

## A connectionist dual-route model for affective priming

PAUL DEN DULK<sup>1</sup> MICHAEL CAPALBO<sup>2</sup> R. HANS PHAF<sup>1</sup>

<sup>1</sup> Psychonomics Department University of Amsterdam The Netherlands

<sup>2</sup> Department of Neurocognition Maastricht University The Netherlands

**Abstract** - Though both LeDoux (1996) and Murphy and Zajonc (1993) claim that the dual-route model for fear conditioning and the affective priming results are mutually supporting, these results have not yet been simulated in a detailed connectionist model. First, the larger suboptimal (i.e., less conscious) than optimal (i.e., conscious) priming by emotional faces on the (indirect) evaluation of neutral targets, and its reversal with the (direct) instruction to evaluate the faces, was simulated in a dual-route network model. Under conditions of stress, the inhibitory effect of the longer route disappeared. Finally, larger suboptimal than optimal affective priming was learned by the network through weight modification in the long route. It is concluded that the simplicity of this kind of model, and the common evolutionary origins, make it improbable that separate networks for fear conditioning and affective priming have evolved.

More than as a specific neurobiological model for fear conditioning, the dual-route model by LeDoux (1986, 1996) may serve as a general view on the relationship between conscious and nonconscious emotion processes (see Buck, 2000; Phaf and Wolters, 1997). Due to the parallel nature of these two types of processes in the model, their contributions to emotions may differ qualitatively. Experimental evidence for such dissociations is not readily available from fear-conditioning studies in animals but can be found in human affective priming research (e.g., Murphy and Zajonc, 1993). Connectionist simulations (Armony, Servan-Schreiber, Cohen, and LeDoux, 1995; den Dulk, Rokers, and Phaf, 1999) have shown that the dual-route model is compatible with a large part of the fear-conditioning results, but a similar 'existence proof' of the model's ability to account for affective priming is absent. The architecture of the present connectionist model was based on LeDoux' conceptual model and on an analysis of the representational requirements for the affective priming task. For the sake of simplicity, elements that were not essential for affective priming, such as learning (i.e., dynamical weight modification), were dropped from the model in the initial simulations. All connections in the model were first set by hand, not by way of parameter fitting, but on the basis of a global analysis. The ensuing model, which was more simple than the fear-conditioning model (e.g., had fewer nodes), revealed differential effects of processing in the

Address for correspondence: R. Hans Phaf, Psychonomics Department, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam. E-mail: hphaf@fmg.uva.nl

direct and indirect routes similar to experimental results. No claim is made that the model even approaches the simulation of consciousness. The differential effects of conscious and nonconscious processing can, however, be obtained even in this relatively simple version of the dual-route model.

The dual-route model of LeDoux (1996) is primarily based on fear-conditioning research in which rat brain regions were systematically lesioned. In an experiment in which the entire input to the cortex was lesioned it was demonstrated that a direct route from the thalamus to the amygdala was sufficient for learning a CS-CR association. Extinction of fear conditioning, however, was not possible without these (indirect) cortical connections. It was concluded that after the extinction procedure a memory trace of the fear response was still preserved in the direct pathway but this response was actively inhibited through a cortical pathway. Extinction may represent a form of learning in which specific contexts are established where the conditioned response is not necessary. Interestingly, the functioning of the indirect route can also be affected by the US alone or other stressful stimulation. Following Jacobs and Nadel (1985), LeDoux (1996) has claimed that stress can bring back responses that were extinguished through the action of the indirect route, but were still preserved in the direct route. Jacobs and Nadel (1985) emphasized the suppression of hippocampal activity, which is presumably responsible for the processing of the context of an event. In humans, a similar phenomenon may be observed when anxiety disorders treated with behavior (extinction) therapy return in stressful situations. The affective priming paradigm also provides situations in which the indirect route counteracts processing in the direct route. Shifts in the balance between the functioning of the two pathways are investigated in the present modelling effort.

The general principle underlying LeDoux' model for affective processing (i.e., fast but global effects through a short pathway and slower but more detailed processing through a long pathway) can also be used as an explanation for the type of affective priming results found by Murphy and Zajonc, 1993 (see also Rotteveel, de Groot, Geutkens, and Phaf, 2001). In these experiments photographs of angry and happy faces, as well as affectively neutral control stimuli served as primes, and Chinese characters (or ideographs), which were unfamiliar to the participants, served as targets that had to be evaluated. A suboptimally presented (non-conscious, or at least less conscious than in optimal conditions) prime congruently influenced (i.e., primed) the affective evaluation of a subsequent stimulus, whereas optimal (fully conscious) presentation showed no influence, and in Experiment 1 even seemed to result in a slight (marginally significant) incongruent priming effect. Murphy and Zajonc (1993) hypothesized that the effects in the sub-optimal condition were due to a diffuse and nonspecific influence outside of awareness due to insufficient activation of higher centers. In the optimal condition, however, higher 'cognitive' processes (e.g., incorporating a context, in which an affective response to an emotional face is not relevant) corrected for the influence of the prime. They also cited the LeDoux dual-route model in support of affective-cognitive independence and primacy of affective processing. Clore and Ortony (2000), in contrast, concluded on the basis of their analysis of the role of

cognition in emotion that LeDoux' research is essentially irrelevant to the findings of Murphy and Zajonc. In their view, suboptimal affective priming reflects incomplete (semantic but not episodic) parsing of the prime in the sequence of processes possibly leading up to a full emotion, whereas processing in the direct route would only constitute rapid response preparation. The response generated in the direct route, moreover, would only be a degenerate, or at the very least nonrepresentative, instance of an emotion.

In our opinion, the discussion as to whether suboptimal affective priming and/or processing in the direct route entails cognition is largely misguided because it entirely depends on how broad one is willing to define 'cognition'. The more interesting question that is hidden in this discussion is whether one thinks that nonconscious and conscious processing are essentially the same (i.e., the identity position, see Mandler, 1985) or assumes that they may differ qualitatively. Both sides of the discussion acknowledge the role of both conscious and nonconscious processes in emotions but differ in their willingness to consider subjective (conscious) reports (e.g., of appraisals) as representative for all emotional processes (Phaf and Wolters, 1997). The structure of, for instance, appraisal processes (e.g., Shaver, Schwartz, Kirson, and O'Connor, 1987; Frijda, Kuipers, and ter Schure, 1987) is based solely on research using subjective report and then extrapolated, according to the identity principle, to also cover nonconscious emotional processes. The results by Murphy and Zajonc (1993) strongly suggest that, at least for emotions, this identity position may be violated.

Murphy and Zajonc (1993) also investigated what they called 'cognitive' (i.e., non-affective) priming. In Experiment 5, for instance, they studied the influence of male and female faces (without a clear emotional expression) on the judgment whether the ideograph represented a masculine or feminine object. In all their 'cognitive' priming experiments (3-5), stronger congruent priming was obtained in optimal than in suboptimal conditions. As can also be seen from the mere-exposure effect (e.g., Bornstein, 1989), the stronger suboptimal-than optimal pattern, thus, seems to be specific for affective processing. This privileged position of affective processing among all forms of 'cognitive' processing does not seem accounted for by the limited-parsing view of Clore and Ortony (2000). Semantic but not episodic parsing of the male and female faces, for instance, should in this account also lead to stronger priming on the masculine/feminine judgments than combined semantic and episodic processing. In this sequential account both affective and non-affective processing should go through the same stages. An account for the priming effects in terms of the dual-route model explains the difference between affective and non-affective processing by the availability of a direct route only for affective processing. The parallel indirect route serves to correct for 'false alarms' to emotional stimuli to which this type of system is prone. This can be likened to parallel accounts for the Stroop task (Phaf, van der Heijden, and Hudson, 1990). The privileged form of processing (i.e., reading a word) interferes with the slower naming of the color in incongruent conditions (cf. 'false alarms'), but facilitates it in congruent conditions. The finding of single cells in the subcortical amygdala responsive to specific facial expressions (Leonard, Rolls, Wilson,

and Baylis, 1985), moreover, further weakens the claim (Clore and Ortony, 2000) that LeDoux' model is irrelevant to the affective priming research.

