Computational approaches to affective processes: evolutionary and neural perspectives

den Dulk, P.

Citation for published version (APA):
den Dulk, P. (2002). Computational approaches to affective processes: evolutionary and neural perspectives
Chapter 3

Connectionist Models of Individual Differences

This chapter focuses on neural network models of individual differences. Two models are presented which deal with two different tasks. On both tasks differences in response latencies have been observed between groups of low and high anxious participants. First, a model of performance on the emotional Stroop task is presented. In this task a high-anxious and low-anxious group show a difference in the way attention is modulated by a threat prime on a color naming task. Second, a model for performance on the attentional deployment task is presented. Here there is a similar difference in the effect of a threat prime on localization of a probe.

Although the architecture of the two models is different, the basic mechanisms are the same, i.e., pre-activation of a node gives this node an advantage in the competition with another node. The response to a subsequent stimulus can thus be hampered or facilitated, depending on which response was required and which stimulus was presented beforehand. On both tasks two different groups of subjects need to be simulated. This was done by varying the connection strengths in the model, while leaving the architecture unchanged.

Individual Differences in a Model for Stroop Performance

Although the Stroop task (MacLeod, 1991; Stroop, 1935) is perhaps the most extensively studied experimental procedure in psychology, relatively little work has been done on individual differences in Stroop performance. A modified version of this task, the emotional Stroop task (for a review which discusses almost all studies with the modified Stroop task, see Williams, Mathews, & MacLeod, 1996) has, however, concentrated on such differences. The classical and the emotional Stroop tasks both show interference in color naming due to the processing of irrelevant words that should be ignored. Whereas in the former version these words are color-related, in the latter version the interfering words are usually related to some concern, which may either be a permanent characteristic of a person or may be temporarily activated. Both trait and state anxiety may, thus, lead to increased interference by threatening words,

---

2 Part of the chapter is based on Phaf, Christoffels, Waldorp, and Den Dulk (1998).
but also other emotional states and traits, such as depression, show similar patterns of interference by concern-related words.

Some progress in understanding the classical Stroop effect has been made in a number of connectionist models (e.g., Cohen, Dunbar, & McClelland, 1990; Phaf, Van Der Heijden, & Hudson, 1990), the issue of individual differences has, however, only been addressed in the framework of one of these models (Matthews & Harley, 1996). The recently suggested 'defective inhibition' hypothesis for individual differences (Fox, 1994) however, seems very well suited for implementation in the other type of connectionist model. Here, we explore the opportunities for modeling individual differences in a highly simplified connectionist model designed according to the same principles as Phaf, et al.'s (1990) model.

In a recent study on the relation of trait anxiety to attentional bias, Fox (1994) suggested the interesting hypothesis that high-anxious participants have difficulty effectively inhibiting distracting information, resulting in a general inability to maintain attentional focus. This hypothesis was based on findings (Fox, 1993) that threatening words flanking color patches interfered with color naming (i.e., a modified Stroop task), even at a visual angle of two degrees and only for the high-anxious participants. Fox (1994) argued that "... a deficit in the effective inhibition of distracting information may be the causal mechanism which produces this preferential allocation of resources to threat-related information in anxious individuals. This may be termed the defective inhibition hypothesis" (p. 191, italics by Fox). In general, the defective inhibition hypothesis seems to imply that Stroop interference decreases with an increase of inhibition. What type of defective inhibition is referred to, however, and whether it is limited to threat-related material or whether it extends to attentional processing of all types of material, is not quite clear.

Fox (1994) extended the defective inhibition hypothesis to include negative priming in which the effects of inhibition are assessed on the trial following the presentation of distracting information, not only on concurrent responding as in the Stroop task. The consequences of defective inhibition may, however, differ in the concurrent and sequential versions of the attentional tasks. If in a priming task the distractor is inhibited poorly on some trial, it may still be active (i.e., in working memory, Phaf, Mul, & Wolters, 1994; Phaf & Wolters, 1997) and processed efficiently on a subsequent trial. An absence of negative priming was indeed observed for high-anxious participants, but not for low-anxious participants (Fox, 1994). For concurrent processing of target and distractor, the effect of defective inhibition on interference is not immediately evident. Since the target and distractor would mutually inhibit each other, defective inhibition would enhance both the activations corresponding to the distractor and the target stimulus. It is not clear how this would affect the competition between target and distractor (i.e., in a Winner-Take-All process). Here we will investigate whether defective inhibition increases or reduces interference in a task involving concurrent processing (e.g., a Stroop task). We will not consider negative priming because it does not seem to raise this question, would require realistic implementations of working memory (e.g., Phaf et al., 1994; Phaf & Wolters, 1997) in the model, and, finally, has also been doubted by Fox (1995) to involve inhibition.

