Perception & Psychophysics
2001, 63 (2), 322-329

The midstream order deficit

ALEX O. HOLCOMBE
Harvard University, Cambridge, Massachusetts

NANCY KANWISHER
Massachusetts Institute of Technology, Cambridge, Massachusetts

and

ANNE TREISMAN
Princeton University, Princeton, New Jersey

The relative order of an auditory sequence can be more difficult to apprehend when it is presented repeatedly without pause (i.e., cycling) than when it is presented only once (Warren, Obusek, Farmer, & Warren, 1969). We find that this phenomenon, referred to as the *midstream order deficit* (MOD), can also occur with visual stimuli. The stimuli need not form separate perceptual "streams," and the effect can occur with presentation rates as slow as five items per second, even though the identification of individual letters is very accurate at this rate. However, if the first item of the sequence is visually very distinct from the preceding items, relative order reports can be as accurate in the cycling condition as in the single-presentation condition. Our results suggest that the MOD is not due to masking, attentional blink, repetition blindness, Reeves and Sperling's (1986) order illusion, memory limitations, or decision criteria. The MOD may reflect an attentional cost to the initiation of order encoding, which is distinct from the allocation of attention is required in order to detect and identify individual items. To initiate order encoding successfully, one's attention must be set for, or captured by, an initial salient event.

Warren, Obusek, Farmer, and Warren (1969) reported a striking inability of listeners to perceive the relative order of a sequence of four sounds presented repeatedly, despite their fairly accurate performance when the sequence was presented only once. Their participants first heard the four sounds separately and learned a name for each. Then they were told that the four sounds would be presented in a particular order repeatedly and that afterwards they would be asked to report the correct relative order. A correct relative order report could begin with any one of the four sounds. For example, if *buzz–beep–hiss–boop* was presented over and over, *hiss–boop–buzz–beep* would be one correct response. The sounds were played for 200 msec each, with no interstimulus interval, and the sequence was repeated continuously without pauses between repetitions. The sequence ended when the participants were ready to respond. However, they felt that they could not apprehend the order, no matter how long they listened, and, in fact, their performance was not significantly different from chance. In contrast, another group of participants to whom the sequence was presented just once, instead of multiple times, performed significantly better than chance.

In the three present experiments, we first demonstrate a similar effect with visual stimuli and then test several candidate explanations for the effect. Our results suggest that poor performance with cycling sequences is not simply due to a greater difficulty in perceiving the individual items; rather, cycling causes a problem that is specific to the apprehension of order. In particular, we found that the accuracy of order judgments was correlated with the degree to which an identifiable item was perceptually distinct from the preceding items. We suggest that this was due to a difficulty in rapidly initiating order encoding without having a salient perceptual event.

EXPERIMENT 1
Visual Letters With Single Versus Multiple Presentation

This experiment documents conditions in which accuracy at judging relative order is lower for repeated presentations than for a single presentation of a sequence. We roughly equated forward and backward masking in the

This research was supported by NIH Grant MH45245 to N.K., NSF Grant SBR-9511633 to A.T., an NIMH NRSA graduate fellowship to A.O.H., and NIH Grant EY092582 to Patrick Cavanagh. We are indebted to Rik Henson for sharing his knowledge of theories of short-term order memory. We thank Etienne Benson, Jared Bush, and Damian Stanley for their help in conducting the experiments, and Charles Stromeyer, Patrick Cavanagh, and two anonymous reviewers for their useful comments on earlier manuscripts. Portions of this research were reported at the Annual Meeting of the Association for Research in Vision and Ophthalmology, 1996, Ft. Lauderdale, FL, the Meeting of the Eastern Psychological Association, 1998, Boston, and the International Conference on Short-Term Memory, 1998, Quebec. Correspondence should be addressed to A. O. Holcombe, UCSD Psychology 0109, La Jolla, CA 92093 (e-mail: holcombe@post.harvard.edu; N.K., ngk@psyche.mit.edu; A.T., treisman@phoenix.princeton.edu).

two conditions by embedding the single presentation in a longer sequence of meaningless line patterns. In both conditions, the first few items were presented at a very high rate, after which the presentation rate gradually slowed until it reached the final rate by the ninth item. Schematic diagrams of the single and repeated (cycling) conditions are shown in Figure 1 and described in detail below.[1]

Method

Participants. Eight Harvard University students, who were naive to the purpose of the experiment and who reported that they had normal or corrected vision, were paid for their participation.

Apparatus. All the subsequent experiments were conducted on a MacOS computer, running custom software created with Vision Shell C libraries (e-mail: raco@wjh.harvard.edu). The participants viewed an Apple 13-in. color monitor from a distance of approximately 30 in. The participants responded by pressing keys on a keyboard.

