Psychological Review
1998, Vol. 105, No. 1, 188–194

0033-295X/98/$3.00

A PDP Approach to Set Size Effects Within the Stroop Task: Reply to Kanne, Balota, Spieler, and Faust (1998)

Jonathan D. Cohen
Carnegie Mellon University and University of Pittsburgh

Marius Usher
University of Kent

James L. McClelland
Carnegie Mellon University

In this reply, the authors point out that the simulations reported by S. M. Kanne, D. A. Balota, D. H. Spieler, and M. E. Faust (1998) did not incorporate mechanisms proposed to explain set size effects in J. D. Cohen, K. Dunbar, and J. L. McClelland (1990). The authors report a new simulation that incorporates these mechanisms and more accurately simulates S. M. Kanne et al.'s empirical data. The authors then point to other factors that could be explored in a more complete test of their model. The use of feed-forward rather than recurrent inhibition is discussed as a potentially important limitation of their original model, and recent work addressing this issue is described. The authors also discuss possible differences between word reading and color naming in the Stroop task. Although such differences may exist, the authors retain their earlier view that such differences do not reflect a dichotomy between automatic and controlled processing.

Kanne, Balota, Spieler, and Faust (1998) reported new empirical data and simulation results that they used to call into question aspects of our model of Stroop task performance (Cohen, Dunbar, & McClelland, 1990). This challenge focuses on the ability of the model to simulate accurately reaction times when the number of stimuli in the task is varied. In particular, they presented data indicating that, as set size is increased, color-naming times increase more than word-reading times for humans but that our model exhibits the opposite behavior. In this reply, we offer three responses to this challenge.

First, we note that Kanne et al. (1998) did not fully incorporate the mechanisms we described in Cohen et al. (1990, pp. 350–353) that were intended to address set size effects. When we incorporate these mechanisms into the simulation, we find that there is a better fit to their data. We also identify additional directions that could be pursued, to assess fully whether our original model can provide an adequate account of set size effects. Overall, we observe that although the simple model presented in the first several simulations of Cohen et al. did not address set size effects, the additional mechanisms we introduced in Simulation 6 may permit an account of such effects.

In the second part of our reply, we discuss fundamental limitations that we believe are inherent in networks using feed-forward inhibition, such as our original model. We have observed such limitations in a variety of behavioral domains, of which the Kanne et al. (1998) findings represent an example. We briefly summarize recent work that considers these limitations and that examines the use of recurrent rather than feed-forward inhibition to address them.

Finally, we consider some of the theoretical claims made by our original model about the relation of word reading to color naming. Although we concur with Kanne et al. (1998) that, in all probability, important differences exist between these processes, we are not convinced that their present results establish this point. Furthermore, we maintain our view—which we believe continues to be supported by our modeling efforts—that the types of differences likely to be discovered will not correspond to the traditional qualitative distinction between controlled and automatic processing (Cohen et al., 1990, pp. 353–354).

Set Size Effects in the Original Model

Kanne et al.'s (1998) study did not incorporate several of the mechanisms for simulating set size effects that we described in our original article (Cohen et al., 1990, Simulation 6, pp. 350–353). These mechanisms implement principles underlying attentional effects central to our model.

First, Kanne et al. (1998) used different networks to simulate different set sizes, each of which was trained on a different number of stimuli. This differs from how we proposed that our model could account for set size effects, which is to alter the allocation of attention among units representing stimuli and responses in the network. In our view, it does not seem plausible that participants come to the experiment with several networks each trained on different numbers of color words. Rather, it seems more reasonable to assume that changing the number of

Jonathan D. Cohen, Department of Psychology, Carnegie Mellon University, and Department of Psychiatry, University of Pittsburgh; Marius Usher, Department of Psychology, University of Kent, Canterbury, England; James L. McClelland, Psychology Department, Carnegie Mellon University.

