DOCUMENT RESUME

ED 079 715 CS 000 676

AUTHOR Weber, Robert J.
TITLE Visual Imagery for Letters and Words. Final Report.
INSTITUTION Oklahoma State Univ., Stillwater Research Foundation.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
BUREAU NO BR-2-0056
PUB DATE Jul 73
GRANT OEG-0-72-3601
NOTE 42p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Cognitive Processes; *Decoding (Reading); Learning; Memory; Orthographic Symbols; Spelling; *Visual Learning; Visual Perception
IDENTIFIERS *Visual Imagery

ABSTRACT

In a series of six experiments, undergraduate college students visually imagined letters or words and then classified as rapidly as possible the imagined letters for some physical property such as vertical height. This procedure allowed for a preliminary assessment of the temporal parameters of visual imagination. The results delineate a number of visual image phenomena: (1) visual image sequencing and formation is guided by implicit speech when long unpronounceable letter strings are used; (2) when short pronounceable words are used, the entire string can be simultaneously represented in visual imagination, and i plicit verbal control is required; (3) the capacity of the visual image system is very limited--between three to five letters for short easily pronounceable letter strings; and (4) visual image and visual percept representations of words are comparable for very short pronounceable letter strings, but as the length of the string increases visual image capacity for simultaneous representation is soon exceeded. (Author/WR)

ED 079715

CS

Final Report

Project No. 20056
Grant No. OEG-0-72-3601

Robert J. Weber
Research Foundation
Oklahoma State University
Stillwater, Oklahoma 74074

VISUAL IMAGERY FOR LETTERS AND WORDS

July 1973

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

National Institute of Education

CS 000 674

ABSTRACT

In a series of six experiments, subjects visually imagined letters or words and then classified as rapidly as possible the imagined letters for some physical property such as vertical height. This simple procedure allowed for a preliminary assessment of the temporal parameters of visual imagination. The results delineate a number of visual image phenomena:

(1) Visual image sequencing and formation is guided by implicit speech when long unpronounceable letter strings (alphabet) are used.

(2) When short pronounceable words are used, the entire string can be simultaneously represented in visual imagination, and no implicit verbal control is required.

(3) The capacity of the visual image system is very limited, between three to five letters for short easily pronounceable letter strings (words).

(4) Visual image and visual percept representations of words are comparable for very short pronounceable letter strings (three letter words), but as the length of the string increases visual image capacity for simultaneous representation is soon exceeded. When this occurs, sequencing to the next set of images again comes under implicit verbal control. In contrast, when visual percept capacity is exceeded by a physical string of letters, sequencing can occur through simple scanning or eye movement.

Final Report

Project No. 20056
Grant No. OEG-0-72-3601

VISUAL IMAGERY FOR LETTERS AND WORDS

Robert J. Weber

Oklahoma State University

Stillwater, Oklahoma 74074

July 1973

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

U.S. Office of Education
National Institute of Education

TABLE OF CONTENTS

INTRODUCTION

Subjective experience is once again becoming increasingly important to psychology. This is shown by the interest of the layman (especially students) in such things as mystical experiences, drug experiences, rock festival experiences, and sensitivity training. Each of these things is in part a reflection of current interest in subjective phenomena. Such interest is not confined to the layman. To take but four general examples, there is a recent book edited by Tart, entitled Altered States of Consciousness (1969), which deals with topics such as hallucinations, dreams, and meditation in terms of their subjective states. A second book edited by Jacobs and Sachs, The Psychology of Private Events (1971), is concerned with a variety of covert states. And two other books, Mental Imagery by Richardson (1969) and Mental Imagery in the Child by Piaget and Inhelder (1971), are concerned exclusively with the topic of imagery. As indicated in several of these books, as well as in numerous articles, one common approach to subjective states is to relate their presence or absence to the presence or absence of physiological indicators. This approach would be even more useful (1) if we could be assured that subjects were in fact having the subjective experiences they claim (i.e., eliminating the possibility that they sometimes respond on the basis of demand characteristics and (2) if we could describe subjective experiences in a richer language than that of simple occurrence and nonoccurrence. For example, it would be of interest to describe subjective experience in terms of certain parameters, such as rate of occurrence. If in fact objective parametric description is possible, subjective phenomena can be profitably studied independently of physiological indicators.

Such an approach is now possible with relatively simple forms of nearly universal subjective experience such as memory imagery (Richardson, 1969), the topic of this proposal. A general definition of memory imagery would be difficult, but it certainly would be a self-initiated, nearly universal, sensory-like experience in the absence of correlated external stimulation (Weber & Bach, 1969; Weber & Castleman, 1970). According to Richardson (1969) memory images would also be under voluntary control, and the content of memory images would correspond to the content of past sensory experience. Such a definition is quite broad but would exclude: (1) eidetic imagery (Haber & Haber, 1964), because it is rare at best, (2) after images (Craik, 1940), because they are not self-initiated, and (3) iconic storage (Sperling, 1960, 1967; Neisser, 1967), also because it is not self-initiated.

The occurrence of memory imagery--subjective sense-like experience--can be grounded in several theoretical views. For example, Paivio (1969) considers visual images to be conditioned sensations. A word like "car" would act as the conditioned stimulus for a host of conditioned responses in the form of visual images of cars with various colors and shapes.