Inhibition can be conceptualized in many different ways. Fox (1994) seemed to define inhibition in terms of outward behavior. When a response is given to some target (e.g., a color in a Stroop task), distracting information (e.g., the color word) has to be inhibited. When there is little interference, according to Fox (1994), then such inhibition is successful. An important question, however, is how this type of
inhibition should be translated into an explanatory mechanism or process. Another type of inhibition has been postulated by Gray (1987), which, contrary to the defective inhibition hypothesis, should increase with level of anxiety. Further, inhibition can also be envisaged on a neural level, representing the tendency of neurons to suppress activity of neighboring neurons. There are indications that neural inhibition increases with the concentration of particular neuromodulators (e.g., norepinephrine, Keeler, Pichler, & Ross, 1989). Again in contrast to the defective inhibition hypothesis, the level of norepinephrine (i.e., of inhibition) seems to be positively related to level of anxiety. Clearly, the concept of inhibition in the defective inhibition hypothesis is in need of further specification, so that a more exact relation between theoretical hypotheses and experimental results can be established. It is not clear to us how Fox' (1994) interpretation of inhibition could function as an explanatory mechanism, therefore, we will present our own interpretation of inhibition and will specify a mechanism which could form an explanation for the behavioral data.

Stroop performance has been explicitly linked to inhibition in the Phaf et al. (1990) model. In the models of Cohen, et al. (1990) and Matthews and Harley (1996), learning processes may result in inhibitory connections, but these are not an explicit architectural feature. Whereas these models focus on learning and automatization processes in homogeneously connected layered networks, Phaf, et al.'s (1990) model concentrates on competitive processes in pre-structured, modularly organized, networks. In the Phaf, et al.'s (1990) model responses are the result of resolution of competition between mutually inhibiting activations due to target and distractor stimuli. We wanted to investigate whether in this type of model (concurrent) Stroop interference decreases with inhibition due to the effective suppression of the activation due to the distractor or whether it increases with increased competition with its activation.

To examine this question, we constructed what we see as the simplest possible connectionist model of Stroop performance, and then varied the strength of all inhibitory connections in the simulation as an experimental manipulation. The model possessed representations for colors and color words only. Though there are consistent ways of incorporating different stimulus positions and other features in the model (see Phaf et al., 1990), this has not been done in the present simulation. The aim of this simulation was to demonstrate the effect of inhibition on Stroop interference in a qualitative manner and not to address all intricacies of the Stroop task.

To be able to compare with Phaf, et al.'s (1990) model, moreover, but also for reasons of simplicity, we used the classical Stroop terminology and representations in these simulations. Alternatively, one may think of the nodes representing color words actually representing threat words. A difference between the classical and the modified Stroop tasks, however, is that the latter does not allow for congruent conditions. The modified Stroop task is, generally, applied in anxiety research because it is believed (e.g., Williams et al., 1996) to rely on similar attentional mechanisms to those in the classical Stroop task. Both a color word and an emotional word lead to activations and to response tendencies which seem to interfere with color naming. In fact, much theorizing on anxiety is done in terms that are borrowed from theorizing on the Stroop task and other attentional tasks (e.g., Fox, 1993; MacLeod, 1991; Williams et al., 1996). The simulation of the defective inhibition hypothesis in terms of the classical Stroop task is also justified when a more general defective inhibition in high anxiety participants (i.e., affecting all inhibition, not only of threatening material) is assumed. This may explain why high-anxious participants
would also be slower than low-anxious participants in naming color patches (Fox, 1994), or in making lexical decisions (MacLeod & Mathews, 1991), even when non-threatening distracting words are presented.

**A model of Stroop performance**

The architecture of the model was modular, like that of the model for visual selection tasks of Phaf, et al. (1990), which in turn was based on the interactive activation model for contextual effects in letter recognition by McClelland and Rumelhart (1981).

Both models decomposed stimuli into their constituent features (e.g., shapes, colors, positions). In subsequent layers different features are represented in different modules. Within a layer the spread of activation due to stimulation results in a competitive relaxation process after which only one node remains active in each module. When many nodes are simultaneously active in a module, competition can only be resolved with some difficulty and interference effects result. Reaction times can be measured in these models by counting the number of calculation steps (i.e., iterations) needed to solve competition in the response module. Modeling excitation-inhibition ‘battles’ in networks allows for simulating interference, facilitation, and disambiguation effects in a number of tasks provided a suitable modular architecture has been designed for the particular task. Modeling the asymmetry in the Stroop task for instance, required the inclusion of a direct word-reading pathway which could suffer from less competition (due to the smaller number of modules) than the color-naming pathway.

A large number of simplifications, in comparison to Phaf, et al.’s (1990) model, were carried out: Representations for shapes of stimuli were, for instance, not included. The basic form of these modular networks, with horizontal inhibition leading to intramodular competition and vertical excitation leading to intermodular spread of activation, was, however, maintained. Intermodular inhibition, present in the former models, can, however, also be obtained by having excitatory connections to competing nodes in another module. A further simplification was that no recurrent connections to lower modules were present in this model. Such connections are useful for implementing top-down disambiguation effects (e.g., for the recognition of masked letters in word contexts), but are not strictly necessary for simulating the basic Stroop effect, as can be seen from Cohen, et al.'s (1990) model which contains only forward connections.