Correspondence concerning this article should be addressed to Jonathan D. Cohen, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent via Internet to jdcohen@cmu.edu.

stimuli and responses in a given experiment (or condition) induces dynamic changes in the allocation of attention to units representing potential stimuli and responses within a large, single network that includes units for all of the stimuli and responses familiar to the participants. From this perspective, the appropriate way to simulate set size effects would be to train a single network on the maximum number of stimuli and responses involved in the experiment (or perhaps an even larger, more realistic number) and then to test the network on subsets of these of varying size while allocating attention to the stimuli and responses being tested. This is the approach we described in Simulation 6 of Cohen et al. (1990).

Kanne et al. (1998) acknowledged this point in Footnote 3 of their article. There, they report the results of a simulation using their 4-4 network, which was trained on four stimuli, and manipulated attention rather than training to simulate set size effects. However, their manipulation of attention differed from the simulation of response set effects that we described in Simulation 6 of Cohen et al. (1990). Kanne et al. implemented set selection at the intermediate unit level and not at the output level as we had done. They also did not ensure that output units were at the most sensitive part of their activation function. As our next point makes clear, we consider these to be critical differences from the approach that we took in Cohen et al., which was based on a fundamental principle about how the allocation of attention affects processing.

The principle in question is the following: Attention acts to ensure that all units in an attended pathway rest in the most sensitive range of their activation function (which, for the logistic function, corresponds to an activity level of 0.5). We suggest that one of the reasons why Kanne et al.'s (1998) simulations did not produce a good fit to their data is because they did not fully conform to this principle. This principle was implemented in the intermediate layer of our original model (described on pp. 338–339) by ensuring that the weights from the task demand units directly offset the negative bias on these units. It was not necessary to implement explicitly this principle at the response layer of our original model because, as Kanne et al. correctly noted, the symmetry of excitatory and inhibitory weights in a 2-2 architecture guarantees that the net input to each unit at rest will be zero, and, therefore, the activation will be 0.5. This is not true for larger (or asymmetric) networks, such as Simulation 6 of Cohen et al. (1990) or those reported in Kanne et al. However, this can be corrected by allowing participants' knowledge of the task demands to have the effect of setting the resting levels of all task-relevant units to 0.5. We emphasize that such adjustments should not be seen as ad hoc because a central principle of the Cohen et al. model was specifically that the role of attention is to place processing units in the attended pathway in the most sensitive part of their activation function.

A more detailed point is worth noting here. Because task-related attentional effects are assumed to be established in advance of stimulus presentation on a particular trial of an experiment, attention can be seen as presetting the state of task-relevant units in a condition-specific way. These effects can be implemented in a simulation in several ways. One way is by explicitly implementing a set of units that represents participants' knowledge about the demands of the different task conditions and that provides activation to units in the attended pathway. We used such task demand units in our original model to represent participants' knowledge about the relevant stimulus dimension for a given task condition (i.e., color naming or word reading) and to preactivate intermediate units in the corresponding pathway. Another way to preactivate processing units is to adjust their biases. This can be done as a "stand-in" for the effects of the task demand units just described (i.e., as a simpler but formally equivalent alternative to actually implementing the relevant-task demand units) or to capture the influence of local factors (e.g., after effects such as residual activation) that may accrue to units previously engaged in processing. In Simulation 6 of Cohen et al. (1990), we used bias adjustments to preset to 0.5 the activity of units in the response set at the output level of the network.

An important question is how task demand units develop and come to be activated, or how processing units come to be biased, so as to allocate appropriately attention for a given task or condition. The situation can become complex when it is considered that the activation of task representations or the setting of unit biases can dynamically be adjusted on the basis of participants' experience with a task. The mechanisms underlying these processes are an important focus for any theory that tries to explain how attention comes to be allocated. However, as we made clear in Cohen et al. (1990), this was not within the scope of our original model:

> Our focus in this article . . . is not on how task interpretation occurs or on how decisions concerning the allocation of attention are made. Rather, we are concerned with how information about the task and the corresponding allocation of attention influences processing in the pathways directly involved in performing the task itself. (p. 338)

Thus, our model was not intended to explain how task demand units or biases come to be chosen on the basis of participants' knowledge about or experience with the task. Rather, it was intended to show how, if such mechanisms are assumed, they can account for attentional effects. The critical principle of our model is that these mechanisms produce attentional effects by placing all task-relevant (i.e., attended) units in the most sensitive part of their dynamic range. We applied this principle to output units explicitly in our consideration of response set effects in Simulation 6 of Cohen et al. (1990):