Another view, that of Neisser (1967), holds that perception and imagery are the result of "analysis-by-synthesis." In this view percepts and images are both constructive acts. Percepts are the result of central operations matching or synthesizing the sensory input. The actions of the synthesizing then determine the subjective experience and memory representation obtained from a given stimulus. That is, percepts are not copies of a stimulus, they are synthetic constructions, activated in order to match the physical stimulus representation. In the same way images are the result of the synthesizing operations being activated in their own right, that is, without an externally present stimulus as would be the case with percepts.

Both of these views are perhaps of heuristic value, but at present they do not seem to provide much deductive mileage. In my view what is now needed is smaller-scale theorizing about selected imagery phenomena, without attempting to explain it all in one theory. If this is the case, then we need to study the basic properties of imagery (e.g., its temporal parameters).

In this proposal temporal aspects of visual memory imagery for letters and words are considered. The visual memory imagery examined here is nearly universal, as determined from subjective reports. In a number of the writer's classes over the last two years the incidence of visual imagery (in the sense of being able to imagine, with eyes closed, the spatial forms of successive letters of the alphabet) is about 90 percent (Weber & Castleman, 1970). Contrary to wide-spread impression, this figure is very close to that obtained by Galton (1880), if we include his subjects with relatively weak imagery (this seems to be reasonable, because even weak imagery should be sufficient to imagine a common form like a letter).

There have been a number of recent approaches to the study of visual imagery. A few pertinent examples follow. Posner, et al. (1969) in their Experiment III have evidence based on reaction times that auditorially presented letter names can be translated into a visual representation prior to matching (on a Yes/No basis) a subsequently presented printed letter. Paivio (1969) reviews a number of his studies indicating that learning in various tasks shows predictable relations to independently scaled imagery values of word stimuli. Segal and Gordon (1969) have shown that the instruction to form an image (visual or auditory) effects the subsequent detectability of perceptually presented signals (visual or auditory). When image and percept are in the same modality, detection is less proficient than when they are in different modalities, possibly due to mutual interference or noise when both tasks appear in the same modality. The work of Brooks (1967, 1968) is particularly impressive for its use of objective indicators. For example, Brooks has used block letters of the form: *F . The asterisk would represent a starting position (upper left hand corner) and the subject's task when presented orally with the letter name "F" is to generate a visual image of the block letter, scan that image clockwise, and indicate Yes/No for each successive corner of the form whether it is an extreme corner (top or bottom) or an intermediate corner. In this case the appropriate responses would be yes, yes, no, no, no, no, no, no, yes, yes. On the basis of various conditions Brooks is able to separate two distinct memory image modes, visual and verbal.

Two recent studies by the writer have emphasized the measurement of visual imagery rates with letters. The first study (Weber & Bach, 1969) used subjective procedures. In the visual imagery condition subjects were asked to visualize the successive letters of the alphabet, with their eyes closed, as they would appear one at a time flashed on an imaginary movie-screen.[1] The rate obtained was an average of about 2.5 letters per second, with considerable variability. This rate was in contrast to implicit and explicit speech rates, both of which are about 6 letters per second (Landauer, 1962). While these results were highly suggestive in pointing to separate imagery systems for visual and verbal representations, the variability associated with measuring visual imagery rates underlined the need for more objective procedures. Weber and Castleman (1970) found an interesting objective procedure for dealing with visually imagined characters. In their Experiment 2 lowercase printed letters of the alphabet were partitioned into two subsets, vertically large letters (b, d, f, g, h, j,...y) and those that were not vertically large (a, c, e, i, m,...z). Subjects were instructed to visually imagine such letters, in alphabetic sequence, and classify them for their vertical height. This task insured that the subject process each letter for its spatial property of vertical size. Subjects in fact reported using visual images of the letters in order to determine whether they were large or small characters. Experiment 3 was a control to determine whether subjects could have rotely learned, during the experiment, the appropriate classification response for each individual letter. The results were strongly contrary to such an interpretation. The possibility that vertical height properties were mediated by a verbal process acquired prior to the experiment did not accord with the subjective reports of visual images being generated nor with other evidence concerning fatigue ratings.

With the visual property condition, a pure measure of the time required to generate visual images was not obtained. In addition to the process of generation, the subject had to abstract from the generated image its height property and respond appropriately; and this would have taken extra time. However, the visual property condition seemed to serve well for establishing what a visual image is and for establishing a minimum standard of clarity (the image must be clear enough to determine whether it is a large letter). In Experiment 2 when the visual property condition was combined with the subjective procedure of Weber and Bach (1969) the rate of generation for the subjective condition (less than 2 letters per second) was even slower than previously found by Weber and Bach. In addition, the shaping of the imagery process by the objective visual property condition was quite effective in reducing variability for the subjective procedure.

In the studies that follow, subjects are instructed imagine letters of the alphabet and to describe for each imagined letter a related objective spatial visual property, for example, to classify it as a vertically large or small letter. The principal response measure is time to process a letter string.

Experiments 1, 2, and 2': Verbal Control of Visual Imagery? (This series of experiments has now been published with the following reference: Weber, R. J., Kelley, J., and Little, S. Is visual image sequencing under verbal control? Journal of Experimental Psychology, 1972, 96, 354-362. Experiment 2' was not part of the original proposal but has been added for completeness.)

EXPERIMENTS 1, 2, AND 2'

Abstract

Three experiments were conducted to examine the conjecture of R. J. Weber and J. Castleman in 1970 that sequencing between visual images of letters if sometimes under the control of verbal imagery (implicit speech) in the sense of saying a letter to oneself before visualizing it. Verbal control of visual image sequencing was investigated in alphabetic and word letter strings. Results with alphabetic letter strings were consistent with a verbal control process. However, results with word letter strings did not provide evidence for verbal control of sequencing, presumably because the name of a short word can serve to represent its several letters as a spatially parallel chunk in visual imagination. It is concluded that the visual imagery system has a limited capacity for spatially parallel representation. When this limited capacity is exceeded, as in alphabetic strings, there is implicit verbal control over visual image sequencing.