The model was constructed in the neurosimulator package MacBrain 3.0 (Jensen, Kennedy, & Blanson, 1991) and consisted of nine nodes, of which five were input nodes, two were intermediate, and two were output nodes (see Figure 8). Activations were iteratively updated. Iterations as a unit of time arise through the implementation of parallel models on a sequential computer. In a parallel network all activations are recalculated simultaneously in one time step, whereas a computer has to calculate them one after another. The input was distributed over three modules (i.e., word-input, color-input, and instruction). Two input nodes represented the input of the two color words 'blue' and 'red' (word-input), and two input nodes represented the input of the color of the letters (color-input). The instruction node served to discriminate between color-naming and word-reading conditions. Activation of this node required the model to respond to the color-input.

The general architecture of the model can also be derived from an analysis of the requirements for correct Stroop performance. It is, for instance, clear that color
and word processing should interact at some level to produce interference effects. The basic Stroop asymmetry (i.e., interference by incongruent distractors only with color-naming) is reproduced in this model due to the additional processing and opportunity for competition with color naming. Cohen, et al.'s (1990) model and Phaf, et al.'s (1990) model offer different explanations for the Stroop asymmetry. Whereas Cohen, et al. (1990) assume different amounts of practice in word reading and color naming, Phaf, et al. (1990) assume that the a priori architecture favors word reading above color naming.

Figure 8. Schematic representation of the model for Stroop performance. An arrow-head depicts excitatory and filled circles inhibitory connections. With these input activations (grey shadings), the output node 'blue' will win the competition.

The architecture of this model provides for two pathways from input to output. The first one leads directly from the word-input module to the output module and the second one has an intermediate module between color input and output. As can be seen in Figure 8, there are two nodes in the intermediate module, depicting internal representations of the colors. The intermediate module represents an additional processing stage in the transformation of colors (input) to color names (output). The node for color-naming instruction raises the activation of the winner of the competition in the intermediate module, so that the activations in the color pathway can dominate over the activations in the word route. More than three layers of modules can be implemented in the model (e.g., additional modules can be included for transforming color-word identities into spoken words), but these were omitted in
this simple implementation. So, the model does not necessarily assume that interference only takes place at an output or response level.

The connections between nodes were fixed and could not be modified by learning. Input to a node was determined by summing all weighted activations to the node. Subsequently, the activation of the node (between 0 and 1) was calculated according to the shunting activation rule (see McClelland & Rumelhart, 1981). The decay of activation in this rule was chosen to be 0.1 and the resting activation level was set at zero. The connection weights were set by hand to obtain correct Stroop behavior. Effective values could vary over a large range and a rather arbitrary choice was made somewhere in the middle of this range. Word input-output connections (a) were chosen to be 0.8, color input-intermediate connections (d) 1.0, intermediate-word output connections (b) 2.0, and finally instruction-intermediate connections (c) 2.0 (see also Figure 8).

To simulate the congruent condition, compatible (i.e., leading to the same output node) colors and words were activated (e.g., simultaneous activation of the word node 'blue' and color node 'blue'). The incongruent condition was created similarly by activating incompatible input nodes (e.g., word 'red' and color 'blue'). A neutral condition was created by activating a color-input node only. All input nodes belonging to an input pattern were set at an activation level of 1.0, whereas all other input nodes remained at zero activation.

**Varying inhibition**

**Method**

The behavior of the model was examined only over the range of inhibitions for which correct Stroop behavior in all conditions, including a Stroop asymmetry, could be obtained (e: -2.3 to -3.5; f: -3.3 to -4.5). The inhibitions in the intermediate module (f) and the output module (e) were varied simultaneously in steps of 0.1. The latter inhibition was always 1.0 higher than the former.

To make the model's behavior more comparable with experimental results, measures of reaction time and error rate were developed. To determine the reaction times in each condition, we calculated 90% of the maximum (asymptotic) activation value of the winning node, representing a state in which sufficient information was available on which node would win the competition but the competition process itself had not yet been fully completed. The reaction time in number of iterations from the start of stimulus presentation at which the activation reached at least 90%. The number of iterations is often transformed linearly to milliseconds on the basis of empirical fitting, but this is not done here as this would suggest a degree of accuracy, which is not actually warranted. The error rate was operationalized in the model by summing the total activation received by the competing non-target node over all iterations. This represents a measure of the probability of the incorrect response being produced and also gives an indication of the interference by the distracting color word.

**Results**

Without instruction (i.e., word reading), processing in the direct pathway dominated over processing in the indirect pathway and output corresponded to the word input. Reaction times for the congruent and incongruent conditions hardly differed. So, the model showed no interference with word reading. The interference in color naming by
the incongruent word, however, was clearly visible in the activation plots of the output nodes (Figure 9). Initially, the (incorrect) word response was more strongly activated than the (correct) color response. Eventually, the color activation, which was helped by activation from the instruction, recovered and succeeded in suppressing the incongruent word.

Figure 9. The activation levels of the two output nodes with inhibition levels $e$: -3 and $f$: -4. As a function of time in iterations which represent the time unit of the model. In the example, the stimulus was the word RED in blue.

