Research Report

STROOP-LIKE EFFECTS FOR MONKEYS AND HUMANS: Processing Speed or Strength of Association?

David A Washburn
Georgia State University

Abstract—Stroop-like effects have been found using a variety of paradigms and subject groups. In the present investigation, 6 rhesus monkeys (Macaca mulatta) and 28 humans exhibited Stroop-like interference and facilitation in a relative-numerousness task. Monkeys, like humans, processed the meanings of the numerical symbols automatically—despite the fact that these meanings were irrelevant to task performance. These data also afforded direct comparison of interpretations of the Stroop effect in terms of processing speed versus association strength. These findings were consistent with parallel-processing models of Stroop-like interference proposed elsewhere, but not with processing-speed accounts posited frequently to explain the effect.

The Stroop effect (Stroop, 1935) is among the most popular and robust phenomena in cognitive psychology. MacLeod (1991) noted that more than 700 Stroop-related articles are now in the literature, with numerous variations on Stroop’s original paradigm having now been documented. In his impressive review and organization of more than 400 of these studies, MacLeod delineated 18 major empirical results and described the effects of numerous variables on Stroop-like interference (e.g., age, sex). The literature review, which summarized more than 50 years of published research, did not, however, mention Stroop-like effects with nonhuman animals. MacLeod did note the particular importance of acoustic articulatory factors in Stroop-like interference, the semantic nature of many Stroop-like effects, and the relative advantage of oral versus manual responding for producing Stroop-like interference. Each of these observations mitigates against finding Stroop-like interference with nonhuman and nonlinguistic animals.

Given recent interest in the numerical competencies of animals, however, a numerical variant of the Stroop task would seem to be appropriate for use with nonhuman species. Windes (1968) described a numerical variation of the Stroop task in which subjects were required to count and report the number of numerals presented in arrays. As in the color-word paradigm, the number of numerals could either be congruent with the identity of the numerals (e.g., five 5s) or incongruent with the identity of the numerals (e.g., five 3s), either case, the subject was to ignore the identity of the numerals and report only the number of numerals. Like Stroop (1935), Windes reported interference in the incongruous condition (see also Howard, 1983).

The present investigation was designed to determine whether Stroop-like effects could be observed in the processing of numerical stimuli by rhesus monkeys.

GENERAL METHOD

Subjects

Human subjects (N = 28) volunteered to be tested in this experiment. The subjects (age range 18–35 years, 18 females) participated in exchange for course credit. Prior to testing, each subject received instructions and practice with the test system.

Six rhesus monkeys (Macaca mulatta, age range 4–11 years, 1 female) were also tested in this investigation. The monkeys were tested in their single-animal home cages with continuous access to water. No animal was reduced in body weight for purposes of this testing.

The extensive training and test histories of these animals have been discussed elsewhere (Rumbaugh, Richard-son, Washburn, Savage-Rumbaugh, & Hopkins, 1989; Washburn, in press). Prior to the present experiment, all the monkeys had participated in a series of experiments on numerical competence, during which they had learned to select the larger of two or more numbers in order to receive a corresponding quantity of rewards (Washburn & Rumbaugh, 1991, 1994). From these data, Rumbaugh and I concluded that the monkeys had learned to respond to the numerals on the basis of the relative—or perhaps, absolute—number of pellets each symbol represented.

Apparatus

All subjects were tested with the Language Research Center’s Computerized Test System (Rumbaugh et al., 1989), an automated, 386-based test apparatus in which software tasks are used for training and assessment. Subjects responded to computer-graphic stimuli by manipulating a joystick. Movements of the joystick were reflected on the color screen as movements of a 1 25-cm × 1 25-cm cursor (“+”) in a direction isomorphic to the angle of joystick displacement. Contact between the cursor and any other stimulus was recorded as a response. Sound feedback generated by the computer was presented via the computer’s internal speaker or through an external speaker-amplifier. For the monkeys, correct responding was followed by delivery of a 97-mg fruit-flavored pellet (Noyes) using a Gerbrands 5120 pellet dispenser with a computer interface (Keithley/Metabyte PIO-12 and ERA-01).