It has been suggested that nonverbal imagery (visual imagery would be an example) and verbal processes (verbal imagery in the form of implicit speech would be an example) are our two primary modes of symbolic representation (Bower, 1970; Paivio, 1971). In addition, one of the important problems of future research in cognitive psychology is seen as determining how visual imagery and verbal processes are related to one another (Bower, 1970; Paivio, 1971). In this paper interrelations of visual imagery and verbal imagery (implicit verbal processes) are examined. By "visual imagery" we mean the ability to deal directly with memorial representations of spatial information, and by "verbal imagery" we mean the use of memorial representations of linguistically related information in the form of implicit speech (Weber & Castleman, 1970). Both imageries may be accompanied by subjective impressions of "seeing" or "speaking", although as Paivio points out this is not necessarily the case.[1]

Experiment 1 deals with possible verbal control of visual image sequencing in an alphabetic list, plus a secondary question involving the degree of generality of scanning rates for different imagined letter forms. Experiment 2 serves as a control for response mode time, which was not considered in the first experiment. Experiment 2' is again addressed to the problem of verbal control of sequencing between visual images, but this time word letter strings as well as alphabet

[1]In the first author's classes over the past several years (N = about 500) about 95% of the undergraduates claim the ability to visually imagine distinctive letters of the alphabet. Fully 100% of the same students claim the ability to name letters in implicit speech.

letter strings are examined. The comparison of word and alphabet strings is made on the conjecture that entire word images, rather than their individual letters, can be represented as a single chunk; i.e., several letters at a time can be generated by the word's name and then held in spatially parallel representation in visual imagination.

Experiment 1

Two recent studies have measured the rate at which visual and verbal imagery occur. The first study (Weber & Bach, 1969) used entirely subjective procedures. In the visual imagery condition, Ss were asked to visualize rapidly the successive letters of the alphabet, with their eyes closed, as they appeared one at a time flashed on an imaginary movie screen. The rate obtained was slightly more than 2 letters/sec. This rate was in marked contrast to verbal imagery rates with implicit and explicit speech, both of which were about 6 letters/sec, in close agreement with values from recent work using more objective procedures (Weber & Blagowsky, 1971). While these studies were highly suggestive in pointing to separate imagery systems for visual and verbal representations, the variability associated with measuring visual imagery rates underlined the need for more objective procedures. Hence, in another study, Weber and Castleman (1970) developed an objective procedure for dealing with visually imagined characters. In their Exp. 2, lowercase printed letters of the alphabet were partitioned into two subsets. vertically large letters (b, d, f, g, h, j, ...y) and vertically small letters (a, c, e, i, m...z).[2] The Ss were instructed to visually imagine letters, in alphabetic order, and to classify each successive letter for its vertical height. This task insured that S process each letter for its spatial property of vertical size. The Ss in fact reported using visual images of the letters in order to determine whether they were large or small characters. When required to abstract from the image and verbally report the visual property of each letter, the rate dropped to about 1 letter/sec. Experiment 2' was a control to determine whether Ss could have rotely learned, during the experiment, the appropriate classification response for each individual letter. The results were strongly contrary to such an interpretation. The additional possibility that vertical height properties were mediated by, say, a rote verbal process acquired prior to the experiment did not accord with the subjective reports of visual images being generated nor with other evidence concerning fatigue ratings.

However, verbal mediation may have been involved in a more interesting sense. In Weber and Bach (1969) and Weber and Castleman (1970), the sequencing between visual images may have been under verbal control. In fact, several Ss reported at the close of the experiment that they

---

[2]Some letters ("i" - NO, "t" - YES) are slightly ambiguous in some type fonts, but once they are designated by E they present no further difficulties (Weber and Castleman, 1970, Exp. 2).

had to speak each letter implicitly before visualizing it. This would be consistent with the findings of Posner, Boies, Eichelman, and Taylor (1969), which indicate that Ss can use letter names to generate visual characteristics of letters. It is not clear how general such a strategy is, nor whether it is optional or obligatory. For long serial lists like the alphabet, perhaps implicit verbal generation is needed to establish the sequential order of items, since the ordering capabilities of memory for unrelated visual materials appear to be limited in comparison to verbal materials (Paivio, 1971). Whatever the case, it seemed necessary to examine directly the possibility of verbal generation and control of visual image sequencing in ordered letter lists like the alphabet.