Figure 10. a. The relation between inhibition and reaction time (in iterations) in the model. b. Incongruent and congruent error rates as a function of inhibition level.
The levels of activation and reaction times in the congruent condition of color naming were exactly the same for all inhibition conditions. The maximum activation was always reached sooner here than in the incongruent and neutral conditions (see Figure 10a). The error rates were effectively zero for congruent and neutral conditions (congruent: between 0.054 for the lowest and 0.012 for the highest level of inhibition). Only in the incongruent condition was an effect of inhibition level found. This resulted in a positive correlation between level of inhibition and reaction time, \( r=0.78, \) \( p<0.01 \). There also appeared to be a second-order polynomial trend in the dependency of reaction time on inhibition (second-order polynomial: \( r=0.96, p<0.01 \)). This may reflect the effective suppression of the distractor with small inhibition, whereas the interference by the distractor increased with larger inhibition. Similar patterns were found for the error rates (see Figure 10b). This was evidenced by the linear \( (r=0.68, p<0.05) \) and second-order polynomial \( (r=0.95, p<0.01) \) correlations.

In conclusion, a predominantly positive relation between interference and inhibition was found in this simulation of the Stroop task, which seems to contradict the central claim of the defective inhibition hypothesis (Fox, 1994), that interference increases for decreasing inhibition. Interference decreased somewhat with increasing inhibition only for small inhibitions. The rise in interference was, however, much larger with stronger inhibition. Thus these results may rather be seen to support an 'effective' inhibition hypothesis for anxiety. Interference even appears to rise quadratically with inhibition. Particularly when inhibition is very high (see Figure 10), as may be the case in clinical populations, interference effects may be observed in the Stroop task. Effective inhibition as a consequence of high anxiety may be a functional mechanism because, in this simulation at least, it inhibited concern-irrelevant but task-relevant (color-) processing and may have facilitated concern-relevant (word-) processing. In general, such a mechanism would enable attentional shifts away from the current task and towards potentially threatening stimuli and situations.

The experimental results corresponding to these simulations also seem to show more interference by incongruent color words on color naming for individuals with high levels of anxiety than with lower levels of anxiety. Whereas Watts, McKenna, Sharrock, and Trezise (1986) found only a small and nonsignificant increase in classical Stroop interference for spider phobics relative to low-anxious controls, Dawkins and Furnham (1989) reported a much larger increase for a non-clinically high-anxious group than for the low-anxious controls.

A clear deficiency of this approach of modeling individual differences in Stroop interference is that it does not seem to explain specific interference by threat-related words. It would only show why high-anxious individuals are more disrupted by all types of distracting material than low-anxious individuals. A general interference effect by anxiety may well be found, but a sufficient account would also need to incorporate attentional biases toward concern-specific material. Neither a general defective inhibition hypothesis, as seems to be suggested by Fox (1994), nor our effective inhibition hypothesis seems to provide an account for such specific interference.

An interesting alternative, though in a different context, has been provided in the work of Dixon, Brunet, and Laurence (1990). They hypothesized that individuals may differ in the general fluency of processing verbal information, possibly due to different amounts of practice (cf. Cohen et al., 1990; see also Matthews & Harley, 1996). Their hypothesis was supported by connectionist simulation work in which they showed that Stroop interference increased with the strength of the connections.
between word input and word responses. According to Dixon, et al. (1990), strong verbal connections would correspond to participants high in hypnotic susceptibility. This hypothesis was corroborated by experimental findings that these participants show larger interference in classical Stroop tasks, both in subliminal and supraliminal conditions, than participants who are not very susceptible to hypnosis.

Research in our group (e.g., Elzinga, de Beurs, Sergeant, Van Dijck, & Phaf, 2000) showed high correlations between hypnotic susceptibility, dissociative style, and trait anxiety. So, the general mechanism postulated by Dixon, et al. (1990) may also be at work with high-anxious participants. The hypothesis regarding the strength of the verbal connections, which we will call the excitation hypothesis, can also be applied specifically to the verbal strength of threat-related words, so that it can explain why interference for this material is larger than for neutral material. This contrasts with the inhibition hypothesis, for which it is not immediately clear how it can be made to work for only one type of material.

In our next simulation, we tested the hypothesis that raising strengths of verbal connections strengths results in increased interference in our model. The variation in strength of verbal connections between word input and word output may reflect different amounts of relevance the words may have to a person, or different amounts of practice a person may have had with them.

### Varying verbal connection strengths

**Method**

In this simulation, the strengths of verbal connections were systematically varied, whereas the values of all other connections were kept constant. For the inhibitory connections, we chose an intermediate value ($e: -2.6; f: -3.6$) from the range in the first simulation. Strengths of verbal connections were again varied over the total range in which the model would show correct Stroop behavior ($a: 0.7$ to $2.3$), including Stroop asymmetry.

**Results**

It should first be noted that in the absence of competition (i.e., distraction) reading times of the words in the model were not strongly affected by increasing the strength of verbal connections. The faster increase in activation at the output level was offset by the rise in asymptotic activation. Similarly, Martin, Williams, and Clark (1991) found no clear effect of trait anxiety on the reading times of threatening words in their Stroop experiments.

