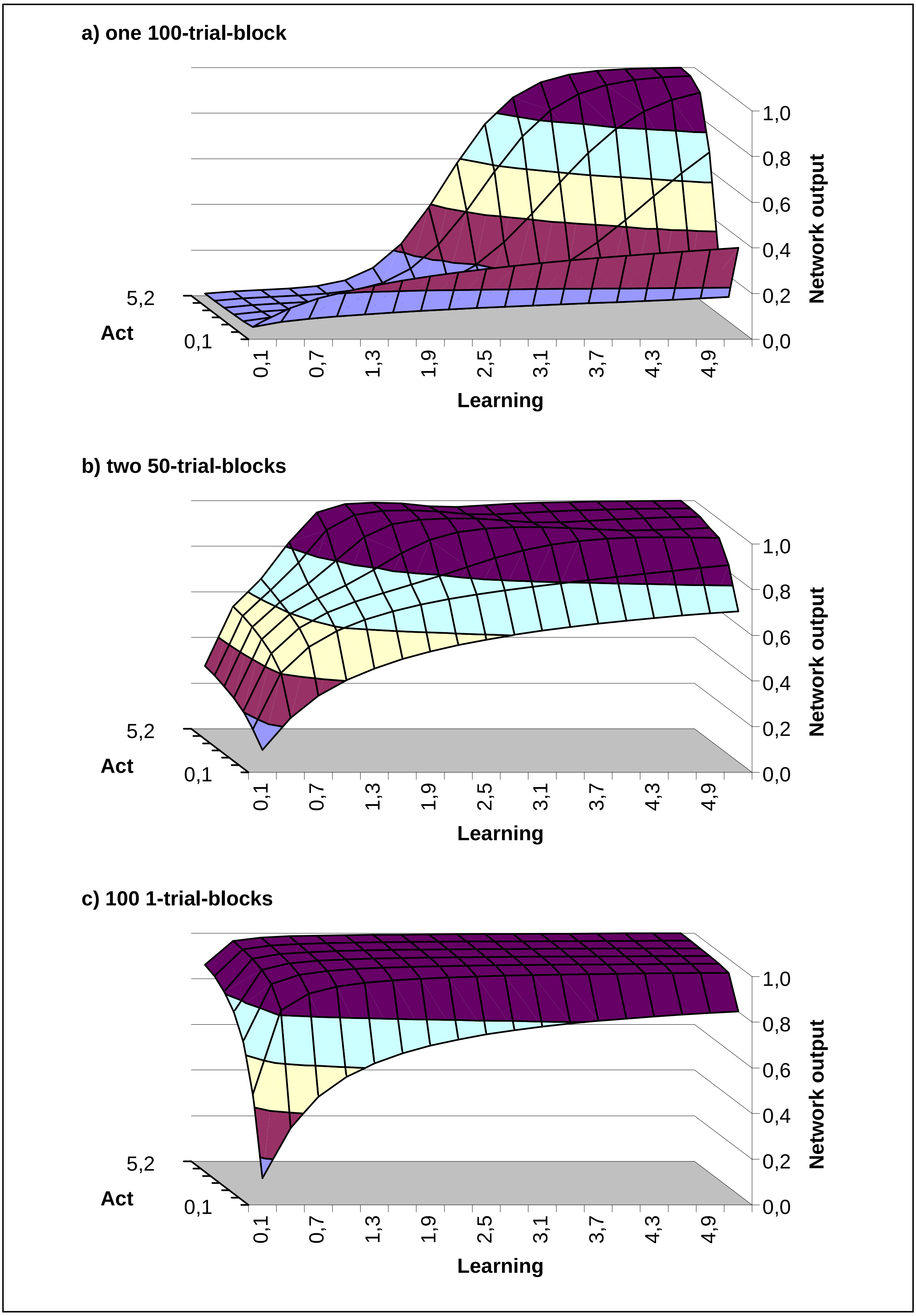

Figure 5. Influence of activation (Act) and learning on network performance

Figure 4 illustrates the interaction of learning and interference. We can see that massed learning (figure 4a) does not result in appropriate performance at all. In network B50 (figure 4b) there is a strong dependence on both learning and interference, particularly with small parameters. The distributed network B1 shows a big plateau indicating that within a broad range of values parameter changes have only a small impact on network performance (figure 4c).

In figure 5 the interaction of activation and learning is illustrated. As can be seen from the size of the plateaus distributed learning is much less subject to the influence of activation and learning than massed learning.

Interaction of activation and interference is similar to the interaction of learning and interference. Because of limited space the diagrams cannot be illustrated in this contribution. Again the distributed network performs best.

Random factor and network performance

Introducing a random factor leads to replicable results.