Examination of visual image and implicit verbal relations is made possible by having an objective indicator for imagery (Weber & Castleman, 1970) and by the fact that implicit and explicit (aloud) speech take place at about the same rate (Weber & Bach, 1969; Weber & Blagowsky, 1970). The experimental question concerns whether an aloud scan mode (saying each letter of a list before reporting its visual property) will result in the same rate of reporting visual properties as an unspecified scan (not aloud), in which S does not say successive letters explicitly but just reports their visual properties. If the rates are comparable for the two kinds of scan, then we would have evidence independent of subjective impressions that implicit speech can accompany the sequencing of successive visual images in the not-aloud scan. This, together with the subjective impression of having to say each letter before visualizing it, would be consistent with the notion of verbal generation and control of visual image sequencing. However, the implicit speech accompaniment might be purely ancillary, that is, without functional significance. So still another way is required to approach the question of verbal sequential control, and to this end we examine also the role of response indicators in denoting the spatial characteristics of visual images. Suppose that visual image sequencing is mediated by saying successive letters prior to visualizing them (the implicit speaking might be necessary to establish the sequential ordering of the letters). Then we would expect a spoken response indicator (YES or NO as to whether the letter has the property in question) to be produced and output before the next letter could be implicitly named and then visualized (saying both YES/NO and speaking the name of the next letter could not take place simultaneously). But if the response indicator were written (a line "/" for YES and a dot "." for NO), then it might be possible to begin saying implicitly the name of the next letter and producing its image while executing the written response classification for the preceding letter. This, of course, assumes that writing and speaking are at least partially independent of one another to the extent that Ss are able to say one thing while writing something else.

Method

Subjects.--The Ss were 48 undergraduate volunteers who received extra course credit for their participation. There were 12 Ss randomly assigned to each of four different between-Ss conditions. Two additional Ss were discarded for not following instructions and/or having extreme difficulty in completing the task.

Procedure.--Upon entering the experimental situation, Ss were checked for their familiarity with the alphabet and given remedial instruction if necessary. All Ss were shown an index card with lowercase, typewritten, letters of the alphabet in random sequence, and then instructed to differentiate between vertically large letters (YES responses) and those that were not vertically large (NO responses). This distinction was similar to that used in the Visual Property conditions of Exp. 2 and 2' in Weber and Castleman (1970), although YES/NO serve as the responses here rather than the LARGE/SMALL of the earlier study. During the experiment proper, S was to imagine visually the letters in alphabetic order and, as rapidly as possible, to classify each imagined letter as YES/NO. At the beginning of each trial, S started a remotely controlled Standard Electric clock when he was ready to begin and stopped it as soon as he had classified the last letter of the alphabet. The S did not receive feedback on his times, but speed and a low error rate were emphasized. If an error did occur, it was pointed out to S at the end of the trial. This constituted the Lowercase (LC) condition. In addition, there was a comparable Uppercase (UC) condition in which Ss distinguished between letters with long vertical lines (B, D, E, F, H, I...) which required YES responses and those without long vertical lines (A, C, G, J, O, Q,...) which required NO responses. Each S was to imagine both LC and UC alphabetic lists alternated over 10 blocks of LC and UC trials. The LC and UC conditions were balanced for serial position across Ss.

There were also four between-S conditions. The first two involved scanning mode. An S scanned the alphabet by speaking it aloud (SA) or by a nonspecified, not-aloud (NA) procedure as he generated in imagination successive letters to examine and respond to their visual properties. In the SA conditions, S would proceed as follows (for LC letters): "a"-NO, "b"-YES, "c"-NO, ..., "z"-NO; that is, speak each letter before responding with YES or NO to describe the vertical height of the imagined letter. In the NA conditions, no instruction was given regarding scanning, and S did not speak aloud the successive letters before responding overtly with YES or NO. That is, he would proceed as follows: NO, YES, NO, ..., NO. It was, of course, possible that S would generate the successive letters in speech imagery before generating their visual images, even though he received no instruction to do so, but this was an empirical question to be answered by the similarity of processing rates for the SA and NA conditions. Two other between-S conditions involving spoken or written responses (SR, WR) were combined factorially with the scanning mode conditions. In the SR conditions, S simply said YES or NO for the visual property of each successive letter. In the WR conditions, S wrote a long vertical line "/" corresponding to a YES response or wrote a dot "." corresponding to a NO response. The S was to classify the alphabetic string as rapidly as possible and to keep his eyes closed, in all conditions, so he would not visually monitor his WRs.

In summary, the design consisted of a factorial arrangement of 2 Scanning Modes (SA, NA) x 2 Response Classes (SR, WR) x 2 Letter Cases (LC, UC), with scanning mode and response class between Ss and letter case within Ss. The principal response measure was time to process the 26 member imagined alphabetic list.

Results and Discussion

Descriptive findings are presented in Fig. 1, in which mean time in seconds is shown as a function of practice. The left ordinate shows mean time per 26 letters and the right ordinate is the same information on a time-per-letter basis. The left-hand panel is for lowercase letters and the right-hand panel, for uppercase letters (the bottom half of the left panel contains results from previous experiments that will be commented on later). The parenthetical number triple next to each function represents the mean, standard error of the mean, and the scan rate in letters per second processed. Each mean is based on 120 observations (12 Ss x 10 Blocks); the standard error of the mean is a between-Ss measure of variability based on each S's mean; and the scan rate is simply the reciprocal of the mean. Errors (omission and commission) occurred for fewer than 2% of the letters (a phase error was counted only once). Those trials on which errors did occur were not temporally disparate from correct trials, so no further distinction is made.

