ON THE CONTROL OF AUTOMATIC PROCESSES: A PARALLEL DISTRIBUTED PROCESSING MODEL OF THE STROOP EFFECT

Technical Report AIP - 40

J. D. Cohen, K. Dunbar & J. L. McClelland

Carnegie Mellon University,
Stanford University &
McGill University

June 1988

This research was supported by the Computer Sciences Division, Office of Naval Research and DARPA under Contract Number N00014-86-K-0678. Reproduction in whole or in part is permitted for purposes of the United States Government. Approved for public release; distribution unlimited.

06 CV. 1

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; Distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AIP - 40	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Carnegie-Mellon University 6b. OFFICE SYMBOL (*If applicable*)	7a. NAME OF MONITORING ORGANIZATION Computer Sciences Division Office of Naval Research
6c. ADDRESS (*City, State, and ZIP Code*) Department of Psychology Pittsburgh, Pennsylvania 15213	7b. ADDRESS (*City, State, and ZIP Code*) 800 N. Quincy Street Arlington, Virginia 22217-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Same as Monitoring Organization 8b. OFFICE SYMBOL (*If applicable*)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0678
8c. ADDRESS (*City, State, and ZIP Code*)	10 SOURCE OF FUNDING NUMBERS p4000ub201/7-4-86 PROGRAM ELEMENT NO: N/A; PROJECT NO.: N/A; TASK NO.: N/A; WORK UNIT ACCESSION NO: N/A

11. TITLE (*Include Security Classification*)
On the control of automatic processes: A parallel distributed processing model of the stroop effect

12. PERSONAL AUTHOR(S)
Jonathan D. Cohen, Kevin Dunbar, and James L. McClelland

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (*Year, Month, Day*)	15. PAGE COUNT
Technical	FROM 86Sept15 TO 91Sept14	1988 June 16	107

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)
Automatic processes, Skill acquisition, Connectionist models, Attention

19. ABSTRACT (*Continue on reverse if necessary and identify by block number*)

See reverse side

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. Alan L. Meyrowitz	22b TELEPHONE (*Include Area Code*) (202) 696-4302 22c. OFFICE SYMBOL N00014

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Abstract

A growing body of evidence suggests that traditional views of automaticity are in need of revision. For example, automaticity has often been treated as an all-or-none phenomenon, and traditional theories have held that automatic processes are independent of attention. Yet recent empirical data suggest that automatic processes are continuous, and furthermore are subject to attentional control. In this paper we present a model of attention which addresses these issues. Using a parallel distributed processing framework we propose that the attributes of automaticity depend upon the strength of a process and that strength increases with training. Using the Stroop effect as an example, we show how automatic processes are continuous and emerge gradually with practice. Specifically, we present a computational model of the Stroop task which simulates the time course of processing as well as the effects of learning. This is done by combining the cascade mechanism described by McClelland (1979) with the backpropagation learning algorithm (Rumelhart, Hinton, & Williams, 1986). The model is able to simulate performance in the standard Stroop task, as well as aspects of performance in variants of this task which manipulate SOA, response set, and the number of competing words in the display. These simulations demonstrate that when two processes are in competition, the weaker process can take on several of the attributes that have previously been associated with controlled processing: susceptibility to interference and a requirement for the allocation of attention. This suggests that the traditional distinction between controlled and automatic processing is in need of reconsideration.

Introduction

At least since Malebranche (1674), it has been observed that some processes appear to be under voluntary control, whereas others do not. Indeed, a central concept in the literature on human information processing has been the distinction between automatic and controlled processes (Cattell, 1886; Posner & Snyder, 1975; Shiffrin & Schneider, 1977). Automatic processes are involuntary and fast, and do not require attention for their execution. In contrast, controlled processes are voluntary and relatively slow, and do require attention. Examples of processes thought to be automatic are word reading (LaBerge, 1980; Posner & Snyder, 1975), activation of semantic associations (Neely, 1977), and target detection after extensive training with a consistent set of targets (Shiffrin & Schneider, 1977). Examples of processes that are assumed to be controlled are the use of arbitrary, novel associations (e.g., Neely 1977), the reading of inverted text (Kolers, 1976), and target detection in situations where the set of targets varies from trial to trail (Schneider & Shiffrin, 1977; Shiffrin & Schneider,1977).

The distinction between automatic and controlled processes is often discussed as though automaticity were all or none. However, it seems fairly clear that automaticity is really a matter of degree. Indeed, we suspect that many discussions that treat automaticity as a dichotomous variable do so only for lack of a theoretical framework in which to formulate a more continuous account. Our goal in this paper is to help fill this gap in cognitive theory. In particular, we hope to provide a more detailed description of the mechanisms that underlie automatic processes, their gradual development with practice, and their interaction with attentional control.

We will begin by illustrating the contrast between automatic and controlled processes in the context of the well-known Stroop interference task. We then present arguments suggesting that the processes involved in this task differ in degree of automaticity, and that automaticity varies continuously with processing experience.[1] In order to make this point of view more explicit, we will describe a theoretical framework for capturing the continuous nature of automaticity and its dependence on processing experience. The main body of the paper will then describe a simulation model that applies this theoretical framework to performance in this task, allowing for an account of several different aspects of performance in the Stroop task and related paradigms.

The Stroop task. The Stroop task actually consists of two subtasks (see Table 1). In one of the two subtasks, subjects have to name the color of the ink in which a word is printed. In the other subtask, subjects have to read the word aloud while ignoring the color of the ink. Usually three types of stimuli are used: conflict stimuli in which the word and the ink color are different (e.g., the word RED in green ink)[2] congruent stimuli in which the word and the ink color are the same (e.g., the word RED in red ink); and control stimuli. The control stimuli for the word reading task are typically color words printed in black ink; for the ink naming task the control stimuli are usually a row of XXXX's printed in a color.

---

[1] Some authors have argued that some automatic processes are innate. For example, Hasher and Zacks (1979) maintain that the encoding of event frequency is an automatic process and that it is innate. Here we will focus on processes that become automatic after extensive practice at a task.

[2] Throughout this paper, references to word stimuli will appear in upper case (e.g., RED), references to color stimuli will appear in lower case (red), and references to potential responses will appear in quotation marks ("red").

Table 1. The Stroop Task.

Condition	*Example Stimulus (where "red" is the response)*
Color Naming:	
Control	XXXX in red ink
Conflict	GREEN in red ink
Congruent	RED in red ink
Word Reading:	
Control	RED in black ink
Conflict	RED in green ink
Congruent	RED in red ink

Detailed reviews of the literature on the Stroop task are available in Dyer (1973) and Glaser and Glaser (1982). As we progress, we will examine many of these details. For present purposes, though, we focus on three basic aspects of the findings: 1) Word reading is faster than color naming. 2) In word reading, ink color has no effect on the time it takes to read the word. 3) In color naming, when the word conflicts with ink color, performance is slower than in the control or congruent conditions. For example, subjects are slower to respond to the color red when the word GREEN is written in red ink, than when the word

RED or a series of X's appear in red ink. Thus, there is dramatic asymmetry between the word reading and color naming tasks in Stroop experiments.

It is natural to account for these findings by assuming three things:

- Word reading is automatic.
- Color naming is controlled.
- If the outputs of two processes conflict, performance based on one of the two will be slowed down.

From these assumptions, together with the basic definitions of automatic and controlled processes, we get an account of the principal results. The finding that word reading is faster than color naming follows from the relatively greater speed of automatic processes. The finding that ink color has no effect on word processing follows from the assumption that color naming is controlled and therefore voluntary; so it will not occur when the task is to read the word and ignore the color. The finding that a conflicting word interferes with color naming follows from the automaticity (i.e., involuntary nature) of word reading, and the assumption that conflicting outputs slow responding.

Traditionally, this interpretation of the Stroop task has suggested a general method for assessing the automaticity of two arbitrary processes A and C. 1) Compare their relative speed. 2) Pit them against each other. If all of the following are true:

- A is faster than C.
- A interferes with C.

- C does not interfere with A.

then A is automatic and C is controlled. Of course, this method requires that processes A and C are in some sense comparable in intrinsic difficulty and number of processing stages.

On the evidence considered so far, automaticity could be considered to be all or none. However, there is further evidence to suggest that automaticity is really a matter of degree: We will see that speed of performance, tendency to interfere with other processes, and susceptibly to interference from other processes all seem to vary continuously as a function of practice. We will argue that all of these attributes can be understood in terms of a common variable: the *strength* of a process.

Speed of performance. Many studies of automatic processes have focussed on speed of processing, and have noted that automatic processes tend to be executed faster than controlled processes. We see this in the Stroop task, where word reading is faster than color naming. However, speed of processing is obviously not a binary variable, and it is a very common observation that practice produces gradual, continuous increases in processing speed (e.g., Blackburn, 1936; Bryan & Harter, 1899; Kolers, 1976; Logan, 1979; Shiffrin & Schneider, 1977). A recent experiment by MacLeod and Dunbar (1988), which we will be examining throughout this paper, illustrates this point. They trained subjects to use color words as names for arbitrary shapes (see Figure 1), and practiced subjects in this "shape-naming" task for 576 trials a day for 20 days. Over the course of training, reaction times to name the shapes gradually improved with practice.

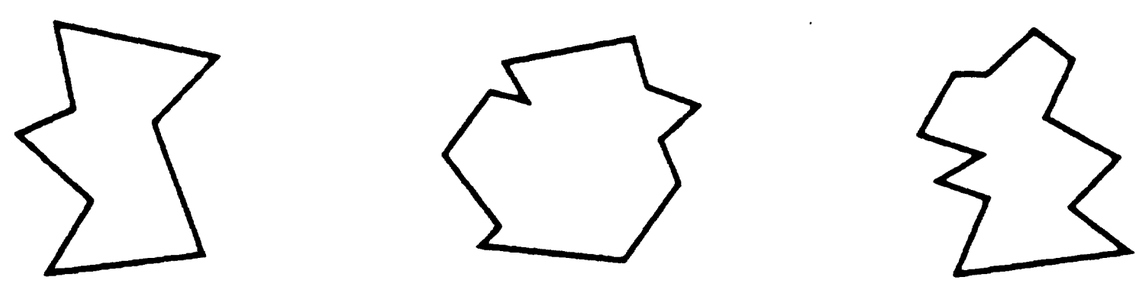

Figure 1. Training stimuli of the type used by MacLeod and Dunbar (1988) for the shape naming task. Each of four shapes was assigned an arbitrary color name, which the subjects had to learn.

The fact that speed of processing varies continuously with practice might not be enough to convince us that automaticity is a continuous phenomenon. For example, one might still suppose that there is a threshold for automaticity. On this view, increased practice might improve the speed of a "controlled" process up to some critical point; additional practice would then cause the process to become automatic, and continued practice would merely lead to increases in the speed of performance, but not to any further changes in the type of processing involved. Contrary to this view, however, there is reason to suppose that increasing speed reflects a continuous strengthening of the process underlying performance, and that it is the strength of a process that is responsible both for its speed of performance, and for the fact that it produces interference. Logan (1980) has presented such a view of processing in the Stroop task, and interference data from the experiment of MacLeod and Dunbar mentioned previously provides strong evidence in favor of this view. They were able to show that practice led not only to a gradual increase in the speed of processing of shape names, but also to a

gradual increase in the tendency of shape names to produce interference with another task. This is discussed below.

Tendency to interfere. As we mentioned above, it is commonly accepted that a central feature of automatic processes is that they can occur outside of voluntary control, and thereby produce interference with a competing task. Gradual development of the ability to produce interference would therefore strongly favor the view that automaticity is continuous in nature. MacLeod and Dunbar (1988) tested for the ability of shape names (which were arbitrary color words) to interfere with naming the color of the ink in which the shapes appeared. Subjects were tested after 1, 5, and 20 sessions of practice associating the shapes with their color word names. After 1 session of practice, there was no effect of a shape's name on the subject's ability to identify the color of the ink in which it appeared; after 5 sessions the shape produced some interference; and after 20 sessions there was a large effect. That is, presenting a shape whose name conflicted with its ink color produced strong interference with the color naming response. The fact that this effect appeared to develop incrementally with practice lends further support to the idea that automaticity is a graded, rather than an all-or-none phenomenon.

Susceptibility to interference from other tasks. We have suggested that there is a gradual strengthening of a process which occurs with practice, and that this leads to an increased tendency to interfere with competing processes. This is consistent with the data just discussed. It also suggests that as the tendency for a process to produce interference increases, its susceptibility to interference might decrease. MacLeod and Dunbar examined this point as well; they looked at the effects of conflicting, congruent, and neutral ink colors on shape naming after 1, 5 and 20 sessions of practice with shape

naming. After one session of practice, conflicting ink color interfered with naming the shape. After 20 sessions, this was no longer the case; if there was any residual effect of color, it was below the level of statistical detectability. Once again, these changes appeared to occur incrementally with practice.

MacLeod and Dunbar's results are incompatible with a dichotomous view of the distinction between automatic and controlled processes. If we assume that automaticity is all or none, we are lead to a contradiction. From the standard Stroop experiment we know that ink color does not influence word reading, while words can interfere with (and facilitate) color naming. Hence we must conclude that color naming is *not* automatic. Yet, in MacLeod and Dunbar's experiment color naming reversed roles. That is, during the first two testing periods ink color was able to interfere with shape naming, while shapes were unable to influence color naming. According to these data, we must conclude that color naming *is* automatic.

A way out of the contradiction is to suppose that interference effects are a reflection not of a dichotomy between automatic and controlled processes, but are a reflection of the relative strength of each of two processes which are placed in conflict. When process A is stronger than process B, A can interfere with B but B can not interfere with A. As B gains strength, there comes a point where the strengths are commensurate and the processes can interfere with one another; as B is strengthened further, to the point where it is stronger than A, B can now interfere with A but A can no longer interfere with B.

Strength versus Speed

The argument just given is consistent with the facts, but it might be supposed instead that the only thing that is really varying among the processes considered above is speed. On this view, we could assume that all the processes in question are automatic but that one process has to finish before another to produce interference.

The view that interference is a matter of relative timing of completion predicts that it should be possible to make color information conflict with word reading, if the color information is presented long enough before the onset of the word. However, this does not occur. This fact was demonstrated by Glaser and Glaser (1982). Glaser and Glaser varied the stimulus onset asynchrony (SOA) of a color patch and a color word[1], and found no interference of the color patch on word reading even when the color preceded the word by as much as 400 msec.

The pattern of results in the Glaser and Glaser study is interesting, and we will return to it below. For now we mention this study to show that it seems to require an appeal to something other than just the relative finishing time of two processes. If color naming and word reading differ simply in how long they take, then we should be able to compensate for this difference by adjusting the onset of the relevant stimulus inputs. But this does not work. It appears instead that color naming really is in some sense weaker, and not simply slower than word reading. Yet it is not so weak that it cannot interfere with another

---

[1] As we will discuss below, the Stroop effect can still be observed even when the two stimulus dimensions are physically disjoint.

task, such as shape naming when this has received only a moderate amount of practice (e.g., less than 1,000 trials).

Overall, then, the findings we have reviewed suggest the following two basic conclusions:

- As a process is practiced, some parameter of that process gradually changes. We will call this parameter *strength*.

- Increasing strength produces increased speed, increased tendency to interfere, and decreased susceptibility to interference.

Automaticity and Attention

Thus far we have omitted one central issue. This concerns the regulation of performance to conform with the assigned task. For example, when faced with the word RED in green ink, the correct response depends on the task. Traditionally, of course, it has been assumed that the role of attention is to determine which processes are to be performed. Indeed, implicit in the notion of controlled processing is the idea that there is a controller. The distinction between automatic and controlled processes implies that some processes are less prone to control than others.

