Journal of Experimental Psychology:
Human Perception and Performance
1992, Vol. 18, No. 3, 872-882

0096-1523/92/$3.00

OBSERVATION

Response Time Distributions and the Stroop Task: A Test of the Cohen, Dunbar, and McClelland (1990) Model

D. J. K. Mewhort, J. G. Braun, and Andrew Heathcote
Queen's University at Kingston, Kingston, Ontario, Canada

Cohen, Dunbar, and McClelland's (1990) model was tested for Strooplike interference tasks by studying the shape of the distribution of response latencies produced by Ss and by the model. The model correctly anticipates changes in mean response latency (M_{RT}) across congruent and incongruent conditions. It does not, however, correctly anticipate changes in the shape of the distributions, even though changes in the shape of the distributions underlie the changes in M_{RT}. Thus, the model predicts M_{RT} successfully but for the wrong reason. It is concluded that the model is not an adequate account of Ss' performance in the Stroop task.

Suppose that you are shown a series of character strings, each printed in either red or green, and that you are asked to indicate the color of the print. Your average response time will be longer when the character string spells a conflicting color word than when it spells a noncolor word or is composed of Xs. The interference implied by the difference in mean response time for the conflicting and neutral conditions is widely known as the *Stroop effect* (Stroop, 1935), and a comprehensive review of its literature has been presented by MacLeod (1991).

How does a color word interfere with subjects' ability to name the color of its print? Cohen, Dunbar, and McClelland (1990) proposed a connectionist model to answer the question. Their account assumes (a) that evidence from all sources of information accumulates in parallel for both potential responses, (b) that activation passes through the system on pathways that vary in strength and not speed, (c) that attention modulates the flow of activation through the system, and finally (d) that subjects have more experience reading words than naming colors. When subjects are asked to name the color of print and the print spells a conflicting color word, evidence for both potential responses accumulates in parallel. As a result, evidence about the word's name—evidence passed through a strong pathway—interferes with the evidence about the color of the print—evidence passed through a weaker pathway. In this way, interference reflects the subjects' relative experience with the two tasks.

Cohen et al. (1990) showed that the model is able to predict changes in mean response latency (M_{RT}) in several variations of the Stroop task (e.g., Dunbar & MacLeod, 1984). In the present study, we test the model by comparing the distributions of response latency generated by the model against distributions produced by subjects. To foreshadow our argument, we show that the model does not correctly anticipate changes in the shape of the latency distributions, even though the changes in shape underlie the changes in M_{RT}. Thus, we contend that the model predicts M_{RT} successfully but for the wrong reason, and we conclude that it does not account for subjects' performance in the Stroop task.

The Cohen, Dunbar, and McClelland Model

The Model's Architecture

The model consists of two evidence accumulators controlled by a 12-node encoder network. The 12 nodes are arranged in three layers: 6 nodes in the input (stimulus) layer, 4 nodes in the middle (hidden) layer, and 2 nodes in the output (response) layer. The six input nodes represent two colors of print (red and green), two color words (RED and GREEN), and two instructions to respond to the color of the print or to the word's meaning. The response nodes tally the activation associated with each response (red and green). Activation flows forward from the input layer to the hidden layer and from the hidden layer to the response layer; there are no connections from one layer back to an earlier layer. Finally, there are no connections within a given layer, and no connections skip a layer.

This work was supported by an operating grant from the Natural Science and Engineering Research Council of Canada (NSERC) to D. J. K. Mewhort, by an NSERC postgraduate fellowship to J. G. Braun, and by a Commonwealth Doctoral Fellowship to Andrew Heathcote.

We thank E. E. Johns for a careful reading of the manuscript and Jonathan Cohen for help in defining the parameters used by Cohen, Dunbar, and McClelland (1990).

Andrew Heathcote is now at the Department of Psychology, University of Newcastle, Newcastle, Australia.

Correspondence concerning this article should be addressed to D. J. K. Mewhort, Department of Psychology, Queen's University, Kingston, Ontario, Canada K7L 3N6. Electronic mail may be addressed to Mewhortd@qucdn.queensu.ca (bitnet) or to doug@vip.psyc.queensu.ca (internet).

Deriving the Weights

Connections among the nodes vary in strength, and the strengths are represented by weights. Cohen et al. (1990) used a back-propagation algorithm to establish appropriate weights by training the network to indicate the correct response to a color word or to the color of the print (see Rumelhart, Hinton, & Williams, 1986). To build stronger connections for reading color words than for naming the color of the print, they gave the network 10 times more training trials in reading color words than in naming the color of the print.

After training, the weights effectively split the network into two subsystems, a reading subsystem and a color-naming subsystem. In the color-naming subsystem, the node that represents an instruction to report the color of the print, and the two color nodes feed activation to two of the four hidden-layer nodes. In the reading subsystem, the two color-name nodes and the corresponding instruction node feed activation to the remaining two hidden-layer nodes (see Cohen et al., 1990, Figure 3).

Running the Model

To simulate a trial with the model, the network is first given instructions. Next, it is given a series of simulated test stimuli. The network responds to each test stimulus in turn.

Instructions. In a Stroop experiment, subjects are required to report either the color word or the color of the print. The corresponding instructions are represented in the model by setting the activation for the appropriate instruction node to 1 (the other stays at 0) and by cycling the network through a few iterations to alter the resting activation of the hidden and response nodes. After a dozen iterations, the network reaches a stable state and is then ready for the test stimuli.

The fact that the model acknowledges a role for instructions is one of its strong points. Experimenters routinely give human subjects instructions, but too few models acknowledge that instructions permit subjects to set the state of internal processing mechanisms in preparation for the task (cf. Mewhort, 1987).

Stimuli. Test stimuli in the Stroop task are created by combining one of three possible colors (e.g., red, green, or black) with one of three possible character strings (e.g., RED, GREEN, or a series of Xs), subject to the constraint that each stimulus must include at least one red or green attribute. For a congruent display, the color word and the color of the print are the same; for an incongruent display, the color of the print requires one response, whereas the color word requires the other. Finally, for a neutral display, the stimulus is a series of colored Xs or a color word printed in black (a neutral color).

The test stimuli are simulated in the model by setting the activation of the appropriate input nodes to 1. Assuming that the network has been given instructions to simulate a red congruent display, for example, the activation of both the red-word node and of the red-print node is set to 1. The activation of each of the other input nodes—except the current instruction node—is set to 0; the activation of the current instruction node (i.e., the node that represents the task) remains at 1. Once set, the activation of the input nodes remains constant throughout the trial.