Procedure and Design. The sequence began at a very fast presentation rate and then decelerated: The first item was presented for one screen refresh (the monitor had a 60-Hz refresh rate), and each succeeding item was presented slightly longer until the final duration was reached by the ninth item. Four final exposure durations were used—120, 165, 210, and 255 msec. In each of these conditions, the first five items were presented for the following durations, from first to fifth: 15, 30, 30, 45, and 75 msec. After the fifth item, the duration of each succeeding item was longer than the duration of the previous item by a constant amount, such that the longest duration was reached by the ninth item. To fit the refresh rate of the monitor, the presentation durations were rounded to the nearest multiple of 15 msec. All of the items were presented at high contrast in a dimly lit room.

In the cycling condition, the four-letter sequence was presented three times after the deceleration. The items were displayed in a thin white rectangle that was filled before the trial began and unfilled when the first item of the stream appeared. For each trial, the four-letter sequence was chosen randomly from the set of all consonants other than Q. In the single-presentation condition, the four-letter sequence was embedded in a stream of line patterns, as is shown in Figure 1. The particular line pattern at a given position in each stream was chosen randomly, with the constraint that successive line patterns always had to be different.

The participants were instructed to fix their eyes on the stimulus for the entire time that it was presented and to report the letters in order, beginning with any of the four. As an example, they were told that if the letter sequence presented was BFGH, the following responses would all be counted as correct: BFGH, FGHB, GHBF, and HBFG. They were to enter the letters by typing them into the computer keyboard. The experimenter guided each participant through six practice trials. The first few practice trials used a very slow presentation rate in order to familiarize the participants with the stimuli and task.

Each participant's initial response was used to calculate the number of letter identities correctly reported. In order to gain supplementary data, when the participant had entered a letter that had not been presented, he/she was presented with the four letters that actually were presented and told to enter the relative order again, using the correct letters. If the participants entered incorrect letters on their first try, their second response was used in the analyses of their relative order accuracy. This allowed for the possibility that the participants might encode the correct order but may still need to be cued with the correct identities to recall the order.

The experimenter guided the participants through six practice trials. The first few trials used a very slow presentation rate in order to familiarize the participants with the stimuli and task. After the practice trials, there were 192 experimental trials, with a short break after half of the trials. The 192 experimental trials comprised the single- versus cycling-presentation factor crossed with the four final-item exposure durations, presented in pseudorandom order.

Results and Discussion

The participants reported the identities of the letters accurately, averaging at least 94% correct in all conditions. Letter identity accuracy was slightly greater in the cycling condition than in the single-presentation condition (98.1% vs. 96.6%) [$F(1,7) = 5.6$, $p < .05$]. The interaction of the presentation condition and the final du-

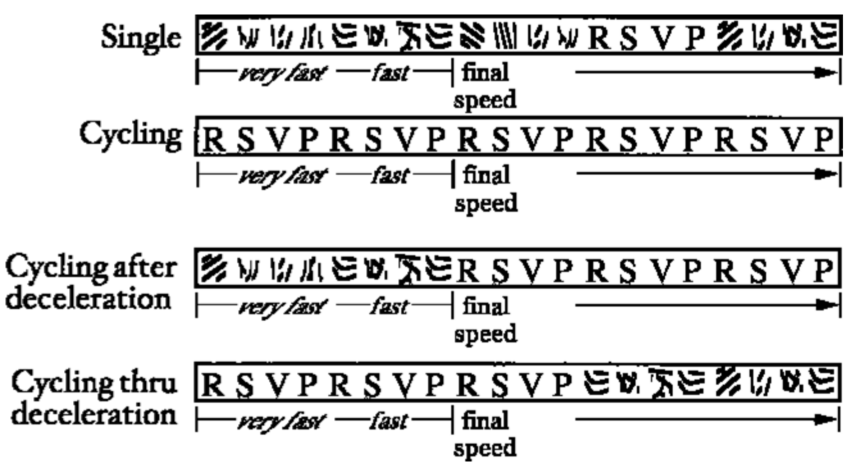

Figure 1. A schematic of the experimental conditions. The first few items were presented very quickly; the rate gradually slowed until the ninth item was presented. All items were presented in the same spatial location. In the single-presentation condition, the letters appeared after the deceleration, in a randomly chosen position. In Experiment 1, only the single and cycling conditions were used. In Experiment 2, the single and cycling conditions, as well as a variant of the single condition in which the line-pattern masks were replaced by digits, were used. In Experiment 3, all of the pictured conditions were used.

ration was not significant. All analyses were carried out with a three-way analysis of variance (ANOVA) with subject, duration, and cycling vs. single presentation as factors; all interactions were included.