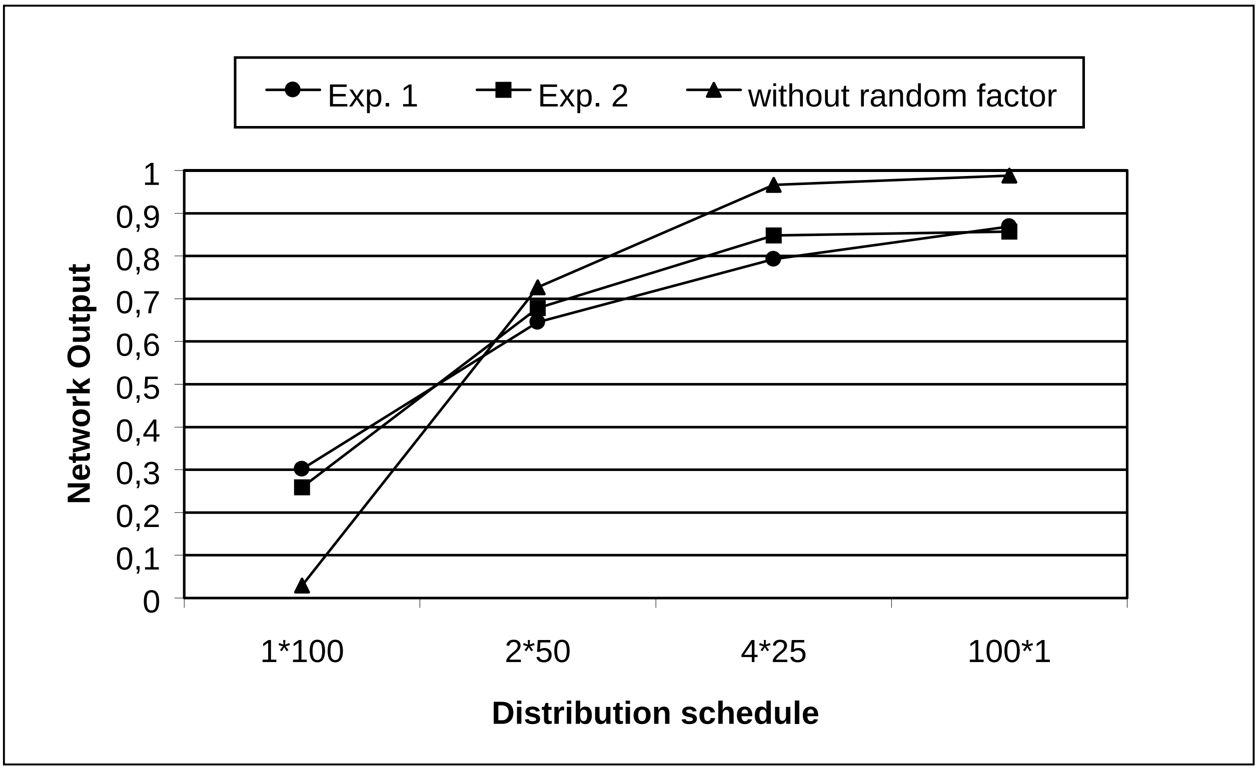

Figure 6. Interaction of random factor and practice schedule

We can find a significant interaction of random factor and practice schedule ($F_{6,684}$=18.8, p<.001 and $F_{6,684}$=14.2, p<.001). Whereas correct stimulus identification improves for massed learning it decreases for distributed learning. With increasing random factor incorrect stimulus identification decreases (see figure 6).

Examining the interaction of random factor and pause length we found a general negative effect of random influence ($F_{2,1197}$=27.1, p<.001). Furthermore we found interactions of random factor and pause length ($F_{12,1197}$=3.7, p<.001) and pause length and retention interval ($F_{12,1197}$=4.1, p<.001) respectively.

In general the results of the simulations can be summarised as follows:

– The simulated delta ANN showed better performance when learning in a distributed fashion as compared to massed learning.
– In specific ranges (particularly small values) the simulated ANNs react quite sensible on changing simulation parameters. Especially when learning in a massed fashion (i.e., one block of 100 learning trials) network performance depends highly on parameters *a, g, l* and *f*. On the other hand when learning in a distributed fashion the ANN is much less subject to parameter changes. This holds also for the interaction of simulation parameters.
– With increasing retention intervals correct stimulus identification decreases and incorrect identification improves.
– Introducing random factors generally produces regression effects.

Conclusion

Our simulations clearly show that even in a simple ANN by choosing appropriate simulation parameters it is possible to produce any network performance. So we can conclude that when producing a desired (e.g., Turingesque) ANN performance we cannot be sure to have found the only true explanation of whatever behaviour. From this follows that ANNs are a very proper tool for specific technical applications (like pattern recognition) but not for valid explanations of human behaviour (cf. also Loeck, 1986).

References

Loeck, G. (1986). Ist Simulation Erklärung? Cognitive Science – wissenschaftstheoretisch betrachtet. *Zeitschrift für allgemeine Wissenschaftstheorie 17* (1) , 14 – 39.

Schmidt, R. (1998). *On the temporal distribution of learning: A unifying approach based on the delta-learning rule*. Darmstadt: Institut für Psychologie.