A connectionist implementation of the dual-route model simulating the affective priming results would show that these results are, at least, compatible with the model. In many cases the relation between conceptual models and experimental results remains vague and open to alternative interpretations. An implementation of the conceptual model in a computational model forces one to specify many other aspects of the model which are sometimes only of secondary relevance to the processes studied. This may lead to a further development of hypotheses and predictions that may be tested experimentally. It has, for instance, not always been easy to replicate the Murphy and Zajonc results (e.g., see Rotteveel et al., 2001) and the modelling work may specify exact conditions when it can be replicated. If we, moreover, succeed in simulating these results in a relatively simple model, this forms a further argument against the claim (cf. Clore and Ortony, 2000) that fear conditioning and affective priming involve different parts of the nervous system. If evolution has provided for a fear-conditioning circuit of which we have shown that it can also perform affective priming, then it would be very unparsimonious to add a new circuit for a similar affective task. The ability to perform both types of tasks, moreover, seems to have evolved from the same selection pressures. The fast detection of threat, or the absence of threat, allows the organism to prepare itself in such a manner that, with fuller processing, it would need less time to act appropriately. A fear preparation to a happy stimulus, for instance, could lead to the organism missing out on advantageous situations, such as additional food or social support. These common origins and the common mechanisms, which we want to show in the present study, make it unlikely that fear conditioning and affective priming are subserved by different neural circuits.

A network model was constructed that had a similar architecture as the fear-conditioning models of Armony et al. (1995) and of den Dulk et al. (1999). Both models had a modular design with within-module inhibition and between-module excitation. A direct labeling of the modules with neuro-anatomical regions, as in the fear-conditioning models was avoided in the present model. Because this work is less neurobiologically motivated than the fear-conditioning work we cannot be completely sure which parts of the cortex are, for instance, involved in the indirect route. Also the visual nature of the stimuli, as opposed to the auditory nature in the conditioning experiments, may implicate other regions in affective priming than in fear conditioning. We would, however, preserve the general distinction between a subcortical direct pathway and a, partly cortical, indirect pathway. The correspondence between the simulation results of the two fear-conditioning models, furthermore, demonstrates that implementational details such as activation rules and learning rules do not matter much. We choose to continue with the rules and parameters of the model by den Dulk et al. (1999, see also Murre, Phaf, and Wolters, 1992).

Due to the within-module inhibition and the learning rules, both fear-conditioning models actually implemented competitive learning in the modules. We disabled in the first three simulations because we think it may play a role in the

long-term development of affective priming but not in a single priming trial. Moreover, if it plays a role in an affective priming experiment, it primarily has a weakening effect. The acquisition of an affective value by a previously neutral Chinese ideograph during the affective rating task will lead to a kind of inertia, so that the next evaluation of the ideograph will be similar to the first one, regardless of the affective prime preceding it. The contribution of learning to the development of affective priming is also not yet fully clear. Both evolutionary 'learning' (e.g., Ohman, 1992) and learning during the lifetime of an organism (e.g., fear conditioning), probably, play a role. We aim at showing, however, that a relatively simple mechanism can, in principle, perform the task. In the first simulation we wanted to demonstrate how the dual-route model performed the affective priming task. In Simulations 2 and 3 we investigated conditions which could interfere with the Murphy and Zajonc (1993) effect. In the final simulation the acquisition of the specific weight settings in the indirect pathway (i.e., through learning) required for the correction effect was simulated.

### **The model**

In Figure 1 the outline of LeDoux' (1996) model can be recognized. From the Input modules there is a direct connection to the Affect-Module, which corresponds to the direct route to the Amygdala in LeDoux' (1996) model. The route through the Indirect modules to the Affect module is longer and involves more elaborate processing than the direct route. Both inhibition and facilitation of the direct affective processing can take place through the Indirect-modules. In the fear-conditioning models the total activation of all nodes of the amygdala module served as output. This summed activation was considered to correspond with some measure of autonomic activation. With affective priming the output consists of a preference rating. Such ratings are clearly not produced at a subcortical level. An additional output module was, therefore, included which not only received input from the Affect module but also from the Indirect modules. These 'indirect' connections make it possible that preference ratings are made without affective involvement (i.e., without activation of the Affect module).

A more detailed representation of the model can be found in Figure 2. The Input-Group, consists of two modules that correspond to the two types of input presented in the experiment of Murphy and Zajonc (i.e., Chinese characters and faces expressing emotions). The input was presented by clamping a quasi-receptor node to a particular activation value. This node was connected to a further input node that would gradually develop its activation on the basis of the clamping of the first node. Only positive and negative faces could be presented to the two nodes in the Face pattern module. There were also two representations for Chinese characters of which one had a slightly positive and the other a slightly negative affective value, representing a small a-priori preference for some of the Chinese ideograph. Because it may be impossible in the actual experiment to obtain perfectly neutral characters, it was considered better to average the results over slightly positive and negative characters. A further type of information presented

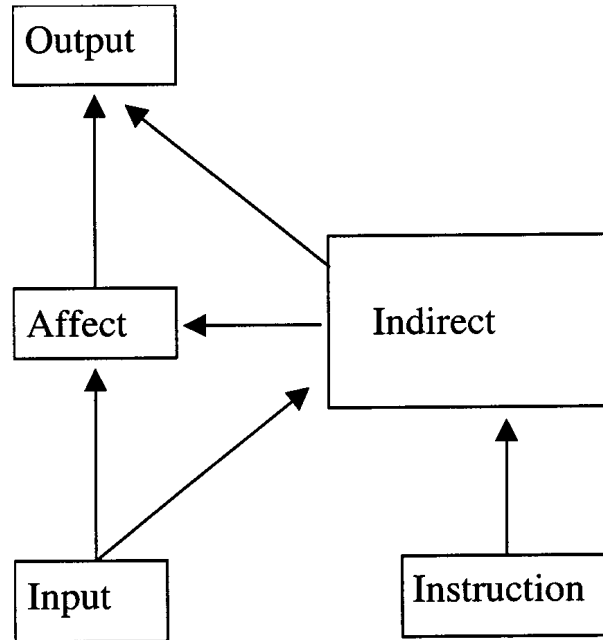


Fig. 1. Global structure of the model. Every box represents a group of modules (e.g., a layer). The group involved in the indirect processing is labeled 'Indirect'. Affective valence is activated in the 'affect' group. The instruction to either evaluate the faces (direct instruction) or the ideographs (indirect instruction) is represented in the 'Instruction' group.

to the network is the instruction to evaluate either the faces or the characters. In line with previous models (e.g., Phaf et al., 1990), the instruction is implemented as the pre-activation of a set of nodes which have representations that are compatible with the instruction. The Instruction-Module has two nodes, which stand for the (indirect) instruction to affectively evaluate the Chinese ideograph, or the (direct) instruction to evaluate the faces. In the modules of the indirect pathway all combinations of representations between instruction and face/ideograph are made leading to four nodes in the indirect face module and four in the indirect ideograph module. The Affect-Module, finally, consisted of two nodes, one node representing positive affect, and the other negative affect.