The participants' high accuracy in reporting the letter identities suggests that they perceived and remembered almost all of the letters in both the cycling and the single-presentation conditions. This result is consistent with work that has shown, at least at the slower rates, that these presentation conditions result in minimal low-level masking, attentional blink (Chun & Potter, 1995; Moore, Egeth, Berglan, & Luck, 1996), and repetition blindness effects (Park & Kanwisher, 1994).

However, the participants often reported the letters in the wrong order. Moreover, accuracy in the report of order was significantly lower in the cycling condition than in the single-presentation condition [$F(1,7) = 5.91$, $p < .05$], as is shown in Figure 2. Most of the participants also reported that the cycling condition seemed to be more difficult.

The interaction of presentation condition and final duration was marginally significant [$F(3,21) = 2.91$, $p = .058$]. However, we suspect that the trend of a smaller difference between the conditions at faster rates was due to random variation, since other similar (unpublished) experiments have yielded the opposite interaction. Furthermore, our informal observations have clearly shown that, as the final-item duration is increased further, the difference between the two conditions rapidly diminishes to zero or reverses; that is, at slow rates, the repetition of the cycling condition becomes a benefit rather than a cost.

For the trials in the single-presentation condition, in which the letter sequence was presented near the beginning of the stream, the participants had to remember the sequence for a few seconds before they could report it. To test whether this led to more forgetting of the sequence, thereby decreasing the accuracy in the single-presentation condition, we examined the effect on accuracy of the letter sequence's position in the stream. The effect of position did not reach significance, either as a categorical variable added to the previous ANOVA [$F(8,64) = 0.69$, $p = .7$] or as a regressor (in an analysis of covariance) [$t(8) = 2.01$, $p = .08$]. Although it nearly reached significance in the second analysis, the regression coefficient was only .01, suggesting that order accuracy was 8% better when the sequence was presented in the last possible position than in the earliest possible position (i.e., eight positions earlier). This effect is too small to explain the difference between the cycling and the single-presentation conditions.

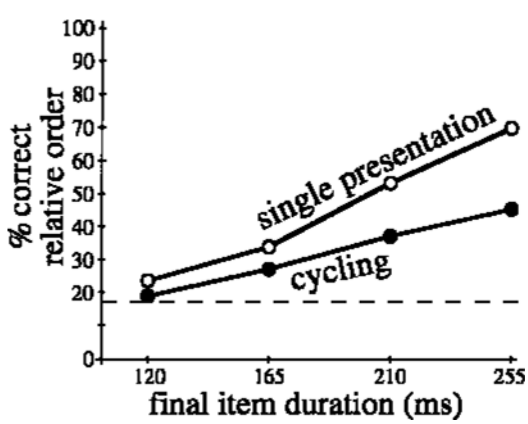

Figure 2. Mean probability of reporting the correct relative order at each duration for the cycling and single-presentation conditions of Experiment 1. If the participants knew which four letters were presented, but not their order, the relative order accuracy would be 16.6%, since there are six possible relative orders for four items (represented by the dotted line).

Within a letter sequence in the single-presentation condition, the first letter was reported in 96% of trials, the second in 98%, the third in 97%, and the fourth in 96%. This pattern provides evidence against the possibility that an attentional blink (Raymond, Shapiro, & Arnell, 1992) contributed to the difficulty of the temporal order reports. If attention to the first letter had produced a blink for its successors, the successors should have been reported less often. But there was no indication that the letters after the first one were identified any less accurately. This result and the fact that the attentional blink has not been observed with unmasked streams presented as slowly as the present rates suggest that the attentional blink was not a factor in Experiment 1.

Overall, the results of Experiment 1 show that, as with auditory stimuli (Warren et al., 1969), repeated, cycling presentation of visual stimuli can lead to lower accuracy on relative order judgments than that found for single presentation. Our use of slow rates and unmasked presentations and the fact that the participants had no difficulty in identifying the items make it unlikely that the poorer relative order accuracy in the cycling condition was due to greater masking, attentional blink, or to the repetition blindness effect.

EXPERIMENT 2
Detection

In the previous experiment, the full report of the sequence was used to assess the perception of temporal order. The deficit in accuracy in the cycling condition could therefore be a reflection of differences in memory load or in decision criteria, rather than differences in perceptual identification or attentional processes. In order to determine more precisely the source of the difficulty in the cycling condition, we used an order detection task in Experiment 2.