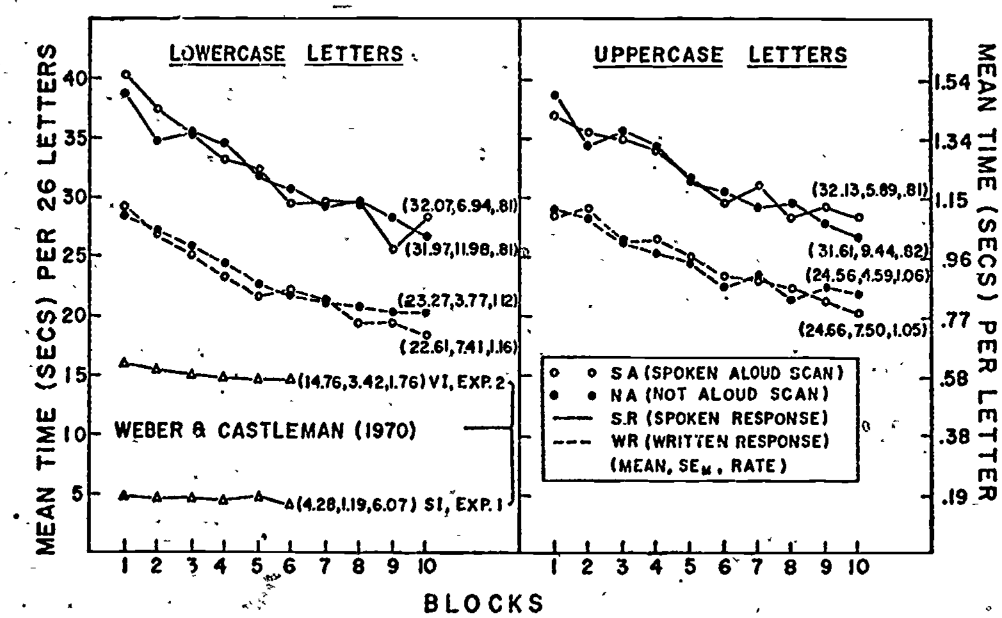

FIG. 1. Response time as a function of scanning mode, response mode, and blocks. (Left panel is for lowercase letters and right panel for uppercase letters.)

Mean processing times are almost identical for spoken scanning (SA) and unspecified scanning (NA), with $F\ (1, 44) < 1$. The absence of a significant scan effect is consistent with the view that implicit speech scanning is involved in the NA conditions. However, the differences in variability for the two scans do suggest at least some ~~difficulty with~~ assuming that completely identical processes are involved.

There is a substantial difference between WR and SR modes, $F\ (1, 44) = 14.55$, $p < .01$. Employing a WR mode takes much less time than an SR mode, even though writing is probably a slower response system than speaking. Again this is consistent with the view that sequencing between visual images is accompanied by implicit verbal processes. SRs appear to delay the hypothesized verbal control processes involved in generation and sequencing between successive visual images of letters.

Finally, the results for upper- and lower-case letters suggest that visual imagery rates are about the same for two quite different classes of visual forms, $F\ (1, 44) = 2.56$, $p > .05$. This would tend to increase the generality of prior findings (Weber & Bach, 1969; Weber & Castleman, 1970) by suggesting that visual imagery rates are constant over fairly broad fluctuations in visual form. Just how large a variation in visual form would be possible before differences in rate occurred is open to question. None of the main effects showed significant interactions with one another.

Large practice effects are also evident in Fig. 1. This is in contrast to some other imagery tasks, as shown in the bottom half of the left panel. The visual imagery (VI) function is from Weber and Castleman (1970, Exp. 2) and did not require S to abstract the height properties of the letters or to respond YES/NO. The procedure was entirely subjective; S simply stopped the clock when he had completed visualizing, one at a time, the letters of the alphabetic string. The large practice effects with the visual property task of the present study would seem, then, to be a result not of visual image representation but of abstraction and/or overt response time. Both response conditions of the present study seem to be stabilizing somewhat above the VI condition, as would be expected if abstracting height properties and responding YES/NO takes additional time over visualizing letters. The bottom function is for verbal or speech imagery (SI) and also from Weber and Castleman (1970, Exp. 1). It simply required S to say to himself, as rapidly as possible, the letters of the alphabet and stop the clock when finished. It provides an estimate of the minimum component time that would be required of a verbal control letter-naming process in a VI condition.

In reviewing the evidence for verbal control in the sequencing of images, perhaps most compelling is the multiple approach to the verbal control issue. The <u>absence</u> of a significant scanning effect (SA, NA) is consistent with an implicit speech letter-by-letter naming process in the NA conditions. The <u>presence</u> of a significant response mode effect (SR, WR) is consistent with an implicit speech process, generating and controlling the sequencing of letters, which is delayed by the overt verbal response of categorizing letters as YES/NO. And once again, there is the strong subjective impression of having to say each letter before visualizing it.

A cautionary note is still in order. The response mode effect would come as no surprise if writing were in fact a more rapid response mode than speaking. This unlikely possibility is considered in Exp. 2. Another related possible explanation of the response mode effect is that the written line and dot responses are initially more compatible with the height properties of visual images than are the spoken YES/NO responses. But if this were the case, we would expect WR and SR mode functions to converge rapidly across blocks of practice, something that does not occur.

Each of the findings and arguments for verbal control may not be completely compelling in its own right, but taken together they provide a strong line of converging evidence pointing to the simple notion that S says each letter before visualizing it. The generality of such a notion is immediately suspect, and Exp. 2 in this series is directly addressed to the question of when verbal accompaniment of visual image sequencing does and does not take place.

Experiment 2

It seems obvious that in most instances "equivalent" materials can be spoken at least as rapidly as they can be written. The findings and interpretations of Exp. 1 depend on such an assumption. However, we could not seem to find evidence on this point in the recent psychological literature. Also it is not completely clear how equivalent spoken YES/NOs are to written lines and dots--perhaps the written responses are more rapid. A direct comparison of writing and speaking rates is called for, in order to strengthen the rationale of Exp. 1. Such a comparison also has intrinsic interest since writing and speaking are, perhaps, our two most important means of signaling information to others about our internal states. Moreover, writing and speaking are widely used as response modes in psychological research, without much regard to their potentially very different output rates and the implications this has for, say, short-term-memory decay work.