The reaction times in the congruent condition were approximately the same over all strengths of verbal connections. Also the error rates in this condition remained close to zero (between $0.006$ and $0.044$), showing that there was hardly any influence from the non-target node in this condition (see Figure 11). Higher strength of verbal connection in the incongruent condition resulted in higher reaction times and error rates. The correlation between strength of verbal connection and reaction time was positive ($r=0.89, p<0.01$), as was the correlation with the error rate ($r=0.88, p<0.01$). This time, there was no indication of an initial decrease of interference. As could be expected, interference increased monotonically with the amount of activation of the distracting word. The increase, however, accelerated for greater strength of verbal connection (reaction time: second order polynomial $r=0.98, p<0.01$; error: second order polynomial $r=0.98, p<0.01$) until the distractor dominated the response.
and 100% incorrect responses were produced. This polynomial trend was caused by the fact that for low activations the distractor was suppressed quite easily, whereas for higher activations, distractor and target had about equal activations and it took a disproportionately long time to solve the competition.

![Figure 11. a. Congruent, incongruent, and neutral reaction times (in iterations) for the simulation in which strengths of verbal connections were varied. b. Error rates in the incongruent and congruent conditions from the simulation of strength of verbal connection.](image)

**Discussion**

Although the model is quite simple, the simulations support the notions that high anxiety may either correspond to a general high inhibition or to high strength of verbal connection to threat words. The current version of the defective inhibition hypothesis (Fox, 1994) was not supported by these simulations. Only for low inhibition was there some indication that the inhibition of the distractor was more important in the interference process than was the inhibition of the target.

For both the inhibition hypothesis and the excitation hypothesis an underlying mechanism seems available. The global increase in inhibition may be due to the production of some neuromodulator. Whereas several candidates for such a global modulatory function on neural networks have been suggested (e.g., norepinephrine, Keele et al., 1989), there are almost no connectionist models that simulate these modulatory effects. Higher strength of verbal connection may develop as a result of greater experience with threat-related words by high-anxious individuals. In a way, threat-related words are very high-frequent for these individuals. An important difference between the two accounts is that the inhibition hypothesis does not specifically address threat-related representations, whereas the excitation hypothesis can explain why this material produces more interference than does neutral material. Of course, there is no logical necessity for deciding between the two accounts and both may be required to provide a full explanation of modified Stroop performance in high-anxious individuals.

Though, in principle, many alternative models for these individual differences seem possible, only few are suggested in the literature. In their comprehensive review of the emotional Stroop task, Williams, et al. (1996) suggested three possible ways of simulating emotional Stroop interference in a connectionist model. Their first proposal, that increased interference effects reflect extended exposure (see also
Matthews & Harley, 1996), corresponds to the excitation hypothesis. This account has been implemented by Dixon, et al. (1990) and by us, and has indeed been shown to produce the postulated increased interference. Williams, et al. (1996), however, rejected this account because emotional Stroop interference appears to decrease after therapy, which constitutes additional practice with threat-related material and so should lead to more interference, not less.

Williams, et al. (1996) suggested two further ways of modeling interference which would not suffer from this criticism. The resting level of activation of input nodes for emotional words may be elevated for concern-related stimuli, so that nodes representing these stimuli effectively have higher activations and interfere more with the production of target responses. This account, however, suffers from the same problem as the excitation hypothesis. Resting levels of activation probably develop as a result of practice and are not reduced by additional practice.

The final suggestion by Williams, et al. (1996) entails that the nodes associated with threat are subject to neuromodulatory control so that in high-anxious states the responsivity of these nodes is increased. This suggestion may be taken to correspond to the inhibition hypothesis, but their connectionist framework (cf. Cohen et al., 1990; Matthews & Harley, 1996) does not explicitly contain modules or intramodular inhibition and competition and, therefore, does not implement the inhibition hypothesis.

In our view, the last suggestion or a similar design may solve the problem raised by Williams, et al. (1996). We propose that anxiety activates (an assembly of) nodes which, either by neuromodulatory control, or by direct network connections, facilitate processing of concern-related words. It represents a modified version of the excitation hypothesis, in which the high strengths are not necessarily brought about by practice with threatening words. High trait anxiety would correspond to high activation levels of this 'anxious' assembly. It can be compared with the instruction node in our present model which favors processing of colors above processing of words. Similarly, 'anxious' instruction nodes may facilitate the processing of threatening material. Therapy could serve to regulate activity in this assembly of nodes, so that there is less facilitation of verbal processing of threat-related words.