> In the preceding simulations . . . [A]ttentional selection occurred at the level of the intermediate units, where information in the two pathways was still separate. However, the attention–allocation mechanism used in this model is a general one and can be applied to other levels of processing as well. In the following simulation, this mechanism was used to select a particular set of responses at the output level of the network. (p. 350)

Although our discussion did not take account of the situation in which there were asymmetric weights, as noted earlier this circumstance can be addressed by assuming that the mechanisms responsible for the allocation of attention provide activation to units in the response set sufficient to zero their net input and produce a resting activity level of 0.5. Again, this abides directly by the principles we used to implement attentional effects in the intermediate layer of the original model (as described on pp. 338–339, Cohen et al., 1990). Therefore, rather

than considering the deviation of output units from a resting level of 0.5 in the Kanne et al. (1998) simulations to be a fundamental problem with the model, we consider this to reflect the absence of mechanisms from their simulations that implement principles central to our theory.

As a preliminary exploration of how these factors may have affected the ability of Kanne et al.'s (1998) simulations to account for their data, we conducted a simulation that more fully implemented the principles just described. Our simulation was similar to the one they reported in their Footnote 3. We used the connection weights from their 4-4 model (Simulation 4). Although we have some concerns about how training influenced this simulation—concerns that we return to shortly—we wanted to know whether augmenting their simulation with the mechanisms discussed earlier would improve its ability to account for the empirical data. We made three modifications to their simulation.

First, we computed the value of the bias parameter that would produce a resting activation level of 0.5 for units at the output level representing items included in the stimulus and response set. We emphasize that these bias values reflect assumed influences of attention. The bias value differed between the word-reading and color-naming tasks because of the different connection weights from the intermediate units to the output units in the pathway relevant to each. Bias values also differed, within each task, as a function of set size. These variations between set size and condition are reasonable given that, in the empirical study, each participant was tested with only one set size and tasks were performed in a blocked fashion (see the Appendix in Kanne et al., 1998). Therefore, participants could have adjusted attention from block to block to optimize sensitivity of units relevant to the task being performed. Table 1 lists the biases for output units used for each task and set size.

Second, we increased the size of the task attentional effect by increasing the negative bias on the intermediate units (to −4.5) and by increasing the weights to these intermediate units from the task demand units (to 4.5). This was done to accommodate the larger network and the somewhat smaller influence of words on colors observed in the Kanne et al. (1998) data, as compared with the data from Dunbar and MacLeod (1984) simulated by our original model. This adjustment is similar to and was done for the same reasons as those described for Simulations 2 and 6 in Cohen et al. (1990).

Finally, units not in the response set were given biases so as to place their resting levels below the most sensitive point of their dynamic range, both at the output level and at the intermediate level. We did this at the output level in exactly the same way as in Simulation 6 of Cohen et al. (1990): Units not in the response set had their biases set to 0.1 below the value required to produce an activation value of 0.5. We also performed a similar manipulation at the intermediate unit layer by adding a negative bias (−0.5) to units not in the stimulus set (this value was arrived at by a limited parameter search).[1]

Table 1
Biases for Output Units at Each Set Size

Set size	Biases	
	Word reading	Color naming
2	1.630	1.010
3	1.955	1.205
4	2.280	1.399

Note. Biases are for output units in the stimulus–response set. Remaining units received a bias of 0.1 less than the value listed.

Table 2
Number of Cycles to Respond by Task, Set Size, and Condition

Task/set size	Condition		
	Congruent	Neutral	Incongruent
Word reading			
2	23.4	24.2	24.3
3	25.1	25.4	26.1
4	26.3	27.2	27.5
Color naming			
2	28.6	32.5	41.3
3	30.1	35.3	44.2
4	31.8	37.4	46.3

Table 2 shows the response times of the simulation for each condition of each task and at each set size. A regression of the model's performance for a set size of two against the corresponding empirical data reveals an r^2 of .96. Figure 1 shows a regression of the model's performance against the data for all three set sizes tested in Kanne et al.'s (1998) empirical study (comparable to their Figure 9). The r^2 for this regression was .91.