In this experiment, then, response times for speaking and writing are assessed. For future use, two different symbolic materials are employed: an alphabetic list and a binary list consisting of spoken or written equivalents of YES/NO. The written analog of YES/NO is, as in Exp. 1, a vertical line and dot, respectively.

Method

Subjects.--The Ss were 10 volunteers from an undergraduate psychology class.

Design and procedure.--A factorial design was used, with 2 Types of

String (Alphabet, Binary) x 2 Response Modes (Written, Spoken). Both factors were within Ss. In the written condition of the alphabet task, Ss were asked to write the alphabet in their normal connected cursive handwriting as rapidly as possible. (Response time was always the dependent variable.) In the spoken alphabet condition, Ss were asked to speak the alphabet aloud as rapidly as possible. In the spoken binary condition, Ss spoke aloud a 26-member series of alternating YES/NOs (YES, NO, YES, NO, ...). In the written binary condition, Ss wrote out a 26-member series of alternating lines and dots ("/./." etc.). The lines and dots correspond to the YES/NO written classification of Exp. 1. In the binary tasks, we did not wish to require Ss to count their responses to have only 26, so a sheet of paper was divided into seven columns 1 in. wide. For the written binary task, S wrote his responses in clusters of 4/column (/./.; /./.; etc.), except for the last column which required only 2 responses to equal the 26. For the spoken binary task, S kept track by moving his finger to a new column on completion of each cluster of four spoken responses.

Procedures for timing were the same as for Exp. 1. Each S was given six trials in each of the four conditions. A trial consisted of a single processing, i.e., writing or speaking an alphabetic or binary list. Following the six trials of the first condition, each S was given the six trials of the next condition, etc. The orders of the four conditions were randomized for each S.

Results and Discussion

Mean response time in seconds for each task is shown in Table 1. The response time for the WR mode is greater than for the SR mode on both the binary and alphabet strings. An analysis of variance shows that the type of response has a significant effect, $F\ (1, 9) = 172.22$, $p < .01$. The type of string does not have a significant effect, $F\ (1, 9) = 1.84$, $p > .05$. However, there is an obvious significant interaction between type of response and type of string, $F\ (1, 9) = 99.92$, $p < .01$. The main comparison of a priori concern is that of the binary spoken and written conditions. Spoken responding is reliably faster than written responding, with correlated $t\ (9) = 3.87$, $p < .01$.

The greater response time for the WR mode as compared to the SR mode for both the binary and alphabet strings reflects in part response times for different motor acts. The significant String x Response interaction can be interpreted as meaning that with written responses, Ss can execute the simple vertical lines and dots required by the binary task faster than the more complex curved lines required by the alphabet task. But when spoken responses were used, Ss could say the alphabet faster than the string of 26 alternating YES/NOs.

These results suggest that the differences between WR and SR modes of Exp. 1 may have been conservative, since in sheer response time (unaccompanied by a visual image task and with a completely regular alternation of YES/NO responses), written responding is slower than spoken responding. It is, of course, possible to argue that sequences other than an alternating YES/NO would produce different results, that different sequential patterns would interact differently with the two response modes, but such effects, if they occur, are likely to be small indeed compared to the basic distinction between writing and speaking.

TABLE 1

MEAN AND STANDARD ERROR OF THE MEAN FOR PROCESSING 26 ITEM BINARY OR ALPHABETIC STRING, EXPERIMENT 2

Statistic	Binary		Alphabet	
	Written	Spoken	Written	Spoken
$\bar{X}$	9.37	7.50	12.87	4.71
SE_M	1.55	1.80	2.07	1.30

Note.--Means averaged over 10 Ss x 6 trials for each condition. SE_M is measured in seconds.

Experiment 2'

Experiment 2 indicated that sequentially generating the visual image representations of an ordered list like the alphabet is under verbal regulation. But it is clearly the case that sequential ordering for at least some visual images is not verbally controlled in the sense of first having to say a verbal counterpart of each imagined item in the list. Likely examples would include dreams, hallucinatory experiences, and visual reverie. Such images seem to have a sequential life of their own, and their content and order often defy verbal description. Another case in which individual images do not seem to be under sequential verbal control would seem to be spatially parallel composite images, that is, images with several different individual components appearing simultaneously but separated in subjective space. An example would be a short word, the letters of which might be imagined at the same time. Thus the letters of the word "cat" might be represented in visual imagination sequentially "c", "a," "t," or simultaneously "cat."

Extracting visual spatial properties from a letter string that is sequentially generated (alphabet) may well be different from abstracting visual spatial properties from a letter string in which all letters might be represented simultaneously in imagination, as in a short word.

In this experiment, the relative rate of abstracting visual properties from different types of letter strings, i.e., from the alphabet and from words, is measured under different scanning and response conditions. By using two different scanning modes, spoken and nonspoken, and two different response modes, written and spoken, the presence or absence of individual verbal letter generation in visual imagery for word versus alphabet strings can be investigated. If visual image formation of the successive individual letters of a word is also governed by saying implicitly each successive letter, then relative scanning times should be the same for spoken and nonspoken scanning, and written responding should be faster than spoken responding. The rationale is the same as Exp. 1. However, suppose the word name (together with the appropriate instructional set) is sufficient to arouse a spatial simultaneously available visual image of the entire word (or at least several letters of it). Then speaking each individual letter of the word would not be necessary prior to visualizing its letters, and an NA scan should take less time than a letter-by-letter SA scan. Also, the difference between the SR and WR modes should diminish for the NA scan in comparison to the SA scan, since in the former condition full advantage could be taken of the simultaneously available image without the additional potentially disruptive process of saying each successive letter one at a time.