The activation rule (see Formula 1 and 2) was the same used as was used by den Dulk et al. (1999, see also Murre et al., 1992).

$$a_i(t+1) = (1-k)a_i(t) + \frac{e_i}{1+e_i} [1 - (1-k)a_i(t)] \quad (1)$$

when the input  $e_i \geq 0$ , and

$$a_i(t+1) = (1-k)a_i(t) + \frac{e_i}{1-e_i} (1-k)a_i(t) \quad (2)$$

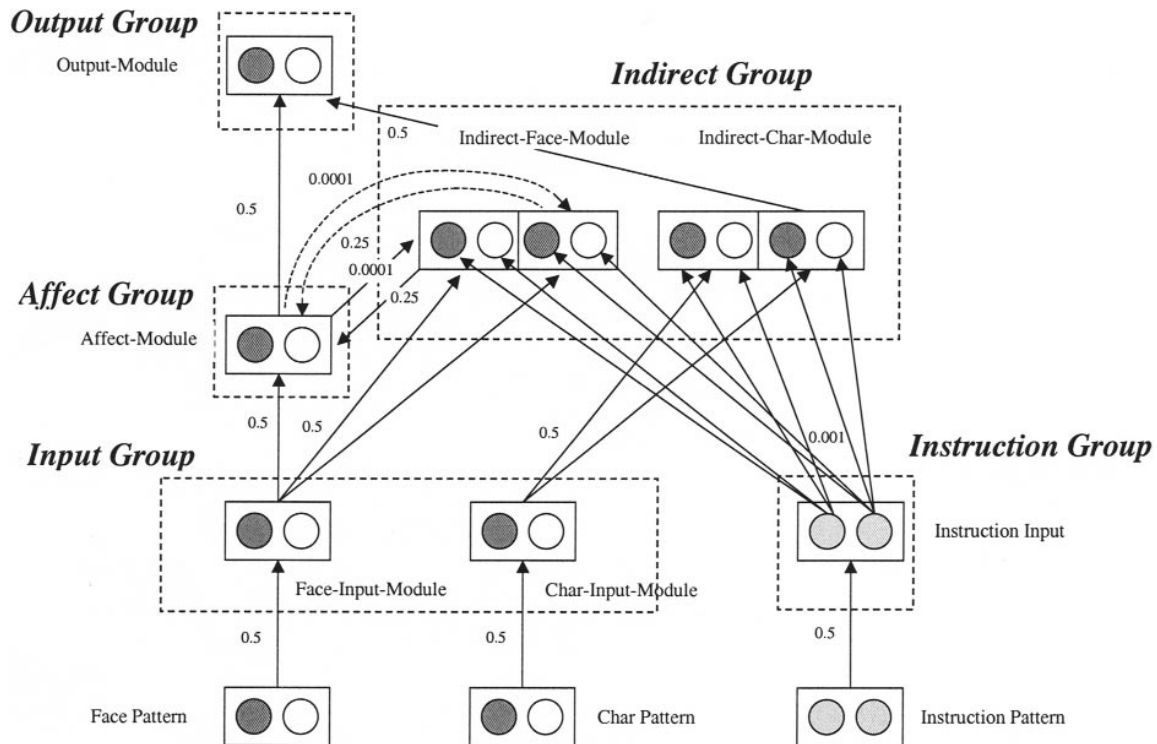


Fig. 2. Detailed structure of the network model. Boxes represent modules (i.e., groups of nodes) and circles represent nodes. Arrows between circles indicate connections between individual nodes. Arrows between boxes indicate one-to-one connections between the corresponding nodes in the two modules. The dotted arrows between modules indicate a crossed connection scheme (i.e., from node 1 to node 2 and from node 2 to node 1). Not shown are the inhibitory connections between all nodes within a module (-5.0 for all modules, except for the Affect and Output Modules, which have an inhibition of -0.1).

when the input  $e_i < 0$

$a_i(t)$  represents the activation of node  $i$  at time  $t$ ;

$k$  represents the decay parameter (0.25 in all simulations);

$e_i$  is the summed weighted input to node  $i$ ;

This rule consists of two terms: a decay term and an input term. The (arbitrary) time unit (i.e., iteration) is defined as the time it takes a node to calculate a new activation. A new activation comprises a part of the previous activation (i.e., decay) and a positive contribution when the net input to the node is excitatory, or a negative contribution when it is inhibitory. In this model it will, for instance, take three iterations after the start of input for the first activation to reach the output nodes. The net input is incorporated in the activation through a sigmoid function, so that the activation always remains in the  $[0,1]$  range. With a stationary input the activation of a node keeps changing over a number of iterations until

it reaches a constant level of activation, when net input to the node and decay of activation are in balance. In each iteration all nodes in the network are updated on the basis of the current input. To prevent deadlocks between competing nodes, a small noise activation was added to the incoming activation at each iteration. The noise was randomly chosen from a uniform distribution between  $-0.0001$  and  $0.0001$ .

### **Connections**

All the weights in the network were first set by hand. Initially, all excitatory weights were set at an arbitrary value of 0.5. The inhibitory connections should be strong enough to allow for the complete resolution of competition (i.e., only one node should remain activated after competition had ended). Inhibitory weights of  $-5.0$  proved sufficient for this purpose. The correction of direct activation through the indirect route needed some additional tuning. First, there had to be less inhibition in the Affect- and Output-Module, because otherwise the node which received the initial direct activation inhibited the other node so strongly that it prevented all further activation of these nodes, and thereby any correction effect. The inhibitory weights in the Affect-Module were, thus, set to the lower value of  $-0.1$ . Second, the weight from the face/indirect instruction nodes to the Affect-Module had to be set at an appropriate value to allow for a correction. At a value of 0.5 there was often no remaining influence of the direct route. At half this value (0.25) a good balance between direct and indirect processing seems to have been struck.

To allow for an evaluation (i.e., with an indirect instruction) of the Chinese ideographs when they are not preceded by affective primes, they should also have some small intrinsic affective value. This valence was assumed to only come about through processing in the indirect route. In the network this was expressed by excitatory connections from the character-nodes to either the positive or negative output node. These weights were set at the default value of 0.5.

There were connections from the Instruction-Module to the Indirect-group to bias attention either to congruent or to incongruent evaluations of the faces. When, for instance, a positive face was presented with an indirect instruction, all positive-face nodes and all indirect-instruction nodes receive activation. The node combining these two representations will receive the highest activation, however, and win the competition. The direct-instruction nodes have congruent connections to the nodes in the Affect module, whereas the indirect-instruction nodes have incongruent connections (the dotted lines in Figure 2). A somewhat similar set-up can be found for the character module in the indirect pathway. The congruent character nodes, however, only have excitatory connections to the corresponding output nodes. Such evaluations are considered to involve only 'cognitive' judgments. The incongruent character nodes have no connections to other modules, because no direct affect is assumed for the characters, that needs to be corrected by later processing. A small connection strength of 0.001 from the Instruction module to the Indirect modules was sufficient to bias competition.

A small connection strength was also given to the feedback connection from Affect-Module to Face-Indirect Module (0.0001). The function of these connections was to make sure that a congruent node would win when no instruction was provided. Without these connections there would be an equal chance of obtaining congruent or incongruent affective reactions to faces. This set-up was deemed necessary because it was assumed that in the absence of any instruction (or context), the processing of optimally presented emotional faces would take priority. No situations were, however, simulated where there was no instruction, so this weight did not contribute much to the results of the simulations presented.