A dichotomous view of automaticity would treat controllability as all or none—that is, automatic processes are *not* subject to control, while controlled processes are. However, there are obvious problems with this view. As Neumann (1984) has noted, behaviors relying on automatic processes must be subject to some form of control, otherwise they

would occur as reflexes. This is apparent in the Stroop task, where subjects are able to respond accurately to color stimuli (i.e., in the color naming task) even when conflicting word information is available. This would not be possible if word reading—even though it is automatic—were not subject to some form of control.

One way to reconcile automatic processing with the control of behavior is to differentiate between control of the behavior required to accomplish a task and control of the individual processes involved in that task. Thus, behavior in a task can be seen as consisting of many component processes. Some of these processes may be automatic, and not subject to control. Nevertheless, the control of behavior could occur when the results of such automatic process are made available to other, controlled processes involved in the task. Shiffrin suggests just such a view: "tasks as a whole are accomplished by complex mixtures of automatic and attentive processes operating in concert" (Shiffrin, 1988, p.23).

In contrast to this view is the view that even individual automatic processes are subject to control. Several theorists have argued along these lines (e.g., Kahneman & Treisman, 1984; Logan, 1980). Their arguments are supported by empirical evidence from the Stroop task. For example, Kahneman and Henik (1981) have shown that the allocation of attention can influence the degree to which word reading interferes with color naming. They demonstrated this in a variant of the standard Stroop task in which they presented two color word stimuli, one surrounded by a square and the other surrounded by a circle. Subjects attended to a point midway between the two stimuli, and were then asked to name the ink color of the word which appeared within one of these shapes. Interference occurred only from the word in the attended shape, and not from the word in the non-attended shape. According to traditional theories of automaticity the word in the non-

attended shape should have produced interference because, on this view, word reading is automatic and therefore independent of the allocation of attention.

On the basis of this and related findings, Kahneman and Treisman (1984) have argued that automatic processes are in fact subject to control by attention, although individual processes may differ in their degree of susceptibility to such control. This coincides nicely with the view that the attributes of automaticity are continuous in nature. On this view would be that control by attention is also an attribute of automaticity that varies by degree. Furthermore, it seems natural to assume that the stronger a process is, the less susceptible it is to control by attention. This can then explain the relationship between strength of processing and interference effects. Thus, earlier we noted that the stronger a process is, the greater will be its tendency to interfere with a competing processes. Now we can say that this is because stronger processes are less subject to control, and can occur more effectively in the absence of attention. From this it follows that increasing the strength of a process will increase its tendency to interfere with other processes, even when that process is not allocated attention.

In summary, we have argued that automaticity is a matter of degree, and reflects a gradual strengthening process that varies continuously with practice. We turn now to a theoretical framework in which it is possible to explore this view in greater detail, and to be specific about the relationships that exist between processing, practice and attentional control. The framework specifies a subclass of models within the more general parallel distributed processing framework described by Rumelhart, Hinton and McClelland (1986); it shares with many cognitive models the basic notion that processing occurs through the propagation of activation through pathways of graded strength.

The Processing Framework

Architectural characteristics. Processing takes place in a system of connected modules. Each module consists of an ensemble of elementary processing units. Each unit is a simple information processing device that accumulates inputs from other units and adjusts its output in response to these inputs. All units are assumed to operate continuously, adjusting their output in response to changes in their inputs.

Representation of information. Information is represented in the pattern of activation over the units in a module. The activation of each unit is a real valued number varying between a maximum and minimum value. Thus, information is represented in a graded fashion, and can accumulate and dissipate with time.

Processing. Processing occurs by the propagation of signals (spread of activation) from one module to another. This occurs via the connections that exist between the units in different modules. In general, there may be connections within as well as between modules, and connections may be bi-directional. However, for present purposes we adopt the simplification that there is a unidirectional flow of processing, starting at modules directly sensitive to sensory input and proceeding "forward" or "bottom-up" to modules whose output governs the execution of overt responses.

Pathways and their strengths. In general a particular process is assumed to occur via a sequence of connected modules that form what we will call a *pathway*. Performance of a task requires that there be some processing pathway such that the pattern of activation in the relevant sensory modules leads to an appropriate pattern of activation in the relevant output modules. The speed and accuracy with which a task is performed depends on the

speed and accuracy with which information flows along the appropriate processing pathway. This, in turn, depends on the connections between the units that make up the modules in that pathway. We will demonstrate this in simulations shortly. We refer to this parameter as the *strength* of a pathway. Thus, the speed and accuracy of performing a task depends on the strength of the pathway used in that task.

Interactions between processes. Inasmuch as individual modules can receive input from and send information to several other modules, each can participate in several different processing pathways. Interactions between processes arise in this system when two different pathways rely on a common module—that is, when pathways intersect. If both processes are active, and the patterns of activation that each would generate at their point of intersection are dissimilar, then *interference* will occur within that module, and processing will be impaired in one or both pathways. If the patterns of activation are very similar, this will lead to *facilitation*.

The intersection between two pathways can occur at any point in processing after the sensory stage. For example, interference at an intermediate stage is consistent with data reported by Shaffer (1975) and by Allport, Antonis and Reynolds (1972). Interference at the output stage would give rise to response competition, such as that observed in the Stroop task (cf. Dyer, 1973). The general view that interference effects arise whenever two processes rely on a common resource, or set of resources has been referred to as the multiple resources view (e.g., Hirst & Kalmar, 1987; Navon & Gopher, 1979; Wickens, 1984). Logan (1985) summarizes this position succinctly: "different tasks may depend on different resources, and dual-task interference occurs only when the tasks share common resources. Thus, the interference a particular task produces will not be an

invariant characteristic of that task; rather, it will depend on the nature of the tasks it is combined with" (p.376). This point will be made explicit in the simulations we present below.

Control. One way to avoid the interactions that occur at the intersection between two pathways is to modulate the information arriving along one of them. This is one of the primary functions of attention within this framework. This view of attention is consistent with the views of several other authors (Kahneman & Treisman, 1984; Logan, 1980; Treisman, 1960). In our system, modulation occurs by altering the responsiveness of the processing units in a pathway. In this way, attention can be used to control individual processes. However, this does not necessarily imply that attention requires a unique, or even distinct component of processing. As we shall see, attention simply represents an additional source of information which provides context for the processing of signals within a selected pathway.

This framework can be used to account for many of the empirical phenomena associated with learning and automaticity. Schneider (1985) has used a similar approach to explain how performance in a category search task changes as a function of practice. Here, we focus on the significance that this approach has for selective attention. In the following section we describe a simulation model based on the processing principles discussed above. Using the Stroop task as an exemplar, we present a series of six simulations that demonstrate that this model is able to account for many of the empirical phenomena associated with automaticity, and for their gradual emergence as a function of practice. The first three simulations examine the basis for the attributes of automaticity evidenced in the Stroop task (viz., speed of processing and interference effects). The remaining

simulations explore the relationship between processing and attention, and the nature of capacity limitations in processing resources.

The Model

The model that we propose embodies the assumptions given above in a specific set of modules and detailed processing assumptions suitable for capturing the Stroop phenomenon and related findings such as those of MacLeod and Dunbar (1988) and of Glaser and Glaser (1982).

Architecture and Representation of Information

The architecture of our model is depicted in Figure 2. The model consists of two processing pathways—one for processing color information, and the other for processing word information—both of which converge on a common response mechanism. Each pathway consists of a set of input units, a set of intermediate units, and a set of output units. Each of the input units in a given pathway projects to all of the intermediate units in that pathway. The intermediate units from both pathways project to all of the output units in the model.

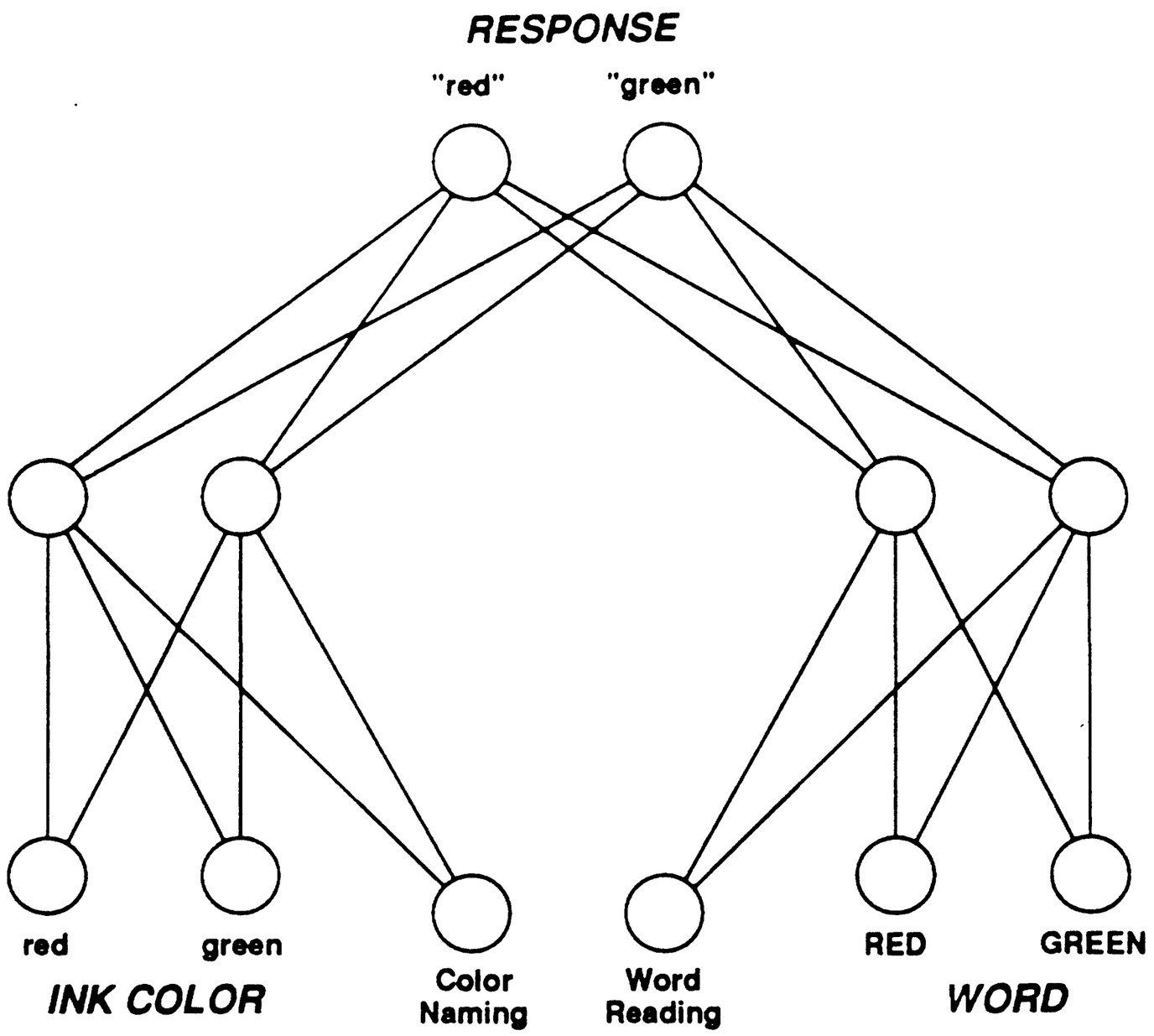

Figure 2. Network architecture. Units at the bottom are input units, and units at the top are the output (response) units.

Processing in this system is strictly feedforward. A stimulus is provided by activating units at the input level of the network. Activation then passes to the intermediate units and, gradually, to the output units. A response occurs when activation of one of the output units exceeds a response threshold. Reaction time is assumed to be linearly related to the number of processing cycles that it takes for this threshold to be exceeded (the response mechanism will be discussed in greater detail below). In addition to the units just described, there are also two task demand (or "attention") units—one for the color naming

task and the other for the word reading task. Each of these is connected to the intermediate units in the corresponding pathway. These are used to allocate attention to information in one of the two processing pathways. Activation of a task demand unit sensitizes processing in the corresponding pathway, as will be explained below.

Information is represented in the network in the following way. Individual stimuli and responses have discrete representations. Each color and each word is represented locally by the activation of a single input unit. Similarly, each output unit represents one potential response. We chose local representations of this kind to keep the model as simple and interpretable as possible. However, nothing in principle precludes the possibility that either inputs or outputs could be distributed over many units. In fact, preliminary investigations suggest that our findings using local representations generalize to more complex systems using distributed representations.

Mechanisms Underlying Learning and the Time Course of Processing

The model is intended to provide an explanation of the relationship between learning and the time course of the psychological processes involved in the Stroop task. To date, PDP models which have addressed the time course of psychological processes have largely been distinct from those which model learning and memory. For example, McClelland (1979) presented a multilevel PDP system which provided an account of the time course of psychological processes, however this system did not include a learning algorithm. The back propagation algorithm described by Rumelhart, Hinton and Williams (1986) was introduced as a general learning mechanism which can be used in multilevel networks. However, PDP systems which have employed this algorithm have not generally simulated

temporal phenomena such as reaction times. Here we describe each of these mechanisms and their limitations in greater detail. We then show how they can be brought together to provide a single system in which both learning and processing dynamics can be examined.

McClelland's cascade model provides a mechanism for simulating the time course of psychological processes. In this system, information is represented as the activation of units in a multilevel, feedforward network. Input is presented as a pattern of activation over units at the lowest level. Information gradually propagates upward, as units at each level update their activations based on the input they are receiving from lower levels. Eventually a pattern of activation develops over the units at the topmost level, where a response is generated. Units in this network update their activations based on a weighted sum of the input they receive from units at the previous level in the network. Specifically, the net input at time (t) for $unit_j$ (at $level_n$) is calculated as:

$$net_j(t) = \sum_i a_i(t) w_{ij} \qquad \text{(Equation 1)}$$

where $a_i(t)$ is the activation of each $unit_i$ (at $level_{n-1}$) from which $unit_j$ receives input, and w_{ij} is the weight of the connection from each $unit_i$ to $unit_j$. The activation of a unit is simply a running average of its net input over time:

$$a_j(t) = \overline{net}_j(t) = \tau\, net_j(t) + (1-\tau)\, \overline{net}_j(t-1) \qquad \text{(Equation 2)}$$

where $\overline{net}_j(t)$ is the time-average of the net input to $unit_j$, $net_j(t)$ is the net input to $unit_j$ at time (t), and τ is a rate constant. This time-averaging function is what establishes the time

course of processing in this model. When τ is small the unit's activation will change slowly; with a larger τ it will change more quickly. One feature of Equation 2 is that, if the net input to a unit remains fixed, the unit's activation will approach an asymptotic value which is equal to this net input. As a result, McClelland demonstrates that with a constant input to the first layer in such a network, all of the units will approach an asymptotic activation value. What is most important is that this value is determined strictly by the pattern of the input to the network and the connections that exist between the units. Thus, given a particular input pattern, and sufficient time to settle, the network will always reach a stable state in which each unit has achieved a characteristic activation value. The speed with which this occurs is determined in large measure by the strength of the connections in the network.

One problem with the type of network used in the cascade model is that it is based on a linear activation function. That is, the activation of a unit is simply a weighted sum of the inputs it receives. It has been shown that networks which rely on linear update rules such as this, even if they are composed of multiple layers, suffer from fundamental computational limitations (*cf.* Rumelhart, Hinton, & McClelland, 1986 for a discussion). Networks which are capable of solving the broad class of computational problems which humans seem able to solve require at least one layer of units between input and output that make use of a non-linear relation between input and output. Another problem with the cascade model, especially within the current context, is that it lacks any mechanism for learning. Both of these problems can be overcome if mechanisms are included which have been used in recent PDP models of learning.

