The time it takes to imagine[1]

ROBERT J. WEBER[2] AND JUDY CASTLEMAN
OKLAHOMA STATE UNIVERSITY

Three experiments were conducted dealing with letter processing in visual and speech imagery. The first two experiments indicated that speech imagery is more rapid than visual imagery (about six letters per second for speech vs about two letters per second for vision). Postexperimental scaling of subjective fatigue also revealed differences between imagery modalities, with visual imagery conditions consistently more fatiguing than speech conditions. The third experiment dealt with error rates in learning to classify letters on the basis of visual image properties or on the basis of arbitrary letter names. Results showed much more efficient performance for classification based on visual image properties. It was concluded that visual and speech imagery modes differ fundamentally in the manner in which they process information.

There is now evidence that visual imagery in serial operation is a slower process than speech imagery. Indirect evidence for this claim comes from mediation studies using either verbal or visual mediators as established by instructional set (Bugelski, Kidd, & Segman, 1968; Paivio, 1966). Direct evidence comes from measurements of speech and visual imagery rates. Landauer (1962) has found virtually identical rates for implicit and explicit speech. Weber and Bach (1969) assessed visual imagery rates and compared them directly with speech imagery rates. They defined imagery as self-controlled sensory-like experience in the absence of correlated external stimulus energy. In their study, Ss processed the letters of the alphabet as rapidly as possible by generating them in implicit or imagistic speech, explicit speech, and visual imagery modes. The two speech conditions produced virtually identical processing rates (about 6.5 letters per second), while visual imagery was much slower (about 2.5 letters per second). The visual imagery rates were, however, quite variable. This may have been because some Ss were poor visualizers and ipso facto used different decision criteria of what constituted an image. It is also possible that Ss in the Weber and Bach study were not visualizing images (even though they reported they were). Hence, while Watson's denial of the existence of visual imagery amounted to the elevation of a personal defect into a universal principle (Blanshard, 1967), Weber and Bach may have committed the opposite error: assuming that because they possessed visual imagery, everyone else did too. In fact, Galton's early surveys (1880) suggest that the development of visual imagery differs widely among people.

The present studies attempt to extend the findings of Weber and Bach through replication procedures that lead to less variability and more objectivity. In Experiment 1, Ss are selected on the basis of possessing visualizing ability. In Experiment 2, an attempt is made to objectively indicate visual imagery processes by having Ss identify spatial properties of letters as they are visualized. Finally, Experiment 3 is conducted to strengthen the interpretation of Experiment 2 through an investigation of classification based on arbitrary or visual properties of letters.

EXPERIMENT 1

The purpose of this study was to replicate Weber and Bach (1969) while at the same time eliminating some of the variability evidenced in that study by selecting Ss according to their claimed ability to visualize and by introducing some improvement in procedure. In the present experiment, each S served under all conditions, i.e., all treatments are within Ss. In Weber and Bach the reverse was true; all treatments were between Ss. The present procedure was viewed as an improvement in that everyone was required to participate in the explicit speech (SE) condition. This insured that every letter was processed, and the S would thereby be supplying himself with a set to process every letter in the unobservable imagery conditions. An additional difference involved the change of a trial from two passes to one pass through the alphabet. Two passes were previously used in an attempt to obtain better reliability than with one pass. But some Ss complained spontaneously of fatigue in the visual imagery (VI) condition, so it may be that two passes actually contributed to variability. Along these lines, the present study incorporated a scaling for fatigue after the termination of the last block of conditions. Also, in an attempt to determine relationships between underlying modalities, a time estimation procedure was introduced during the last block to allow comparison of subjective time to process the alphabet with observed time. It was hypothesized that different imagery modalities might show substantial differences between observed and estimated time and this would be indicative of different underlying processes.

Method

Subjects. The Ss were 16 undergraduate psychology volunteers who received extra course credit for participation. Only those Ss who said they could project, with eyes open, an imaginary visual image of letters of the alphabet on a blank screen were used. About 51% of the class of 81 claimed to have this ability. All of these Ss claimed to have the additional ability of visualizing letters with their eyes closed. This latter ability was evidently more widespread, being claimed by 90% of all students.