### **Simulation 1**

Four conditions from Experiment 1 of Murphy and Zajonc (1993) were simulated: positive and negative suboptimal primes and positive and negative optimal primes. Within a trial, first the prime was presented by clamping one of the Face-Input Module nodes at 1.0. This presentation lasted 5 iterations in the suboptimal condition and 20 iterations in the optimal condition. The presentation time of 5 iterations for the suboptimal conditions was chosen so that it could activate the Affect Module through the direct connections, but did not lead to resolution of competition in the Indirect modules. Immediately after the prime, the target was presented by clamping one of the character-input-nodes to 1.0 for 20 iterations. The activation of the output node was recorded after the presentation of the target. The instruction to evaluate the character was activated during the whole trial.

The evaluation was determined by taking the difference in activation of the two output nodes (positive output activation minus negative output activation). The activation of the nodes could vary between 0.0 and 1.0, so the affective evaluation could vary between -1.0 and 1.0. Because Chinese characters also had some intrinsic affective value, in every condition both positive and negative characters were presented. For all four conditions output was averaged over 11 replications with a slightly positive character and 11 replications with a slightly negative character.

A regression fit (the slope was 14.6 and the intercept 3.1) was performed between the activations of this simulation and the ratings from Experiment 1 of Murphy and Zajonc (1993). This has no theoretical significance but facilitates comparison between simulation and experiment. All other simulation results in the paper were transformed according to the same regression fit.

### **Results**

The evaluations by the model are shown in Figure 3. The congruent priming effect from suboptimal conditions is reversed in optimal conditions. A *t*-test revealed a significant congruent difference in the suboptimal condition ( $t(10) = 94.6, p < 0.0001$ ). The incongruent difference in the suboptimal condition was also significant ( $t(10) = -46.1, p < 0.0001$ ). Such tests should, of course, be treated with some caution because there is no guarantee that the variance over different simulation runs corresponds to the variance over different participants in an experi-

ment. Overall, the simulations produced results quite similar to the results of the experiment, demonstrating that this simple dual-route architecture is, in principle suitable for explaining the Murphy and Zajonc results.

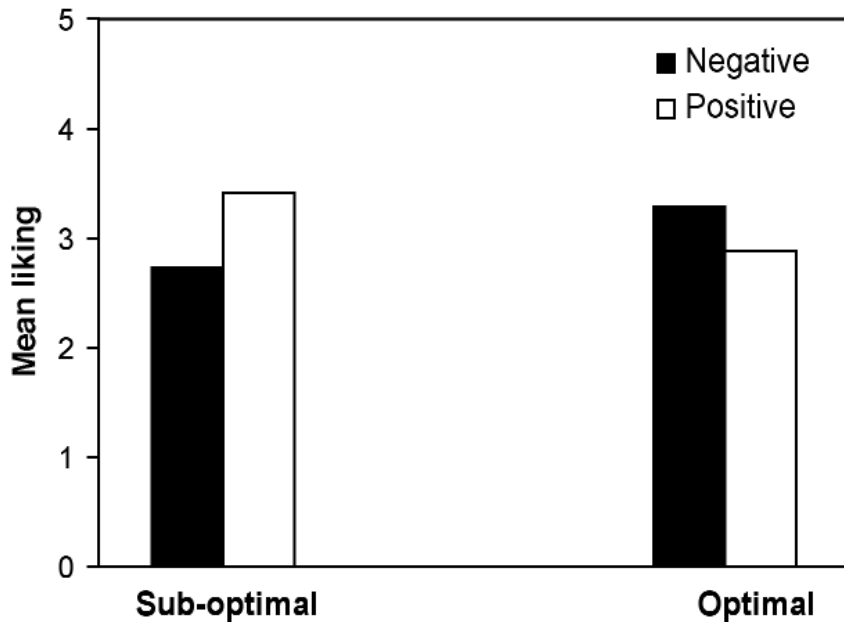


Fig. 3. Ratings of the Chinese ideographs in the four priming conditions of Simulation 1, obtained from the network after a linear transformation of the activation difference.

The functioning of the model can be understood quite easily. Processing in the direct pathway invariably leads to affectively congruent activations in the Affect module. Only with optimal presentation and with an indirect instruction are these activations overruled by the incongruent activations from the Indirect-Face module. The competition is never completely solved in the Affect module, so that the combined activations of the face and the instruction get the opportunity to reverse activations in this module when the competition in the Indirect-Face module has been solved. In the model, the indirect pathway is only fully involved in affective processing when there is sufficient time and a sufficiently 'sharp' contrast in activation to solve the competition. With suboptimal presentation, thus, only the direct route is taken, whereas with optimal presentation processing in the indirect route overrides direct processing.

It is important to design not only models that capture unexpected aspects of behavior, but to also address the more common aspects of expected behavior. In the following simulation the network received the (direct) instruction to evaluate the prime and to ignore the target. Under these instructions Geutskens en Rotteveel (1996), for instance, obtained larger congruent evaluations in optimal than

in suboptimal conditions. The finding of congruent direct evaluations of the sub-optimally presented faces means that even when the primes may not have been perceived consciously they may be available for evaluations when attention is directed to them. From the observation that ‘perception without attention’ has parallel effects to ‘perception without awareness’ (i.e., through masking; Merikle and Joordens, 1997), it is not a big step to assume that directing attention towards a stimulus may reduce the efficacy of masking (see also Enns and Di Lollo, 2000, for a view which entails a similar interaction between masking and attention). It is important to note that in the present model attention affects processing only in the indirect route (by enhancing the activations of a selection of nodes) but not in the direct route, much in the same way as processing in the direct route was not affected by masking.

## **Simulation 2**

This simulation was completely similar to the first, with the exception that now the name-face node (direct instruction) of the Instruction Module was clamped at 1.0, instead of the name-character node (indirect instruction). It was expected that the congruent effect in the suboptimal conditions would be enhanced due to the congruent contribution from the indirect route. This contribution is expected to increase even further in optimal conditions, so that a larger congruent effect is expected in these conditions.

## **Results**

The affective evaluations, after applying the same regression as in Simulation 1, are shown in Figure 4. As expected, the judgments were congruent with the valence of the faces in both optimal and suboptimal conditions. The congruent priming effect was, moreover, larger in optimal than in suboptimal presentation conditions and also much larger than with indirect instructions (c.f., Pessoa, McKenna, Gutierrez, and Ungerleider, 2002). Direct and indirect pathways appear to work in the same direction with this instruction.