The first step is to introduce non-linearity into processing. This has been done most commonly using the logistic function. Typically this function has been used to calculate the activation of a unit, based on its instantaneous net input:

$$a_j(t) = \text{logistic}(net_j(t)) = \frac{1}{1 + e^{-net_j(t)}} \qquad \text{(Equation 3)}$$

where the $net_j(t)$ is given by Equation 1. The logistic function introduces non-linearity by constraining the activation of units to be between the values of 0 and 1 (see Figure 3). The difficulty with this formulation for us is that it does not exhibit a gradual buildup of activation over time. The full response to a new input occurs in a single processing step at each level, and so the effects of a new input can be propagated through the network in a single sweep through all of its levels from input to output. However, we can reintroduce the dynamic properties of the cascade model if we assume—as the cascade model did—that the net input to a unit is averaged over time before the activation value is calculated (see Equation 2). This gives us the following activation rule:

$$a_j(t) = \text{logistic}(\overline{net}_j(t)) \qquad \text{(Equation 4)}$$

where $\overline{net}_j(t)$ is defined as above. The only difference between this activation rule and the one used in the cascade model is that the time-averaged net input to a unit is passed through the logistic function to arrive at its activation. We are still assured that the activation value will approach an asymptote which depends only on the input pattern and the connection weights in the network. In fact, this asymptote is the same as the

activation that the unit would achieve without the use of time-averaging (to see this, consider the limiting case in which $\tau = 1$).

Learning in multilevel networks can be accomplished using the generalized delta rule, or back propagation learning algorithm described by Rumelhart, Hinton and Williams (1986). Learning occurs by adjusting the connection strengths so as to reduce the difference between the output pattern produced by the network and the correct pattern that should be produced for the current input. This difference is essentially a measure of the error in the performance of the network. Error reduction occurs by repeatedly cycling through the following steps: (a) presenting an input pattern to be learned; (b) allowing the network to generate its asymptotic output pattern; (c) computing the difference between this output pattern and the one desired; (d) propagating information derived from this difference back to all of the intermediate units in the network; (e) allowing each unit to adjust its connection strengths based on this error information. By repeatedly applying this sequence of steps to each member of a set of input patterns, the network can be trained to approximate the desired output pattern as closely as desired for each input.

The non-linearity of the activation update rule discussed above is compatible with the back propagation algorithm, which only requires that the activation function be monotonic and continuous (i.e., differentiable). The logistic function satisfies this constraint. Furthermore, so long as units are allowed to reach their asymptotic activation values before error information is computed at the output level, then learning in this system is no different from systems which do not include a time-averaging component.

Throughout training, the error signals represented the discrepancy between the actual asymptotic activations of the output units and their "correct" target values. These were taken to be 1.0 for the unit representing the correct response and 0.0 for all other response units. Note, however, that response selection is assumed to occur based on less extreme calculations (see below), and that the network is not actually able to generate activations of 1.0 and 0.0 with finite inputs. Thus there is always some "error" and therefore always some additional strengthening of connections that is possible, even when response selection is perfectly correct.

Response Selection Mechanism

By combining the cascade model's assumption of time-averaged net input with the logistic activation function and the back propagation learning algorithm, we are able to explore the interaction between processing dynamics and learning within a single system. However, in order to model reaction time data, a response selection mechanism must be included. We assume that response selection is a competitive process which is influenced by the activations of all of the available alternatives. To capture this competition, we assume that the response strength of each possible output at a particular time is equal to the ratio of its activation to the sum of the activations of all of the output units at that time:

$$r_j(t) = \frac{a_j(t)}{\sum_i a_i(t)} \qquad \text{(Equation 5)}$$

A response is assumed to occur when this value exceeds a specified threshold for one of the output units. The response chosen is assumed to be the one that exceeds this threshold.

The processing and response choice assumptions that we have made are deterministic; these are simplifications. This means, for example, that the model in its present form does not simulate error rates or reaction time distributions. Instead, it focuses on predictions for mean reaction times. Detailed modeling of these other aspects of performance is under investigation, but is beyond the scope of this paper.

Attention

The role of attention in the model is to select one of two competing processes on the basis of the task instructions. In order for this to occur, one of two task demand specifications must be provided as input to the model: "respond to color" or "respond to word." We assume that this information is available as the output from some other system which is involved in encoding and interpreting the task instructions, and which is outside the scope of the current model. However, we assume that this information is available in a form which is the same any other type of information in our model: as a pattern of activation over a set of processing units. Accordingly, task information is introduced into our system by activating one of the task demand units. This has the effect of driving units in the appropriate pathway into a more responsive portion of their dynamic range. To see how this occurs, let us first consider the activation function in greater detail.

As described by Equation 4, the activation of a unit is determined by the logistic of its net input. From Figure 3 it can be seen that the logistic function can be divided roughly into three regions. In the middle region—when the net input is close to zero—the relationship between net input and activation is more or less linear, with a slope close to 1. In this region, the activation of a unit is very responsive to changes in its net input. That is, changes in the net input will lead to significant changes in the unit's activation. In contrast, at each end of the logistic function its slope is dramatically reduced. In these regions—when the magnitude of the net input is large, either in a positive or a negative direction—changes in the input to a unit have a small effect on it's activation.

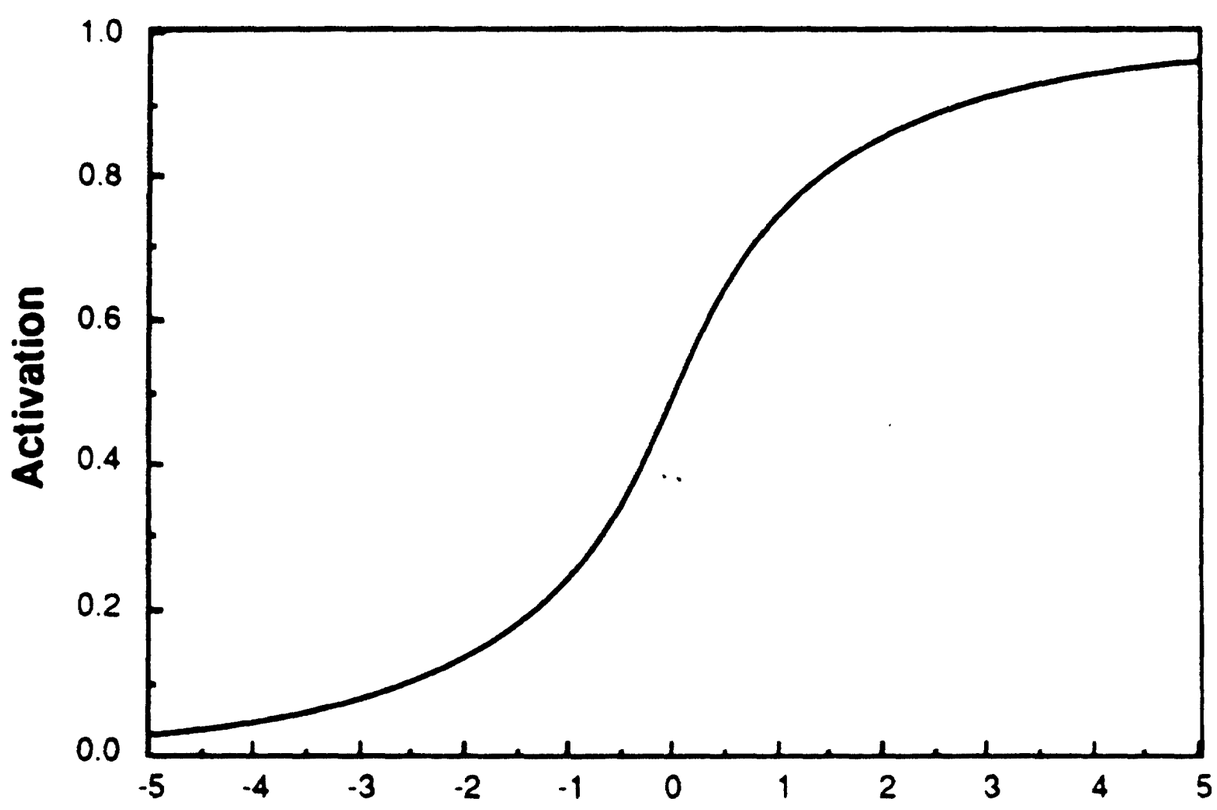

Figure 3. The logistic function. Note that the slope of this function is greatest when the net input is 0.0, and decreases when the net input is large in either the positive or negative directions.

In our model we assume that processing units have a large negative bias. This means that, when they are at rest, their activation is near 0. In this range, units are relatively unresponsive to input. However, the connections from each task demand unit to units in the corresponding pathway are assumed to counteract this bias. Thus, activating a task demand unit brings the activation of units at the intermediate level of the corresponding pathway up to 0.5, placing them in the most responsive region of their dynamic range.

Simulations

We implemented the mechanisms described above in a specific model of the Stroop task. In the following sections we describe how the model was used to simulate human performance in this task. We start by describing some of the general methods used in the simulations. We then describe three simulations which provide an explicit account of the attributes of automaticity and how they relate to practice. These are followed by three additional simulations which extend our consideration to issues concerning the relationship between attention and automaticity.

Simulation Methods

All simulations involved two phases: a training phase and a test phase.

Training phase. During the training phase, the network was trained with each of several input patterns. The network was trained on all of the patterns which included only input to the relevant (attended) processing pathway, and none to the irrelevant pathway (see

Table 2a). Thus, conflict and congruent stimuli were omitted from the training set: each pattern was made up of a task specification and a single input to the corresponding pathway. For example, one input pattern was "red-color-NULL," which activated the red input unit in the color pathway, the "respond to color" task demand unit, but did not activate any word input units. The network was trained to activate the red output unit as its response to this input pattern. The choice of training stimuli reflects the assumption that in ordinary experience we rarely encounter the kind of conflict between word and color that is introduced in the Stroop task.

Table 2a. Training Stimuli.

	Task demand	*Color input*	*Word input*	*Output*
a)	color	red	——	"red"
b)	color	green	——	"green"
c)	word	——	RED	"red"
d)	word	——	GREEN	"green"

Table 2b. Test Stimuli.*

Stimulus Type	*Task demand*	*Color input*	*Word input*
Color naming:			
task specification	color	——	——
control	color	red	——
conflict	color	red	GREEN
congruent	color	red	RED
Word reading:			
task specification	word	——	——
control	word	——	RED
conflict	word	green	RED
congruent	word	red	RED

* Only those stimuli for which "red" was the correct response are shown. The network was also tested with the corresponding stimuli for which "green" was the correct response.

A training trial consisted of the following steps. An input pattern was presented to the network, and all of the units were allowed to reach their asymptotic values. Difference terms were then computed by comparing the actual output of the network to the desired output. These were used to back-propagate difference signals to units at earlier levels of processing. Units used the error signals they received to calculate the change in connection strengths that should be made to decrease this error. Changes to connection strengths were summed over a series of stimulus presentations, and then made at the end of each series.[1] A series is called a training epoch: the composition of each epoch is discussed below. All of the connections along the word and color processing pathways were modifiable, and their values were set using the learning procedure just described. However, the connections from the task demand units to the intermediate units and the bias terms that establish the resting activations of these units were assumed to be unmodifiable. Training proceeded until the network was capable of correctly processing all of the test stimuli (see below: "Test Phase").

One of the purposes of the model is to demonstrate how the processing mechanisms it employs can account for the relationship between practice effects and automaticity. In the context of the Stroop task, it is commonly accepted that word reading is more highly

---

[1] This method for changing connection strengths was used in order to eliminate any effect of stimulus presentation order during training. It can be shown that if the learning rate is small enough, this procedure is equivalent to a procedure in which stimuli are presented in random order and connection strengths are updated after each stimulus presentation.

practiced than color naming. In order to model this discrepancy, we gave the network differential amounts of training on the word and color patterns. This was done by modifying the frequency with which each type of pattern was presented: each of the color patterns was presented with a probability of 0.6 in a given epoch, while all of the word patterns were presented in every epoch. Thus, on average, the color patterns were seen 3/5 as often as the word patterns, and at any given point during training, the network had received a greater amount of practice on word patterns than color patterns.

Figure 4 displays the weights on all of the connections in the network at the end of training. As expected, they are stronger in the word pathway than in the color pathway, due to the greater frequency of word training.[1]

---

[1] We focus on frequency of training as the primary difference between word reading and color naming because this has been the emphasis in the literature. However, other differences between these tasks might also be important. For example, it seems likely that word reading is also a more consistently mapped task than color naming: a particular sequence of letters is almost invariably associated with the word they represent (even if the word itself has an ambiguous meaning); however colors are often associated with words other than their name (e.g., red is associated with heat, embarrassment and "stop"). While this point has not been emphasized with regard to the Stroop task, it is a well established finding that consistent mapping leads to the development of automaticity, while variable mapping impedes it (e.g., Logan, 1979; Shiffrin & Schneider, 1977). Our model captures this fact: the more consistently a stimulus is related to a particular response, the stronger will be the connections for processing that stimulus. Although in this paper we will focus on frequency (i.e., *amount* of practice) as a determinant of pathway strength, it should be kept in mind that consistency of practice is an equally important variable that may be a significant factor underlying the Stroop effect.

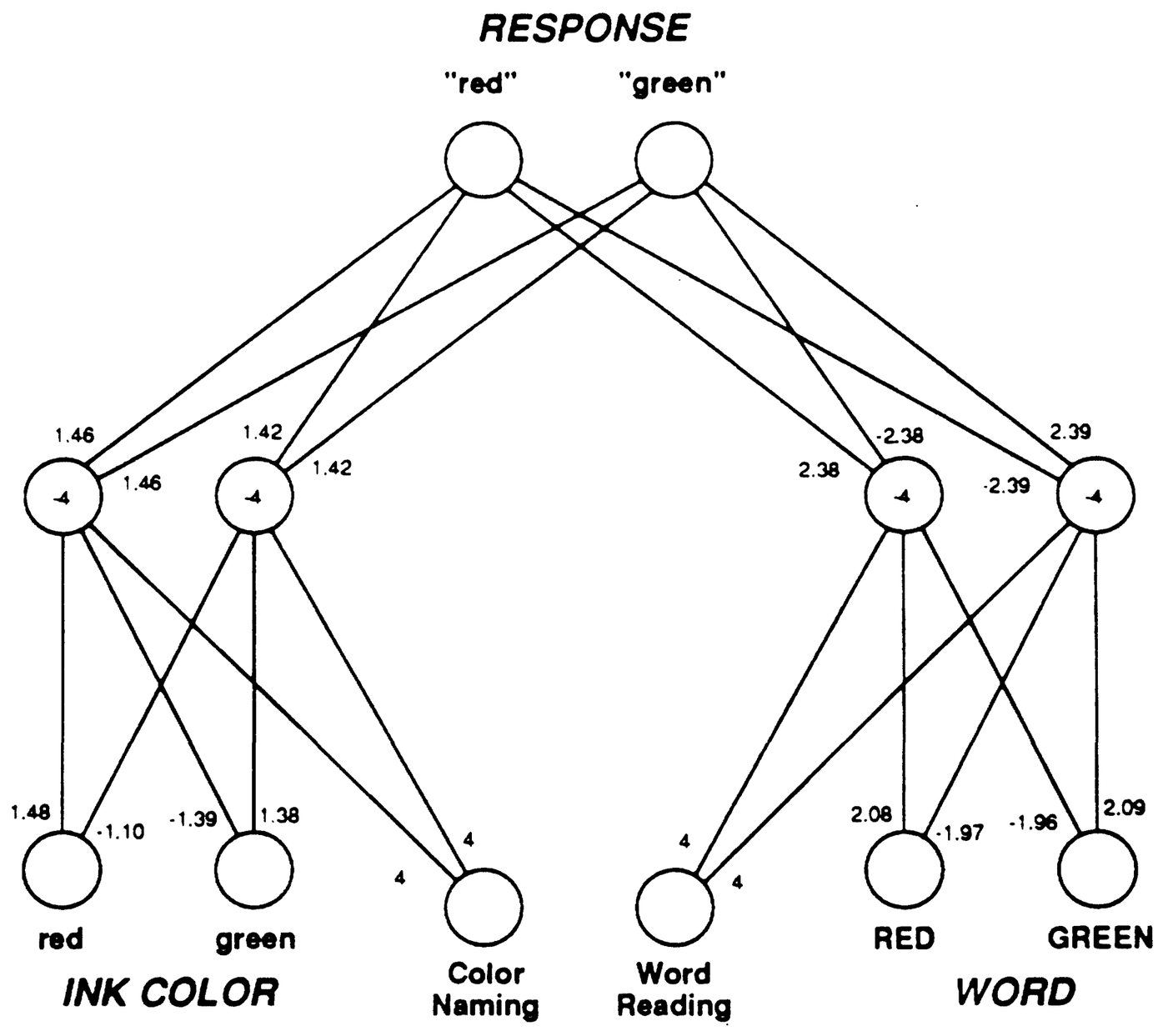

Figure 4. Diagram of the network showing the value of the weights after training on the word reading and color naming tasks.