Method

Subjects.--The Ss were 40 undergraduate college students who participated in order to fulfill an introductory psychology course requirement. Two potential Ss were not used because they could not recite the alphabet correctly. One other potential S was not used because he could not learn to represent imagined large and small lowercase, typed letters of the alphabet with YES and NO, respectively. The Ss were assigned at random to conditions, with 10 Ss per between-Ss condition.

Design and procedure.--The experimental tasks used in this experiment involved the use of the alphabet and 18 four-letter words; all words were of one syllable and were high-frequency-usage words (defined as A or AA by Thorndike and Lorge). Each of the words contained both vertically large and small letters, e.g., "girl." From images of the two different letter strings (words and alphabet), Ss were to abstract visual properties of the successive letters. Thus the image "girl" would give rise to the response sequence YES, NO, NO, YES. The design and procedure were similar to that of Exp. 1 and consisted of a factorial arrangement of 2 Response Modes (Written, Spoken) x 2 Scans (SA, NA) x 2 String Types (Alphabet, Word). The first two factors were between Ss and the last factor was within Ss.

Each S was given several preliminary practice trials for abstracting visual properties from the letters of the alphabet and for abstracting visual properties from the letters of a practice word. A spoken ready signal was given, and E started the remotely controlled clock with the onset of the spoken cue ("alphabet" or "X", where X was one of the stimulus words). The S was to respond as rapidly as possible by classifying the successive letters as YES/NO. The E also stopped the clock on completion of the visual property "spelling" of the word or alphabet, i.e., after the response to the fourth or twenty-sixth letter. This was a departure from the previous two experiments in which Ss controlled the clock. The departure was necessary because pilot data showed that E was more accurate than Ss in measuring the short response times involved with words in this experiment. The response times were measured from the onset of the alphabet or word cue to the response to the last letter in a string. To provide greater accuracy in measuring response times, E also tape-recorded all responses and later played back the tapes at half-speed. The S's spoken responses were recorded as usual. To tape-record S's written responses, a soft lead pencil was used to write on an aluminum sheet to which a microphone was taped. The distinctive sounds made by S's pencil on the aluminum sheet were readily picked up and served to demarcate response onsets. Again, S did not visually monitor his responses.

Each S was given 10 trial blocks. Each block was composed of 6 stimulus words of 4 letters each and one alphabet string. This allowed for roughly the same amount of work per block for both types of letter string (actually 24 letters for word strings and 26 letters for the alphabet string in a given block). The 6 words for each S were randomly selected from the pool of 18, 4-letter words. The same 6 words appeared in each of the 10 trials given to a particular S, but different Ss received different sets of 6 words to increase generality. The order of appearance for words and the alphabet was randomized from block to block.

Results

Each S's mean response time across trials was calculated. In order to compare directly S's responses in processing the word and the alphabet letter strings, his response times in each condition were computed on a time-per-letter basis. Therefore, each S's mean processing time per letter was the unit used in the subsequent analyses.

As can be seen from Table 2, the processing times are slightly faster in the NA scan conditions than in the SA scan conditions, F (1, 36) = 5.53, $p < .05$. Processing times are also faster in the word condition than in the alphabet condition, F (1, 36) = 97.99; and faster in the WR condition than in the SR condition, F (1, 36) = 46.01, both $ps < .01$.

TABLE 2

MEAN, STANDARD ERROR OF THE MEAN (SE_M), AND RATE AS A FUNCTION OF STRING, SCAN, AND RESPONSE MODE, EXPERIMENT 2

String and Scan	Response Mode: Spoken M	Spoken SE_M	Spoken Rate	Written M	Written SE_M	Written Rate
Alphabet						
SA	1.01	.14	.99	.61	.16	1.64
NA	.96	.27	1.04	.60	.15	1.67
Words						
SA	.86	.10	1.16	.49	.12	2.04
NA	.53	.09	1.89	.47	.12	2.13

Note.--Means are determined by averaging over 10 Ss and 10 trials per condition. SE_M is measured in sec/letter and rate is measured in letters/sec.

Of particular interest are the interaction effects in this experiment. There is a significant Scan x String interaction $F\ (1, 36) = 11.20$, $p < .01$, indicating the differences between spoken and nonspoken scans were greater in the word condition than in the alphabet condition. There is also a significant Response x String interaction, $F\ (1, 36) = 15.27$, $p < .01$, showing that the differences between SRs and WRs were greater in the alphabet condition than in the word condition. Finally, there is a significant three-factor Response x Scan x String interaction, $F\ (1, 36) = 11.20$, $p < .01$, indicating the relatively unique character of word strings with SRs and NA scanning.

Practice effects across trial blocks warrant comment. They were relatively flat for word strings, with no indication of a Condition x Practice interaction. The alphabet strings had practice effects similar to the corresponding conditions of Fig. 1, Exp. I, again with no indication of a Condition x Practice interaction.