A similar design also seems to be favored by Matthews and Harley (1996). They postulated an additional task demand unit for threat monitoring in their networks. The network was subsequently trained to respond to emotion words only during which the threat-monitoring node was activated. When this node had a low level of activation during testing, color naming of emotion words was impaired to a larger degree than color naming of neutral words. According to Matthews and Harley (1996) therapy may reduce the likelihood of activation of the threat-monitoring node and thus reduce interference. The extant models on individual differences in Stroop interference seem to converge on the presence of (an assembly of) nodes, which get their activation either due to prolonged personality traits or due to momentary states. The nodes subsequently have modulatory effects on processing of neutral stimuli but even larger effect on the processing of concern-related stimuli. We agree with Matthews and Harley (1996) that this probably provides the most promising mechanism for explaining emotional Stroop effects but would also like to note that this hypothesis is not incompatible with either the inhibition or excitation hypothesis. In fact, both may have played a role in their simulations, but they cannot be distinguished very clearly because inhibition and excitatory connections are mixed in the back-propagation framework and may even interchange
The simulations with the present model by no means prove that the actual system performs these tasks in the same way. They merely help to clarify the relation between theory and empirical data, but do not exclude other implementations or models. The first simulation, for instance, showed that, in contrast to expectations derived from a conceptual model, defective inhibition does not necessarily lead to more interference in a computational model. Of course, it can be argued that inhibition should be envisaged in some other manner, but then it should be specified how this type of inhibition relates to interference and simulations with a model implementing this type of inhibition should be done. A conceptual model may not be sufficiently detailed to provide exact predictions for experimental results. On the other hand, implementation of a model requires many further specifications which are not prescribed by the conceptual model. From the fact that the excitation hypothesis has been simulated in different connectionist frameworks (i.e., by Dixon et al., 1990), it can be deduced that many details of implementation (e.g., activation rules, etc.) do not matter much. In so far as they do matter, they may provide further hypotheses for experimental research and so stimulate further development. The present simulation may give rise to research in which the roles of inhibition and competition are investigated. Simple lexical decision, for instance, does not seem to involve much competition. Enhanced processing of threat-related material due to stronger verbal connections would only facilitate lexical decision. If inhibition indeed plays a role in anxiety, the introduction of competition in the experimental design (e.g., by presenting two letter strings at different positions) would be required to show differential effects of high and low anxiety on lexical decisions involving threat words. Such results have indeed been obtained by MacLeod and Mathews (1991), which may be interpreted as showing that an excitation hypothesis alone is not sufficient for explaining attentional effects of anxiety. We feel that the research into information-processing effects of emotion provides many more opportunities for specifying a more exact relation between theory and data.

**Individual Differences in a Model of Attentional Deployment**

One of the few hypotheses explicitly linking anxiety to cognitive processes is the 'attentional bias' hypothesis, which assumes 'that anxiety is associated with the assignment of high processing priorities to related information from memory' (MacLeod & Mathews, 1991, p. 599). The effects of such priorities can be measured most purely in the attentional deployment task where there is a competition between the processing of different stimuli at different locations of the visual field.

The task we will focus on in this research was first performed by MacLeod, Matthews and Tata (1986). In this task two words were presented simultaneously above and below the center of the screen. The subject was instructed to always read the top word. In some cases a probe appeared right after the presentation of the words at the position of either the upper or the lower word. The subject had to press a button as soon as the probe was noticed. One of the words in each pair, however, was emotionally threatening. The presentation of such a threat word influenced the reaction to the probe. If there was a fast reaction this was considered as an indication the preceding word had directed the attention of the subject towards that location.

Results from the first experiment with the attentional deployment task (MacLeod et al., 1986) indicated that there was a difference between two groups of subjects. The high-anxious participants seemed to direct their attention towards
threatening material, while low-anxious participants tended to direct their attention away from threatening material.

Here we present a connectionist model, which simulates the results from this experiment, addressing how the differences in trait-anxiety between participants may lead to differences in attentional performance. The models for the two types of participants will have the same architecture and will differ only in the value of some of the connection strengths.

The experiment

In the experiment by MacLeod et al. (1986) a group of 16 high-anxious participants and a group of 16 low-anxious participants were selected on the basis of prior testing. For the experiment 48 word pairs were used, consisting of one threatening word and one non-threatening word. Each word pair was presented for 500ms. One word was presented above the center of the screen, and one below. The threat word could appear in each of the two locations with equal probability. After the presentation of the word pair a dot probe appeared in 33% of the trials, which was presented either in the upper location or the lower location. Participants were asked to always read the upper word aloud and only press a button as quickly as possible whenever the probe appeared. Response latencies were recorded (see Table 1).

<table>
<thead>
<tr>
<th>Trait</th>
<th>Location probe</th>
<th>Threat in upper area</th>
<th>Threat in lower area</th>
</tr>
</thead>
<tbody>
<tr>
<td>High anxiety</td>
<td>Upper area</td>
<td>592.82</td>
<td>651.98</td>
</tr>
<tr>
<td></td>
<td>Lower area</td>
<td>694.53</td>
<td>662.85</td>
</tr>
<tr>
<td>Low anxiety</td>
<td>Upper area</td>
<td>540.41</td>
<td>523.63</td>
</tr>
<tr>
<td></td>
<td>Lower area</td>
<td>583.65</td>
<td>615.99</td>
</tr>
</tbody>
</table>

Table 1: Response latencies to probe presentation in the experiment by MacLeod et al. (1986)

Latencies for probe presentation in the upper area are lower than the latencies of probe presentation in the lower area, because of the instruction to always read the upper word. The probe-above and probe-below conditions can, therefore, not be compared. Comparing latency pairs within each location shows that for high-anxious participants the shortest latencies are in the condition where the probe appears in the location of the threat word. For low-anxious participants this pattern is reversed. This indicates that high-anxious participants direct their attention toward the location of the threat word, and low-anxious participants direct their attention away from the threat word.