The software written for this experiment was inspired by the version of the Stroop task described by Windes (1968) and other investigators. At the start of each trial, a 1 25-cm circle was presented in the middle of the screen, or with the cursor located in a random position outside the circle. The subject was required to manipulate the joystick so as to bring...
Stroop Effects by Monkeys and Humans

Fig 1 A monkey reaches through his cage to manipulate the joystick so as to bring the cursor into contact with one of two arrays of stimuli. On this trial, the subject should touch the seven 0s rather than the three 2s.

the cursor into the circle to initiate a trial. Subsequently, two arrays of letters were presented on the screen, one array to the right of the cursor and the other to the left of the cursor (see Fig 1). Each array consisted of one to seven letters, either As, Bs, Cs, or Ds. Any left or right movement of the joystick caused the cursor to jump immediately to the corresponding array of letters. Note that, unlike Winde's task, subjects were not required to count the stimuli in the arrays. Rather, subjects were rewarded (with a tone and, for the monkeys, with pellets) for choosing the array with the most elements irrespective of the identity of the elements.

Twenty percent of the trials were probe trials, in which numerals (0, 1, 2, or 3), rather than letters, were presented in both arrays. Half of these trials were congruent trials, in which the larger array of numerals was composed of the larger of the two numbers (e.g., five 3s vs two 0s). On incongruent trials, the larger array was composed of the smaller of the two numbers (e.g., seven 1s vs six 2s). As on the baseline trials (in which letters were used), the subject was rewarded for selecting the larger array, and the specific identity of the items composing the arrays was irrelevant to responding. Selection accuracy and response time (RT) were recorded on each trial.

**Procedure**

Because no monkey had ever experienced a task in which it was required to select the more numerous of two arrays, each monkey initially received 500 trials of the task using only the baseline condition. That is, the monkeys were trained to respond to the more numerous array using only letter stimuli. Subsequently, each rhesus monkey was tested on 2,000 additional trials (20 blocks of 100 trials each), of which an average of 200 trials were congruent and 200 trials were incongruent probe trials.

Human volunteers were instructed to select as quickly and as accurately as possible the array of stimuli that contained the most items, and that the identity of the particular items was irrelevant to the correct response. After 3 practice trials, each human subject completed 100 trials, of which an average of 10 trials were congruent and 10 trials were incongruent probe trials.

**STROOP EFFECTS**

**Results**

The rhesus monkeys learned the new task quickly, with mean accuracy of 67% observed in the initial 100 trials (z = 3.40, p < .05). Accuracy levels for the entire 500-trial training session averaged 75%.

For the test data, a separate analysis of variance of the RT data was conducted for each species using probe condition (baseline, congruent, incongruent) as a within-subjects manipulation. Sig-

---

**Table 1 Mean response times and accuracy levels as a function of species and condition**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Accuracy (%)</th>
<th>Response time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>Rhesus monkeys (N = 6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent numerals</td>
<td>92</td>
<td>3</td>
</tr>
<tr>
<td>Baseline (letters)</td>
<td>86</td>
<td>4</td>
</tr>
<tr>
<td>Incongruent numerals</td>
<td>73</td>
<td>8</td>
</tr>
<tr>
<td>Human subjects (N = 28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent numerals</td>
<td>99</td>
<td>1</td>
</tr>
<tr>
<td>Baseline (letters)</td>
<td>99</td>
<td>1</td>
</tr>
<tr>
<td>Incongruent numerals</td>
<td>97</td>
<td>1</td>
</tr>
</tbody>
</table>

1 Had the task required counting and reporting the number of digits, then trials like seven 3s and four 2s would be considered incongruent. Because the task required only selecting the larger array, such a trial is congruent.
Discriminant Stroop-like effects were found in the data from human subjects, $F(2, 54) = 25.56, p < 0.01$ RTs from incongruent trials were significantly longer than those from baseline or congruent trials, and responses in congruent trials were significantly faster than those in the baseline condition ($HSD = 26$ ms, see Table 1). Similar findings were observed in the rhesus monkeys data. RTs were significantly longest in the incongruent condition and significantly shortest in the congruent condition, $F(2, 10) = 27.74, p < 0.01$ ($HSD = 57$ ms). A similar pattern of results was seen in the accuracy data for rhesus monkeys, $F(2, 10) = 8.58, p < 0.01$, but only in the incongruent condition ($HSD = 11$%), no accuracy effect was obtained with humans.

**Discussion**

Rhesus monkeys and human alike were susceptible in their relative-numerosness judgments to the influence of numerical meanings. The pattern of results reported here is comparable to that which has been observed in hundreds of other Stroop-effect experiments. Relative to baseline, performance was significantly compromised by incongruent trials, and was significantly facilitated by congruent trials. Note also that the degree of facilitation was slightly less than the degree of interference, as is typically seen in Stroop-related research.