The instruction only affects processing in the indirect pathway. Without an indirect pathway the congruent priming effect with suboptimal presentation would have the same size for direct and indirect instructions. This simulation, thus, also shows that, in the model at least, all indirect processing is not exhaustively excluded by suboptimal presentation, which is a longstanding issue in consciousness research (see Merikle, 1992). So, optimal and suboptimal presentation cannot be exclusively identified with either conscious or nonconscious processing but probably represent different mixtures of them. This exhaustiveness criterion (Merikle, 1992) is, however, not necessary to usefully investigate consciousness in masking studies. Experimental research has revealed some qualitative differences in processing between optimal and suboptimal processing indicating that manipulations of level of consciousness are not simply a matter of scale. If it were, for

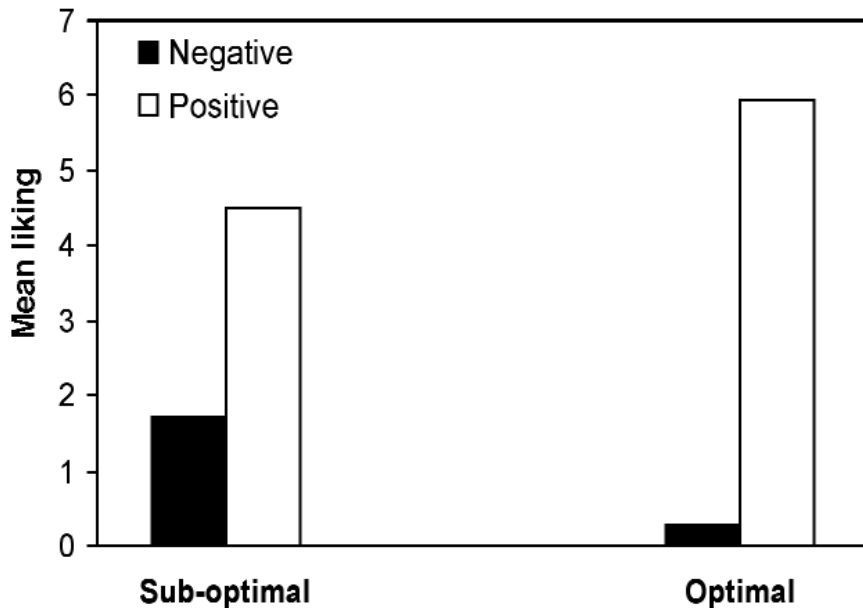


Fig. 4. Direct evaluations of the face stimuli in Simulation 2 after optimal and suboptimal presentation.

instance, assumed that conscious processes were just nonconscious processes rising above some threshold, this kind of dissociation would not be possible.

This simulation has also shown that the dissociation may be rather elusive, because in the right context, when the valence of the stimulus is relevant to the task, both direct and indirect processing cooperate to produce congruent effects. A further implication of these simulations is that, if the participants, or only some of them, accidentally follow a direct instruction instead of an indirect instruction, this could reduce the incongruent effect or even cause a congruent effect. Contamination by direct evaluations, thus, reduces the chance of finding a larger-suboptimal-than-optimal pattern of results.

The model allows for the investigation of further factors that may dilute the larger suboptimal than optimal pattern. It has been argued before (Jacobs and Nadel, 1985; LeDoux, 1996) that stress disrupts the working of the indirect pathway. Similarly, the indirect correction of the evaluations may be attenuated or even eliminated by stress. This would, for instance, be expected when stress leads to the direction of attention towards affective stimuli. The famous Easterbrook (1959) hypothesis, for instance, assumes that attention is narrowed only to the most relevant features of the environment under stressful conditions. If the affective prime is relevant, congruent priming would be enhanced under this hypothesis due to experimental stress. If attentional selection is operationalised in a network model as the resolution of competition between alternative representations (cf. Phaf et al., 1990), it seems straightforward to associate the narrowing of attention with an increase of lateral inhibition. Though the role of stress in the affective

priming task still has to be investigated experimentally, it may constitute another factor that reduces the chances of finding the Murphy and Zajonc results.

### **Simulation 3**

An affective priming experiment with some kind of stress-arousal in the participants was simulated by the model. Arousal is assumed to lead to the release of neuromodulators, which cause a regionally higher level of lateral inhibition (see Izquierdo and Medina, 1997; Mintz, Gotow, Triller, and Korn, 1998, for data from neurochemistry, and see Keeler, Pichler, and Ross, 1989; Phaf, Christoffels, Waldorp, and Den Dulk, 1998; Rumelhart, 1997, for a connectionist perspective). In fact, raising lateral inhibition in an array of nodes also has the effect of enhancing contrast between competing stimulus features. This leads to a higher activation of the 'central' (winning) features and a lower activation of the peripheral (loosing) features, which may correspond to some form of the Easterbrook (1959) hypothesis. We only varied inhibition in the Affect module here, because the other modules already had a sufficiently high inhibition level to solve competition completely. Because direct processing always arrives first at the affect module, the variable inhibition will have little effect on direct processing. Only indirect processing which has to overcome the direct activations will be affected by this manipulation. The inhibition level of the Affect-Module was multiplied by either a factor 10, 100, or 1000. The indirect instructions of the first simulation were again applied in this simulation. The expectation was that affective priming in suboptimal conditions would remain the same as in Simulation 1, but that in optimal conditions the incongruent priming effect would reverse into a congruent effect due to the strengthening of within-module inhibition.

### **Results**

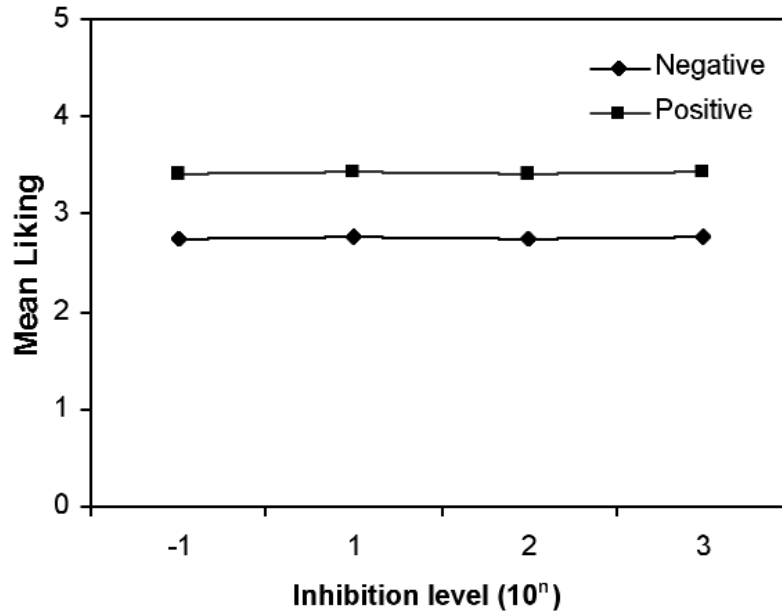
There was a congruent priming effect in both optimal and suboptimal condition for all new inhibition levels (see Figure 5). The direct effect in the suboptimal conditions was little affected by the increase in inhibition. The indirect effect that previously led to a reversal in the optimal conditions, was no longer able to override the direct effect. The (congruent) effect in the optimal priming conditions was now also determined primarily by direct processing. Though the claim that the action of the indirect pathway is disturbed by stress seems well supported experimentally (LeDoux, 1996), to our knowledge, this is the first model that simulates such a phenomenon. The account in terms of variable within-module inhibition is, of course, rather simple and readily leads to the expected behavior, but it still cannot be excluded that some other mechanism in the nervous system is actually responsible. These results could also give some further indication why the Murphy and Zajonc (1993) results are sometimes difficult to replicate. Only when all participants follow the indirect instructions and are relatively relaxed may the larger suboptimal than optimal priming effect of Murphy and Zajonc (1993) be found.

From the previous simulations it is clear that the (over-)correction effect through the indirect pathway is rather fragile, whereas the direct effect is very robust. This may reflect the way these two effects have come about. The latter type of processing may be largely innate and only slightly modified by learning processes. The ability to react affectively to evolutionary relevant stimuli (e.g., see Öhman, 1992), even immediately after birth, probably has great survival value. Of course, the direct route is not completely fixed. The conditioning experiments of LeDoux showed that it is capable of forming some associations. It is also known, however, that some stimuli can be more easily conditioned than others (Öhman, 1992). More extensive learning processes, however, seem to be involved in the indirect pathway. Context-dependent conditioning, for instance, where the predictive value of the stimulus depends on the specific context, also requires an indirect pathway in order to be learned. The hippocampus, which is one of the 'modules' forming part of the indirect pathway is the primary structure known to be involved in this form of conditioning (LeDoux, 1996).