Test phase. The network was tested on 12 input patterns which corresponded to all possible stimuli in a Stroop task when there are two potential responses (i.e., "red" and "green"): a control stimulus, a congruent stimulus, and a conflict stimulus for each of the two inputs (red or green) in each of the two tasks (word reading and color naming) (see Table 2b). Presentation of a particular pattern consisted of activating the appropriate input unit(s) and task demand unit. For example, one of the conflict stimuli in the color

naming task (the word GREEN in red ink) was presented by activating the red color input unit, the "attend to color" task demand unit, and the GREEN word input unit.

A test trial consisted of the following sequence of events. First, the appropriate task demand unit was activated, and activations were allowed to reach asymptote. This put the network in a "ready" state corresponding to the appropriate task. At this point, the intermediate units in the selected pathway had a resting activation level of 0.5 (that is, they were in the middle of the logistic function), while units in the competing pathway were quite inactive (activations of approximately 0.01). Once this ready state had been achieved, the test pattern was presented, and the system was allowed to cycle until one of the output units exceeded the response threshold. A value of 0.6 was used for the response threshold in all of the simulations we report.[1] The number of cycles required to exceed this threshold was recorded as the "reaction time" to that input. The system was then reset, and the next trial began.

Simulation 1. The Basic Stroop Effect

The purpose of the first simulation was to provide an account for the set of empirical findings that comprise the basic Stroop effect. These are displayed in Figure 5a, and are described below.

---

[1] This was the lowest value that permitted a correct response to be generated in all conditions of all of our simulations.

- Word reading is faster than color naming. The time to read a color word is about 350-450 msec, whereas the time to name a color patch or a row of colored X's is 555-650 msec. Thus word reading is about 200 msec faster than color naming (cf. Cattell, 1886; Dyer, 1973; Glaser & Glaser, 1982).

- The ink color has virtually no effect on the amount of time that it takes to read the word. That is, reaction times to read the word in the conflict and congruent conditions are the same as in the control condition. This phenomenon was originally discovered by Stroop (1935) and can be seen in the flat shape of the graph for word reading in Figure 5a. This finding is extremely robust and has been very difficult to disrupt. Even when the ink color appears before the word it does not interfere with word reading (Glaser & Glaser, 1982). It is only when the task is changed radically that the ink color will interfere with word reading (Dunbar & MacLeod, 1984; Gumenik & Glass, 1970).

- The word can effect the time it takes to name the ink color. In the conflict condition, the word interferes with ink naming producing a substantial increase in reaction time relative to the control condition. The amount of interference is variable, but is usually around 100 msec (e.g., Dunbar & MacLeod, 1984; Glaser & Glaser, 1982; Kahneman & Chajczyk, 1982). This finding is also extremely robust and nearly all subjects show the effect. Even when the word and the ink color are presented in different spatial locations (e.g., the word is placed above a color patch) the word still interferes with ink color naming (Gatti & Egeth, 1978; Kahneman & Henik, 1981). In the congruent condition the word facilitates ink naming, producing a decrease in reaction time relative to the control condition

(Hintzman et al., 1972). The amount of facilitation can range from about 20 msec (Regan, 1978) to about 50 msec (Kahneman & Chajczyk, 1982).

- The amount of facilitation is less than the amount of interference. Congruent stimuli have not been used as extensively as conflict stimuli, but the general finding is that the amount of facilitation obtained is much less than the amount of interference (Dunbar & MacLeod, 1984).

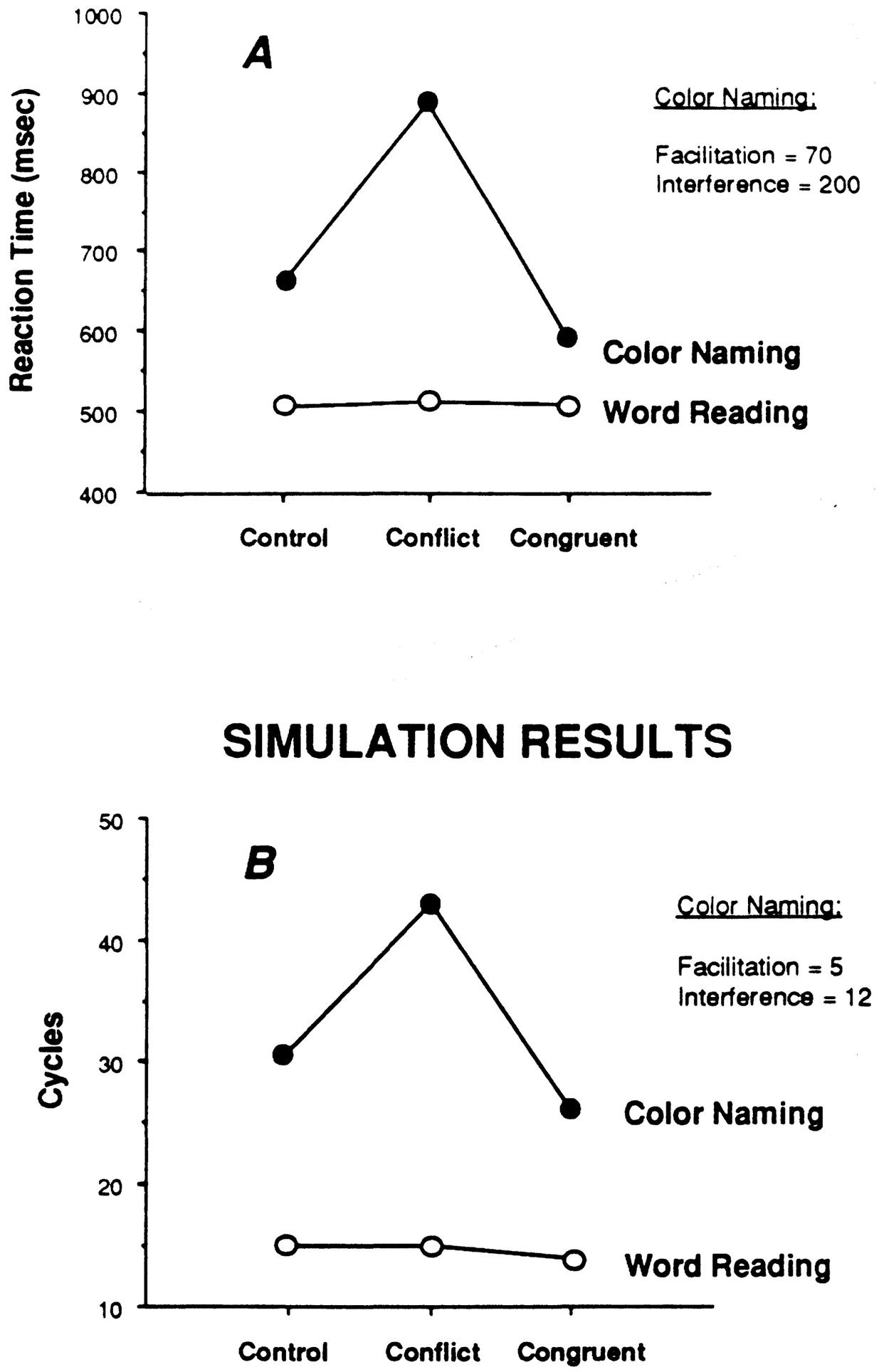

Figure 5. Performance data for the standard Stroop task. Panel A shows data from an empirical study (after Dunbar & MacLeod, 1984). Panel B shows the results of the model's simulation of this data.

Figure 5a represents the findings in a standard Stroop experiment (Dunbar & MacLeod, 1984). Figure 5b presents the results of our simulation of this study. This figure shows that the model reproduces all of the effects described above. By examining the mechanisms which underlie its behavior, we can see how the model provides an explanation for these effects.

Word reading was faster than color naming in the simulation because differential amounts of training led to the development of a stronger pathway for the processing of word information. The fact that the network was trained more extensively with word stimuli than with colors meant that units in the word pathway had a greater number of trials in which to increment their connection weights in the direction necessary to generate the correct response. This resulted in stronger connections in the word pathway than the color pathway (see Figure 4). As a consequence, activation could develop more rapidly in this pathway, allowing a response to be generated in fewer cycles. Recall that the activation of a unit depends on its net input. This, in turn, depends on the activation of each of the other units to which it is connected *times its strength of connection* to those units (see Equation 1). This means that stronger connections to a unit increase the magnitude of its net input. A larger net input causes larger changes in activation to occur in each processing cycle (see Equation 2) and therefore a more rapid change in activation. The faster the correct response unit accumulates activation (and competing units become inhibited), the faster the response threshold will be exceeded. Thus, strength of a pathway determines speed of processing.

The difference in the strength of the two pathways also explains the difference in interference effects between the two tasks. Let us first consider the lack of an effect of

color information on word reading. In this task, activation of the word reading task demand unit offsets the negative bias of the intermediate units in that pathway, so that information flows vigorously along the word pathway. In contrast, because no attention is allocated to the color pathway, units in this pathway remain in an unresponsive part of their dynamic range, and the accumulation of information at the level of the intermediate units is severely attenuated. Furthermore, because the weights from the intermediate units to the output units are weaker in the color pathway, what information does accumulate is transmitted to the output level more weakly than information flowing along the word pathway. Both of these factors serve to diminish the impact that color information has on the production of a response to a word. As such, reaction time in the word reading task is only very slightly affected by the presence of either congruent or conflicting color input.

A very different pattern of results occurs when the task is color naming. In this task, information *does* flow along the competing (word) pathway which affects the time to respond. Attention is allocated to the color naming pathway, so that units in this pathway are placed in a responsive part of their dynamic range, and information flows unattenuated to the output level. It is now the units in the word pathway that are relatively unresponsive. However, because of the larger weights in the word pathway, activation does build up at the intermediate unit level. The amount of this accumulation is more than

was the case for units in the color pathway.[1] Furthermore, the weights are also larger from the intermediate units to output units in the word pathway (as compared with the color pathway), so that what information accumulates is partly amplified and transmitted to the output level. Thus, some information flows along the word pathway even in the absence of the allocation of attention. Although this flow of information is only partial, and is not sufficient to determine the response, it is enough to affect the color naming process. The fact that information is processed in the word pathway without the allocation of attention accounts for the "involuntariness" of word reading, and it is this involuntariness that underlies the interference and facilitation effects that are observed. All of these effects are attributable to the fact that the word reading pathway is stronger (i.e., has greater connection weights) than the color naming pathway.

The fourth finding that we mentioned above, and one that is a highly consistent in standard Stroop experiments, is that the amount of interference is larger than the amount of facilitation. Figure 5a shows that the model reproduces this effect (interference = 12, facilitation = 5). While there are several factors in the model which contribute to this

---

[1] As an example, consider the case in which the red word input unit is activated. This has an excitatory connection to the leftmost hidden unit in the word pathway, with strength 2.38. In the absence of input from the word reading task demand unit, this hidden unit receives a net input of (2.38) + (-4_{bias}) = (-1.62). After passing this through the logistic function, we arrive at an asymptotic activation of .13 for this hidden unit. This will be the amount contributed to the net input of the red output unit. Now consider the situation for the color naming pathway. There, the strength of the connection from the red input unit to the corresponding hidden unit is only 1.39. In the absence of attention, the hidden unit will have a net input of -2.61 ($1.39 - 4_{bias}$) which, when passed through the logistic function, amounts to an activation of 0.7. This is half the activation of the corresponding word pathway unit.

behavior, the most important one is the non-linearity of the activation function. This imposes a ceiling on the activation of the correct response unit, which leads to an asymmetry between the effects of the excitation it receives from the irrelevant pathway in the congruent condition, and the inhibition it receives in the conflict condition. To see this more clearly, consider the idealized situation depicted in Figure 6a.

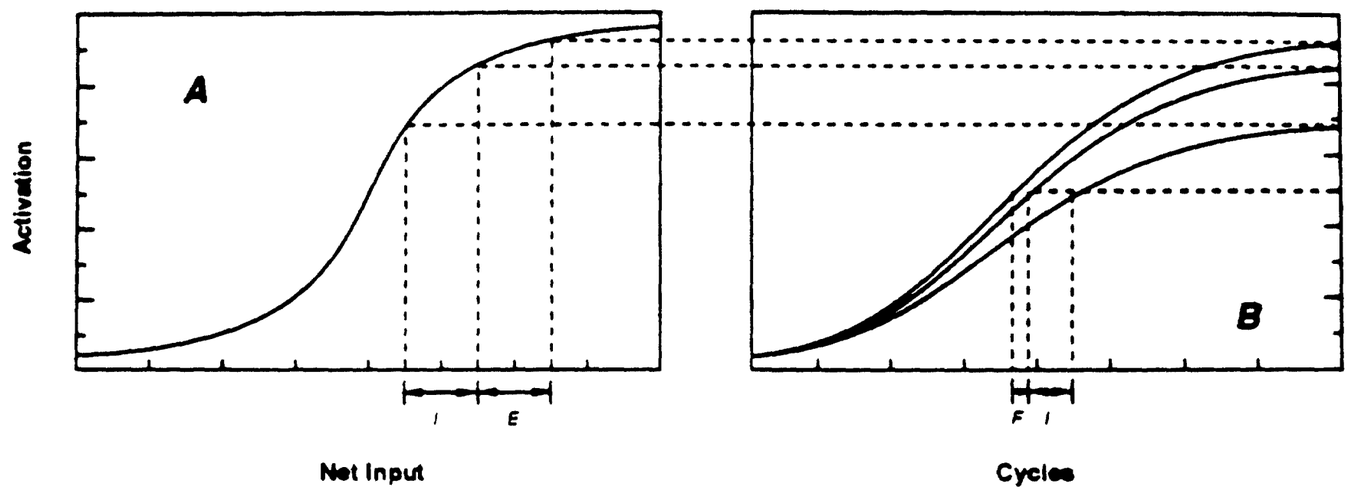

Figure 6. Mechanisms underlying the asymmetry between interference and facilitation effects. Panel A shows the effects that equal amounts of excitation (E) and inhibition (I) from a competing pathway have on the asymptotic activation of an output unit. Panel B shows the effects that these different asymptotic levels of activation have on the time to reach a particular level of activation (F = facilitation; I = interference).