Instruction and procedure. Upon entering the experimental situation, the Ss were checked for their familiarity with the alphabet and sometimes given relearning if necessary. They were then given common instructions and practice on a series of four tasks. These tasks entailed going through the alphabet a single letter at a time as follows: For the speech explicit (SE) task the S was instructed to say the alphabet *aloud* without stopping, for example, "abc . . . z." For the speech imagery (SI) task the S was instructed to say the alphabet *silently*, talking to himself. In the visual imagery (VI) task the S was instructed to *close* his eyes and visualize or imagine the letters of the alphabet passing before him as if successively flashing on the same spot on a movie screen one letter to a frame. The letters were to be visualized as single, black, upper case, typed letters on a white background. In the visual explicit (VE) task, the S was instructed to go through the alphabet in a manner very similar to the VI task, except to keep his eyes *open* and visualize or imagine the letters of the alphabet being projected successively on the same spot on a sheet of 8½ x 11 in. white typing paper. Since the Ss had been preselected for their claim to this ability, the task requirement was readily understood.

For each trial the examiner placed an index cue card with one of the instructions, aloud, silent, eyes closed, or eyes open, before the S. The S then activated a standard electric clock's remote

 Copyright 1970, Psychonomic Journals, Inc., Austin, Texas

Table 1
Summary Statistics for Experiments 1 and 2

	Condition	Time in Seconds Per 26 Letters M	SE_M	Mdn	Letters Per Sec	Pearson r Block 6	Mean Rank	Mean Rating
Experiment 1 (N = 16)	SE	4.18	0.50	4	6.22	0.22	1.62	
	SI	4.28	1.19	4	6.07	0.66	1.44	
	VI	12.85	6.45	11	2.02	0.72	3.00	
	VE	13.80	6.22	13	1.88	0.72	3.94	2.87[a]
Experiment 2 (N = 20)	SE	5.18	1.84	4	5.02		1.10	1.05
	VI	14.76	3.42	14	1.76		2.00	1.90
	VP	27.42	9.16	24	0.95		2.90	3.20

[a] N = 15

switch at the onset of each trial and stopped the clock remotely when he finished the alphabet. He was not allowed to see the clock face and he was not given information on his response times.

Data were collected for six blocks of the four conditions, with each S receiving a different random order of conditions. There was an interval of about 20 sec from the end of one trial to the beginning of another; during this time the E reset the clock and wrote the time for the trial on a data sheet.

At the close of the fifth block, the S was instructed that following completion of each condition in the next block he would be asked to make an estimate, in seconds, of how long he took to go through the alphabet. After completion of the sixth block, the E placed the index cue cards before the S and asked him to rank the tasks in order of their difficulty. Using the task that the S selected as most difficult, the E then asked him to rate it on a 1-5 scale according to fatigue, No. 1 being "not fatiguing at all" and No. 5 being "the most fatiguing thing you've ever done."

Results

Results collapsed over the 16 Ss and six blocks are shown in the top half of Table 1. The two speech condition means are quite close to one another, each requiring about 4 sec for processing one pass through the 26 letters of the alphabet. Similarly, the two visual conditions yield comparable times of about 13 sec each. There is a substantial difference between processing times for visual and speech conditions. A significance test is not required since each one of the 16 Ss has faster mean processing times for speech than for vision. The difference between means for VE and VI is not significant [correlated $t(15) = 1.92$, $p > .05$], nor is there a significant difference between the speech conditions ($t < 1$). In the fourth column of Table 1 we find the same information expressed in terms of rates, that is, in terms of letters processed per second. Processing rates for speech are about six letters per second and for vision about two letters per second.

In the next column, Pearson rs are shown relating observed and subjectively estimated times for Block 6. The visual conditions show a higher correlation than do the speech conditions, but this may be due to the more restricted range of variation occurring for the speech conditions. The same explanation may account for the difference in rs for SI and SE. One might add that estimated times were usually overestimates of real time with 12 or 13 out of the 16 comparisons for each condition being overestimates.

In the next column of Table 1, the average rankings of fatigue are shown. A value of 1.0 would be least fatiguing and 4.0 most fatiguing. The speech conditions are perceived as less fatiguing than the visual conditions. Condition VE was judged as most fatiguing by 15 of the 16 Ss. The average rating given to that most tiring condition was 2.87, which corresponds closely to a judgment of "moderately fatiguing" on the 5-point scale.