Notwithstanding these similarities, the present investigation is unique in that a nonlinguistic, nonhuman species exhibits results that are virtually indistinguishable from human responding (at least in the chronometric data). Of course, basic response competition is nothing new in either the cognitive or comparative literatures. What makes the classic Stroop effect special among instances of response competition is that color-word meanings, although irrelevant to the response demands of the task, are automatically processed and compete favorably against color naming. Knowledge about color-word meanings is presumably built upon knowledge about colors themselves, yet the semantic dimension influences processing of the more primitive color knowledge. Similarly, knowledge about numeral meanings is founded on the capacity to discriminate numbers of things. However, performance on the basic relative-numerosness task was influenced significantly by the representations associated with numeric symbols—for both species. The present findings are termed "Stroop-like" only in deference to the fact that Stroop's color-word task was not used, because the automatic processing of symbol meanings appears to be comparable across test paradigms and species.

**SPEED OF PROCESSING OR STRENGTH OF ASSOCIATION?**

The Stroop effect is typically interpreted in terms of differential processing speed for competing stimulus dimensions (see MacLeod, 1991, for a detailed discussion of theoretical accounts). The meaning of a color word (or the meaning of an Arabic numeral, in the present experiment) is processed automatically and quickly, whereas naming the color of ink (or determining the number of the items) is processed more slowly. When subjects are required to respond on the basis of the slowly processed attributes, competition from the highly practiced dimensions results in interference or facilitation. Given this interpretation, the magnitude of the Stroop effect should be sensitive to variations in processing speed on the automatic dimension (e.g., color-word or numeral meanings). The faster these meanings can be processed, the larger the facilitation and interference should be. In contrast, increases in the processing speed of the other dimension (ink color or number of items in the arrays) should result in less interference and facilitation.

---

**Table 2 Hypotheses from the speed-of-processing and strength-of-association accounts for variations in symbolic distance and array size**

<table>
<thead>
<tr>
<th>Type of difference</th>
<th>Speed of processing</th>
<th>Strength of association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbolic difference between numerals presented (e.g., 4 vs 2)</td>
<td>Bigger absolute differences between numerals result in faster processing of the automatic dimension, producing greater interference and facilitation (and smaller absolute differences produce less interference and facilitation)</td>
<td>Bigger absolute differences between numerals exaggerate congruence or increase facilitation and interference (and smaller differences decrease facilitation and interference)</td>
</tr>
<tr>
<td>Differences in sizes of arrays (e.g., five stimuli vs four stimuli)</td>
<td>Bigger absolute differences between arrays result in faster processing of the attended dimension, producing reduced interference and facilitation (and smaller absolute differences produce increased interference and facilitation)</td>
<td>The absolute difference between arrays does not affect the congruence or incongruence, or the magnitude of facilitation or interference</td>
</tr>
</tbody>
</table>

---

2 Note that Washburn and Rumbaugh (1994) reported the mean RT for selecting the greater of two digits to be 573 ms, or about 22% faster than the numerosness-judgment times recorded in the present baseline condition.
Stroop Effects by Monkeys and Humans

Such variations in the speed of processing each dimension can be examined in the present data. Symbolic distance effects have been observed for speeded judgments between numerals by humans and by monkeys (Moyer & Landauer, 1967, Washburn & Rumbaugh, 1994). Therefore, it is reasonable to expect more rapid processing of the numeral-meaning dimension when the digits are disparate (e.g., 0s vs 4s) than when they are similar (e.g., 2s vs 3s). A similar prediction can be made for disparate arrays of stimuli (e.g., five numerals vs two numerals will be judged faster than five numerals vs four).

In contrast to this speed-of-processing view, MacLeod (1991) has reviewed several parallel-processing accounts of the Stroop effect that are based on strength of association. According to this interpretation, each response is determined by the accretion of evidence from various sources. The relative numerosness of the two arrays suggests which should be selected as largest, however, numeral meanings within the arrays also contribute to the decision. Each source of information is weighted by strength of association. On incongruent trials, there is disagreement regarding the response, on congruent trials, concordance between the sources of information results in facilitation.

Both the speed-of-processing and the strength-of-association accounts predict exaggerated interference and facilitation with large symbolic distances, albeit for different reasons (see Table 2). That is, both generate the hypothesis that seven 3s versus two 6s will produce more interference than seven 3s versus two 4s. However, different predictions are generated for the manipulation of array-size differences in the relative-numerosity task. According to the speed-of-processing interpretation, reducing the processing time for numerosity judgments (e.g., trials with seven vs. two stimuli compared with seven vs six stimuli) should generally attenuate interference and facilitation. The strength-of-association view, however, would predict no effect of differences in the ratio of numerals in the two arrays. The important factor for these parallel-processing, strength-of-association views is the degree of congruity or incongruity, and these degrees are unaltered by array sizes. That is, seven 2s versus two 3s is just as incongruous in the present task as seven 2s and six 3s.