In this figure, the activation function for the correct response unit is shown. Its asymptotic activation is plotted for each of the three experimental conditions in a color naming trial. Note that its activation is highest in the congruent condition and lowest in the conflict condition. This is because in the congruent condition, the competing pathway contributes activation to the response unit, increasing its net input. In the conflict condition, the competing pathway inhibits the unit, decreasing its net input. Note that,

although the net input is incremented and decremented by equal amounts, the effects that these influences have on the activation of the response unit are not symmetric: inhibition has a greater effect than excitation. This is because the activation of this unit is in a non-linear region of the logistic function. In this region, increasing the net input has less of an effect on activation than decreasing it.[1]

The activation values plotted in Figure 6a represent the asymptotic levels approached by the response unit in each of the three conditions. Figure 6b plots the rise in activation, over time, of the response unit toward each of these asymptotic values. It can be seen that, with respect the number of cycles required to reach a particular level of activation, the difference between the control and conflict conditions is substantially greater than the difference between the control and congruent conditions. Thus, the non-linearity of the logistic function, and its interaction with the dynamics of processing help to produce the observed asymmetry between the size of the interference and facilitation effects.

The logistic function was not included in the model specifically to produce this asymmetry. Rather, it was included in order to introduce non-linearity into processing for fundamental computational reasons (see above: "The Mechanisms Underlying Learning and the Time

---

[1] The reason that output activations fall in this region has to do with the nature of representations and training in this system. It can be seen from the logistic function that, at rest (i.e., with net input = 0), units have an activation of 0.5. If the correct response to a particular input pattern requires that an output unit have an activation value of 1.0, then learning will progressively adjust its connections such that its activation shifts from 0.5 to a value closer to 1.0 when that input pattern is present. The region between 0.5 and 1.0 (for units whose output should be 1.0) is precisely the region of the logistic function that produces the asymmetry between interference and facilitation observed in our simulations.

Course of Processing"), and to allow attention to modulate the responsiveness of units in the processing pathways. The fact that this lead to an asymmetry between interference and facilitation is one that we discovered only in the course of conducting simulations.

It is worth noting that most theories have been unable to account for this asymmetry in terms of a single processing mechanism. In fact, several authors have argued that separate processing mechanisms are responsible for interference and facilitation effects (e.g., Glaser & Glaser, 1982; MacLeod & Dunbar, 1988). Although this remains a logical possibility, our model demonstrates that this is not necessarily the case. We believe that the failure of previous theories to account for this asymmetry in terms of a single mechanism has been due to their reliance, either explicitly or implicitly, on linear processing mechanisms.

Simulation 2. Stimulus Onset Asynchrony Effects: Speed of Processing versus Pathway Strength

The results of the previous simulation demonstrate that strength of a pathway determines both the speed with which processing occurs, and whether or not one process will influence (interfere with or facilitate) another. In this simulation we demonstrate that it is pathway strength and not speed of processing *per se* that underlies interference and facilitation effects.

As we discussed in the introduction, there is empirical evidence which contradicts the simple speed of processing account. Thus, consider an account of the Stroop effect in which it is the faster finishing time of the word reading process that is assumed to be

responsible for the asymmetry of interference effects. On such an account, it should be possible to reverse the Stroop effect by presenting color information before the word (this requires spatial separation of color and word stimuli, but this does not eliminate the effect in and of itself—see Gatti & Egeth, 1978). Glaser and Glaser (1982) tested this assumption and found no support for it: color information failed to interfere with word reading even when color information preceded the word by 400 msec. Indeed, they found no effect of colors on words over stimulus onset asynchronies ranging from -400 msec (color preceding word) to 400 msec (word preceding color). This is shown in Figure 7a, which presents data from the word reading condition of one of their experiments.

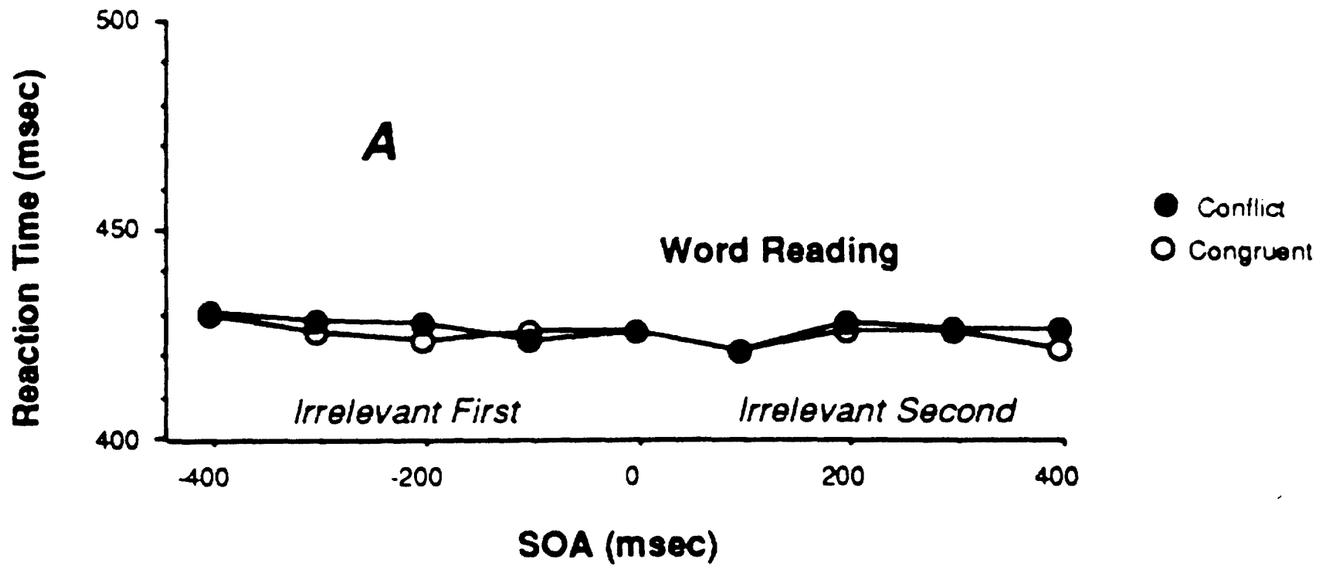

Figure 7. Effects of varying SOA between word and color stimuli on time to read the word. Panel A shows data from the empirical study (after Glaser & Glaser, 1982). Panel B shows the results of the model's simulation of these effects.

We simulated their experiment by activating the color input unit before and after the word input unit with various SOA's. The results of this simulation are presented in Figure 7b. Neither in the empirical study nor in our simulation did presenting the color at any SOA prior to or after the word have any effect on word reading. The model explains this finding by attributing interference effects to the pathway strength rather than to the speed of processing. That is, when no attention is allocated to the weaker pathway, it is simply unable to produce sufficient information at the output level to have an influence on the timing of a response generated through the stronger pathway. According to the model, then, differential speed of processing is not itself the *cause* of interference. Speed is associated with interference only insofar as both result from the strength of the pathway involved. Processes relying on stronger pathways are both faster and have a stronger tendency to produce interference.

We want to emphasize that stronger pathways have a *tendency* to produce interference, but that this is not absolute. Rather, the *relative* strength of a pathways, in comparison to a competing one, is an important factor in determining the kinds of effects that are observed. This is demonstrated in the next simulation.

Simulation 3. Practice Effects: Relative Strength of Pathways

This simulation was conducted to explore, in greater detail, the relationship between training and the effects observed in Simulation 1. In empirical studies, MacLeod and Dunbar (1988) have shown that both speed of processing and the ability of one process to interfere with (or facilitate) another is affected by the relative amounts of training that subjects have received on each. In their experiments, subjects were taught to associate a

different color name to each of four different shapes (see Figure 1). During the training phase, the shapes were all presented in a neutral color (white) and subjects received extensive practice at naming the shapes. After 1, 5, and 20 days of practice with shape naming the subjects were given a set of test stimuli to respond to. In this test phase, subjects were presented with the following types of stimuli: (a) shapes in a neutral color (control condition), (b) shapes in a color that was inconsistent with the shape name (conflict condition) and, (c) shapes in a color that was consistent with the shape name (congruent condition). The results they obtained are shown in Figure 9a, and can be summarized as follows:

- After 1 day (288 trials) of practice on shape naming, this was still more than 100 msec slower than color naming. The shapes had no effect on the time to name the ink colors. However, the ink colors produced both interference and facilitation in the shape naming task. The amount of interference was greater than the amount of facilitation.

- After 5 days (2,304 trials) of practice, shape naming was significantly faster than on day 1. In addition, the shapes now interfered with color naming, although they did not produce facilitation. The colors continued to produce both interference and facilitation in shape naming.

- After 20 days (10,656 trials) of practice, shape naming was slightly faster than ink naming. The shapes produced a large amount of interference and a small amount of facilitation in naming colors. The colors now produced much smaller amounts of facilitation and interference in shape naming.

MacLeod and Dunbar argued that these data contradict the idea that the attributes of automaticity are all or none. Rather, they argued for a continuum of automaticity, in which it is the *relative* amount of training on two tasks that determines the nature of the interactions between them. Our model captures this idea.

To simulate the MacLeod and Dunbar experiments, we preserved the network from the previous simulations (which had been trained on color naming and word reading), and added a new pathway which was used for shape naming. This pathway was identical in all respects to the two pre-existing ones, except that it had not received any training (see Figure 8).

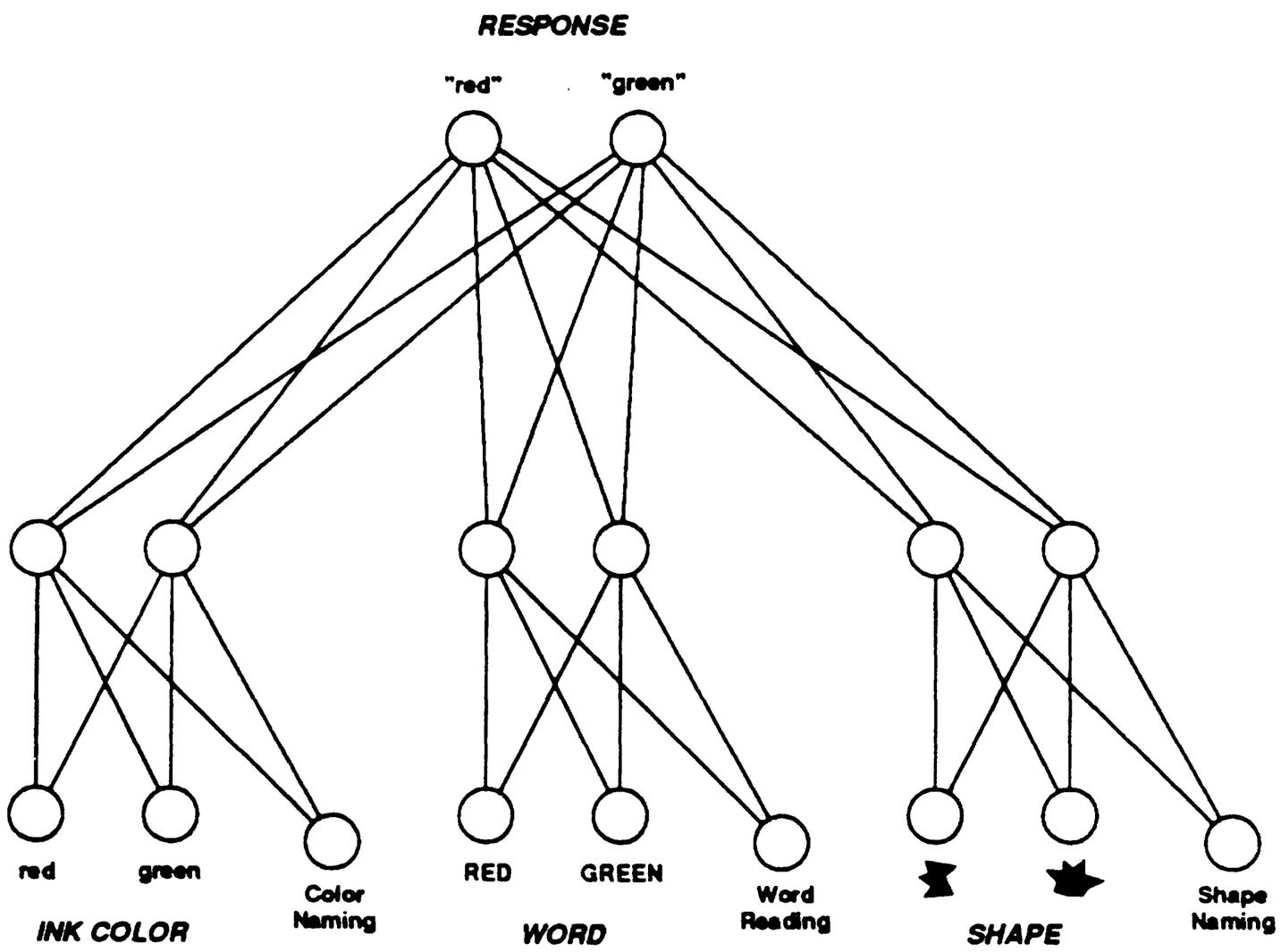

Figure 8. The network architecture used to simulate the shape naming experiments conducted by MacLeod and Dunbar (1988).

This arrangement allowed us to compare performance on a novel task—at various points in training—with one that had already received a moderate amount of training (color naming). No stimuli were presented to the word pathway in this simulation. All other parameters of the model (including the attention weights) were the same as in Simulation 1.

The simulation involved a series of alternating phases of training and testing, as occurred in the empirical study. During each training phase, the network was presented with only the two control shape patterns. Both these patterns were presented in every training epoch. The network did not receive any further training in the color pathway. In all other respects training epochs were identical to those in Simulation 1. After every 10 epochs of training, the network was tested on the 12 relevant test patterns: three for each of the two shape names (control, conflict and congruent) and three for each of the two colors.