Figure 2 shows the mean response times by task and set size for both the empirical data and our simulation (comparable to Kanne et al.'s, 1998, Figures 8 and 7, respectively). The simulation produces a substantially better fit to the data than those reported by Kanne et al. In particular, it now shows that color-naming times increase rather than decrease with set size. These findings suggest that placing all task-relevant processing units—at both the intermediate and output levels—in the most sensitive part of their dynamic range can overcome at least some of the problems that Kanne et al. encountered.

Nevertheless, the simulation still exhibits some problematic effects. Although color-naming times increase with set size, this effect is not as large as in the empirical data, and word-reading times may increase too much (see Figure 2). Kanne et al. (1998) observed similar effects in their simulations and attributed this to the fact that the resting activation level of output units was lower than 0.5 in the larger networks. However, these effects persist in our simulations, even after the output units are placed in their most sensitive range, suggesting that Kanne et al.'s analysis of these effects may be incomplete. Their analysis focused on the nonlinearity of the activation function. However,

[1] Although we did not do this in our original model, we considered it to be justified in the current case because Kanne et al. (1998) manipulated both stimulus and response set sizes simultaneously (unlike the study simulated in Cohen et al., 1990, which manipulated only response set size).

there are also other factors that govern the dynamics of processing in the model, including the effects of the cascade (time-averaging) function and interactions between these and the activation function (i.e., the difference between the accumulated activation of the most active and next-most active units). A full analysis of these effects, and how they might relate to set size effects, is beyond the scope of the present discussion. We should note that there are also other problems with the behavior of our simulation. For example, as Kanne et al. pointed out, it produces a greater amount of facilitation in the congruent condition than is observed in their data, and this increases somewhat with set size.

These shortcomings may or may not reflect fundamental problems with our model. Determining this will require a more complete exploration than we or Kanne et al. (1998) have conducted. The purpose of our current simulation was simply to point out that incorporating mechanisms that implement central principles of our original theory improves the model's ability to account for the data. There are other implementational factors influencing the model's performance, factors that we identified in Cohen et al. (1990, Appendix, pp. 360–361) and used to fit the model to the empirical data that might also be explored profitably in a more complete evaluation of the model. These include the specific ratio of training of colors to words, the stopping criterion for training, and the size of attentional influences at the intermediate and output levels. Furthermore, as we noted in our original article (Cohen et al., 1990, pp. 360–361), there are also complex interdependencies among these parameters.

For example, consider the relative amounts of training given to words and colors. In Cohen et al. (1990, pp. 340, 360), we assumed that participants have had more experience with word reading than color naming. However, because there are no direct empirical data regarding this difference, the specific ratio of word-reading to color-naming training trials was treated as a free parameter, and a value was determined that produced the closest fit to empirical data concerning the relative speed of word reading versus color naming in the basic Stroop paradigm. This turned out to be 5:1 and was fixed across simulations of task variants. However, the ratio producing the best fit for a 2-2 network may not be best for larger networks. Furthermore, this parameter interacts with the strength of the attentional effects at both the intermediate and output levels. We did not extensively explore these parameters in our current simulation. It may be that the remaining disparities concerning set size effects reflect the fact that, for larger networks, a value of 5:1 is too small. A larger ratio may be needed to capture the finding that word reading is less affected by manipulations of set size than is color naming. Furthermore, we have not explored the effects of asymmetries in the number of colors and words on which the network is trained, which was one motivation for Kanne et al.'s (1998) studies. These too may interact with set size effects (e.g., training on a greater number of words than colors may help dilute set size effects for words).[2]

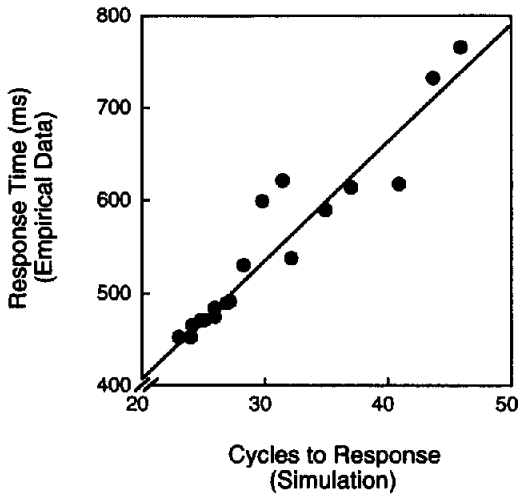

Figure 1. Regression plot of our simulation results and the empirical data reported by Kanne, Balota, Spieler, and Faust (1998). The equation for the regression is response time = 12.6 cycles + 154, r^2 = .91.