Processing and decision. At the start of each test trial, activation from the input nodes feeds to the hidden nodes and then to the response nodes. Changes in the activation at the hidden and response nodes are achieved in a series of steps or cycles and depend on the running average of the node's previous input and its current input (cf. McClelland, 1979).

Evidence for both potential responses is derived on each cycle from the difference in activation between the two response nodes. The evidence is tallied in evidence accumulators, and when enough evidence has accumulated to trigger one of the two responses, the trial is complete. Response latency is a linear function of the number of cycles needed to trigger the response.

To simulate the variability of human performance, Gaussian noise is added independently whenever the activation of each hidden- and response-layer node is changed and whenever evidence for one response or the other is added to the corresponding evidence accumulator. Adding noise admits the possibility of response error and introduces considerable uncertainty about when the network will respond.

Summary. At the start of each simulated trial, the network is given an instruction and permitted to modify its internal state in response to the instruction. The evidence accumulators are set to 0, and inputs representing the appropriate stimuli and instructions activate the corresponding input nodes. On each cycle of the trial, activation from the input nodes feeds forward, and evidence for the two responses is tallied in the response accumulators. When evidence in one of the two accumulators reaches threshold, the trial is complete.

Generality of the Model

To facilitate the exposition, we have cast the model in terms of color versus word interference. In fact, the model provides a general framework with which to account for performance in Strooplike tasks. In the color versus color-word example, the six input nodes represent two colors of print (red and green), two color words (RED and GREEN), and two instructions (one to report the name of the color and one to report the word).

For other tasks, the input and response nodes represent different kinds of information. For example, consider the conflict between local and global information demonstrated by Navon (1977). He used small letters to construct large letters and found that when subjects were asked to report the small letter embedded in the large letter, the large letter interfered with subjects' report of the small letter in much the same way a color word interferes with subjects' report of the color of print.

To apply the model to Navon's (1977) stimuli, the six input nodes must be relabeled so that the nodes represent the two letters at the global level, the same letters at the local level, and instructions to respond to information either at the global level or at the local level. In addition, the asymmetric weights must be mapped to acknowledge the asymmetric character of the two tasks: A color word interferes with naming the color of print but not vice versa, and global information interferes with local information but not vice versa. As before, congruent displays require the same response to information at both levels, and incongruent displays require conflicting responses; for neutral displays, subjects are required to respond to information at one of the two levels, and no relevant information is provided at the other.

Performance of the Model

To study the model, we implemented it in a computer program. With the posttraining weights and the biases provided by Cohen et al. (1990, Figure 3), we simulated 10,000 trials for each of the congruency conditions (i.e., the congruent, neutral, and incongruent conditions) at each of 135 combinations of the model's parameters. In each case, we set the task so that the nontask information would dominate, that is, color naming in the standard Stroop task or reading the small letters in Navon's (1977) version of the task. For each simulation, we generated a frequency distribution of the cycles to respond for correct responses, and then we characterized each of the distributions in terms of the ex-Gaussian distribution (see Heathcote, Popiel, & Mewhort, 1991).

The ex-Gaussian distribution is the convolution of a normal and an exponential distribution; it is skewed to the right and has three parameters, μ, σ, and τ. The first two parameters (μ and σ) describe the mean and standard deviation of the normal component of the distribution; μ reflects the leading edge and mode of the ex-Gaussian distribution. The third parameter (τ) is derived from the exponential distribution and is a measure of skew. The mean of the distribution equals $\mu + \tau$. The ex-Gaussian distribution provides a good fit to latency distributions (see Hockley, 1984; Ratcliff & Murdock, 1976); we found that it also provides a good fit to the simulated cycles-to-respond distributions (examples are shown in Figures 1 and 2).

The 135 conditions tested in our simulations were derived from the factorial combination of 4 parameters: (a) the standard deviation of the Gaussian noise added to the evidence accumulators (σ_D), with values of 0.005, [0.01], 0.05, 0.1, and 0.15; (b) the standard deviation of the Gaussian noise added to the activation when updating each hidden and response node (σ_P), with values of [0.1], 0.5, and 1.0; (c) the cascade rate (called τ by Cohen et al., 1990), with values of 0.05, [0.1], 0.2; (d) the rate of information accumulation (called α by Cohen et al.), with values of 0.05, [0.1], and 0.2. The values in brackets are the values used by Cohen et al.[1]

Table 1 summarizes our simulations: For each simulated distribution, the table shows the mean number of cycles to respond, the percentage of error responses, the standard deviation of the distribution, the maximum frequency of the distribution (f_{max}), the median of the distribution, and the three parameters of the ex-Gaussian distribution (μ, σ, τ) fitted by maximum likelihood estimation.

The 15 sets of simulations summarized in Table 1 are a subset of the 135 that we examined. We selected the subset of 15 because it illustrates the fundamental characteristics of the model's performance. Except as noted, the simulations in the table used the same parameter values reported by Cohen et al. (1990); in particular, for all the simulations in the table, the cascade rate was 0.1, and the rate of information accumulation was 0.1.

As is clear in Table 1, the means for the congruent, neutral, and incongruent conditions were remarkably stable across the 15 combinations of noise parameters at approximately 30.5, 35.9, and 49.7, respectively. Clearly, within each congruency condition, neither processing noise nor decision noise had much systematic effect on the mean number of cycles to respond. The noise parameters did affect performance, however.

First, they controlled the accuracy predicted by the model. The model was error-free when σ_D was 0.05 or less. As σ_D increased, the errors increased.

Second, they controlled the variance and the shape of the distribution of cycles to respond. When the noise parameters were small, the distributions were very narrow and almost symmetrical. For example, when $\sigma_D = 0.01$ and $\sigma_P = 0.5$ (the parameters Cohen et al., 1990, used), the *SD* was less than 10% of the mean, and the most frequent cycle to respond in the congruent case represented 26.4% of the whole distribution. As the noise parameters increased, however, the variance increased, and the distributions became more skewed. One can see the amount of skew by examining τ or the difference between the mean and the median.

Third, for all combinations of parameters, τ increased monotonically across the congruency conditions (i.e., from the congruent to the neutral condition and from the neutral to the incongruent condition). The behavior of μ depended on the σ_D: When σ_D was larger than about 0.1, μ decreased monotonically across the congruency conditions, but when σ_D was smaller than 0.1, μ increased monotonically across the same conditions. Thus, the relative contribution of μ and τ to changes in the mean depended on σ_D. Nevertheless, across the congruency conditions, the variance always increased from the congruent to the neutral conditions and from the neutral to the incongruent conditions.