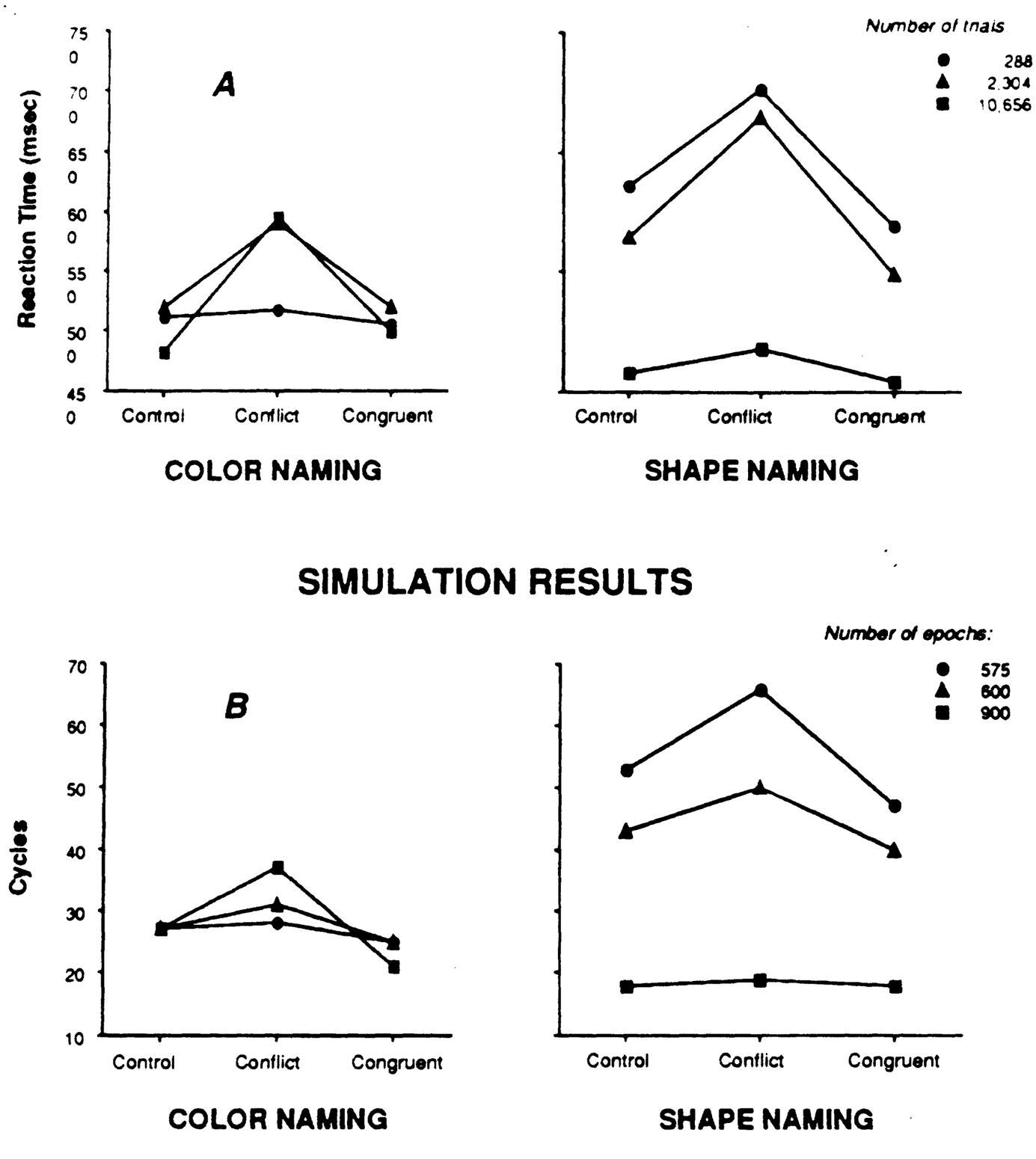

Figure 9. Interference and facilitation in the shape naming and color naming tasks after varying amounts of training on shape naming. Panel A shows the results obtained in MacLeod and Dunbar's Experiment No. 3 (1988), after subjects had received 1, 5, and 20 sessions of practice. Panel B shows the simulation results at three different points in training.

Figure 9a depicts the data of Experiment 4 from MacLeod and Dunbar (1988). Figure 9b presents the results of our simulation at each of three points in training on shapes. The

initial point represents the point at which the model first began to perform accurately in the shape naming task. The second point was chosen to represent an intermediate stage of training, corresponding to the fifth day of shape naming practice in the MacLeod and Dunbar experiment. Finally, the network was tested after the amount of shape training had exceeded the amount of color training in Simulation 1.

These figures reveal an overall correspondence between the behavior of the human subjects and of our model at the initial and final stages of training. In both cases, shape naming is much slower than color naming at the first point in training. Shape naming also shows interference and facilitation from colors at this point, while color naming is not affected by shapes. At the final point in training, the relationship between the two processes has been reversed: shape naming has become the faster process, and is much less sensitive to interference, while it is able to produce interference with color naming. The reason why shape naming starts out as the slower process is because, early in training, the size of the weights in this pathway are still much smaller than those in the color pathway. The relationship between the two processes at this point is directly analogous to the relationship between word reading and color naming in Simulation 1. Note, however, that in this simulation color naming has switched roles: it is now the process which is insensitive to interference or facilitation, and which is able to produce these effects. This switch in roles occurred without any change in the strength of the color pathway. This makes it clear that it is not just the absolute strength (i.e., magnitude of the weights) of a pathway, but also its relative strength compared with a competing pathway that determine whether or not a process will produce, or be subject to interference in a Stroop-like task. This claim is further substantiated by the patterns of

performance at the end of training in the simulation. At this point, shape naming has improved to the point where it is slightly faster than color naming. Accordingly, shape naming is now insensitive to colors, while shapes have come to produce both facilitation and interference in color naming.

Based on their data, MacLeod and Dunbar suggested that the Stroop effect can be understood in terms of the relative position of competing tasks along a continuum of automaticity, and that the position of the tasks along this continuum can be influenced by training. The results of our simulation are consistent with this view, and demonstrate that the observed effects are explained by increases in pathway strength that accompany training.

However, there are some aspects of human performance which this simulation does not capture. In our simulation, the network required an extended period of initial training before it was able to respond correctly to the full set of test stimuli.[1] In contrast, human subjects are able to perform accurately on a novel task such as shape naming after just a small number of practice trials. We will return to this issue in the general discussion. For present purposes, we simply point out that the model focuses on processes that have received moderate to large amounts of training. This is consistent with the empirical paradigm, in which subjects received a considerable amount of practice (more than 250 trials) before they were first tested on the "novel" shape naming task.

---

[1] Approximately 500, although the exact number of epochs is dependent on the learning rate parameter, and the number and type of patterns that must be learned.

There is also another respect in which the behavior of our model differs qualitatively from that of human subjects. This can be seen by examining performance at the intermediate stage of training. At this point, the empirical data show that each task interferes with the other. In the simulation, however, shape naming produced only a small amount of interference for color naming, despite the fact that shape naming showed overall improvement (reaction times were faster in all conditions). This turned out to be a robust property of the model: we were unable to get a process in a weaker pathway to interfere with a process using a stronger one without sacrificing other aspects of performance. It is not clear at this time whether this represents an inherent limitation of our approach, or of the particular model we have implemented.

Simulation 4. Attention and Processing

The first three simulations point out that speed of processing and interference effects all vary continuously with practice, as a result of changes which occur in the strength of the underlying processing pathways. Speed of processing and interference effects have played a central role in empirical research on automaticity. At the theoretical level, however, discussions have focused largely on the extent to which attention can influence or control processing. Automatic processes have been distinguished from controlled processes by their independence from requirements for attention. To the extent that automaticity has been treated as if it were all or none, this has implied that automatic processes are not influenced by attention whatsoever. However, as we pointed out in the introduction, there is a growing body of evidence which indicates that automatic

processes are, in fact, subject to some degree of control by attention. Our model is entirely consistent with this view.

While the simulations we have presented so far show how processing can occur in absence of attention—capturing the "involuntariness" of automatic processes—it is clear that all of the processes in the model are nevertheless influenced by attention. Although processing in the absence of attention was able, under some conditions, to produce interference, in no case was it able to determine a response. That is, even the strongest processes required the allocation of attention in order to occur successfully.

In the following set of brief simulations, we directly examine the requirements that different processes have for the allocation of attention, and the factors which affect these requirements. Like the other attributes of automaticity that we have discussed, requirements for attention are found to depend on the strength of processing in a pathway, as well as the strength of a competing pathway.

First we examined the requirements that different processes had for attention when they were not in competition with another process. Requirements were considered in terms of the minimum amount of activation that was needed on the task demand unit in order for a process to generate an overt response, regardless of reaction time.[1] As expected, stronger pathways required less attention than weaker pathways. Thus, for example,

---

[1] Of course, this will depend to a large extent on the response criterion (see above: "The Response Selection Mechanism." For example, lowering the response criterion will lower the requirement for attention. The value that was used in these simulations was the same as the one used in all other simulations–see earlier note.

successful word reading required that the word reading task demand unit have a minimum activation of 0.11, whereas color naming required a task demand activation of at least 0.48. Note, however, that although stronger pathways require less attention for successful processing, they still exhibit a requirement for attention. Without any activation of the task demand unit, even the word reading pathway was unable generate a response. Indeed, the fact that even processing in stronger pathways is controlled by attention is what allows the model to perform a task using the weaker of two competing pathways. If, for example, word reading had been unaffected by the withdrawal of attention, then the model would not have been able to perform in the conflict condition of the color naming task: the word information would always have overridden information in the color pathway.

Requirements for attention are not a simple function of the absolute strength of a pathway. They are also affected by the circumstances under which a process occurs. In an additional set of simulations, we tested the attentional requirements of processes when conflicting information was present in a competing pathway. In this condition, the attentional requirements of a process were significantly affected by the strength of the competing pathway. For example, color naming in the presence of conflicting shape information (when this pathway was still the weaker one) required that the activation on the color task demand unit be increased from 0.48 to 0.52. However, the minimum activation was substantially higher (0.74) when color naming had to compete with information in the much stronger word reading pathway. Thus, like the other attributes of automaticity we have discussed, the requirements for attention that a process exhibits depend heavily on the context in which the process occurs.

Thus, the model shows that, although processing can occur in the absence of attention, all processes are influenced by, and exhibit requirements for the allocation of attention. Like the other attributes of automaticity, requirements for attention vary according to the strength of the underlying pathway, and the context in which the process occurs.

Simulation 5. Response Set Effects: Allocation of Attention at the Response Level

In the preceding simulations, we focused on the role that attention plays in selecting information from one of two competing pathways. Attentional selection occurred at the level of the intermediate units, where information in the two pathways was still separate. However, the mechanism we used for the allocation of attention is a general one, and can be applied to other levels of processing as well. In the next simulation we demonstrate how this mechanism can be used for the selection of a particular set of responses at the output level of the network, and how this accounts for response set effects that have been observed in empirical studies. Before we describe this simulation, let us first consider the empirical findings.

Response set effects reflect the fact that information related to a potential response leads to more interference (and facilitation) than information which unrelated to the task. In the standard Stroop experiment, information in the irrelevant dimension always supports a potential response—that is, one which is appropriate for some of the trials in the experiment. Potential responses are said to make up a "response set." For example, in the color naming task, when the word RED is written in green ink, although "red" is an incorrect response in that particular trial, it will be a correct response on other trials (e.g.,

when the word GREEN is written in red ink). Thus, both "red" and "green" are in the response set. However, a word such as DOG is not in the response set, since it is never a response in this task. Several studies have shown that stimuli such as this produce less interference and facilitation than those which are in the response set (e.g., Dunbar, 1985; Klein, 1964). An explanation that is commonly offered for this effect is that members of the response set are primed, either directly by instructions for the task (i.e, by informing the subjects of the stimuli they will have to respond to) or through direct experience with the stimuli in the course of the task itself (e.g., Kahneman & Treisman, 1984). Our model provides a related account of response set effects, in terms of the general mechanism that it uses for the allocation of attention.

The same mechanism that we used to allocate attention to a particular pathway in previous simulations can also be used to allocate attention to a particular response or set of responses at the output level. In the previous simulations, allocation of attention to a processing pathway placed the intermediate units in that pathway on a more responsive part of their activation curve. This occurred through the activation of a task demand unit which offset the negative bias on intermediate units in that pathway. The same mechanism can be implemented at the response level, by adding a negative bias to each

of the output units, and having the allocation of attention to a response offset the negative bias on the appropriate output unit.[1]

To demonstrate how this mechanism can account for empirically observed response set effects, we extended the model used for Simulation 1 in the following ways. First, we added three new units to the word pathway at each level of processing (see Figure 10). One of these represented an additional member of the response set (YELLOW). The remaining two were used to represent stimuli and responses that were not part of the response set. These stimuli were labelled DOG and TABLE. We also added three units to each level of the color naming pathway. As with the word naming pathway, one of these represented an additional member of the response set (yellow). However, we did not assign any meaning to the other two units (i.e., those corresponding to DOG and TABLE in the word pathway). These were used strictly to maintain symmetry between the two pathways in the network, so differences in processing between them could not be attributed to architectural asymmetries.

---

[1] In fact, this mechanism was implicit at the output level in the first two simulations. To see this, imagine that a negative bias was associated with each of the output units, just as it was with the intermediate units. However, because both output units were always in the response set, we can assume that attention was maximally allocated to each. This would offset the negative bias on both of them. That is, the bias terms on the output units would always be equal to zero. For simplicity, we simply omitted the bias terms at this level of the network.

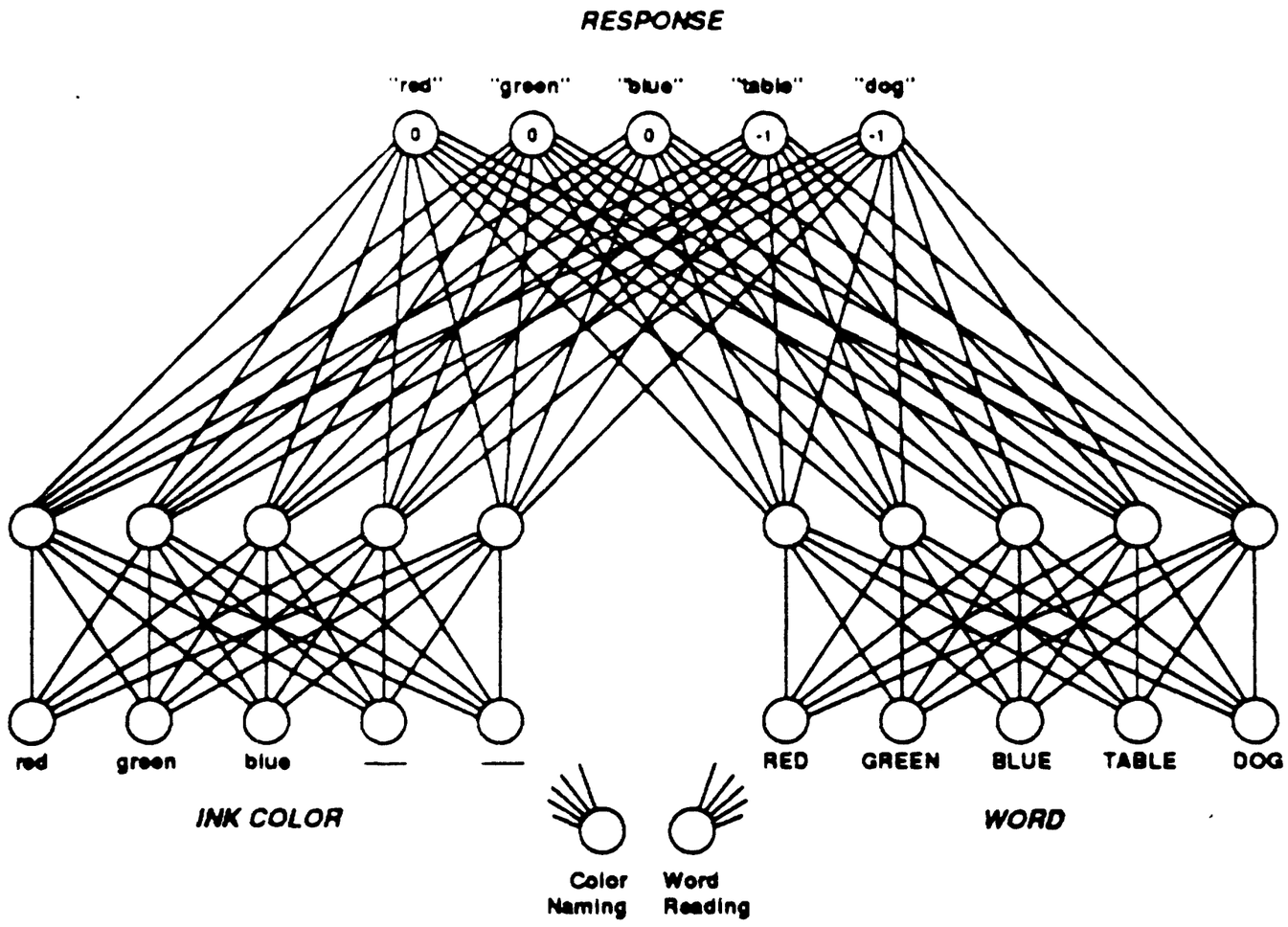

Figure 10. The network used to simulate response set effects and the dilution effect. The numbers that appear with each output unit are the bias terms that were assigned to these units during testing (see text for an explanation).