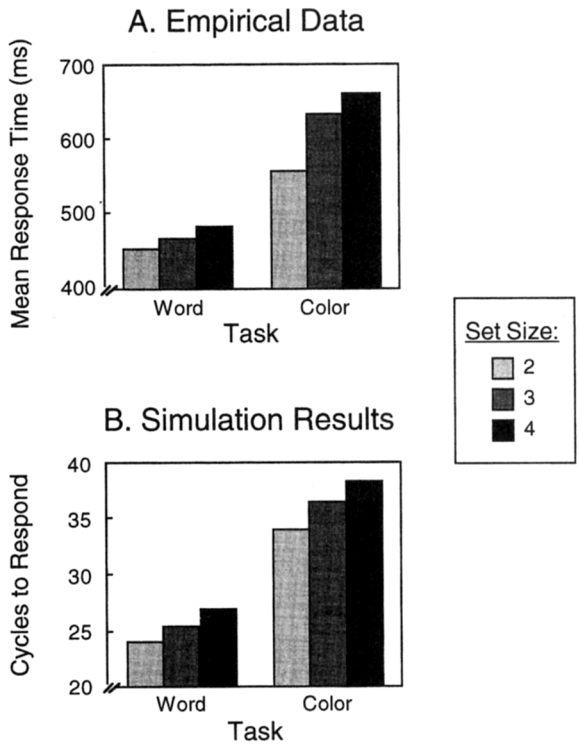

Figure 2. Influence of set size on response times for word reading and color naming. (A) Empirical data reported by Kanne, Balota, Spieler, and Faust (1998, Figure 8). (B) Results of our simulation. For both, values plotted are the means of the congruent, neutral, and incongruent conditions at each set size for each task.

[2] This raises an interesting methodological challenge—how best to determine the correct ratio when this depends on the size of the network, which itself is not precisely known for human subjects. However, this parameter dependency should not, by itself, be counted against the model. The relevant question is whether a reasonable training ratio (reflecting greater word-reading than color-naming experience, but not absurdly large) can be found for a network of a given size that produces the proper pattern of effects in the Stroop task and that can simulate performance across the range of task variants that has been studied.

Kanne et al. (1998) reported exploration of some of these implementational factors (e.g., training ratio, stopping criterion, and asymmetry in the number of colors and words). However, they did not explore others (e.g., the strength of the attention effects). More important, their simulations did not include attentional mechanisms that implement a central principle of our theory (i.e., that all units in an attended pathway rest in the most sensitive range of their activation function). Therefore, we believe it may be premature to reject the model on the basis of these studies.

Our concern that Kanne et al. (1998) did not implement mechanisms relevant to set size effects that we proposed in Simulation 6 of Cohen et al. (1990) might raise the reasonable question of why we did not use same mechanisms in Simulations 1–5 from our original study. Specifically, those simulations ignored issues of set size, implementing a 2-2 model of empirical studies that involved more than two stimuli in each dimension. The reason is that we chose the simplest model that could implement the principles we considered to be relevant to the phenomena of interest. We did this for the purposes of both analytic and descriptive simplicity. Because none of the empirical studies addressed by Simulations 1–5 manipulated set size, we did not consider this to be a relevant factor. Implementing a larger network in these simulations (to simulate a larger but fixed set size) would simply have required different but still constant values for the biases and attentional weights. Matters become more complicated, however, when set size is varied. Indeed, the purpose of Simulation 6 was to show that set effects can be accommodated by the model, that this requires some additional mechanisms (e.g., bias adjustments), but that these mechanisms are consistent with the central principles of our theory. Simulation 6 used a larger network, and addressed response set effects, but did not directly examine the effects of changing set size. Kanne et al. (1998) examined this directly and more thoroughly by manipulating stimulus as well as response set size. This highlights some of the complexities involved in attentional allocation when set size is varied. While acknowledging these complexities, our current simulation shows that by including the additional mechanisms proposed in Simulation 6, which are consistent with our original theory of attention, our model is better able to account for Kanne et al.'s data. At the same time, this account is not perfect. Therefore, it is worthwhile to consider whether there are fundamental limitations to the model as we originally proposed it. In fact, we believe this is the case. In the section that follows, we briefly review recent work that considers these limitations.