The participants viewed a specific cue sequence at the beginning of each trial and were asked to determine whether the subsequent stimulus sequence had the same relative order. The demands on perceptual identification and memory should be significantly lower in the detection task than they were in the full-report tasks. The participants knew which letters were to be presented, which made it easier for them to identify the individual letters during presentation. Memory load for the stimulus sequence should be lower because, instead of having to remember the entire stimulus sequence up to the response, the participants needed only to remember whether or not the stimulus sequence matched the cue. Although the participants had to remember the cue sequence, this should be less demanding than remembering the stimulus sequence, because the participants could commit it to memory before each trial. The detection paradigm also has the advantage of yielding separate measures of the stimulus information available and the participants' decision criteria.

Another goal of Experiment 2 was to investigate the possibility that the difficulty in apprehending order is not specific to cycling sequences, but rather is a reflection of a more general difficulty in apprehending order from the middle of a relatively undifferentiated stream. To test this, we embedded the single cycle of letters within a stream of digits and compared this with embedding it within the stream of line patterns that had been tested in Experiment 1. Although the line patterns should have been about as effective as the digits in masking, digits are more similar to the letters at a conceptual level.

Experiment 2 also offers a test of another explanation of the difficulty with cycling sequences. Perhaps the relative order of the four letters was encoded equally well in both the cycling and the single-presentation conditions, but the subsequently presented letters in the cycling condition disrupted memory. If the additional letters do harm retention of the stimulus sequence, they may also harm retention of the cue sequence, resulting in poorer memory for its order in the cycling condition than in the single-presentation condition. To test this hypothesis, the participants were asked to report the cue sequence after each trial.

Method

Participants. Thirty Harvard University students, who did not know the purpose of the experiment and who reported that they had normal or corrected vision, were paid for their participation or received course credit.

Procedure and Design. At the beginning of each trial, a four-letter cue sequence was displayed (simultaneously in one row) until the participant made a keypress. The participants' task was to determine whether the stimulus sequence was in the same relative order as the cue sequence. The four-letter stimulus sequence was drawn randomly from the set of all consonants other than Q. For half of the trials, the cue sequence was identical to the stimulus sequence. For the other half, the cue sequence was formed by scrambling the stimulus sequence—that is, by transposing a random pair of the letters, excluding those pairs that included the first letter. (Transpositions involving the first letter were avoided in order to prevent the participants from simply checking the first letter—a strategy that would benefit the single-presentation condition.) The first letter of the cue sequence was always the same as the first letter of the stimulus sequence (although the very first letter of the stimulus was unidentifiable in the cycling condition since it was presented very briefly).

When the participants had memorized the cue sequence, they pressed a key, and a filled white rectangle was presented until another key was pressed. The stimulus sequence was then presented in the same fashion as in Experiment 1, except that the four-letter sequence of the single-presentation condition could begin only at the 9th, 13th, or 17th position of the 20-item stream. The choice of position was distributed randomly across trials and counterbalanced across conditions. The participants were encouraged to wait for the first letter of the cue and to then check whether the rest of the letters appeared in the same order as they had in the cue sequence. After the presentation of the stimulus sequence, the participants pressed a key to indicate whether the stimulus was the same as the cue or different. Finally, the participants were prompted to enter the cue sequence into the keyboard.

The two fully crossed factors of this experiment were final duration of the letters (165 msec or 180 msec) and cycling versus single presentation. In addition, the single-presentation condition contained either digits or line-pattern masks. One block consisted of cycling presentation trials and single-presentation trials with digit masks; the other block consisted of cycling-presentation trials and single-presentation trials with line-pattern masks. The order of blocks was counterbalanced across participants.

After the participants were guided through eight practice trials, they each participated in a 96-trial block. After the first block, each participant took a short break, after which the experimenter explained that the masks would be changed (from digits to line patterns, or vice versa). Each participant then participated in eight more practice trials, followed by the second 96-trial block.

Results and Discussion

The pattern of results was equivalent for the two presentation durations, so the data were collapsed across duration. A signal-detection analysis was conducted in order to obtain estimates of performance independent of guessing or response biases. A high-threshold alpha model was used to estimate sensitivity, since its assumptions were found to fit the data of previous RSVP detection experiments better than the Gaussian d' model (Kanwisher, Kim, & Wickens, 1996). In addition, the alpha model is more in accord with the subjective reports of most of the participants that they knew the order on some of the trials but on most they did not and had to guess. All of the analyses reported here were also done with uncorrected percent correct as the dependent measure; the same patterns of results were found, with the same comparisons statistically significant.

Table 1 shows the mean percent correct in each of the three conditions (cycling, single presentation with line-pattern masks, and single presentation with digit masks), as well as the means of the sensitivity measure (α) and the estimated guessing rate (g). These measures were figured separately for each participant before the means were calculated. The percent correct numbers were higher than in Experiment 1, because chance in this detection experiment was 50%, whereas chance in Experiment 1 was no higher than 16.6% (16.6% is an upper bound on chance relative order accuracy since it was assumed that the par-

Table 1
Mean Estimated α, g, and Percent Correct for the Three Conditions of Experiment 2

	Single-Line Pattern Masked	Single-Digit Masked	Cycling
α	.85	.76	.71
g	.2	.26	.32
% correct	84	78	74

Note—α, sensitivity; g, percent of trials in which the participant guessed that the target had been presented, though it had not been apprehended.

ticipants knew which four letters were in the sequence on every trial).