The new network was trained in a way that was analogous to training in Simulation 1. Both pathways were trained on all five of their inputs, with stimuli in the color pathway receiving 0.6 the amount of training received by those in the word pathway. During training, none of the five response units was given a negative bias. That is, we assumed that during training attention was maximally allocated to all responses.[1] During testing, however, the

---

[1] The activation required to offset the negative biases was assumed to come from a set of units which represented subjects' explicit or implicit knowledge about the composition of the response

(note continued on next page)

non-response set units ("dog" and "table") were given a partial negative bias (-1.0). This corresponded to the hypothesis that, during testing, subjects allocate attention maximally to relevant responses and disattend to irrelevant responses, though not completely (cf. Deutsch, 1977; Kahneman & Treisman, 1984).

Table 3 presents data from an empirical study that examined response set effects (Dunbar, 1985) as well as the results of our simulation. In both cases, stimuli that were not in the response set produced less interference (and facilitation) than response set stimuli. In the model, this difference was due to the partial negative bias on the non-response set output units, which simulated failure to maximally attend to corresponding responses. This negative bias led to partial inhibition of these units, reducing the degree to which they were activated by input stimuli. As a result, these units could not compete as effectively with the correct response unit (e.g., "red") as could output units that were in the response set (e.g., "green" or "yellow").

---

set. These units would be analogous to the task demand units used to allocate attention to the intermediate units in the color and word pathways. For the sake of computational efficiency, we chose not to actually implement these "response selection units" (and their connections to the output units). The results that we report are identical to those that would have occurred if we had included these units, and allowed them to remain active throughout the simulation.

Table 3. Response Set Effects.

Condition	*Dunbar (1985)*	*Simulation*
Conflict	781	61
Related	748	49
Unrelated	721	45
Control	657	43
Congruent	634	33

It is important to note, here, that the model does not explain how potential responses come to be primed and irrelevant ones do not, in just the way that it did not explain how attention comes to be allocated to one pathway and not the other in our previous simulations. In both cases, mechanisms are involved which are sensitive to subjects' knowledge about the overall demands of the task. We have not modelled these mechanisms, but simply assumed that their output is available in a form that can be used to allocate attention in our model: i.e., as a pattern of activation over a set of units. What the model *is* able to explain—in terms of a single attentional mechanism applied to two different levels of processing—is how the allocation of attention can affect the patterns of interference that arise from competing processes.

Simulation 6. The Dilution Effect: Interference Within a Pathway

In general, interference arises in our model when two different sources of information provide conflicting input to a single module. This is because the two inputs lead to competing representations within the module (i.e., different patterns of activation). From this point of view, the informational capacity of a module is limited. That is, there is a trade-off between the number of disparate signals that can be processed at once and how effective the processing will be for each (see Schneider, 1985 and Schneider & Detweiler, 1987 for related arguments). In the simulations described above, we observed Stroop-like interference when competing signals arose in different pathways and converged at the response level of processing. However, there is no reason why the same kind of interference could not occur at other stages of processing. For example, within the word reading pathway, activating more than one input unit should lead to interference at the intermediate unit level, and thereby reduce the effectiveness of processing in that pathway.

This is consistent with an empirical effect reported by Kahneman and Chajczyk (1983). They used a modified form of the Stroop experiment in which words were printed immediately above and/or below a color patch. With this technique, they were able to reproduce the pattern of interference and facilitation obtained when standard, integrated stimuli are used (i.e., words printed in colored ink). However, when two words were presented (a color word and a neutral non-color word), there was a consistent reduction in both the amounts of interference and facilitation produced by words. They referred to this as a dilution effect, because the presence of the neutral word "diluted" the impact of the color word on the color naming task. The implication of this finding was that the

presence of an additional word decreased the effectiveness with which the color word was processed. Our model suggests that one source of this effect is interference between the two inputs within the word processing pathway. To demonstrate this, we conducted a simulation of Kahneman and Chajczyk's main experiment.

For our simulation of this experiment, we used the network that was developed for Simulation 5 (see Figure 10). There were five input units and five intermediate units in each of the two processing pathways, with both pathways converging on a set of five response units. Three of the response units were used for colors ("red", "green" and "yellow") which corresponded to the color-related inputs in each of the two pathways. The two additional input units in the word reading pathway were used to represent neutral stimuli. Kahneman and Chajczyk used the term neutral to refer to words which were not in the response set; that is, word which were not related to any of the responses involved in the color naming task. We used the same method described in Simulation 5 to differentiate response set items from neutral (nonresponse-set) items: we added a partial negative bias to the output units associated with the neutral stimuli. As in Simulation 5, the two additional units in the color naming pathway were included only to maintain symmetry between the two pathways, and were not used during the simulation. Connection weights were also the same as those in Simulation 5. Each trial consisted of turning on a color input unit, the task demand unit for color naming, and various combinations of input units in the word pathway. For example, one of the trials corresponding to Kahneman and Chajczyk's conflict-neutral (CN) condition consisted of turning on the red color unit, the "attend to color" task demand unit, and the GREEN and DOG word units. As in previous simulations, reaction time was the number of cycles

required for an output unit to exceed the response threshold. In all cases, the network was able to produce the correct response (e.g., "red" in the example above).

Table 4 presents the results of Kahneman and Chajczyk's main experiment (Experiment 1, Kahneman & Chajczyk, 1983) and the results of our simulation. From this table it can be seen that the model produces a dilution effect which is qualitatively similar to the empirical effect. The addition of a neutral word reduced both the amount of interference produced by a conflicting color word and the amount of facilitation produced by a congruent word. The cause of this effect has to do with interactions that occur within the word processing pathway. Note that each input unit has an excitatory connection to a corresponding intermediate unit, and inhibitory connections to the intermediate unit for other inputs. As a result, inputs can interfere with one another: when two input units are activated, each inhibits the intermediate units for the other. This mutual inhibition reduces the amount of information that arrives at the output level from the word pathway, thereby reducing the impact that word information has on the color naming process.

Table 4. The Dilution Effect.

Condition	*Kahneman & Chajczyk (1983) Experiment 1*	*Simulation*
Conflict	682	48
Conflict-Neutral	650	39
Dual Neutral	614	35
Neutral	610	35
Congruent-Neutral	585	29
Congruent	561	26

The match between our simulation results and the empirical data is not perfect. The dilution effect we obtained was larger than that observed by Kahneman and Chajczyk. We also found a dilution effect when two conflicting color words were presented (e.g., GREEN and YELLOW when the correct response was "red"), which did not appear in the empirical data (CC condition, Experiment 4).

These discrepancies require further explanation. Nevertheless, our results suggest that an important cause of the dilution effect may be interference between different sources of information within the word processing pathway. This interference is directly analogous to the Stroop interference that occurs at the response level of the network when conflicting information arrives from the color and word pathways. As we pointed out earlier, these interference effects arise due to limitations in the informational capacity of the processing modules involved. In this view, then, both Stroop interference and the

dilution effect are due to capacity limitations of "local" resources—i.e., modules which are more or less specific to the processing pathways involved in the particular task. This view is closely akin to the notion of multiple resources, which we will discuss more fully below. This can be contrasted with an alternative account, in which interference is seen to arise from the capacity limitations of a more general, central processing mechanism. On this view, the dilution effect would be explained by assuming that word processing relies on the processing capacity of this central mechanism, and that dilution arises as a consequence of the extra demands placed on this mechanism when two or more words are processed at once. Our model does not rule out this possibility. Interestingly, however, the model suggests that the same general principles which govern capacity limitations in local resources, such as the modules within the word processing pathway, may also apply to capacity limitations in more central mechanisms. This idea will be explained more fully in the general discussion below.

General Discussion

We have shown that a simple network-based account can capture many of the empirical phenomena associated with the Stroop effect. Among these are the fact that interference effects are typically larger than facilitation effects (Dunbar & MacLeod, 1984); that presenting the color before the word does not succeed in producing interference in word reading (Glaser & Glaser, 1982); that words which are not in the response set produce less interference with color naming than words which are (Dunbar, 1985; Klein,

1964); and that the presence of additional words in the display can reduce the amount of interference produced in color naming (Kahneman & Chajczyk, 1983).

The model provides a common explanation—in terms of pathway strength—for the properties most frequently associated with automaticity: speed of processing, resistance to interference, the ability to produce interference, and the ability of processing to occur without the allocation of attention. Our account of these phenomena goes beyond many other theories of automaticity by describing an explicit set of processing mechanisms from which these phenomena are shown to arise. These mechanisms provide a basis for learning, the time course of processing, and the allocation of attention. Several important features of automaticity emerge from this account, including the fact that the properties of automaticity are continuous, and that their emergence depends largely on the strength of the pathway underlying a process *relative to* the strengths of the pathways used by competing processes.

The model we have presented is not perfect in its present form. For example, it falls short of precise quantitative predictions regarding some of the response set effects and the dilution effect discussed above. In addition, it is does not capture the fact that weaker processes seem, at times, to be able to interfere with stronger ones, as evidenced by the performance of MacLeod and Dunbar's (1988) subjects at the intermediate point in training on the shape naming task. The model also fails to capture performance early in training on a novel task. Some of these shortcomings may be due to the fact that the model does not include mechanisms for the processing of strategic components in a task (e.g., interpretation of task demands, or evaluation of the response set). Further research and development are needed in order to capture these and other aspects of

performance. Nevertheless, the successes of the model to date impress us with the usefulness of the general approach. In the remainder of this discussion we consider the implications of the approach for issues beyond those directly addressed in our simulations.

Redefining Controlled and Automatic Processing

A theme which runs through most research on controlled versus automatic processing is that continued practice leads to automaticity. Thus, Bryan and Harter (1899), Shiffrin and Schneider (1977), Logan (1980) and others have provided evidence that performance on novel tasks often exhibits the attributes of controlled processing, while highly practiced performance (in consistently mapped tasks) appears to be automatic. Our model provides a set of mechanisms which account for the link between practice effects and the attributes of automaticity. Put simply, practice results in an increase in connection strengths, which lead to faster processing as well as many of the other attributes of automaticity. The model shows how these attributes develop gradually with training, reflecting an underlying continuum of automaticity.

However, at the extreme low end of the automaticity continuum, where there is no pre-existing pathway to perform a task, it is clear that processing must occur in a very different way. To be concrete, consider a subject who is told to say "red" when a particular random figure is presented, and to say "green", "blue", etc. for each of several other shapes. While the details could be quibbled with, something like the following is probably happening when the subject is initially asked to perform this task. Early in training, the subject will simply lack the relevant connections for performing the task. Under these

circumstances, the task might at first be performed with the assistance of feedback provided by the experimenter (i.e., by being reminded of the color word that corresponds to the shape on the screen). The subject might try to learn each correspondence as she goes along, using verbal associations to the shapes (e.g., orange is the name of the shape that looks like Florida) or other mnemonics, as well as by getting feedback regarding her response. As this process proceeds, our assumption is that the subject would also be starting to build up connections in a pathway that will ultimately allow the shape to directly activate an internal representation corresponding to the desired response, without recourse to verbal and/or mnemonic mediation.

At some point in this process—let us say, after the feedback from the experimenter is no longer needed, but perhaps before the mnemonics have been dispensed with—the subject can respond correctly. At this point, the subject will be able to perform the task, but the pathway that will ultimately be sufficient—here called the direct pathway—will not. During this phase, we suppose that the activation of a response would be based on the combination of information from both sources, with the relative importance of the contribution of the direct pathway growing steadily over trials, and the relative importance of the other source diminishing. In summary, what we see during the early phases of practice is a gradual transition from what we might call indirect to direct performance.

There is a partial correspondence between our indirect/direct distinction and the traditional distinction between controlled and automatic. We imagine that initial, indirect performance would have all the earmarks of what is typically called a controlled process; it would be slow, it would consist of a series of more or less discrete steps which were subject to disruption or interference, and it would tend to depend on declarative (verbal)

memory (e.g., "Florida is orange"), requiring the allocation of attention. At high extremes of practice, direct performance would closely correspond to what has typically been called automatic; processing would be faster, less susceptible to interference, more capable of producing interference, and less influenced by the allocation of attention. In between, however, the correspondence between these distinctions breaks down. As we have seen in our simulations, a process which is completely direct can—under some circumstances—evidence all the usual criteria for being a controlled process. For example, color naming was simulated as a direct process in our model. However, when competing with word reading, it was found to be a slower process, susceptible to interference, unable to produce interference, and to have a greater requirement for the allocation of attention. Thus, we propose that processes which have previously been classified as controlled might more profitably be segregated into those which are direct and those which are indirect. Within the range of direct processes, there would be a continuous spectrum of pathway strengths which would span the range of different degrees of automaticity. Table 5 illustrates the correspondence between traditional usage and the terms we propose here.

Table 5. Types of Processing.

Novel Tasks				Highly Practiced Tasks
Indirect	→	(Indirect with very weak direct)	→ Weak Direct →	Direct
Controlled	→		Controlled →	Automatic

Whether a direct process exhibits the traditional attributes of automaticity or of controlled processing depends upon the context in which it occurs. If it must compete with a process occurring in a stronger pathway, it will appear to be controlled; if it competes with processing in a weaker pathway, it will appear to be automatic. This view suggests that the criteria traditionally used to distinguish between controlled and automatic processes—speed of processing and interference effects—may not always represent qualitative differences in processing. While two tasks may show radically different performance using these criteria, the same type of processing may underlie the performance of each. Thus, the criteria of speed of processing and interference effects cannot be used as evidence for qualitatively different types of processing. In terms of our distinction between direct and indirect processes, differences in performance according to these criteria do not tell us whether a process is direct or indirect.

Our model provides an explicit account of direct processes, and shows how changes in the strength of these processes—which result from practice—can lead to seemingly qualitative changes in performance. However, the model does not provide mechanisms for indirect processing, nor does it describe in any detail the evolution of a process from direct to indirect processing. As we have suggested, this may explain our inability to model performance during the early stages of training, as well as our inability to capture some of the strategic effects that are associated with performance in the Stroop task (e.g., the decision processes which direct the allocation of attention). We should note, however, that others have begun to address these issues using PDP techniques (e.g., Schneider, 1985; Schneider and Detweiler, 1987; Hinton & Plaut, 1987).

Attention and the Control of Processing

We have pointed out that an important difference between our direct/indirect distinction and the traditional dichotomy between controlled and automatic processing is that in our distinction, processes of either type can exhibit performance characteristics traditionally associated with controlled processing—such as slower speed and susceptibility to interference. The same difference can be found between these two approaches concerning their claims about the attentional control of processing.