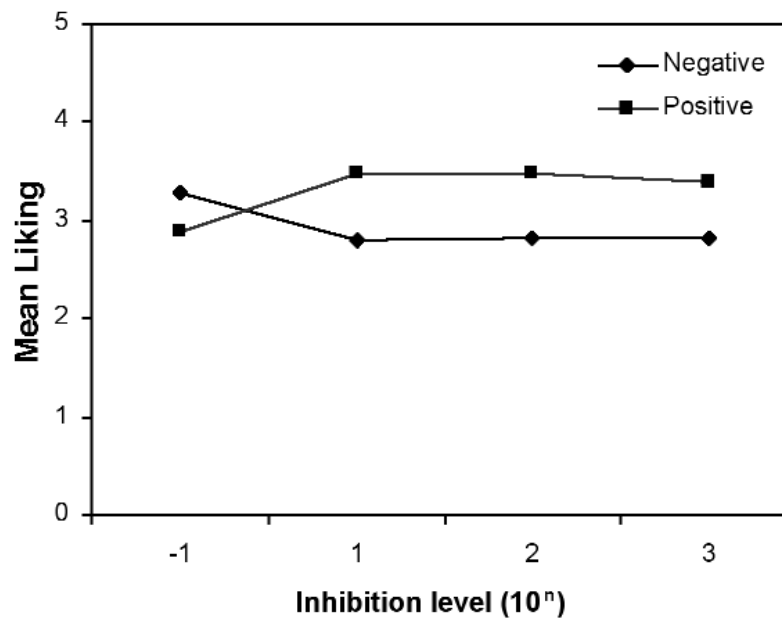
Darwin (1872/1965) already conjectured that there is a genetic basis, not only for the production, but also for the recognition of facial expressions. Darwin's observations of his baby daughter were, for instance, supported in a modern study by Serrano, Iglesias, and Loeches (1992). In a habituation-recovery procedure 4-6 months old babies already appeared to be able to discriminate and recognize expressions of fear, anger, and surprise. They, moreover, spent more time looking at expressions of anger and surprise than of fear. Such a genetic basis, of course, does not exclude further specification and modulation of face recognition by later learning, but this learning probably takes place largely in the indirect pathway. Thus, it must be possible to learn in the indirect pathway that under some contexts angry faces have a meaning which differs from a 'natural', direct, interpretation. When these faces are presented on a computer screen or television in the laboratory, for instance, they do not have the same significance as when encountered in real life. During learning in the next simulation, therefore, two different types of context were specified in which the evaluation could be performed. In one context the emotional expressions were relevant and the affective reactions should be enhanced. In the other context these expressions were not relevant and their influence, which already has taken some effect through the direct pathway, should be counteracted. The simulation, therefore, also entails the prediction, yet to be tested experimentally in humans, that the indirect correction effect develops through experience with situations where the affective response has to be suppressed.

#### **Simulation 4**

In this simulation we attempted to demonstrate that the weights in the network necessary for correcting the direct congruent priming effect can be formed through a learning process. First, there was a learning phase, and subsequently the resulting network was tested in an affective priming experiment, identical to Simulation 1. Only the weights from the Indirect Module to the Affect Module were



(a)



(b)

Figure 5. a) Evaluation of the Chinese Ideographs with optimal priming in Simulation 3 as a function of inhibition in the Affect Module. The first data-points (with inhibition -0.1) are from Simulation 1. b) Evaluation of the Chinese Ideographs with suboptimal priming in Simulation 3 as a function of inhibition in the Affect Module. The first data points (with inhibition -0.1) are from Simulation 1.

adjusted by the learning algorithm, but all other weights remained fixed. The learning rule (see Formula 3) from the CALM procedure, which is a competitive learning procedure (Murre et al., 1992), was applied in this simulation.

$$\Delta w_{ij}(t+1) = \mu a_i [(K - w_{ij}(t)) a_j - L w_{ij}(t) \sum_{f \neq j} w_{if}(t) a_f] \quad (3)$$

$\Delta w_{ij}(t+1)$  represents the weight change of the weight from node  $j$  to  $i$ .

The size of the weight change per iteration is controlled by the learning parameter  $B$  which was set to 0.005. The first term determines the increase of the weight, whereas the second term leads to a decrease. The term  $\sum_{f \neq j} w_{if}(t)$  is the summed activation coming from all other learning weights to node  $i$ . This term serves to normalize the weights (i.e., keep the sum of the weights to a node more or less constant).  $K$  and  $L$  are constants (both set to 1.0) determining the balance between increase and decrease. The learning connections were all initialized to 0.0, so that before learning starts the network responds congruently to faces in both conditions.

In the learning phase four different input patterns were presented. The two types of input were combined with the two types of instruction input. The instruction input should in this experiment be interpreted as a context, indicating whether the face inputs are relevant or not and should be corrected or not. Supervised learning was achieved by clamping one of the affect nodes in the Affect-Module. In the relevant instruction condition the congruent affect node is clamped, and in the irrelevant instruction condition the incongruent affect node is clamped. Such clamping could be achieved in the real world by having an angry face, for instance followed by a rewarding US indicating that in this context the initial reaction tendencies should be corrected. In the irrelevant conditions the associative learning mechanism will, thus, magnify the weights from the incongruent nodes to the opposite affect nodes. Each of the four input patterns was presented a thousand times for 4 iterations in a random order (with a clamping activation of 0.9). Between presentations of the patterns all activations (but not the weights) were reset to zero.

## Results

The weights (rounded to the second decimal) between the Indirect and Affect-Module developed after learning are shown in Figure 6. The weights reached the same configuration as in the hand-designed network. Weights that were not present in the hand-designed network also stayed close to zero after learning in this simulation. The most crucial weights, responsible for the correction effect, approximated the asymptotic value of 0.25. The supervised learning procedure appears the most simple way to achieve the correction effect in the indirect pathway. Extensions to the model would, however, enable the learning of this correction without applying clamped activations 'deep' in the network. The network would for this purpose have to be confronted with situations in which emotional faces are followed by stimuli with the opposite valence. Though activations persist for some

time due to the limited decay in this model, the (indirect) associations formed between the faces and the opposite valence would only be small and would probably be smaller than the associations with the same-valence node (which has a high activation due to the direct connections). To form strong associations between the faces and consequences of the opposite valence, there probably should be some ability to maintain and strengthen the activations of the faces for a longer time than they are actually presented (i.e., in working memory; Phaf and Wolters, 1997). We expect that the presence of a working memory ability in the network model would enhance the modulatory powers of the indirect pathway. If the absence of a negative stimulus can, for instance, be contrasted in working memory with a previous situation in which it was present, this can give rise to positive affect (i.e., 'relief'). Only when such abilities come into play may the affective processes studied here develop into full-blown emotions (e.g., see LeDoux, 1994).

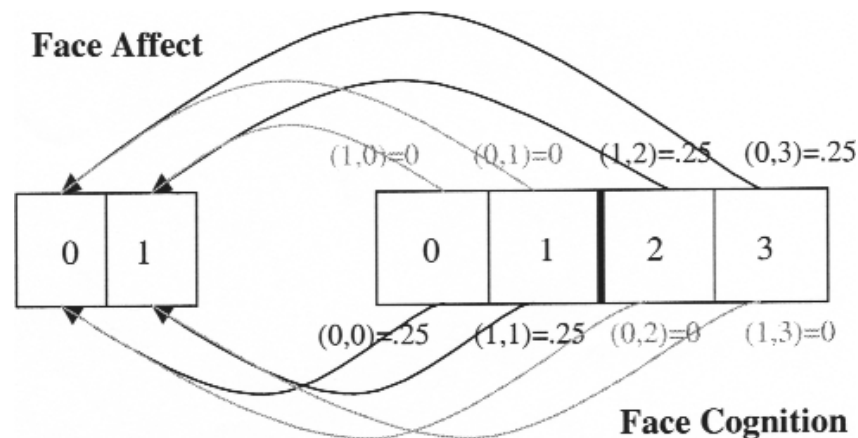


Fig. 6. The individual weights between the Indirect-Face Module and the Affect Module after learning in Simulation 4.