Sign tests revealed that all the pairwise α differences shown in Table 1 were significant. The participants performed less accurately in the cycling condition than in both the single-presentation line-pattern mask condition ($p < .001$) and the single-presentation digit-mask condition ($p < .01$). All participants also reported that the cycling condition seemed the most difficult.

Thus, even though the participants knew exactly what they were looking for and knew exactly which letters would be presented on each trial, their accuracy was still significantly lower in the cycling condition than in the single-presentation condition.

The digit-mask condition was more difficult than the line-pattern condition ($p < .05$), yet it was easier than the cycling condition ($p < .05$). This suggests that the low accuracy in the cycling condition may be due not to repetition of the sequence per se, but rather to a more general difficulty in apprehending order in an undifferentiated stream.

Experiment 2 was also designed to test the hypothesis that subsequent repetitions of a sequence disrupt memory retention. If this is true, one would expect the participants to forget the cue four-letter sequence more often in the cycling condition. To the contrary, the participants' accuracy in reporting the cue was equal in both conditions, at 94% correct. However, it remains possible that the additional letters of the cycling condition disrupted short-term retention of the stimulus sequence without its having disrupted the retention of the cue sequence, since the cue sequence was more securely committed to memory.

EXPERIMENT 3
Effect of Initial Deceleration

In the preceding experiments, the cycling condition differed from the single-presentation condition in two important respects. First, the letter sequence was presented during the deceleration phase in the cycling condition, whereas it was not presented during the deceleration phase in the single-presentation condition. Second, in the cycling condition, the letter sequence was presented multiple times after the deceleration phase instead of just once. In Experiment 3, these two factors were varied independently in order to determine which one accounted for the low order accuracy in the cycling condition. To preview the results, performance in the cycling presentation condition was as good as performance in the single-presentation condition when the sequence was not presented during the deceleration phase, suggesting that cycling does not impair accuracy if the beginning of the sequence is salient.

Method

Participants. Eight Harvard University students who had normal vision and were naive to the purpose of the experiment were paid for their participation.

Procedure and Design. The procedure was the same as that in Experiment 2, except as noted. Instead of detection procedure, observers reported the four-letter sequence as in Experiment 1. The four display conditions of this experiment are schematized in Figure 1.

The cycling and single-presentation displays were identical to those of Experiment 1. But in the new *cycling during deceleration* condition, the letter sequence was presented throughout the deceleration phase and then was presented once immediately after the deceleration phase. Line-pattern masks were presented thereafter. In the *cycling after deceleration* condition, line-pattern masks were presented during the deceleration phase, and the letter sequence was presented thereafter. If the lower accuracy in the cycling condition was due simply to the cycling of the sequence, regardless of whether this cycling occurred during the deceleration phase, then accuracy in both of the new conditions should be lower than accuracy in the single-presentation condition. If, alternatively, it is critical that cycling begin during the deceleration phase, lower accuracy should occur only in the cycling during deceleration condition.

The experimenter diagrammed and verbally described the deceleration phase and the four display conditions, then guided the participants through eight practice trials, which exposed them to each of the conditions twice. Each trial began by informing each participant of the condition that was about to be presented. Each observer participated in 192 experimental trials; there was a short break after half of the trials. The four display conditions were fully crossed with the two final exposure durations of 165 and 180 msec. This resulted in 24 trials for each of the resulting condition × duration combinations, which were presented in pseudorandom order.

Results and Discussion

As in the previous experiments, accuracy in reporting the letter identities was generally high, as is shown in Figure 3. For the statistical tests, least-significant differences tests were used in the context of an ANOVA (with duration, stimulus type, subject, and all interactions as factors), which is equivalent to testing contrasts on the pairs of means. Interactions with presentation duration were not significant, and the pattern of results for the two presentation durations was the same, so the data were collapsed across duration. Letter identity accuracy was higher