[1] There is an error in the Cohen, Dunbar, and McClelland (1990) article concerning the parameter values. When describing how noise is added to the evidence accumulators, they state that "the amount added is random and normally distributed, with mean μ based on the output of the network, and with fixed standard deviation σ. The mean is proportional to the difference between the activation of the corresponding unit and the activation of the most active alternative: "$\mu_i = \alpha(\text{act}_i - \text{max\_act}_{j\neq i})$" (p. 338). In addition, "throughout [their] simulations, the value of α was 0.1, the value of σ was 0.1, and the value of the [evidence] threshold was 1.0" (p. 338). With two responses, the means μ_1 and μ_2 are $\mu_1 = \alpha(a_1 - a_2)$ and $\mu_2 = \alpha(a_2 - a_1)$, where a_1 refers to the activation of one response node and a_2 refers to the activation of the other. Given the description in the article, the amount added on each cycle to each of the two evidence accumulators should be $[\mu_1 + N(0, \sigma)]$ and $[\mu_2 + N(0, \sigma)]$, respectively, where $N(0, \sigma)$ is a random sample from a normal distribution with mean 0 and standard deviation σ. In fact, however, Cohen, Dunbar, and McClelland (1990) weighted the noise as well as the difference in activation by α so that the effective standard deviation of the generator was $\alpha \times \sigma$: with $\alpha = 0.1$, the true value of σ_D was 0.01 (J. D. Cohen, personal communication, June 12, 1991).

Table 1
Percentage Errors and Measures on the Distribution of Cycles Produced by the Model as a Function of Decision Noise, Processing Noise, and Congruency Condition

	Measure							
Noise parameter	*M*	% error	*SD*	f_{max}	Median	μ	σ	τ
$\sigma_D = 0.15$, $\sigma_P = 1.0$								
Congruent	30.5	4.38	14.6	295	28.0	16.8	6.9	13.8
Neutral	36.3	6.16	21.2	244	32.0	14.9	6.2	21.5
Incongruent	50.1	10.34	39.2	189	38.0	11.3	4.2	38.8
$\sigma_D = 0.15$, $\sigma_P = 0.5$								
Congruent	30.3	4.22	14.8	310	28.0	16.3	6.7	14.0
Neutral	35.7	5.66	19.9	240	32.0	15.8	6.8	19.8
Incongruent	49.4	9.63	38.4	181	38.0	11.7	4.6	37.7
$\sigma_D = 0.15$, $\sigma_P = 0.1$								
Congruent	30.6	4.57	14.6	330	28.0	17.1	7.2	13.5
Neutral	36.1	5.38	20.2	267	32.0	16.2	7.0	19.9
Incongruent	49.5	9.94	38.2	181	38.0	11.4	4.1	38.1
$\sigma_D = 0.1$, $\sigma_P = 1.0$								
Congruent	30.6	0.18	10.0	440	30.0	23.0	6.6	7.7
Neutral	36.5	0.48	14.1	343	34.0	24.0	7.6	12.4
Incongruent	51.4	1.50	28.4	211	45.0	22.4	7.2	29.0
$\sigma_D = 0.1$, $\sigma_P = 0.5$								
Congruent	30.5	0.15	9.7	449	30.0	23.4	6.7	7.2
Neutral	35.8	0.40	13.4	353	34.0	24.5	7.7	11.4
Incongruent	50.5	1.12	27.3	230	44.0	23.0	7.6	27.5
$\sigma_D = 0.1$, $\sigma_P = 0.1$								
Congruent	30.6	0.17	9.8	447	30.0	23.2	6.6	7.4
Neutral	36.1	0.36	13.3	361	34.0	24.8	7.7	11.4
Incongruent	49.6	1.01	26.0	227	43.0	23.6	7.7	26.0
$\sigma_D = 0.05$, $\sigma_P = 1.0$								
Congruent	30.7	0.00	5.4	766	30.0	27.6	4.4	3.1
Neutral	36.3	0.00	7.7	574	36.0	31.0	5.6	5.3
Incongruent	51.2	0.00	16.5	301	48.0	36.1	7.8	15.1
$\sigma_D = 0.05$, $\sigma_P = 0.5$								
Congruent	30.5	0.00	5.0	816	30.0	27.9	4.2	2.6
Neutral	35.9	0.00	6.8	617	35.0	31.5	5.3	4.4
Incongruent	49.6	0.00	14.0	335	48.0	37.4	7.9	12.2
$\sigma_D = 0.05$, $\sigma_P = 0.1$								
Congruent	30.5	0.00	4.9	863	30.0	27.6	4.0	2.9
Neutral	35.9	0.00	6.8	621	35.0	31.5	5.2	4.4
Incongruent	49.0	0.00	13.0	259	47.0	38.5	8.2	10.5
$\sigma_D = 0.01$, $\sigma_P = 1.0$								
Congruent	30.5	0.00	2.5	1580	30.0	29.1	2.1	1.4
Neutral	36.0	0.00	3.8	1106	36.0	33.6	2.9	2.4
Incongruent	50.4	0.00	9.1	519	49.0	43.0	5.5	7.5
$\sigma_D = 0.01$, $\sigma_P = 0.5$								
Congruent	30.3	0.00	1.5	2639	30.0	29.7	1.4	0.6
Neutral	35.6	0.00	2.2	1844	36.0	34.7	1.9	0.9
Incongruent	48.8	0.00	4.9	857	48.0	45.6	3.8	3.1
$\sigma_D = 0.01$, $\sigma_P = 0.1$								
Congruent	30.3	0.00	1.0	3690	30.0	29.9	1.0	0.4
Neutral	35.5	0.00	1.4	2689	35.0	34.9	1.3	0.6
Incongruent	48.2	0.00	2.7	1430	48.0	46.9	2.4	1.3
$\sigma_D = 0.005$, $\sigma_P = 1.0$								
Congruent	30.5	0.00	2.3	1712	30.0	29.2	2.0	1.3
Neutral	36.1	0.00	3.5	1151	36.0	33.8	2.8	2.3
Incongruent	50.5	0.00	8.8	521	49.0	43.2	5.3	7.4
$\sigma_D = 0.005$, $\sigma_P = 0.5$								
Congruent	30.3	0.00	1.3	3072	30.0	29.8	1.2	0.6
Neutral	35.6	0.00	1.8	2105	36.0	34.8	1.7	0.8
Incongruent	48.7	0.00	4.3	936	48.0	46.0	3.4	2.7
$\sigma_D = 0.005$, $\sigma_P = 0.1$								
Congruent	30.2	0.00	0.6	6027	30.0	30.1	0.6	0.2
Neutral	35.5	0.00	0.8	4234	35.0	35.2	0.8	0.3
Incongruent	48.2	0.00	1.5	2588	48.0	47.6	1.4	0.6

Across the congruency conditions, there was also a modest interaction of the two noise parameters. Decreasing σ_P narrowed the distribution of cycles to respond and shifted its parameters. The effect of changing σ_P, however, was masked by decision noise, except when σ_D was very small.