The top panel of Fig. 1 shows processing time as a function of practice. The left ordinate is time to process one pass through the alphabet, and the right ordinate is the same information expressed in time per letter. The abscissa shows a break between Blocks 5 and 6 because the last block was procedurally different, containing the time estimation series. However, this procedure did not seem to alter any trends in the figure. There is the suggestion of possible significant practice effects for the visual conditions. Each S's times for Trials 1-3 were compared with his times for Trials 4-6. For VE there is a significant change, correlated $t(15) = 51.5$, $p < .001$, and also for VI, with correlated $t(15) = 15.9$, $p < .001$. The practice effects are small but reliable. The differences between the visual and the speech condition means are very consistent across blocks in that no reversals of relative processing time occur.

EXPERIMENT 2

The results of Experiment 1 made sense internally, as well as providing a replication of the original study of Weber and Bach. However, there was still considerable variability in the visual conditions and always a remote possibility that Ss were doing something other than visualizing letters. An objective method of requiring the S to visualize was needed. The admirable objective methods of Brooks (1968) used in his studies of imagery and memory suggested the present method.

Individual lower-case, typed letters have an interesting spatial property, vertical size, that readily divides them into two classes, large and small. Large letters include "b," "d," "f," "g," "h," "j," . . . , and small letters include "a," "c," "e," "i," "m," "n," . . . . This classification results in a total of 12 large and 14 small letters. If an S were required to process letters of the alphabet seriatum and appropriately call out "large" or "small" as he processed each individual letter, we would be reasonably assured that the letters were being processed in a visual imagery mode. This would not, however, give a pure measure of visual imagery time, because, in addition to generating the image, the S would then have to read or abstract from it the property of large or small, and that would take extra time. We would expect, though, that in order to abstract that property each image would have to possess some minimal standard of clarity, i.e., there would be a lower bound in establishing a decision criterion of what constitutes an image. Given suitable instructions, it is not unreasonable to expect that decision criterion to carry over to a regular VI condition.

The present experiment may be viewed as an attempt to increase the objectivity of

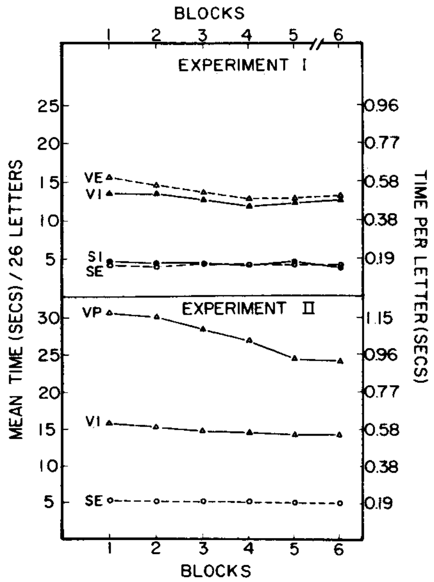

Fig. 1. Mean processing time in seconds as a function of practice and modality. Top panel, Experiment 1, and bottom panel, Experiment 2.

measuring visual imagery processing and to further reduce the variability associated with it.

Method

Subjects. The Ss were 20 summer school volunteers between the ages of 18 and 30 years. They received extra course credit for their participation, and were not selected on the basis of any predetermined ability to visualize.

Instruction and procedure. Upon entering the experimental situation the Ss were given instructions and a practice trial for each of three tasks. These tasks included going through the alphabet by saying the letters explicitly (SE), by visually imagining the consecutive letters with the eyes closed (VI), and by visualizing consecutive letters with the eyes closed and at the same time calling out each letter's visual property (VP) of large or small. For the SE task, the S was instructed as in the previous experiment. In the VP task, the S was shown a typed list of 26 alphabetic letters in random order; each letter was described once by the E as "large" or "small." Then the S was shown another random array of letters and to insure his understanding he was asked to name each letter according to its size. For the VI task, the S was instructed to close his eyes and visualize or imagine the letters just as he had in the VP condition, except that he was not to report their size properties.

Procedural details involving cueing and response times paralleled those of Experiment 1. There were six blocks of the three conditions, each S receiving a different random order of conditions. On completion of the sixth block, the S was asked to both rank and rate the three tasks for fatigue. The rating categories were identical to those in Experiment 1, but this time all conditions were rated.