The model

To simulate the experiment a connectionist network was constructed. The network had no learning connections and was based on the competitive learning paradigm (McClelland & Rumelhart, 1981). The activation and learning rule were taken from Murre, Phaf, & Wolters (1992). In designing the model we set out to construct a minimal architecture with which simulation of all conditions in the experiment would be possible. Figure 12 depicts the network schematically.
Figure 12. Schematic representation of the model

The bottom layer consists of two modules with nodes that correspond directly to the different stimuli used in the experiment. The two middle modules have more abstract representations, which become activated by certain features present in the stimuli. The bottom and top layer have bi-directional connections, allowing feedback from the feature modules back to the stimulus modules, so that it can facilitate the processing of specific stimuli. Attention for a location can be simulated by feedback activation from the Location Module to stimuli at specific locations in the stimulus layer.

Inputs: There were six different inputs to the network. The three types of input, threat word, neutral word and probe, were represented separately for the two locations, upper and lower. The inputs were presented to the network from an input layer that had one to one connections to the stimulus layer of the network.

Processing modules: The 6 nodes in the two bottom modules of the model, the Word Module (4-nodes) and the Probe Module (2-nodes), are direct representations of the 6 possible inputs in the experiment. The nodes in the middle two layers are representations for location and affect of the stimulus. The Location Module represented upper and lower locations on separate nodes, and the Affect Module represented threat and neutral words on separate nodes.

Output: In the experiment the dependent variable was latency of probe identification. In our model activation of a single node represented this response. When this activation exceeded a threshold the model produced the response. The threshold for probe-identification was simulated by way of a response-inhibition-node. This node had bi-directional inhibiting connections with the probe-identification-node, and had a bias to its activation, i.e., each iteration 1.0 was added to its incoming activation. Activation of the probe-identification-node was now only possible with enough activation coming from the probe-stimulus-nodes to inhibit the response- inhibition-node and to overcome the inhibition by this node.

Instruction bias: In the experiment, participants were instructed to read the above word aloud, causing systematically faster reactions to probes presented in the top location. We simulated this by adding a bias of 1.0 to the incoming activation of the node representing attention to the above location in the Location Module (node AA in Figure 13).
Connectionist models of individual differences

Figure 13. Complete network and connection strengths for high-anxious participants.

**Connection scheme:** The connection scheme for high anxiety networks was such that the nodes of the bottom layer had connections to the nodes in the middle layer which represented the corresponding features. Thus, the Probe Above Node in the Probe Module was connected to the above location in the Location Module, and the Threat Word Above Node in the Word Module is connected to the Above Node in the Location Module and the Threat Word Node in the Affect Module. Both nodes in the Probe Module had connections to the Probe Identification Node in the Output Module. All connections were bi-directional. Within all modules there were inhibiting connections between all nodes. The exact connection scheme is depicted in Figure 13. The connections to the Location Module were 1.0 for threat words and 0.1 for neutral words, representing the difference in capacity to direct attention to them.

The low-anxious networks differ from the high-anxious networks in that they divert their attention away from the location of the threat word. Therefore, the model for low-anxious participants had connections from the Threat Above Word to the Below Location and from the Threat Below Word to the Above Location. For the remainder the model was identical to the model for high-anxious participants.

**Decay of activation:** The activation rule of the model specifies a gradual decay of activation. This is essential for the simulation because the general mechanism of attention is simulated by pre-activation of certain nodes. When a node is activated by a stimulus, it will have retained some trace of previous activation. When a second stimulus activates the node while the decaying activation is still present, its activation will be higher than without pre-activation, and this will be an advantage in the competition with other nodes. Only a small residue of activation can be decisive in a competition process between two mutually inhibiting nodes.

**Method**

Because the network was completely deterministic no replication or randomization procedure was necessary, but only a single presentation for each condition was
examined. There were four conditions, two types of word input, threat above or threat below, which could be followed by a probe in the below or above location. The word input always consisted of a threat word accompanied by a normal word in the other location. Within a trial, first the word pair was fed to the network for 60 iterations (word input activations were clamped at 0.5). Subsequently, no input was presented for 3 iterations. Then the probe input was fed to the network for 60 iterations. Before each trial all activations in the network were set to zero. The procedure was repeated for all four possible input patterns and for both networks (high-anxious and low-anxious).

Response latencies were established by recording the number of iterations before a stable level of activation was reached, i.e., when there was no change in activation (as determined by the MacBrain simulation software, Jensen et al., 1991) of the output node. To evaluate our model we compared the simulated results to the results from the experiment by MacLeod et al. (1986), as presented in Table 1. We did not attempt to produce a quantitative correspondence, however, but only the qualitative direction of the effects was compared to the actual results.

For most parameters the initial choices were sufficient. It was not possible, however, to find one level of inhibition for both intermediary modules which was strong enough to produce a clear effect of previous presentation in the Location Module without interfering with normal processing of stimuli in the Probe and Word Module. Thus, different inhibition levels had to be chosen for the Location Module and the Probe and Word Module. Also, the connections to and from the Affect Module were set to 0.5 instead of our initial choice 1.0, this is because high connection strengths interfered with processing within the Word Module.