At the heart of the theoretical distinction between controlled and automatic processing are two basic assumptions: controlled processing depends on the allocation of attention; automatic processing occurs independently of attention. Kahneman and Treisman (1984) refer to the latter of these two assumptions as the "strong automaticity" claim—that is, that automatic processes are not influenced by the allocation of attention. In dispute of

this claim, they and their colleagues (e.g., Kahneman & Henik, 1981; Kahneman & Chajczyk, 1983; Treisman, 1960) have reported a large body of data which suggest that few processes, if any, can be considered to be entirely independent of attention. For example, as we mentioned in the introduction, Kahneman and Henik have demonstrated that the amount of interference produced by an automatic process such as word reading is reduced when attention is not allocated to this process. This indicates that attention does affect processes that are putatively automatic. Our model is consistent with this viewpoint: it shows that direct processes are subject to control by attention. In the simulations we presented, even the strongest pathways—in which processing exhibited all of the other attributes of automaticity—processing was affected by the allocation of attention. For example, although processing in the word pathway occurred without the allocation of attention—leading to interference with color naming—this processing was only partial, and was insufficient to determine the response. Indeed, as we pointed out in Simulation 4, if word reading had not been degraded by the lack of attention, the color naming task could not have been performed.

We should distinguish our claim from similar, but non-identical claims. Others (e.g., Shiffrin, 1988) have attempted to explain the fact that automatic processing tasks such as word reading are subject to attentional control by arguing that behavior relies on numerous processes, some of which are automatic, and some of which may be controlled. In this view, control over the performance of a task could be explained by the allocation or withdrawal of attention from the controlled processes involved, preserving the independence of the automatic processes from the effects of attention. While we do not dispute the claim that behavior is composed of many component processes, our

model asserts that all of these processes are subject—in varying degrees—to control by attention.

Attention as the Modulation of Processing

As we have argued above, all processes can be subject to attentional control. However, the question arises as to how this control is achieved. Attention is implemented in the model as the modulation of processing in a pathway. This occurs by shifting the responsiveness of units in a processing pathway. When attention is allocated to a particular pathway, that pathway becomes more responsive to incoming information. The change in the responsiveness of a pathway is due to input from the attention (task demand) units.

It is worth noting, in this context, that the implementation of attention in our model uses exactly the same processing mechanisms as all of the other components of the model. The connections from the attention units to the units in a processing pathway are no different than the connections within the pathway itself, and attentional information is represented like any other information in the network: as a pattern of activation over a set of units. As such, the input that a pathway receives from the attention units is qualitatively identical to the input received from any other source of information in the network. Attention can be viewed simply as an additional source of information which provides a sustained context for the processing of signals within a particular pathway. Thus, attentional mechanisms are not given special status in the model and, in general, an attentional module can be thought of as any module which has a set of connections that allow it to modulate processing in another pathway. There may be many such modules

within a system, a given module may modulate one or many pathways, and it might even participate directly in one set of processing pathways, while it serves to modulate others.

The notion of attention as a modulator, together with the idea that processing is continuous and that the resulting activations are graded in strength, has a long history in the attention literature; the idea is essentially the same as that suggested by Treisman (1960). Treisman claimed that messages outside of the focus of attention were not completely shut out; rather, the idea was that the flow of information was simply "attenuated" on the unattended channel. This is exactly what happens in our model. Indeed, the very same mechanisms of pathway modulation that we have used to implement task selection in the Stroop task (color or word) can be used to implement channel selection in dichotic listening or other selective attention tasks.

Continuous Nature of Processing

The assumptions that we make that information is graded and is propagated continuously from one level to the next are worth examining in their own right. With respect to these assumptions, our model differs significantly from discrete stage models, in which processing must be complete at one stage before information becomes available to others. In the model we have presented, information at one level is continuously available to subsequent levels. As such, a process need not be completed in order for it to affect performance. It is precisely the partial processing of information in the stronger of two pathways that produces interference and facilitation effects.

In this respect our model is similar to one proposed by Logan (1980). Both models make use of continuous processing mechanisms, and explain interference (and facilitation) effects in terms of the relative strength of the pathways used by competing processes. According to Logan's model, "evidence is assumed to accumulate over time in some composite decision process until a threshold is exceeded and a response is emitted" (p. 528, Logan, 1980). Different sources of evidence (e.g., different stimuli, or different stimulus dimensions) are weighted so that evidence accumulates from each at different rates. This model accounts for the Stroop effect by assigning stronger weights to word reading than to color naming. The strength of the connections in a pathway in our model are analogous to the weights assigned to a process in Logan's model. In both models, attention acts by modulating the effectiveness with which information accumulates from each process in a manner that is responsive to the demands of the current situation.

However, our model differs from Logan's in several important respects, some of which lead to significant differences in performance. First, Logan's model is a linear model, with respect to the way in which information accumulates both as a function of time and as a function of pathway strength. The processing mechanisms in our model are non-linear in both of these respects. This allows it to account for the asymmetry between interference and facilitation. Logan's model can not account for this finding. However, perhaps the most important difference between the two models is that Logan's model does not include any mechanisms for learning. The weights associated with automatic processes are fixed. We consider it to be one of the primary strengths of our model that it can directly address the relationship between training and automaticity in terms of an integrated set of learning and processing mechanisms.

Resources and Capacity

A final issue that we would like to address concerns the notions of processing resources and capacity limitations. The perspective that we offer on these issues is very similar to the "multiple resources" view that has been expressed by others (e.g., Allport, 1982; Logan, 1985; Navon and Gopher, 1979; Wickens, 1984), and to the notion of functional cerebral distance described by Kinsbourne and Hicks (1978). These theories share the view that performance of a task typically involves a number of different processes, which in turn depend on a multiplicity resources. They predict that two behaviors will compete for processing capacity, and may interfere with one another to the extent that they rely on the same resources for different purposes. Our approach suggests ways in which we can begin to think, in more specific terms, about the nature of these resources, and how limitations in their capacity can affect performance. Thus, the modules which make up processing pathways can be thought of as a set of resources within the system. These resources are shared by two or more processes to the extent that their pathways intersect—that is, they rely on a common set of modules. We have noted that when two signals to be processed by a particular module are disparate (i.e, they involve different patterns of activation), they will compete for representation within that module. In this sense, the processing capacity of that module, or resource can be thought of as being limited—that is, it cannot fully support the processing of both signals at once. Schneider (1985) has presented a very similar view of limitations in processing capacity (also see Schneider & Detweiler, 1987).

Although we have not pursued a quantitative analysis of capacity in this paper (see Rosenfeld & Touretsky, 1987 for an example of how this can be done in PDP systems),

the model has allowed us to consider some of the qualitative effects that might be attributed to capacity limitations. For example, our simulation of the dilution effect (Simulation 6) showed that interference can arise when two stimuli (e.g, two words) are processed concurrently within the same pathway. A similar argument could be made with respect to the Stroop effect. In this case, interference would be attributed to capacity limitations with the module shared by the word reading and color naming pathways (i.e., the response module in our network).

Like the multiple resources view, our account of interference effects focuses on capacity limitations that are more or less specific to particular processing pathways. However, the model also suggests ways in which to think about the capacity limitations of other components of processing. Earlier (in our discussion of Simulation 6) we mentioned that limitations in the capacity of more central, general purpose processing mechanisms are often used to explain interference effects, rather than attributing these to capacity limitations in "local" resources. The attention module in our network might be thought of as one kind of central mechanism. Simulation 4 showed that all of the processes in our model relied on the allocation of attention in order to be performed successfully. Allocation of attention to a pathway required that a particular pattern of activation be present in the attention module. To the extent that this pattern is different for different pathways, then the situation could arise in which these patterns come into competition with one another. From this perspective, the capacity of the attention module can be thought of as being limited—it may not always be possible to allocate attention maximally to all processes at once. Because stronger processes have weaker requirements for the

allocation of attention (see Simulation 4), such processes may be less susceptible to capacity limitations in the attention module.

The significance of different sources of capacity limitations (e.g., those arising within a particular pathway, or within a more central, attentional module) are in need of further clarification, both theoretically and empirically. However, we would like to re-emphasize, in this context, that attentional information is not qualitatively different from any other information in our framework. The competition between patterns of activation within the attentional module (that are required for the allocation of attention to particular pathways) is analogous to the competition that can occur between patterns of activation in any other module. This suggests that some central resources, such as the attentional module in our model, may be governed by the same sorts of principles and constraints that govern more local resources within the system.

Conclusion

The model that we have presented provides not only an account of the empirical data on the Stroop effect, but also a more general model of processing in highly practiced tasks and its relation to attention. Like other theorists (e.g., Kahneman & Treisman, 1984; Logan, 1980; Schneider, 1985), we have noted that there are many problems with the traditional views of automatic and controlled processing. Our model suggests that a more useful way to consider tasks is in terms of a continuum of automaticity based on strength of processing. We have outlined a set of mechanisms that can produce gradual and continuous strengthening, and we have shown how these mechanisms can account for a variety of empirical phenomena that are difficult to explain using the controlled/automatic

distinction. In particular, these mechanisms capture the continuum that appears to exist in the attributes of automaticity, and relate this continuum directly to the effects of practice. Differences in practice lead to differences in the strength of processing, and this makes it possible to capture asymmetries of performance such as those observed in the Stroop task. The model also suggests ways in which it may be possible to characterize the notion of capacity in greater detail than has been possible up to now.

The mechanisms used in this model show how the principles of continuous processing—expressed in terms of the PDP framework—can be applied to the study of attention and the control of automatic processes. A problem for our model, however, and for the PDP approach in general is to characterize the mechanisms underlying indirect processing. We see this as an important challenge for future research.

Author notes

We gratefully acknowledge the helpful suggestions made by Colin MacLeod, David Rumelhart and Walter Schneider regarding the ideas presented in this paper.

This work was supported by a research grant from the Scottish Rite Schizophrenia Research Program, N.M.J., U.S.A. and a Physician Scientist Award (MH00673) to the first author; by funding from the Department of Psychology of Carnegie-Mellon University to the second author; and by ONR Contract N00014-82-C-0374, NR442a-483, and a Research Scientist Career Development Award (MH00385) to the third author.

Correspondence concerning this paper should be addressed to Jonathan D. Cohen, Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA 15213.

References

Allport, D.A., (1982). Attention and performance. In G. I. Claxton (Ed.), *New directions in cognitive psychology*. London: Routledge and Keagan-Paul.

Allport, D.A., Antonis, B. & Reynolds, P. (1972). On the division of attention: A disproof of the single-channel hypothesis. *Quarterly Journal of Experimental Psychology, 24,* 225-235.

Blackburn, J.M. (1936). Acquisition of skills: An analysis of learning curves. IHRB Report No. 73.

Bryan, W.L. & Harter, N. (1899). Studies of the telegraphic language. The acquisition of a hierarchy of habits. *Psychological Review, 6,* 345-375.

Cattell, J.M. (1886). The time it takes to see and name objects. *Mind, 11,* 63-65.

Deutsch, J.A. (1977). On the category effect in visual search. *Perception & Psychophysics, 21,* 590-592.

Dunbar, K. (1985). The roles of multiple sources of interference in a picture-word analogue of the Stroop task. Unpublished PhD thesis, University of Toronto.

Dunbar, K., & MacLeod, C.M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 622-639.

Dyer, F.N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory and Cognition, 1,* 106-120.

Gatti, S.V. & Egeth, H.E. (1978). Failure of spatial selectivity in vision. *Bulletin of the Psychonomic Society, 11,* 181-184.

Glaser, M.O., & Glaser, W.R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8, 6,* 875-894.

Gumenik, W.E. & Glass, R. (1970). Effects of reducing the readability of the words in the Stroop color-word test. *Psychonomic Science, 20,* 247-248.

Hasher, L., & Zacks, R.T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 106,* 356-388.

Hinton, G.E. & Plaut D.C. (1987). Using fast weights to deblur old memories. In *Program of the Ninth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum.

Hintzman, D.L., Carre, A., Eskridge, V.L., Owens, A.M., Shaff, S.S., & Sparks, M.E. (1972). "Stroop" effect: Input or output phenomenon? *Journal of Experimental Psychology, 95,* 458-459.

Hirst, W. & Kalmar, D. (1987). Characterizing attentional resources. *Journal of Experimental Psychology: General, 116, 1,* 68-81.

Hirst, W., Spelke, E.S., Reaves, C.C., Caharack, G. & Neisser, U. Dividing attention without alternation or automaticity (1980). *Journal of Experimental Psychology: General. 109,* 98-117.

Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 497-509.

Kahneman, D. & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J.R. Pomerantz (Eds.), *Perceptual organization.* Hillsdale, N.J.: Erlbaum.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, R. Davies, & J. Beatty (Eds.), *Varieties of attention.* New York: Academic Press.

Keele, S.W. (1972). Attention demands of memory retrieval. *Journal of Experimental Psychology, 93,* 245-248.

Kinsbourne, M. & Hicks, R.E. (1978). Functional cerebral space: A model for overflow, transfer and interference effects in human performance. In J. Requin (Ed.), *Attention and Performance VII.* Hillsdale, NJ: Lawrence Erlbaum.

Klein, G.S. (1964). Semantic power measured through the interference of words with color naming. *American Journal of Psychology, 77,* 576-588.

Kolers, P.A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 554-565.

LaBerge, D. (1980). Unitization and automaticity in perception. In *Nebraska Symposium on Motivation.* Lincoln, NE: University of Nebraska Press.

Logan, G.D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 189-207.

Logan, G.D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12,* 523-553.

Logan, D.G. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology, 39, 2,* 367-386.

MacLeod, C.M. & Dunbar K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology, 14,* 126-135.

Malebranche, N. (1674). *Recherche de la verite.* Translated by T. Lennon & P. Olscamp. (1980) as *The search after truth and elucidations of the search after truth.* Columbus: Ohio State University Press.

McClelland J.L. (1979). On the time-relations of mental processes: An examination of systems of processes in cascade. [Psychological Review, 86,] 287-330.

Navon, D. & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86,* 214-255.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226-254.

Neumann, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz and A.F. Sanders (Eds.), *Cognition and motor processes.* Berlin: Springer-Verlag.

Posner, M.I., & Snyder, C.R. (1975). Attention and cognitive control. In R.L. Solso (Ed.), *Information processing and cognition.* Hillsdale, NJ: Erlbaum.

Regan, J. (1978). Involuntary automatic processing in color naming tasks. *Perception and Psychophysics, 24,* 130-136.

Rosenfeld R. & Touretsky, D.S. (1987). Scaling properties of coarse-coded symbol memories. In *Proceedings of IEEE Conference on Neural Information Processing Systems-Natural and Synthetic.*

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart, J.L. McClelland, and the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1.* Cambridge, MA: MIT Press.

Rumelhart, D.E., Hinton, G.E., & McClelland, J.L. (1986). A general framework for parallel distributed processing. In D.E. Rumelhart, J.L. McClelland, and the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1.* Cambridge, MA: MIT Press.

Schneider, W. (1985). Toward a model of attention and the development of automatic processing. In M.I. Posner & O.S.M. Marin (Eds.), *Attention and performance XI* (pp.475-492). Hillsdale, NJ: Lawrence Erlbaum.

Schneider, W. & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. Bower (Ed.), *The psychology of learning and motivation, Vol. 21.* New York: Academic Press.

Schneider, W. & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1-66.

Shaffer, W.O. (1975). Multiple attention in continuous verbal tasks. In P.M.A. Rabbitt and S. Dornic (Eds.), *Attention and performance V.* New York: Academic Press.

Shiffrin, R.M. (1988) Attention. To appear in R.C. Atkinson, R.J. Herrnstein, G. Lindzey, and R.D. Luce (Eds.), *Steven's handbook of experimental psychology, 2nd edition.* New York: John Wiley & Sons, Inc.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Solomons, L. & Stein, G. (1896) Normal motor automatism. *Psychological Review, 3,* 492-512.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

Treisman, A.M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12,* 242-248.

Wickens, D. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention.* Orlando, FL: Academic Press.