Results

The principal results are shown in the bottom portion of Table 1. The shortest processing times are for SE and the longest for VP. Conditions SE and VI show mean time slightly longer than in the first experiment, but the variability of VI is considerably less than in the first study. All 20 Ss displayed the same increasing order of processing times, SE, VI, and VP, so a significance test is once again not required. In terms of fatigue rankings, SE is the least fatiguing and VP the most fatiguing. The agreement on this ranking was almost identical for each S. The values for rated fatigue reveal the same pattern. However, the value of 1.90 for VI is lower than the comparable value of 2.87 from the first study.

The bottom panel of Fig. 1 shows response time as a function of practice. When comparisons of response times for the first three blocks vs the last three blocks are made for the visual conditions, a significant practice effect is found: for VP, correlated $t(19) = 86.06$, $p < .001$; for VI, correlated $t(19) = 50.57$, $p < .001$. Again the magnitude of the practice effects is small but reliable.

EXPERIMENT 3

In Experiment 2 it was assumed that Ss were visualizing spatial properties in the VP condition, but it was remotely possible that the large-small labels were learned separately for each letter during the instruction phase when the E pointed out which letters were small and which large. To eliminate this possibility, a measure of learning facility in assigning alphabetical letters to two categories is needed. It would seem that this could be tested readily by assigning the labels large-small at random to the letters of the alphabet. But this would result in a crossed or disordered mapping, something that has been shown to be quite difficult (Weber, Love, & Goldstein, 1967). A better solution would be to take two highly learned responses, such as the digits "1" and "2," and use them in a random pairing with the letters. Learning measures on such a task could then be compared directly with similar learning measures on a task requiring a conventional pairing of the responses large-small with letters (as in the VP condition, Experiment 2).

Method

Subjects. The Ss were five college student volunteers. Each S served in both experimental tasks, first arbitrary and then visual properties.

Procedure. In the arbitrary part of the experiment, each S was told he was participating in a learning task. He was first asked to repeat the alphabet to insure his knowledge of it. He was then given a study trial in which he was presented with an index card that had an arbitrary alphabetic sequence typed on it in lower-case letters. Each letter had been paired randomly with either the digit "1" or the digit "2," which was juxtaposed with it, i.e., f-2, y-1, c-1, etc. There were 14 occurrences of one digit response and 12 occurrences of the other digit response to match the corresponding number of large-small responses in the VP condition of Experiment 2. The S was to repeat each letter and the correct response as the E pointed to them at a rate of about one pair every 2 sec. Then the card was removed and the S was to repeat the alphabet, in serial order, beginning with "a," assigning either a "1" or a "2" to each letter as he proceeded in a self-paced manner. Each S was given six trials of this study-recall sequence. While the order of letter-digit pairs differed from one study trial to the next to keep the S from learning a fixed sequence of responses, the mapping relation remained constant in that the letter-digit pairing remained the same for a given S.

In the visual property part of the experiment, the S was given a study-recall series the same as the above except that each letter was systematically assigned the label "large" or "small" as in Condition VP of Experiment 2. Once more, six study-recall trials were alternated.

The E recorded all errors on a prepared data sheet and alternately presented study trial cards, each of which contained a different sequence of pairs.

Results

For the arbitrary condition, the mean number of errors averaged over the five Ss and six trials was 7.6. For the six trials, the mean number or errors per S ranged from 5.8 to 9.3, so that at least a moderate error rate occurred for all Ss. Finally, there was a steady decline in mean errors from 10.0 on Trial 1 to 5.0 on Trial 6. The second part of the experiment, the visual property condition, is easily summarized. Not one S committed a single error.

It is difficult to believe that the same manner of processing could have been used for both the arbitrary and visual property tasks. The evidence is strongly in favor of a visual imagery process at work here and in the previous experiment.

DISCUSSION

The rates of imagery in this study correspond closely to those found in the earlier study (Weber & Bach, 1969), which relied heavily on subjective procedures. The visual imagery rates of about two letters per second obtained here are slightly less than in the earlier study (2.5/sec). The overall agreement between the disparate methods is nonetheless gratifying. It seems likely that S-selection procedures will give way to objective assessments of visual imagery such as the visual property method of Experiment 2. Indeed, the VP condition seemed to stress so clearly what a visual image was that the VI condition showed less variability in Experiment 2 than in Experiment 1—and this without any S selection. The relatively large variability for the VP condition is probably indicative of combining several underlying processes such as image generation and abstraction of spatial properties from the image. One difference between the present study and the earlier paper by Weber and Bach concerns practice effects. The earlier study revealed no significant practice effects for the VI condition, whereas a small but significant decrease in processing time is found to occur with practice for all visual conditions in this study. No obvious explanation suggests itself for this discrepancy between the two studies.