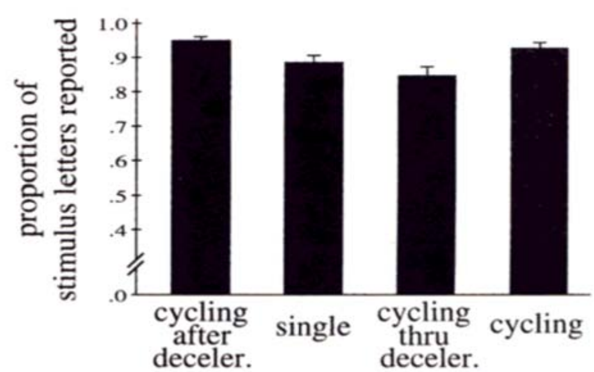

Figure 3. Proportion of stimulus letters reported for each condition of Experiment 3. "deceler." = deceleration. Standard error bars are based on subject means (and thus different from the analysis of variance error term used for the statistical tests).

in the cycling condition than in the single-presentation condition ($p < .01$). It was slightly lower in the cycling through deceleration condition than in each of the other conditions (ps $< .05$). This may be explained by the possibility that in the cycling through deceleration condition, the participants tried to apprehend the sequence during the deceleration phase and that the extremely rapid presentation of the letters during the deceleration occasionally led to misidentification of the letters.

The percentage of correct relative order responses in each condition is plotted in Figure 4. Statistical comparisons were performed using the same statistical analysis as described above, but relative order accuracy was substituted for identity accuracy as the dependent variable. Interactions with presentation duration were not significant, so the data were collapsed across duration. Performance differed significantly ($p < .01$) in each of the conditions from each of the other conditions. Relative order reports were less accurate in the cycling condition than in the single-presentation condition, replicating the patterns found in previous experiments. In the cycling through deceleration condition, accuracy was also lower than in the single-presentation condition.

The differences in order accuracy between the conditions were much larger than the differences in identity accuracy—in particular, the 5% difference between identity accuracy in the cycling through deceleration condition and the single-presentation condition does not explain the 30% difference in relative order accuracy. Comparison of Figures 3 and 4 shows that, in general, the differences in number of stimulus letters reported were small compared with the differences in correct order reports.

Relative order accuracy in the cycling after deceleration condition was higher than in the single-presentation condition. However, when only trials in which the participants reported all four stimulus letters correctly are considered, the difference between relative order accuracy in the cycling after deceleration condition and that in the single-presentation condition diminishes to insignificance. This suggests that the higher order accuracy in the cycling after deceleration condition was due to better memory for the items in the cycling after deceleration condition, which was perhaps due to repeated exposure. Nevertheless, relative order accuracy is clearly at least as good in the cycling after deceleration condition as in the single-presentation condition. In other words, when the letters were presented only after the fade-in period (the cycling after deceleration condition), performance was as good as that after single presentation. This is probably the most revealing finding of the experiment.

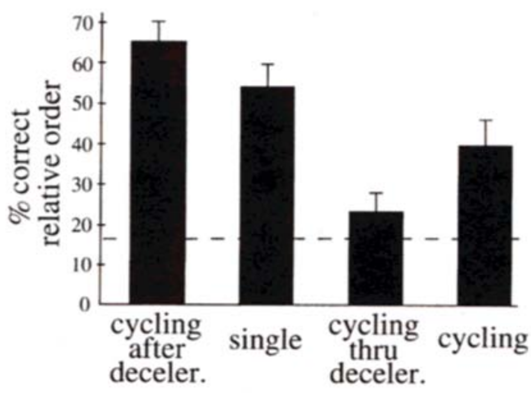

Figure 4. Proportion of correct relative order reports for each of the conditions of Experiment 3. "deceler." = deceleration. Standard error bars are based on subject means (and thus different from the analysis of variance error term used for the statistical tests). The dotted line represents accuracy when the letters presented were known but their order was guessed.

This effect can be explained in terms of the degree to which an identifiable letter in the stream differs perceptually from the preceding items. In the cycling condition, the first several items of the stream were presented too rapidly to be identified. Gradually, the presentation rate decreased such that the sequence seemed to fade in without its having a differentiated initial letter. In contrast, in the single-presentation and cycling after deceleration conditions, the first identifiable letter was preceded only by line patterns, which are very different from the letters. This resulted in high order accuracy. We term the lower relative order accuracy found in the cycling condition *midstream order deficit* (MOD), because it occurs when observers are forced to apprehend the order of a rapidly presented sequence from midstream—that is, when there is little distinction between the beginning of the sequence and the preceding items.

Although the items preceding the sequence play a large role, by contrast, the presence or absence of an end to a sequence different from the subsequent items seems to have had little, if any, effect. Compare the single-presentation condition, in which line patterns immediately followed the presentation of the sequence, with the cycling after deceleration condition, in which letters followed the sequence. Accuracy was comparably high in both conditions.