Set Size Effects and the Mechanisms of Inhibition

One feature of our model that we have long felt requires modification is the use of a strictly feed-forward processing architecture. Such an architecture suffers from several computational limitations. Indeed, early models such as those of Grossberg (1978, 1980) and McClelland and Rumelhart (1981) used between-layer excitatory connections and within-layer inhibitory connections. Although McClelland and Rumelhart allowed feed-forward inhibition, subsequent variants of the interactive activation model (McClelland, 1985, 1993; McClelland & Elman, 1986) eliminated such connections entirely.

Figure 3. A display that shows that three ambiguous characters can each constrain the identity of the others.

Between-layer inhibition creates problems in connectionist networks containing "localist" units—units that serve as detectors for specific cognitive entities such as letters, words, or concepts such as the color green, and so forth. The problem arises whenever partial activation of several alternatives is possible at a particular layer of a processing system: Feed-forward inhibition can prevent exploiting these partial activations at the next layer, even when they are highly constraining at that layer. To see the problem, one can consider a simple version of the interactive activation model of letter perception (McClelland & Rumelhart, 1981) with three positions so that words of three letters can be processed. Now suppose that an ambiguous stimulus like the one in Figure 3 is presented, activating two alternatives in each position: R or P in the first position, E or F in the second, and D or B in the fourth. For these inputs, there is only one word that fits one of the alternatives in each position. However, if activation of each alternative at a given set of units excites alternatives at the next level that it is consistent with, but also inhibits to the same extent alternatives at the next level that it is inconsistent with, there is no net excitation of any alternatives.

The problem is easily overcome by eliminating the bottom-up inhibition and by using recurrent or lateral inhibition instead. This is implemented by providing mutual inhibitory connections from each unit within a layer to every other unit in the same layer. In this case, a number of words will receive net bottom-up excitation, but only the unit for RED will be excited by one of the possibilities in all three positions. As shown in many simulations, lateral inhibitory influences will allow this best fitting alternative to win out, even when there are many competing alternatives. For these and other reasons, McClelland (1993) and Usher and McClelland (1995) have begun to develop a general alternative to the cascade model of McClelland (1979), which is a strictly feed-forward model of the dynamics of information processing. In this new model, between-level connections are only excitatory, and within-level connections are used to carry out a competitive, lateral inhibition process.

Given the earlier discussion, one may wonder why, in Cohen et al. (1990), we used feed-forward rather than recurrent inhibition. There are two reasons. First, the simplest form of our model focused on the case in which there were only two diametrically opposed alternatives under consideration. In this situation, "evidence" (represented by activation) supporting one alternative counts as equally strong evidence against the other alternative. Second, at the time we did our initial work, learning algorithms for connectionist networks were suitable for training strictly feed-forward networks, but not networks that exploit lateral or recurrent inhibition. We felt then, and still believe today, that

significant insight into the effects of gradual strengthening of connection weights could come from the analysis of the effects of learning in feed-forward networks.