Although the interaction is not shown in Table 1, the noise parameters also interacted with both the cascade rate and the rate of evidence accumulation. For example, increasing the latter parameter reduced errors, but the same manipulation also increased the relative contribution of τ to changes in the mean across the congruency conditions. To anticipate the argument to follow, changes in the cascade rate and the rate of evidence accumulation did not help the model to fit the empirical data; in particular, the monotonic change in μ and τ across the congruency conditions was not disturbed by changes in either parameter.

Figures 1 and 2 show the distribution of cycles to respond across the three congruency conditions for two of the combinations of noise parameters illustrated in Table 1. Both figures also show the ex-Gaussian distribution fitted by maximum likelihood estimation. Figure 1 shows the distribution when the decision noise was relatively small: $\sigma_D = 0.01$ and $\sigma_P = 0.5$. The noise parameters are the same values used by Cohen et al. (1990) in their simulations. As noted earlier, most of

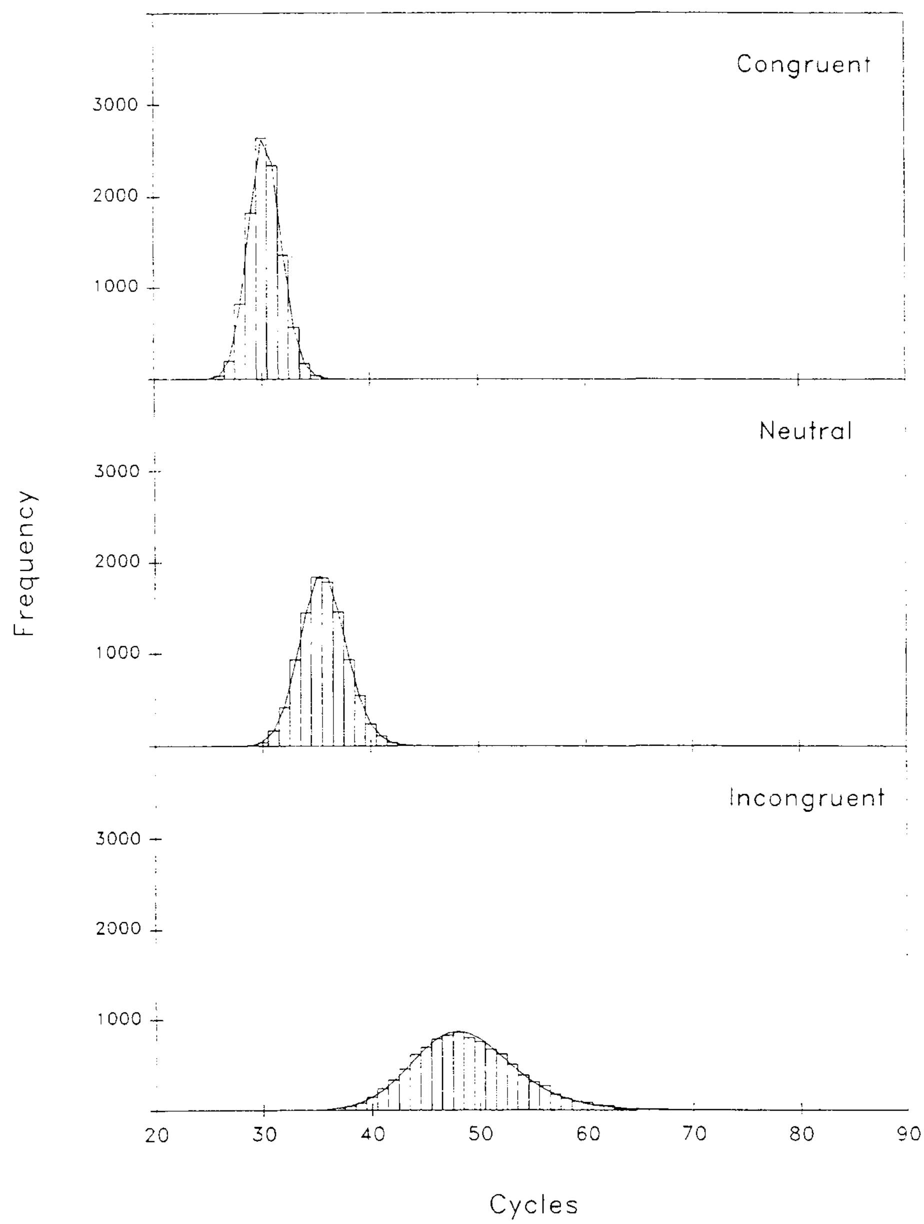

Figure 1. Frequency histograms of simulated cycles to respond as a function of congruency condition. (The solid curve shows the ex-Gaussian distribution fitted to the raw data by maximum likelihood estimation. Parameters: $\sigma_D = 0.01$ and $\sigma_P = 0.5$.)