Results

RTs were analyzed as a function of condition (baseline, congruent, incongruent) and symbolic distance (the absolute difference between the values of the numerals in the arrays). For human subjects, a significant condition-by-difference effect was observed, $F(1, 27) = 30.72, p < .01$, and $F(1, 5) = 12.78, p < .01$, respectively. Judgments were made more rapidly with disparate (e.g., seven vs. two stimuli) arrays than with comparable arrays (e.g., seven vs. six stimuli). However, the variable did not interact with condition for either species ($p > .05$), as can be seen in Figure 2. The amount of facilitation was approximately equal across differences in array size (mean facilitation was 29 ms for humans, 61 ms for monkeys). Similarly, interference was comparable across differences in array size (mean interference was 50 ms for humans, 90 ms for monkeys).

Table 3 Mean response times and standard deviations (in milliseconds) as a function of species, condition, and difference in array size

<table>
<thead>
<tr>
<th>Condition</th>
<th>Difference in array size</th>
<th>1-3</th>
<th>4-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhesus monkeys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent numerals</td>
<td>Mean: 686, SD: 26</td>
<td>645</td>
<td>59</td>
</tr>
<tr>
<td>Baseline (letters)</td>
<td>Mean: 749, SD: 56</td>
<td>704</td>
<td>53</td>
</tr>
<tr>
<td>Incongruent numerals</td>
<td>Mean: 851, SD: 41</td>
<td>781</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Human subjects</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent numerals</td>
<td>Mean: 598, SD: 57</td>
<td>558</td>
<td>52</td>
</tr>
<tr>
<td>Baseline (letters)</td>
<td>Mean: 623, SD: 63</td>
<td>591</td>
<td>68</td>
</tr>
<tr>
<td>Incongruent numerals</td>
<td>Mean: 670, SD: 62</td>
<td>643</td>
<td>59</td>
</tr>
</tbody>
</table>

Discussion

As predicted by both theoretical accounts, Stroop-like interference and facilitation were increased as the symbolic distance between numbers increased. However, analysis of the effects of array-size difference in the relative-numerosity task failed to confirm the speed-of-processing hypothesis. This explanation had predicted that array differences of four to seven numerals would...
produce less interference and facilitation than differences of one to three numerals. Rather, inhibition and facilitation were approximately equal across levels of disparity between arrays of numerals. Consequently, a strength-of-association account, like those reviewed favorably by MacLeod (1991), was supported by these data. It appears that this is the first direct evidence for parallel-processing, strength-of-association versus sequential, speed-of-processing interpretations of the Stroop effect.

SUMMARY

In this experiment, Stroop-like effects have been demonstrated for humans and monkeys on a relative-numerousness task, and these effects have been shown to conform best to parallel, strength-of-association accounts of the Stroop interference and facilitation. Contemporary notions of semantic interference and priming would seem to require broadening to include these Stroop-like effects from a nonlinguistic, nonhuman species. Whereas it remains clear that auditory and articulatory factors modulate the Stroop effect, Stroop-like interference and facilitation cannot depend on such “human” specializations as reading, language, oral responding, or even counting. The monkeys in this experiment cannot be said to possess any of those abilities. They do, however, possess the elaborated cortices and the cognitive complexity to learn symbolic relations—and to have subsequent performance influenced by these relations. Further, the compulsory processing of the semantic dimension suggests that the mechanisms of attention and automaticity may be quite similar across animal species (see also Riley & Leth, 1976, Rotblat, 1987). This similarity affords the opportunity for further study, including the use of animal models in the cognitive science of attention (e.g., to study the effects of practice on Stroop-task performance, see MacLeod, 1991).

The present procedures and results should also interest researchers of numerical competence in animals, as the behaviors elicited here would seem to reflect a rather impressive array of basic knowledge about numerosness. However, full consideration of these issues is beyond the scope of the present report.

Acknowledgments—This research was supported in part by a grant (NAG2-438) from the National Aeronautics and Space Administration to Georgia State University, and by the College of Arts and Sciences of Georgia State University. Additional support was provided by the Army Research Office (DAAL03-92-G-0382) to the Center of Excellence for Research on Training, Morris Brown College.

REFERENCES


Rotblat H L (1987) Introduction to comparative cognition New York Freeman


Washburn D A & Rumbaugh, D M (1994) Symbolic distance in numerical judgments by rhesus monkeys: Number versus amount Manuscript submitted for publication

Windes J D (1968) Reaction time for numerical coding and naming of numerals Journal of Experimental Psychology 78 318–322

(Received 11/29/93, Revision accepted 5/18/94)