The fatigue scaling in Experiments 1 and 2 serves to show in still another way that what we have called visual imagery and speech imagery are indeed different psychological processes. The learning effects of Experiment 3 lend further convincing evidence to this point. If the arbitrary task is so much more difficult to perform than the visual property task, it is very likely that different underlying processes are at work. It is, of course, remotely possible that Ss at one time learned to associate the verbal labels large-small (or a pair of semantic equivalents) with the lower-case letters. Against this, we have many subjective reports that Ss did in fact visualize the letters in the visual property task. Also, if prior verbal learning were a satisfactory explanation, we might expect the fatigue scaling of Experiment 2 to produce comparable values for both SE and VP conditions or at least fatigue values no greater for the VP than the VI condition. While a mediational theorist might counter that the larger fatigue values could be due to a more complex set of implicit verbal processes, it is difficult to square such an account with the subjective reports. Again, the most consistent explanation requires the existence of at least two kinds of imagery, one speech and the other visual.

Several concluding notions are worthy of mention. First, the visual imagery studied here is probably serial in nature (instructions emphasize that only one letter at a time is to be processed). The conclusion that visual imagery is slower than speech imagery assumes serial processing. However, it seems clear to us that it is possible to visualize at least several letters at a time while it does not seem to be possible to speak, implicitly or explicitly, more than one letter at a time. Hence it is conceivable that some tasks could be processed as fast or faster with visual imagery than with speech imagery, provided that the task allowed the visual system to operate in parallel.

Second, visual imagery may be under verbal control when long serial lists like the alphabet are being processed. In fact, several Ss reported that they implicitly spoke each letter before visualizing it. If this is generally the case, then the measurement of VI time is made up of at least two components, the verbal control process and the generation of the image. But if we assume that it takes 13-14 sec to go through the alphabet in a VI mode and about 4 sec in a speech mode, it still means VI is slower than speech, even if 4 sec of the VI time is attributable to verbal control. Also, it seems clear subjectively that at least some visual imagery is not under verbal control. Whatever the case, there are fundamental differences in the rates and ways in which information is processed in visual and speech imagery modes.

Third, a few academic people (J. Watson, for example) have expressed difficulty in experiencing visual imagery. This may be because of the use of different defining and decision criteria for the nature of an image. Or perhaps Galton's (1880) finding that the prevalence of visual imagery decreases with increasing age and educational level has bearing on this point. Galton accounts for this by saying that those who engage in abstract activity tend to suppress visual imagery and develop instead verbal representations of the world—with which it may be easier to represent and manipulate abstractions. Irrespective of the level of facility for visual imagery, it seems apparent that the logic and method of its assessment used here do not differ from the logic and method for assessing implicit speech rates, an accepted undertaking (Landauer, 1962; Weber & Bach, 1969).

REFERENCES

BLANSHARD, B. The problem of consciousness: A debate. Philosophy & Phenomenological Research, 1967, 27, 317-324.

BROOKS, L. Spatial and verbal components of the act of recall. Canadian Journal of Psychology, 1968, 22, 349-368.

BUGELSKI, B., KIDD, E., & SEGMAN, J. Image as a mediator in one-trial paired-associate learning. Journal of Experimental Psychology, 1968, 76, 69-73.

GALTON, F. Statistics of mental imagery. Mind, 1880, 5, 301-318.

LANDAUER, T. Rate of implicit speech. Perception & Motor Skills, 1962, 15, 646.

PAIVIO, A. Latency of verbal associations and imagery to noun stimuli as a function of abstractness and generality. Canadian Journal of Psychology, 1966, 20, 378-387.

WEBER, R., & BACH, M. Visual and speech imagery. British Journal of Psychology, 1969, 60, 199-202.

WEBER, R., LOVE, W., & GOLDSTEIN, M. Numerically varied S-R mapping. Perception & Motor Skills, 1967, 25, 361-373.

NOTES

1. The work reported herein was supported in part by the Research Foundation, Oklahoma State University and in part by a grant from the U.S. Office of Education, Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by the U.S. Office of Education should be inferred.

2. Address: Department of Psychology, Oklahoma State University, Stillwater, Oklahoma 74074.

(Accepted for publication December 11, 1969.)