GENERAL DISCUSSION

Our results show that accuracy in perceiving the relative order of rapidly presented items in a cycling sequence is much lower than when the sequence is presented just once. This effect is quite general, given that it occurs with sounds (Warren et al., 1969), visual letters, shapes, and colors (O'Brien & Treisman, 1971). The accuracy of order reports in our experiments is correlated with the degree to which an identifiable letter is visually distinct from preceding items. Accuracy was highest when the first identifiable letter constituted the onset of the stimulus (as shown in an unpublished experiment), and it became progressively lower when the letter sequence was preceded by line patterns (Experiments 1 and 2), by digits (Experiment 2), and by other letters (the cycling conditions).

The effect occurred at relatively slow presentation rates, for which the participants' performance in reporting the letter identities was high, so it could not have been due to a problem in identifying the individual items. Furthermore, masking was approximately equated in the cycling and single-presentation conditions by the line-pattern masks, so the MOD is unlikely to have been due to masking. In addition, the MOD occurred even when a detection task was used, so it was probably not due to a memory limitation or to differences in decision criteria.

Previous Theories and Other Attentional Phenomena Do Not Explain the MOD

Warren (1974) and Teranishi (1977) provided evidence that the threshold duration for reporting the order of cycling sound sequences is determined by the length of time required to name the individual items as they are being presented. Warren (1982) explained that lower threshold durations occur for singly presented sequences by suggesting that after single presentation, persistence of the sequence in short-term memory allows naming at a slower rate. Although this account may be partly right, it does not explain why the difficulty with cycling sequences disappeared when the sequence had a clear beginning (i.e., when it was not presented during the deceleration phase, Experiment 3).

Indeed, in our experiments, the deceleration or fade-in period was critical to obtaining lower relative order accuracy in the cycling condition than in the single-presentation condition. Yet, in Warren et al.'s (1969) experiments with auditory stimuli, there was no fade-in or deceleration period, and still the observers were much less accurate in the cycling condition than in the single-presentation condition. This discrepancy may have been due to a difference in observer strategy rather than to the difference of auditory versus visual stimuli. In Warren et al.'s study, each observer participated in only one trial. Thus, the observers did not realize the importance of preparing to encode the order starting with the very first item. They probably waited for the stimulus to begin and then tried to encode the order, which left them in the position of trying to apprehend the order midstream, without there being an item distinct from the preceding items.

Audition researchers have sometimes attributed the difficulty in perceiving the order of cycling stimuli to *auditory streaming* (Bregman & Campbell, 1971). Auditory streaming (Miller & Heise, 1950) refers to the tendency of alternating acoustic stimuli to be perceived not as alternating, but rather as being two continuous streams of sound, particularly when the sounds are cycled (Bregman, 1978). Bregman and Campbell (1971) found that perceiving order across streams was very difficult, since streaming items are not perceived to occur successively; this may explain part of the difficulty of perceiving the order in Warren et al.'s (1969) experiment.

In our experiments, however, subjective reports strongly suggest that streaming did not occur. The participants reported that they did not experience the letters as being segregated into multiple streams. Rather, the letters were experienced as successively presented, even at the fastest rates. Furthermore, in unpublished experiments conducted by O'Brien and Treisman (1971) with visual stimuli, there was no effect of hetereogeneity of items, which would have been expected to modulate the stream segregation if it had occurred (Bregman & Campbell, 1971). Accuracy for cycling with three colors or three shapes was identical to accuracy for cycling two colors and one shape or two shapes and one color.

The attentional blink (Raymond et al., 1992) is unlikely to play a part in the MOD. First, the attentional blink is not typically found at presentation rates as slow as those in the present experiments, and extrapolation from published data suggests that it would not occur at these rates (e.g., Moore et al., 1996). Furthermore, the lack of a serial position effect in the single-presentation condition is strong evidence that a blink did not occur. The temporal order illusion documented by Reeves and Sperling (1986) is also unlikely to be a factor in the MOD, since Reeves and Sperling's errors occurred only at faster rates than those used in the present experiments and were accompanied by an attentional blink. Similarly, the rate of repetition was too slow for repetition blindness to play a role (Park & Kanwisher, 1994).

In sum, we have rejected accounts based on masking, memory limitations, decision criteria, naming time, perceptual streaming, the attentional blink, Reeves and Sperling's (1986) order illusion, and repetition blindness.

Theory and Conclusions

Our data indicate that the accuracy of order reports is correlated with the degree to which an identifiable letter is visually distinct from preceding items. The presence of a salient end to the sequence is not important and neither is the distinctiveness of the items occurring after the first item. This last point is evident in the single-presentation condition: The transitions from letter to letter were no more distinct than those in the cycling condition, yet order for the subsequent letters was accurately reported in the single-presentation condition. The overall lesson of the MOD appears to be that observers cannot easily begin encoding by using an arbitrarily selected item in a cycling stream. But why should this be?