The results of the affective priming simulation after learning are shown in Figure 7. The pattern is very similar to Simulation 1. Though little is known about affective priming in children, it suggests that congruent priming can be found in all ages, but that a context-dependent incongruent priming effect may only develop at a later age. The preferential development of inhibition and correction of affect may well be a consequence of the propensity to 'false alarms' due to the biological preparation for direct affective processing. The correction mechanism may, thus, not itself be innate but may be acquired to supplement direct affective processing. Again this suggests that the Murphy and Zajonc results may not replicate equally well in all individuals. Considerable individual differences in the indirect regulation of emotion would be expected.

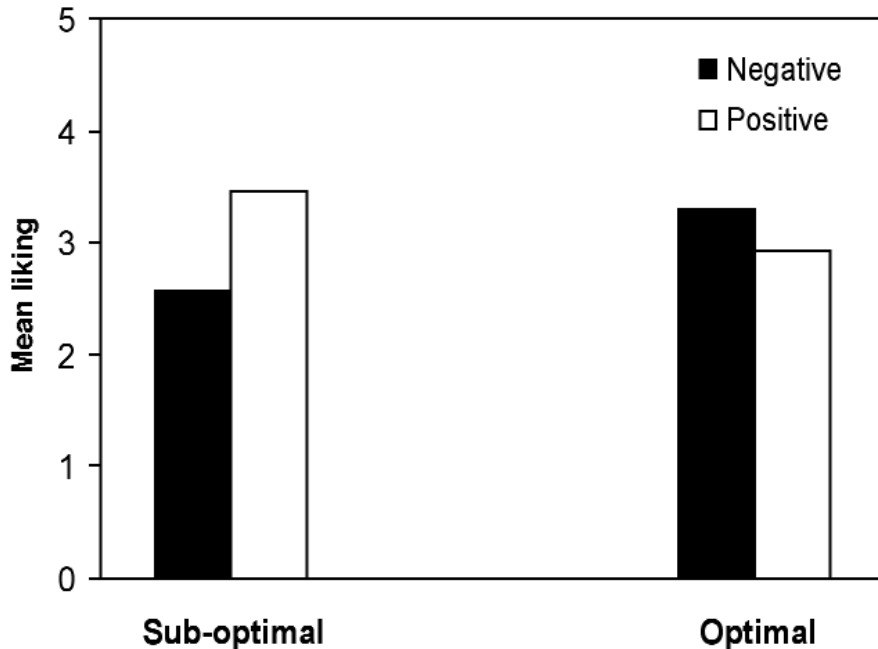


Fig. 7. Affective evaluations of the network after learning in the indirect pathway in Simulation 4.

## Discussion

Simulation 1 showed that the basic Murphy and Zajonc (1993) effect could be obtained within a dual-route architecture. In Simulations 2 and 3 this effect was disturbed by more direct instructions and by high arousal, respectively. Simulation 4 demonstrated that the weight configuration of the indirect pathway in the network could also be formed as a result of learning. These simulations supported the idea that the underlying mechanisms of the Murphy and Zajonc (1993) effect are to be found in a model with a LeDoux' dual-route architecture. Though such simulations in no way prove that affective priming is performed by the dual-route architecture, they make the assumption that fear conditioning and affective priming take place in different parts of the system (Clore and Ortony, 2000) rather implausible. The model also proposes an explanation for why the congruent priming effect is more robust than the incongruent priming effect. Essentially, this boils down to the fact that congruent priming is largely an innate phenomenon, whereas incongruent evaluations require elaborate learning and processing which is much more easily disturbed by a number of different factors. It should also be noted that the correction effect only forms in the presence of a biologically prepared direct reaction. This striking difference between affective and non-affective processing (i.e., in the conscious-nonconscious contrast; Murphy and Zajonc, 1993), which seems to have been ignored by Clore and Ortony (2000), can, thus, easily be accommodated in the present framework by assuming that the biologically prepared dual-route architecture is available only for affective processing.

An alternative explanation of the incongruent effect in the optimal condition could be that the participants suspected that the purpose of the experiment was to bias their judgements of the ideographs by the emotional prime, and they explicitly tried to counteract any influence of the prime. Such a more rule-based strategy is different from the learned correction effect in our model because it does not necessarily assume biologically prepared action tendencies. There are, however, findings contradicting this explanation. Rotteveel et al. (2001), for instance, found that the larger-suboptimal-than-optimal pattern can also be obtained with more central measurements than a Likert scale, such as facial EMG (i.e., of the *musculus zygomaticus major* and *musculus corrugator supercillii*). If higher cognitive strategies played a role, we would not expect to also see them play a role in these facial muscles.

An even more central measure is provided by the recent neuro-imaging methods that have already been used frequently to study the neural processing of emotional faces. The neuro-imaging (i.e., PET) study that came closest to obtaining the larger-suboptimal-than-optimal pattern was performed by Morris, Öhman, and Dolan (1998, 1999). The difference in activity between angry faces that had either been previously conditioned or not was larger in suboptimal than in optimal presentation conditions for the right amygdala, whereas the reverse was true for the left amygdala. Covariation techniques showed that, only with masked (suboptimal) presentation, right amygdala activity correlated positively with thalamic activity (i.e., the direct route), but correlated negatively with orbitofrontal activity (i.e., the indirect route). Also with masked presentation of fearful and happy faces that were not conditioned beforehand, the right amygdala was clearly implicated in a fMRI study (Whalen, Rauch, Etcoff, McInerney, Lee, and Jenike, 1998). This evidence not only strongly supports, at least for the right amygdala, the LeDoux model, but also supports the idea that the primary function of the indirect pathway is to inhibit affective reactions.

The present model may be one of the very few computational models to actually simulate the fate of suboptimally presented stimuli. These stimuli are not processed completely in the indirect route because the competition cannot be solved before the activations have decayed due to the short presentation time. Small activations suffering from the inhibition by competing representations, can, however, still exert some influence over subsequent processing (e.g., priming). Processing in the direct pathway is little hampered by suboptimal presentation and this processing is primarily responsible for the congruent affective priming. Another somewhat similar account of suboptimal presentation is the object substitution theory by Enns and Di Lollo (2000). In this theory the function of lateral inhibition is taken over by re-entrant (or recurrent from a higher level to a lower level) connections that allow for a comparison of activity at two different levels. Presentation is suboptimal in this view when there is a mismatch between the re-entrant signal and the lower-level activity. A mismatch can, for instance, occur when either the lower-level activity has decayed due to the short duration of presentation or has been replaced by a mask.

It is likely that the two types of accounts and also the two types of connections

(i.e., lateral inhibition and recurrent excitation; see also Phaf et al., 1990) should both be used to explain suboptimal processing. Both can serve to implement the type of route selection as a function of presentation condition that has been simulated here. The inclusion in the present model of recurrent connections only in the indirect pathway, would probably have enabled more accurate simulations at the expense, however, of making the model more complex. In a broader sense, however, the dual-route architecture itself can be envisaged as a (dual-route) implementation of object substitution. A strong congruent reaction is only observed when direct and indirect processing match. A weaker congruent reaction occurs when there is no re-entrant (i.e., indirect) signal, and, finally, an incongruent reaction (i.e., only the 'mask' is seen) results in non-matching conditions.