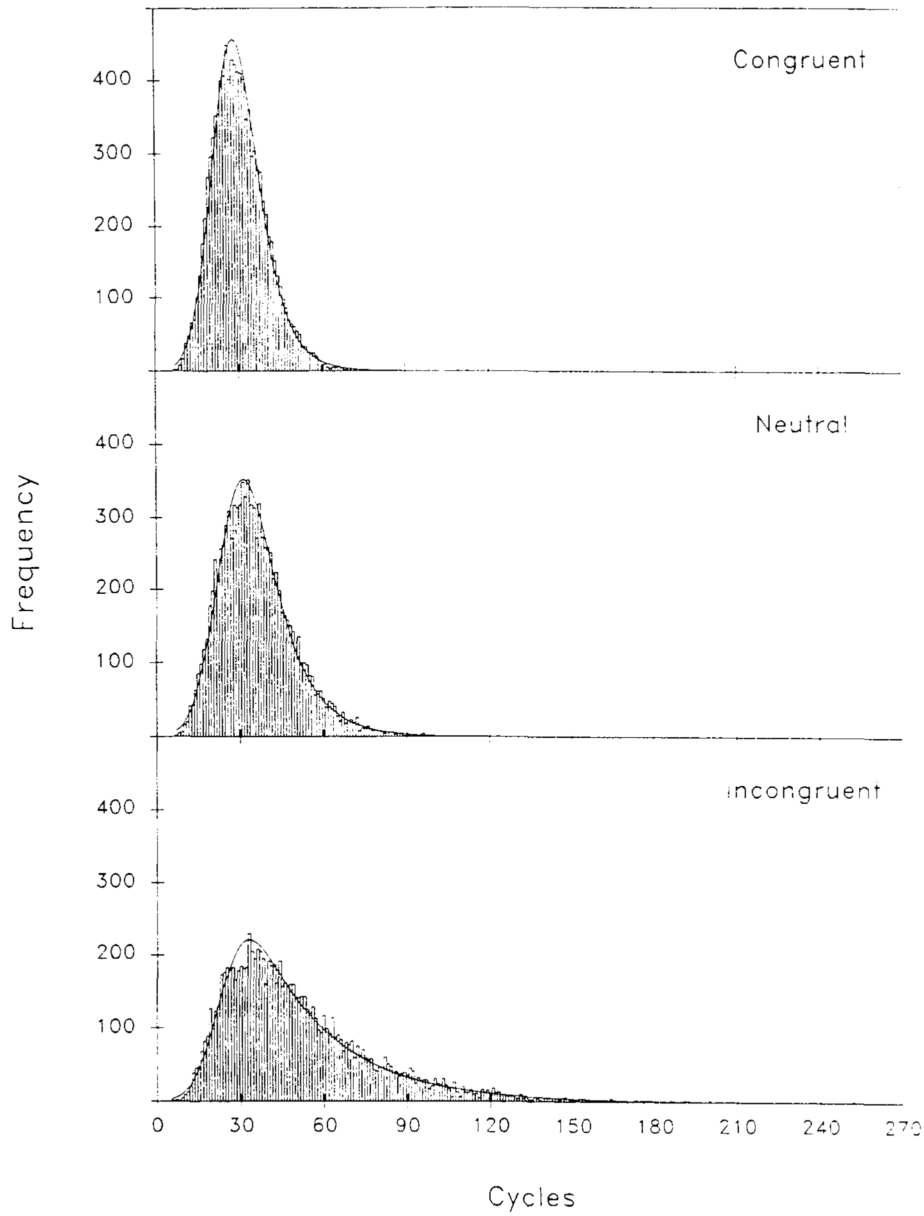

Figure 2. Frequency histograms of simulated cycles to respond as a function of congruency condition. (The solid curve shows the ex-Gaussian distribution fitted to the raw data by maximum likelihood estimation. Parameters: $\sigma_D = 0.1$ and $\sigma_P = 0.5$.)

the shift in the mean across the congruency conditions reflects an increase in μ. Figure 2 shows the distributions when the decision noise was relatively large: $\sigma_D = 0.1$ and $\sigma_P = 0.5$, the values at which the model starts to predict errors. Here, in contrast to the pattern in Figure 1, most of the shift in the mean across congruency conditions reflects an increase in τ.

Tests of the Model

Color-Word Interference

Although a monotonic shift in both μ and τ across the congruency conditions is characteristic of the model's behavior, it is not consistent with subjects' behavior. Heathcote et al. (1991) provided one illustration of the point. They conducted a Stroop experiment with the character strings BLUE and GREEN printed in blue or green. The neutral condition involved a row of colored Xs. They characterized the latency distributions in terms of the ex-Gaussian distribution, and the results showed that in contrast to the Cohen et al. (1990) account, changes in M_{RT} were often associated with a tradeoff in μ and τ. Table 2 summarizes the Heathcote et al. data.

As is shown in Table 2, responses in the incongruent condition were, on average, 116 ms slower than in the neutral condition. Analysis of the shape of the distribution of response times showed that the interference effect (i.e., the difference

Table 2
Summary of the Heathcote, Popiel, and Mewhort (1991) Data (in Milliseconds) for Each Congruency Condition

Congruency condition	M_{RT}	SD_{RT}	μ	σ	τ
Empirical results					
Congruent	624	136	497	55	127
Neutral	617	117	524	64	92
Incongruent	733	179	596	112	136
Simulated results: $\sigma_D = 0.1$, $\sigma_P = 0.5$[a]					
Congruent	608	58	565	40	43
Neutral	639	80	572	46	68
Incongruent	727	163	563	45	164
$\sigma_D = 0.01$, $\sigma_P = 0.5$[b]					
Congruent	607	10	603	9	4
Neutral	641	14	635	12	6
Incongruent	726	32	706	24	20

[a] Simulated response time = 5.97 (cycles) + 425.43. [b] Simulated response time = 6.45 (cycles) + 411.54.

between the incongruent and neutral conditions) reflected an increase in μ of 72 ms plus an increase in τ of 44 ms (see Heathcote et al., 1991, for details of the procedure and of the statistical analysis). On average, responses in the congruent condition were not significantly faster than in the neutral condition. In fact, there was a negative advantage of 7 ms for congruency; analysis of the shape of the latency distributions showed that the negative advantage reflected a decrease in μ of 27 ms accompanied by an increase in τ of 35 ms.

Note that the change in shape of the latency distribution suggests why a congruency effect is less frequently noted than the corresponding interference effect (see MacLeod, 1991). Most experimenters report M_{RT}, and that measure averages μ algebraically with τ. A change in the shape of the distribution can push μ and τ in opposite directions, and because the mean equals the sum of μ and τ, an advantage (decrease) in μ can be masked by an increase in τ.

Table 2 also shows data from the model fit to Heathcote et al.'s (1991) results by using a linear equation to convert from cycles to respond to milliseconds. We fit the data twice, once with $\sigma_D = 0.01$ and $\sigma_P = 0.5$ and once with $\sigma_D = 0.1$ and $\sigma_P = 0.5$, the same combinations of noise parameters illustrated in Figures 1 and 2. We used the first combination because it had been used by Cohen et al. (1990); we used the second because it yielded an error rate close to that in the Heathcote et al. study (1.7%).

We used the same procedure as Cohen et al. (1990) to derive the linear equation: We minimized the sum across congruency conditions of the squared deviation of the simulated mean and empirical mean to obtain optimal parameters for the slope and intercept of the equation. The equations are given in Table 2.

Next, we generated distributions of simulated response times for each congruency condition. Specifically, we applied the appropriate linear equation to each score in the corresponding distribution of cycles to respond. Finally, to consider the shape of the distribution of simulated data in relation to the empirical results, we characterized the simulated distributions in terms of the ex-Gaussian distribution.