Ultimately, however, we do aspire to a more fully adequate model that makes use of recurrent rather than feed-forward inhibition, and we have taken several steps in this direction. In an initial effort, we developed a simplified recurrent version of the Stroop model and showed that it could account for the basic Stroop effects (Cohen & Huston, 1994). In more recent work (Usher & Cohen, 1997), we have developed a more refined version of this model to address the set size effects reported by Kanne et al. (1998) as well as other concerns that have been raised about our original model (e.g., Mewhort, Braun, & Heathcote, 1992). Furthermore, we have used the same basic framework to account for performance in a variety of other attentional tasks, including the Eriksen flanker paradigm (Cohen, Servan-Schreiber, & McClelland, 1992), a spatial-cued reaction time task (Cohen, Romero, Farah, & Servan-Schreiber, 1994), and a variant of the continuous performance test (Braver, Cohen, & Servan-Schreiber, 1995). However, as Kanne et al. pointed out, an important remaining challenge for this work is to incorporate a learning mechanism that allows this new framework to provide the same integrated account of attentional and learning phenomena as our original model.

Word Reading Versus Color Naming

On the basis of the work just referred to, we remain strongly committed to the idea that connectionist models, especially those that include recurrent rather than feed-forward inhibition, can provide important insights into the mechanisms underlying attentional phenomena. Our account of performance in the Stroop task in which this form of model is used has the virtue not only of being explicitly mechanistic but also of being parsimonious. One of the central points made in Cohen et al. (1990, pp. 353–354) was that findings widely thought to reflect the operation of qualitatively distinct processes can in fact be accounted for by qualitatively identical mechanisms. We recognize that this is a strong claim, particularly with regard to tasks such as word reading and color naming that would seem to differ in important ways. We also recognize that parsimony is not an infallible guide to the truth. We suspect, much as Kanne et al. (1998) did, that there are indeed important differences between word reading and color naming. However, for all of the reasons just described, it may still be possible to account for these processes in terms of qualitatively identical processing pathways in a connectionist model. This possibility has not fully been explored, either by using the original Cohen et al. architecture or by using network architectures that incorporate recurrent inhibition.

This is the most literal interpretation of our claim: that Stroop effects "can be explained by differences in the strength of two processes that use qualitatively identical mechanisms" (Cohen et al., 1990, p. 353). However, even assuming this account fails, there is still an interesting and potentially important claim that could be made about the qualitative similarity of the processing pathways involved, at least with respect to the distinction between automatic and controlled processing. Suppose, for example, that to explain the difference in Stroop effects between color naming and word reading, an extra stage of processing is required for color naming (e.g., an extra layer of intermediate units, representing the "meaning pathway" for colors, as suggested by Kanne et al., 1998, in their General Discussion section). Although the pathways involved in each task would no longer be identical, the mechanisms underlying both, and used to account for their behavior in the Stroop task, would still be qualitatively very similar. That is, no unique or qualitatively distinct mechanisms would have been invoked for color naming. Indeed, this is just the sort of account that has been proposed by Phaf, Van der Heiden, and Hudson (1990). This contrasts with the traditional theory that color naming is controlled, whereas word reading is automatic and that this difference involves qualitatively distinct mechanisms.[3]

There are also other possible differences between word reading and color naming. For example, it has been proposed that word reading is a task that can exploit the quasi-regularity of the mapping between spelling and sound (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Plaut & Shallice, 1993; Seidenberg & McClelland, 1989), whereas color naming lacks this systematicity. Accordingly, the types of representations in each pathway may differ. Exactly how attentional influences of the sort we have proposed interact with such representational factors has yet to be explored. However, it seems possible that such interactions will not be relevant to, or at least sufficient to account for, the pattern of effects observed in the Stroop task, especially because such effects are observed even when other types of stimuli are used, such as pictures or spatial locations.

Perhaps the most important point about Stroop-like effects made by our original model is that relative strength of processing is the critical factor in accounting for such effects, even when there may be other differences between the pathways involved. Strength of processing is a continuous variable, and our model suggests that quantitative differences in this variable may be sufficient to account for Stroop effects, in contrast to more traditional theories that have interpreted such effects in terms of a qualitative distinction in processing. Thus, although we believe that there may well be important, and even qualitative differences in color naming and word reading, these differences may not be relevant to, and should not obscure, the explanation that our model offers for Stroop effects.

Finally, we would like to make it clear that we do not believe that performance in all tasks can be explained in terms of qualitatively similar mechanisms. For example, we believe that distinctly different mechanisms come into play when a task involves highly novel associations between stimuli and responses (see discussion of strategic processes and direct vs. indirect pathways in Cohen et al., 1990, pp. 347 and 354). Our point is simply that we do not believe that word reading and color naming differ in this particular way because both involve highly familiar sets of associations.