**Results**

The simulation results (see Table 2) were similar to the experimental results. For the high-anxious model responses were faster when the probe appeared on the same location as the threat word. For the probe-above condition this was 64 instead of 68 iterations, for the probe-below condition this was 85 instead of 97 iterations. For the low-anxious model these effects were reversed. For the probe above condition this was 21 instead of 25 iterations, and for the probe below condition this was 48 instead of 51 iterations. For both models, responses to probes in the upper area were faster than probes in the lower area because of the instruction bias, as was the case in the experiment. The correlation between the experimental data en the simulated results was 0.96, t(7) = 8.69, p < 0.0001.

<table>
<thead>
<tr>
<th>Trait</th>
<th>location probe</th>
<th>Threat in upper area</th>
<th>Threat in lower area</th>
</tr>
</thead>
<tbody>
<tr>
<td>High anxiety</td>
<td>Upper area</td>
<td>64</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Lower area</td>
<td>97</td>
<td>85</td>
</tr>
<tr>
<td>Low anxiety</td>
<td>Upper area</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Lower area</td>
<td>48</td>
<td>51</td>
</tr>
</tbody>
</table>

Table 2: Number of iterations between probe presentation and reaching stable probe-identification-node activation.
Overall, the results obtained were in accordance with the experimental results. The mechanism responsible for this effect is the pre-activation of the location node. The pre-activation caused faster convergence if the probe was presented at that location of the active node. The high and low-anxious models only differed in which location node was activated by the threat word.

Discussion

A remarkable aspect of the original findings of MacLeod et al. (1986) is that normal participants apparently divert their attention away from threat stimuli. This does not seem to be adaptive behavior. If threat stimuli matter at all, we would rather expect normal people to direct their attention towards a threatening stimulus. Whether the subsequent reaction is withdrawal or approach is a second matter, but in order to make an appropriate choice attention seems necessary.

One explanation is that the threat stimulus is no real threat. It is just a word with a negative meaning, and only overly anxious people will treat it as a threat for pathological reasons. It remains mysterious, however, why low-anxious people should direct their attention away, and not simply ignore the threat words. MacLeod et al. (1986) hypothesize that this is to protect the cognitive system.

Mildly threatening stimuli are extremely common in the everyday environment. ... One could reasonably speculate that the perceptual bias in normal subjects may be protective in limiting increases in anxiety by excluding such minor threatening stimuli from the cognitive system at a very early stage. (MacLeod et al., 1986, p. 18)

Thus, because the environment is full of threatening stimuli the higher cognitive systems have to be protected from it, which is done by diverting attention away at a very early stage of processing. There are at least two assumptions underlying this argument. First, there is the assumption that too much threat information poses a problem for higher level cognition. Although we experience threat as unpleasant, it is not self-evident that higher cognitive systems should have a problem with processing threatening information. A second assumption is that more primitive systems are capable of discriminating between seriously threatening and minor threatening stimuli, and apparently on a very early stage. This is not a trivial task because if a truly dangerous stimulus is mistaken for a harmless stimulus this may have serious consequences. What is least likely, however, is that both assumptions can be true at the same time. It seems peculiar that the threat information at the higher level poses a problem, and yet more primitive systems can filter out irrelevant threats just fine. Rather one would expect this to be the other way around. Of course the human cognitive system is not designed for beauty, and there might be unknown reasons for such an organization, but at first hand it does not seem to be a very likely evolutionary solution.

Our model was not an attempt to make a mechanism as complex as the one discussed above, but we think that by building a simple model it has already become clearer what the implications would be for model based on the explanation of MacLeod et al (1986). Such a model would discriminate between two stages, an early stage, which performs a fast discrimination, and a later stage, which is supposed to be more sophisticated. We discuss a similar distinction between two cognitive systems in the following chapters. Although that distinction does not discriminate between
stages, but between routes, which operate in parallel. In that case the function of the more primitive system is not to do something the higher cognitive simply cannot handle, but to do it faster than the higher cognitive system could do.

An advantage of modeling is that it requires the explicit formulation of implicit assumptions. An important assumption in the interpretation of the effect of the attentional employment task is in the way attention functions. MacLeod et al. (1986) assume that focused attention means focussed at a certain location, and its consequence is facilitation of processing at that location. This could make sense as an evolutionary adaptation, since we could expect more relevant stimuli at a location where we have already seen something. However, we could also imagine attention to focus on a certain concept or type of stimulus. For example, when we have seen one sparrow, we could expect more, and thus our attention could become focussed on sparrows, as a general concept, and a subsequent encountered animal might be faster recognized if it happens to be another sparrow. This could make sense as an evolutionary adaptation just as well, because when we have seen some type of stimulus we could expect more stimuli of that type. A mechanism like that is present in our model in the form of the Affect Module, and the connections between the Word and the Affect Module. The connections between the Word and Affect Module, however, have been made weaker than between the Word and Location Module, and this is for a reason. With stronger connections presentation of a threat leads to activation of both nodes in the Affect Module, which feed back their activation to the Word Module and this is relayed to the Location Module. Both locations will be pre-activated, however, and competition between the locations will cause a longer delay in the response, which is at odds with the behavioral data. This indicates that this is not the way the brain functions, i.e., attention is directed to a location, and not to a type of stimulus.