The means derived with the two sets of noise parameters were very similar, and the simulated means fit the data moderately well. In particular, they matched the large interference effect associated with the incongruent condition. The simulated mean did not match the empirical mean in the congruent condition, however: The model predicted a small congruent advantage in the mean that did not appear in the empirical results.

Although the predicted means were similar for both sets of noise parameters, the variance of the distribution that was based on $\sigma_D = 0.01$ and $\sigma_P = 0.5$ (the parameters used by Cohen et al., 1990) was much too small, a point clear in *SD* in relation to the corresponding empirical value. The *SD* predicted with $\sigma_D = 0.1$ for the incongruent case was close to the empirical value, but the corresponding measures for the other conditions were much too small and were ordered incorrectly.

The predicted means matched the empirical means moderately well, but the model did not correctly anticipate the shape of the distributions on which the means were based. For example, the model predicted a congruency advantage in the mean on the basis of an advantage in either μ or τ, depending on σ_D. The empirical congruency effect (i.e., the difference between the neutral and congruent conditions) involved a decrease in μ accompanied by an increase in τ, changes in the shape of the distribution that canceled when measured in mean response time. No combination of the model's parameters provides a way of shifting μ to the left while shifting τ to the right (with a large σ_D, however, the model can predict the reverse pattern!). The model always predicts monotonically increasing values for τ across the congruent, neutral, and incongruent conditions, and although the model can predict a wide variety of shapes for the distribution of response times, it cannot accommodate the empirical differences in shape across conditions.

The model can explain the interference effect in the means, but it does not correctly predict the shape of the latency distributions. In particular, because it predicts the wrong shape for the response time distributions across the congruency conditions, it predicts a congruency advantage in the mean that is not found in the data.

Local–Global Interference

To extend the evidence provided by Heathcote et al. (1991), we studied a second illustration of Strooplike performance. Instead of conflict between the color of print and a color word, we studied conflict between local and global information obtained from the same stimulus display.

Method

Subjects. Six undergraduates at Queen's University (Kingston, Ontario, Canada) served as subjects. Participation permitted the subjects to earn bonus marks in an introductory psychology course.

Apparatus and stimuli. The stimuli were presented on a Tektronix 608 display monitor driven by a high-speed plotter (see Finley, 1985, for a description of the plotter). The plotter was controlled by a Zenith Z-241 computer; the computer recorded responses and calculated the latency for each response (see Heathcote, 1988, for the timing algorithm).

The stimuli, generated by brightening dots on the display monitor, involved letters of two sizes. Small uppercase letters (*H* and *Z*) were created by brightening the appropriate dots in a matrix of 17 rows and 13 columns. The dot matrix subtended a visual angle of about 0.18° × 0.14°. Large examples of the same letters were created by placing small letters at the appropriate positions in a letter matrix of seven rows and five columns. The matrix subtended a visual angle of about 1.65° × 0.72°.

Congruent displays involved a large letter constructed by using the same small letter (i.e., H constructed with small Hs and Z constructed with small Zs); incongruent displays involved a large letter constructed by using the other small letter (i.e., H constructed with small Zs and Z constructed with small Hs). In addition to the large letters, two neutral displays were created by placing the small letters in a diamond: One neutral display was composed of small Hs, and the other was composed of small Zs.

Procedure. At the start of each trial, a small fixation cross appeared in the center of the display monitor. Subjects pushed a button to continue the trial. When the subject pushed the start button, the fixation cross was replaced, after an interval of 200 ms, by a target stimulus. The target was centered on the monitor and consisted of a set of small letters (either Hs or Zs) arranged to form a large H, a large Z, or a diamond. The subjects were required to identify the small letter.

Subjects responded by pressing one of two buttons. Subjects 1, 3, and 5 used the index finger of their dominant hand to indicate H and the index finger of the nondominant hand to indicate Z. Subjects 2, 4, and 6 used the index finger of their dominant hand to indicate Z and the index finger of their nondominant hand to indicate H.

If the subject pressed the wrong button (or did not respond within 5 s), the response was considered to be an error. Response latency was calculated from the onset of the target, and the target remained on the display monitor until the subject responded (or until the trial was declared to be an error).

Error trials were replaced later in the session. In addition, to eliminate slow posterror responses, the trials following an error trial were ignored until the subject made a correct response; data were taken following that response (see Rabbitt & Rogers, 1977).

Before starting the experiment proper, subjects received 20 practice trials to introduce them to the task and to the stimuli. When the subjects had completed half of the experimental trials, they were given a short rest. No feedback on performance was provided.

Design. Each subject received 480 trials arranged in 8 blocks of 60 trials. Within each block, the subject received 20 congruent displays, 20 incongruent displays, and 20 neutral displays. For half of the trials of each display type, the displays were constructed with small Zs; the remaining displays were constructed with small Hs. The order in which the displays were shown was randomized independently within each block. Error trials and posterror dummy trials were replaced within the same block.

Results

About 2.1% of the congruent trials, 2.4% of the neutral trials, and 6.7% of the incongruent trials were in error.

Each subject completed 160 trials for each congruency condition, and for each subject the ex-Gaussian distribution was fitted to the resulting distributions through maximum likelihood estimation. Table 3 summarizes the results; the table presents M_{RT}, SD_{RT}, and the three ex-Gaussian parameters for each subject and congruency condition.

Figure 3 presents a graphical summary of the empirical latency distributions. The figure was derived through Vincent averaging, a technique that permits one to obtain the average shape of a distribution across subjects (see Ratcliff, 1979). For each congruency condition, each subject's distribution was divided into 16 quantiles, with each quantile holding 6.25% of the cases. The mean latency within each quantile was calculated, and the mean across subjects was calculated for corresponding quantiles. Figure 3 presents equal-area histograms that are based on the Vincent averages. The figure also shows the ex-Gaussian distribution fitted to the average

Table 3
Results (in Milliseconds) for Each Congruency Condition and Subject

	Congruency condition														
	Congruent					Neutral					Incongruent				
Type of result	M_{RT}	SD_{RT}	μ	σ	τ	M_{RT}	SD_{RT}	μ	σ	τ	M_{RT}	SD_{RT}	μ	σ	τ
Empirical															
Subject 1	597	154	465	40	132	603	129	481	48	121	625	103	542	56	83
Subject 2	432	91	347	45	85	449	101	371	56	78	488	100	438	86	50
Subject 3	523	114	425	37	99	517	89	433	29	84	586	117	482	43	103
Subject 4	562	146	424	49	138	569	147	444	66	125	675	167	527	55	148
Subject 5	505	95	409	34	95	531	101	435	48	96	608	92	534	56	73
Subject 6	444	127	343	36	101	463	123	361	37	102	518	100	435	54	83
M	510	121	402	40	108	522	115	421	47	101	583	113	493	58	90
Simulated															
$\sigma_D = 0.01$[a]	506	6	504	6	3	528	9	524	8	4	582	20	569	15	13
$\sigma_D = 0.10$[b]	507	36	480	25	27	527	50	488	29	43	582	103	479	29	103
$\sigma_D = 0.15$[c]	506	59	451	27	55	528	79	449	27	78	582	152	433	18	149

[a] Simulated response time = 4.08 (cycles) + 382.52. [b] Simulated response time = 3.76 (cycles) + 392.05. [c] Simulated response time = 3.95 (cycles) + 386.57.