[3] Changing the number of layers in a given pathway might alter its susceptibility to control, but this is expected and explained by the principles used in our original model (see discussions on pp. 354–355 and, in particular, p. 357 of Cohen et al., 1990).

Conclusion

In summary, we find the data presented by Kanne et al. (1998) to be interesting and of significant value in calling into sharper focus the principles central to our original model and issues concerning their implementation. However, we differ in our conclusion about the ability of the principles we described in Cohen et al. (1990) to capture these data and find that augmenting the simulations reported by Kanne et al. with mechanisms that adhere to these principles results in a much improved fit to their data. The most important of these is the allocation of attention at all levels of the network that act to place task-relevant units at the most sensitive part of their activation function. Kanne et al. did not implement this principle at the output level of their network. At the same time, we recognize that the results of our simulations are not perfect. We point to additional implementational factors that could be explored (such as the specific ratio of training of colors to words, the stopping criterion for training, and the size of attentional influences at the intermediate and output levels) and suggest that a more thorough evaluation of these might reveal whether our original model can fully account for the phenomena in question. In addition, we identify a principle of the original model—feed-forward inhibition—that may face a fundamental limitation, and we discuss how our framework might be extended to address this through the use of recurrent inhibition.

Finally, Kanne et al.'s (1998) data highlight an important challenge that a more general elaboration of our theory faces, which concerns the mechanisms that determine the allocation of attention. Although such mechanisms are beyond the scope of our original model, they relate closely to the kinds of attentional effects that our model predicts and, therefore, warrant careful consideration in future work. We believe that the current exchange has brought these issues into sharper focus and that it demonstrates how the scientific process is well served by the close interaction between carefully designed experimental studies and efforts to capture the results of such studies in explicit simulation models.

References

Braver, T. S., Cohen, J. D., & Servan-Schreiber, D. (1995). A computational model of prefrontal cortex function. In D. S. Touretzky, G. Tesauro, & T. K. Leen (Eds.), *Advances in neural information processing systems* (Vol. 7, pp. 141–148). Cambridge, MA: MIT Press.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 1–19). Cambridge, MA: MIT Press.

Cohen, J. D., Romero, R. D., Farah, M. J., & Servan-Schreiber, D. (1994). Mechanisms of spatial attention: The relation of macrostructure to microstructure in parietal neglect. *Journal of Cognitive Neuroscience, 6,* 377–387.

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology, 105,* 239–269.

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.

Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. *Progress in Theoretical Biology, 5,* 233–374.

Grossberg, S. (1980). How does the brain build cognitive code? *Psychological Review, 87,* 1–51.

Kanne, S. M., Balota, D. A., Spieler, D. H., & Faust, M. E. (1998). Explorations of Cohen, Dunbar, and McClelland's (1990) connectionist model of Stroop performance. *Psychological Review, 105,* 174–187.

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

McClelland, J. L. (1985). Putting knowledge in its place: A scheme for programming parallel processing structures on the fly. *Cognitive Science, 9,* 113–146.

McClelland, J. L. (1993). Toward a theory of information processing in graded, random, and interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655–688). Cambridge, MA: MIT Press.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

Mewhort, D. J. K., Braun, J. G., & Heathcote, A. (1992). Response-time distributions and the Stroop task: A test of Cohen, Dunbar, and McClelland's (1990) parallel distributed processing model. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 872–882.

Phaf, R. H., Van der Heiden, A. H. C., & Hudson, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22,* 273–341.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377–500.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Usher, M., & Cohen, J. D. (1997). *A connectionist model of the Stroop task revisited: Reaction time distributions and differential effects on facilitation and interference captured by a single set of mechanisms.* Manuscript in preparation.

Usher, M., & McClelland, J. L. (1995). *On the time course of perceptual choice: A model based on principles of neural computation* (Tech. Rep. No. PDP.CNS.95.5), Department of Psychology, Carnegie Mellon University.

Received March 4, 1997
Revision received September 23, 1997
Accepted September 24, 1997 ■