Though a model like the present one can only model nonconscious processes, important aspects of the contrast between conscious and nonconscious affective processes are captured by the model. Nonconscious processes in the direct route are largely fixed due to the biological preparation and do not seem to allow for much modulation by other nonconscious processes. Nonconscious processes also occur in the longer indirect route before competition is solved (i.e., before relaxation, see Kihlstrom, 1987) and the winning activations can reach working memory (Phaf and Wolters, 1997). The latter nonconscious processes differ from the first type by their ability to support all possible parsings (e.g., both congruent and incongruent responses) of the stimuli. Conscious processes are characterized by a definite choice from all these alternatives (e.g., only congruent-negative or incongruent-positive). Modelling of the kind reported here, therefore, supports experimental efforts to contrast conscious and nonconscious conditions (Merikle, 1992) by providing a more detailed analysis of the component processes, so that the properties of conscious processes can eventually be combined in a bottom-up fashion. Two types of dissociations between conscious and nonconscious processes are distinguished here. Dissociations within the indirect pathway can occur both for affective and non-affective processing. Dissociations between direct and indirect pathways may be specific for affective processing. In the question what distinguishes emotions from other mental phenomena, this study takes the position that only the core components of emotions, such as autonomic and bodily states and states of action readiness, can be (pre-)activated through direct nonconscious processes in a subcortical pathway.

#### *Acknowledgements*

The work was supported by a grant (number 575-23-006) from the Dutch Organisation for Scientific Research (NWO).

## References

- Armony, J.L., Servan-Schreiber, D., Cohen, J.D., and LeDoux, J.E. (1995). An anatomically constrained neural network model of fear conditioning. *Behavioral Neuroscience*, 2, 246-257.
- Bornstein, R.F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968-1987. *Psychological Bulletin*, 106, 265-289.
- Buck, R.W. (2000). The epistemology of reason and affect. In J.C. Borod (Ed.), *The neuropsychology of emotion*, pp. 31-55. Oxford, UK: Oxford University Press.
- Clore, G.L., and Ortony, A. (2000). Cognition in emotion: Always, sometimes, or never? In: R.D. Lane and L. Nadel (Eds), *Cognitive neuroscience of emotion*, pp. 24-61. Oxford, UK: Oxford University Press.
- Darwin, C. (1872/1965). *The expressions of the emotions in man and animals*. Chicago: University of Chicago Press.
- Den Dulk, P., Rokers, B., and Phaf, R.H. (1999). Connectionist simulations with a dual route model of fear conditioning. In B. Kokinov, *Perspectives on Cognitive Science*, vol. 4, pp. 102-112. Sofia, BU: NBU.
- Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 66, 123-201.
- Enns, J.T., and Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, 4, 345-352.
- Frijda, N.H., Kuipers, P., and ter Schure, E. (1987). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, 57, 212-228.
- Geutskens, A., and Rotteveel, M. (1996). *Affective priming with direct and indirect measures*. Unpublished Master's Thesis, University of Amsterdam, Amsterdam.
- Izquierdo, I., and Medina, J. H. (1997). The biochemistry of memory formation and its regulation by hormones and neuromodulators. *Psychobiology*, 25, 1-9.
- Jacobs, W.J., and Nadel, L. (1985). Stress-induced recovery of fear and phobias. *Psychological Review*, 92, 512-531.
- Keeler, J. D., Pichler, E. E., and Ross, J. (1989). Noise in neural networks: Thresholds, hysteresis and neuromodulation of signal-to-noise. *Proceedings of the National Academy of Sciences*, 86, 1712-1716.
- Kihlstrom, J.F. (1987). The cognitive unconscious. *Science*, 237, 1445-1452.
- LeDoux, J.E. (1986). Sensory systems and emotion: A model of affective processing. *Integrative Psychiatry*, 4, 237-248.
- LeDoux, J.E. (1994). Emotional processing, but not emotions, can occur unconsciously. In: P. Ekman and R.J. Davidson (Eds), *The nature of emotion*, pp. 291-292. Oxford, UK: Oxford University Press.
- LeDoux, J. (1996). *The emotional brain*. New York: Simon and Schuster.
- Leonard, C.M., Rolls, E.T., Wilson, F.A.W., and Baylis, G.C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behavioural Brain Research*, 15, 159-176.
- Mandler, G. (1985). *Cognitive Psychology: An essay in cognitive science*. Hillsdale NJ: Lawrence Erlbaum.
- Merikle, P.M. (1992). Perception without awareness. *American Psychologist*, 47, 792-795.
- Merikle, P.M., and Joordens, S. (1997). Parallels between perception without attention and perception without awareness. *Consciousness and Cognition*, 6, 219-236.
- Mintz, I., Gotow, R., Triller, A., and Korn, H. (1998). Effect of serotonergic afferents on quantal release at central inhibitory synapses. *Science*, 245, 190-192.
- Morris, J.S., Öhman, A. and Dolan, R. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, 393, 467-470.
- Morris, J.S., Öhman, A. and Dolan, R. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences, USA*, 96, 1680-1685.
- Murphy, S. T., and Zajonc, R. B. (1993). Affect, cognition and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, 64, 723-739.
- Murre, J. M. J., Phaf, R., H., and Wolters, G. (1992). CALM: Categorizing and learning module. *Neural Networks*, 5, 55-82.
- Öhman, A. (1992). The psychophysiology of emotion: Evolutionary and nonconscious origins. In: G. d'Ydewalle, P. Bertelson and P. Eelen (Eds.), *Current advances in psychological science: An international perspective*, pp. 197-227. Hove, UK: Erlbaum.

- Pessoa, L., McKenna, M., Gutierrez, E., and Ungerlicher, L.G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Science, USA*, 99, 11458-11463.
- Phaf, R.H., Christoffels, I.K., Waldorp, L.J., and den Dulk, P. (1998). Connectionist investigations of individual differences in Stroop performance. *Perceptual and Motor Skills*, 87, 899-914.
- Phaf, R.H., van der Heijden, A.H.C., and Hudson, P.T.W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology*, 22, 273-341.
- Phaf, R.H., and Wolters, G. (1997). A constructivist and a connectionist view on conscious and non-conscious processes. *Philosophical Psychology*, 10, 287-307.
- Rotteveel, M., de Groot, P., Geurtskens, and Phaf, R. Hans. (2001). Stronger suboptimal than optimal affective priming?. *Emotion*, 1, 348-364.
- Rumelhart, D. E. (1997). Affect and neuromodulation: A connectionist approach. In J. D. Cohen and J. W. Schooler (Eds.), *Scientific approaches to consciousness*, pp. 469-477. Mahwah: Lawrence Erlbaum.
- Serrano, J. M., Iglesias, J., and Loeches, A. (1992). Visual discrimination and recognition of facial expressions of anger, fear and surprise in 4- to 6- month old infants. *Developmental Psychobiology*, 25, 411-425.
- Shaver, P.R., Schwartz, J., Kirson, D., and O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52, 1061-1086.
- Whalen, P.J., Rauch, S.L., Etcoff, N.L., McInerney, S.C., Lee, M.B., and Jenike, M.A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *The Journal of Neuroscience*, 18, 411-418.

*Received:* May, 2001

*Accepted:* August, 2002