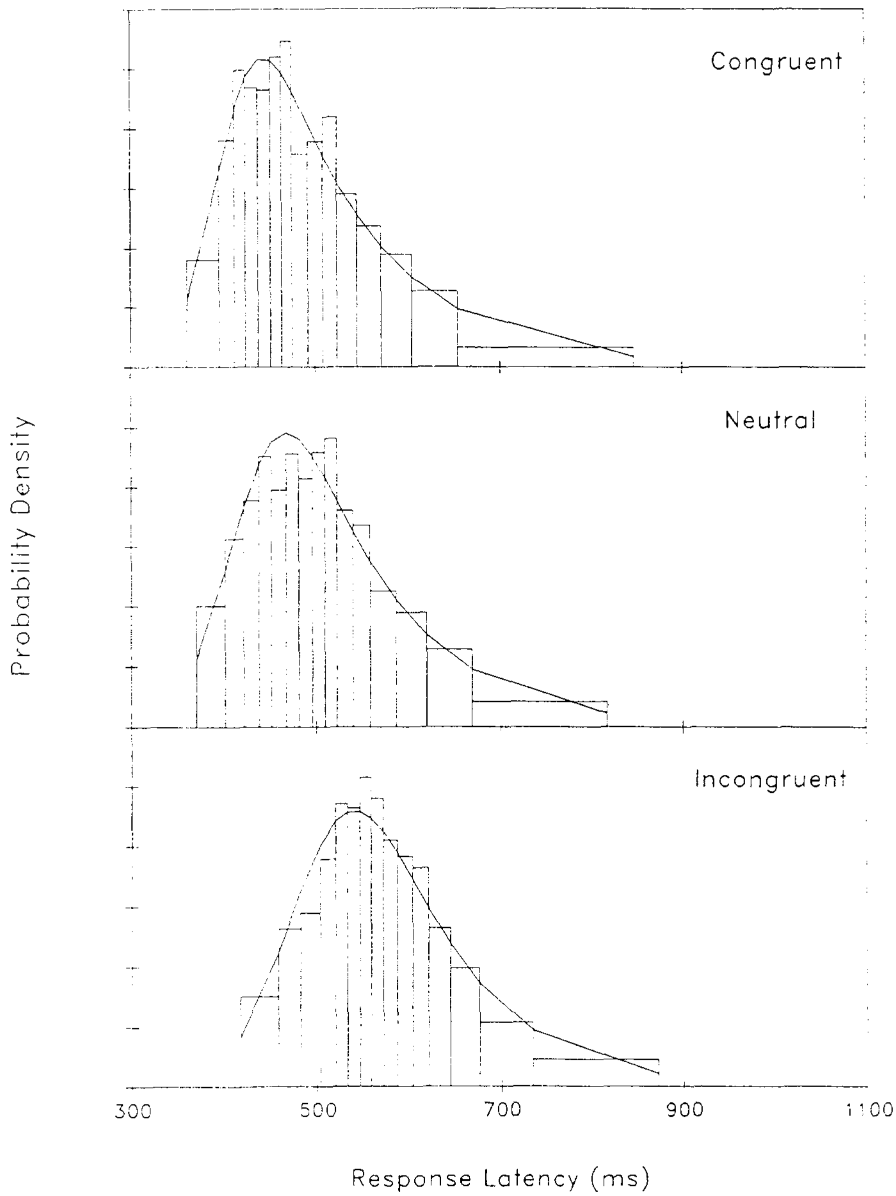

Figure 3. Probability density obtained by Vincent averaging across subjects as a function of congruency condition. (The solid curve shows the ex-Gaussian distribution fitted by maximum likelihood estimation.)

curve.[2] As is clear in the figure, the ex-Gaussian distribution provides a good fit to the empirical latency distributions.

Analysis of M_{RT}. As Table 3 shows, in relation to the neutral condition, M_{RT} increased by about 61 ms in the incongruent condition, $F(1, 5) = 26.04$, $p < .01$. In contrast, mean reaction time decreased by about 11 ms in the congruent condition, $F(1, 5) = 5.79$, $.05 < p < .10$. The data are similar to those reported by Heathcote et al. (1991). In particular, note that the M_{RT} data exhibit strong interference in the incongruent condition and weak facilitation in the congruent condition.

M_{RT} is a summary of the data. When the underlying distribution is symmetrical, the mean is an adequate measure. When the underlying distribution is skewed, however, the mean is ambiguous and is therefore not an adequate measure

[2] The ex-Gaussian parameters obtained by fitting the Vincent average shown in the figure are close, but not identical, to the parameters obtained by averaging the parameters in Table 3. The difference reflects inaccuracy in Vincent averaging introduced by forcing the data into 16 quantiles.

by itself. The ambiguity arises because two distributions can generate the same mean even if they take different shapes and are caused by different underlying processes. Hence, to understand performance when the underlying distribution is skewed, one must examine the shape of the distribution as well as its mean.

Interference, facilitation, and the shape of the latency distributions. The 61-ms interference effect in M_{RT} was generated by a 72-ms increase in μ, $F(1, 5) = 102.94$, $p < .001$, accompanied by a modest 11-ms decrease in τ, $F(1, 5) = 1.09$, $p > .30$. Thus, in contrast to Heathcote et al. (1991), most of the interference reflects an increase in μ.

Next, consider the weak (11-ms) facilitation effect in M_{RT}. The effect reflects a 19-ms decrease in μ, $F(1, 5) = 52.98$, $p < .01$, accompanied by an 8-ms increase in τ, $F(1, 5) = 6.87$, $p < .05$. Again, as in the Heathcote et al. (1991) study, the congruency advantage in μ was offset by a loss in τ. Although it is modest when measured in M_{RT}, note that the congruency effect is not a weak effect when considered on a subject-by-subject basis: All 6 subjects showed the decrease in μ, and 4 of the 6 showed the increase in τ. The congruency effect is weak when measured in terms of M_{RT}, because M_{RT} averages μ and τ.