Short-term memory is capacity limited, and encoding into explicit short-term memory requires attention (Rensink, 2000; Sperling, 1960), hence there is a need to actively initiate order encoding. In the case of our letter streams, this initiation act is distinct from the act of allocating spatial attention, since the participants presumably allocated their spatial attention to the letter stream from its onset. When an item in a sequence is marked by a perceptual or categorical discontinuity, an automatic or exogenous segmentation occurs, just as a visual onset or change can pull exogenous spatial attention to a new location. We believe that such a discontinuity allows for rapid initiation of order encoding. However, in the cycling condition, the engaging of attention on an arbitrary item requires an act of endogenous segmentation, which appears to have a low temporal resolution (relative to exogenous attention and other visual processes), just as endogenous attention has a coarse *spatial* resolution (He, Cavanagh, & Intriligator, 1996). If the engaging of endogenous attention on an item within a stream takes 100– 200 msec from the onset of the target item, as it does with a spatially distinct item (Nakayama & Mackeben, 1989), the target must last long enough for it to be processed, once attention has reached it. If, instead, it is succeeded by another item in the same location, the whole attempt to initiate order encoding will have to be restarted and will likely fail again for the same reasons. Experiments on sequential task switching certainly suggest long delays in the resetting of attention (Allport, Styles, & Hsieh, 1994; Rogers & Monsell, 1995), although these tasks are not directly comparable.

It is possible that midstream deficits are not confined to order judgments. In an experiment by Holcombe and Cavanagh (2001) that also presented stimuli in a rapid visual stream, two conditions were compared. In one, a leftward-tilted Gabor adjacent to a red patch rapidly alternated with a rightward-tilted Gabor adjacent to a green patch. In the other condition, the pairing was reversed—leftward tilt and green alternated with rightward tilt and red. The fastest rate at which the observers could correctly discriminate between these conditions in 75% of trials was approximately 180 msec/stimulus. After correcting for the difference in chance level, this threshold is similar to the temporal order threshold in the cycling condition and may be explained by the long latency of endogenous attentional engagement. In this case, a voluntary attention shift between the color patch and the Gabor stimulus may have been necessary within each frame of the sequence in order to encode both of the adjacent items before the next stimulus was presented. If the stimuli were replaced before the switch was completed, the participants would have been unable to determine the color–orientation pairing.

Outside the laboratory, the MOD may occur during the viewing of modern television programs and commercials that utilize rapid successive cuts. The present work suggests that at rates faster than about three per second, observers will find it difficult to accurately apprehend the order of the scenes, even though they may momentarily comprehend each individual scence (Potter, 1976, 1993). Studies on the phenomena of the attentional blink and change blindness (Rensink, O'Regan, & Clark, 1997) have revealed severe limitations in our ability to remember what we see. The MOD (as well as possible variants of it) merits further study, since it illustrates yet another limitation in our ability to retain what we see.

REFERENCES

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 421-452). Hillsdale, NJ: Erlbaum.

Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 380-387.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, **89**, 244-249.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 109-127.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, **383**, 334-337.

Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience*, **4**, 127-128.

Kanwisher, N. G., Kim, J. W., & Wickens, T. D. (1996). Signal detection analyses of repetition blindness. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1249-1260.

Miller, G. A., & Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America*, **22**, 637-638.

Moore, C. M., Egeth, H., Berglan, L. R., & Luck, S. J. (1996). Are attentional dwell times inconsistent with serial visual search? *Psychonomic Bulletin & Review*, **3**, 360-365.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, **29**, 1631-1647.

O'Brien, T., & Treisman, A. (1971). *The perception of order with recycling visual stimuli.* Unpublished manuscript, Oxford University.

Park, J., & Kanwisher, N. (1994). Determinants of repetition blindness. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 500-519.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 509-522.

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, **2**, 156-161.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 849-860.

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review*, **93**, 180-206.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, **7**, 17-42.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, **8**, 368-373.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, **124**, 207-231.

Sperling, G. (1960). The information available in brief presentations. *Psychological Monographs: General & Applied*, **74**, 1-29.

Teranishi, R. (1977). Critical rate for identification and information capacity in hearing system. *Journal of the Acoustical Society of Japan*, **33**, 136-143.

Warren, R. M. (1974). Auditory temporal discrimination by trained listeners. *Cognitive Psychology*, **6**, 237-256.

Warren, R. M. (1982). *Auditory perception: A new synthesis*. New York: Pergamon.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science*, **164**, 586-587.

NOTE

1. A demonstration of the difficulty of apprehending the order of cycling sequences can be viewed at http://www-psy.ucsd.edu/aholcombe/ or obtained by contacting the first author.

(Manuscript received June 7, 1999;
revision accepted for publication June 2, 2000.)