We do not contend, however, that the change in shape associated with the congruent condition reflects two independent factors that cancel each other out, although we leave that possibility open. The shape of the latency distribution is different across the congruency conditions. We have used two parameters of the ex-Gaussian distribution to characterize the change in shape, but it is not clear which factors will be needed to explain the difference. Our point is methodological: Because the change of shape is not reflected in M_{RT}, the mean can be a misleading measure for response time distributions.

Predictions from the model. To apply the model to our task, we assume that global information interferes with local information in much the same way reading interferes with naming the color of print. In terms of the model, our assumption is that global information is carried on stronger pathways than local information; it is parallel to Cohen et al.'s (1990) assumption that information about color names is carried on stronger pathways than information about the color of print.

We fit the model to the new data by applying the same procedures as before, and as before we characterized the resulting distributions in terms of the ex-Gaussian distribution. To fit the model, we used results from the same combinations of noise parameters as before (i.e., the cases illustrated in Figures 1 and 2) and from a third combination selected to fit the error data from the experiment. Table 3 also shows the predicted means, the *SD*s, and the corresponding ex-Gaussian parameters. The linear equations are presented as well.

Comparison with the model. As before, the means predicted with the three sets of noise parameters were very similar, and the simulated means fit the data very well: They matched both the large interference effect associated with the incongruent condition and the modest facilitation effect associated with the congruent condition. Because there was an advantage for congruency in the local–global means, the model provided a better fit to the local–global experiment than to the color-word experiment.

Although the predicted means were similar for all sets of noise parameters, the variance of the distribution that was based on $\sigma_D = 0.01$ and $\sigma_P = 0.5$ (the values used by Cohen et al., 1990) was much too small, a point clear in the *SD*s in relation to the corresponding empirical values. Moreover, performance with Cohen et al.'s parameters was error-free. When the decision noise was increased, the variance of the simulated distributions increased. (Of course, variance can be adjusted by altering the slope parameter when mapping cycles to milliseconds.)

With $\sigma_D = 0.15$ and $\sigma_P = 0.5$, the model predicted the correct pattern and roughly the right number of errors as well as the correct mean latency. The *SD*s and, more important, the pattern of *SD*s across the congruency conditions did not match the empirical results. Whereas the empirical values were roughly the same across conditions, the model predicted an increase from the congruent to the neutral condition and from the neutral to the incongruent condition.

The reason is clear: The model does not correctly anticipate changes in the shape of the distributions across conditions. For example, when $\sigma_D = 0.01$ and $\sigma_P = 0.5$, the model predicted that both the congruency advantage and the interference effect would reflect a change in μ. By contrast, when $\sigma_D = 0.15$ and $\sigma_P = 0.5$—parameters that provide a good fit to the error data—the model assigned both the congruency advantage and the interference effect to changes in τ. The empirical interference effect, however, reflected an increase in μ, whereas the empirical congruency effect involved a decrease in μ offset by an increase in τ.

Note that *SD* for the empirical distributions was roughly constant across conditions; the model can hold the *SD* constant, but only with very small values of σ_D. Such values yield unrealistically narrow distributions and unrealistically good accuracy. When the noise parameters are adjusted to yield realistic variance and accuracy, the model yields unrealistic response time distributions: It changes mean response time by shifting the tail of the distribution.

The model can predict a wide variety of shapes for the distribution of response times, shapes that are based on a shift in the whole distribution or a shift in the tail. When it is constrained to match accuracy, it cannot accommodate the empirical differences in shape across conditions.

Discussion

Our argument may be summarized as follows: The Cohen et al. (1990) model is able to account for the M_{RT} data quite well. The mean, however, is an ambiguous score: Two distributions can generate the same mean even if they take different shapes and are caused by different underlying processes. Close examination of the distribution predicted by the model suggests that it does not account for subjects' performance. The model assigns changes in mean across the congruency conditions to monotonic changes in either μ or τ, depending on the decision noise. Subjects alter the shape of the distribution in more complicated ways. Thus, although the model can account for mean performance, it does so for the wrong reasons, and we conclude that the model is not an adequate account of subjects' behavior.

We find many aspects of the model attractive. In particular, we like the idea that information is carried on pathways that vary in strength. The main difficulty with the model is that its decision component is too separate from the processing component. The thrust of our empirical work is that processing is intimately associated with the shape of the response time distribution. Yet in the model, we could obtain the full range of distribution shapes by manipulating the decision noise alone. We think that processing and decision are more intimately linked than the model allows.

To illustrate our point, consider the shape of the distributions in the Heathcote et al. (1991) experiment. The neutral condition involved a series of colored Xs and yielded the most symmetrical of the three distributions. One possible reason is that Xs are very easy to discriminate from color words. Given that subjects can discriminate the neutral stimulus early enough, we presume that they can alter processing to take advantage of the discriminability. To fix the model, we need a modification that will allow discriminability factors to influence the decision at an early point in processing.

Finally, although we have pointed repeatedly to difficulties in using the mean with skewed distributions, we do not wish to banish it from the response time literature. Instead, our point is that the mean can mislead when the underlying distributions change shape. We recommend checking the shape of the distributions before calculating the mean: If the shape is stable across conditions, the mean is an adequate measure. If the shape changes across conditions, the mean is dangerous and should be supported by an analysis of the change in shape. A null difference in the means is particularly dangerous because distributions of widely different shape can yield the same mean value. Hence, an experimental manipulation may have a large effect on the shape of the distribution without affecting its mean. Likewise, because an interaction involves a comparison of two differences, its interpretation is ambiguous unless the shapes of underlying distributions remain constant.

References

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.

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.

Finley, G. (1985). A high-speed plotter for vision research. *Vision Research, 25,* 1993–1997.

Heathcote, A. (1988). Screen control and timing routines for the IBM microcomputer family using a high-level language. *Behavior Research Methods, Instruments, and Computers, 20,* 289–297.

Heathcote, A., Popiel, S., & Mewhort, D. J. K. (1991). Analysis of response-time distributions: An example using the Stroop task. *Psychological Bulletin, 109,* 340–347.

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 598–615.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

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

Mewhort, D. J. K. (1987). Information stores and mechanisms: Early stages in visual processing. In H. Heuer & A. Sanders (Eds.), *Perspectives on perception and action* (pp. 335–357). New York: Erlbaum.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383.

Rabbitt, P., & Rogers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology, 29,* 727–743.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461.

Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83,* 190–214.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

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

Received March 28, 1991
Revision received October 23, 1991
Accepted October 25, 1991 ■