Discussion

The equivalence of SA and NA scanning for alphabetic strings agrees with the findings of Exp. 1 and of Weber and Blagowsky (1970). The present results are also consistent with the notion that the visual imaging of individual letters in an alphabet string is under verbal control in the sense that each letter is implicitly spoken prior to imagining it. In contrast, the results for word strings are not equivalent for spoken and nonspoken scanning and are therefore inconsistent with the notion that S implicitly speaks each letter of a word when using a nonspoken scan mode. Verbal control over the sequential generation of visual images of the letters is not needed with the word letter-strings used in this study. Instead, on the basis of a word's spoken name, S may visually generate the word as a whole, or at least with more than one letter at a time simultaneously represented in visual imagination. The S could then scan left to right for the visual properties of the simultaneously available letters without having to fall back on implicit verbal processes. This is indeed the subjective impression of what is done. The time advantage with nonspoken scanning would arise because the names of the individual letters constituting the word would not need to be verbalized, since the visual image representation of the word would be simultaneously available prior to scanning individual letters. Hence, with word strings the verbal YES/NO responses should be comparable in rate to the written YES/NO equivalents, as is the case. Along these same lines, the sequential generation problem disappears, even for alphabet strings, if the string is perceptually available. Weber and Kelley (1972) have found that the spoken/written difference for the visual property task disappears when the alphabet string is visually present in front of S. They hypothesize that the sequential ordering problem is handled in that case by left-to-right eye movements over the string rather than through the use of implicit verbal processes as suggested here.

While the processes suggested for word and alphabet strings in the present series of experiment may seem dichotomous, it is likely that scanning a long polysyllabic word for the visual properties of its letters would be part way between the limiting cases of short familiar words versus long serial alphabetic lists. We would expect spatially parallel visual image representation to be possible within familiar letter clusters with a one- or perhaps two-syllable name, but sequencing between syllables for longer words might well be under verbal control, as with the alphabetic strings. Thus, the visual property scanning of the imagined lowercase word "oklahoma" might allow for the spatially parallel visualization of letters within a syllable, while the sequencing between syllables would involve verbal processes. This would certainly seem to be the case with a word like "supercalifragilistic-expialidocious."

The picture that emerges is of a visual imagery system with a capacity for limited, spatially parallel representation (Paivio, 1971). When a familiar letter string with a short name falls within that capacity, visual image sequencing is not verbally generated. But when a letter string exceeds that capacity, sequential generation and ordering become necessary and visual imagery comes under the control of verbal imagery in the form of implicit speech.

References

Bower, G. Analysis of a mnemonic device. American Scientist, 1970, 58, 496-510.

Paivio, A. Imagery and verbal processes. New York: Holt, Rinehart & Winston, 1971.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. Journal of Experimental Psychology Monograph, 1969, 79 (1, Pt. 2).

Weber, R. J., & Bach, M. Visual and speech imagery. British Journal of Psychology, 1969, 60, 199-202.

Weber, R. J., & Blagowsky, J. Metered memory search with implicit and explicit scanning. Journal of Experimental Psychology, 1970, 84, 343-348.

Weber, R. J., & Blagowsky, J. Metered memory search and concurrent chanting. Journal of Experimental Psychology, 1971, 89, 162-170.

Weber, R. J., & Castleman, J. The time it takes to imagine. Perception & Psychophysics, 1970, 8, 165-168.

Weber, R. J., & Kelley, J. Aspects of visual and acoustic imagery. Psychonomic Science, 1972, 27, 121-122.

EXPERIMENT 3: IMAGE AND PERCEPT REPRESENTATION OF LETTERS

The purposes of this study are to compare processing rates for visual and acoustic imagination, to compare rates of abstracting visual and acoustic properties from imagined vs perceptual representations of letters, and to show by manipulation of response mode and representation the possibility of verbal control of visual image sequencing. Rates of processing letters in visual imagination have been previously investigated (Weber & Bach, 1969; Weber & Castleman, 1970), and the sequencing problem has been dealt with at some length by Paivio (1971). The sequencing problem pertains to how, in a serial list of visual images, the S gets from one visually imagined form to another. In the Weber et al studies, Ss claimed, subsequent to the experiment, that it was necessary to say implicitly each letter before visualizing it. It is hypothesized here that the series of visually imagined letters of the alphabet is verbally ordered in the sense of S's saying implicitly each successive letter before visualizing it. If so, then we have a potentially important interrelationship between our two most important symbolic codes, visual imagery and verbal processes (Paivio, 1971). Investigation of these foregoing problems seems to require an objective method of assessing the existence of visual imagery.

Following Weber & Castleman (1970, Experiment 2), it is possible to employ objective criteria of imagining letters and to assess the rate at which letters are processed in imagination. These investigators had Ss classify imagined lowercase letters on the basis of height. Some letters are vertically small (a, c, e, i, ..., z) and other letters are vertically large (b, d, f, g, ..., y). The Ss were instructed to imagine visually the successive alphabetic letters and to classify each successive letter for its vertical height. The Ss reported visually imagining and were able to correctly classify the letters, using a verbal response ("small," "large") at a rate of about 1 letter per second. The rate differences for the objective and subjective procedures require comment. If Ss did, in fact, say implicitly each successive letter before visualizing it, then we would expect that a verbal classification response would compete for processing capacity with the covert verbal sequencing. If this is the case, then processing rate ought to be faster with a nonverbal classification response. Thus, if Ss are to give either a verbal YES/NO or a written line/dot equivalent (/,.) in responding to each imagined letter, we would expect slower rates for the verbal response mode than for a written response mode if verbal control of the sequence is required. This would be due to competition for verbal processing space on the part of both the overt verbal YES/NO response and the implicit verbal control of sequencing.

The general technique of abstracting from images the correlates of physical stimuli also can be applied to the study of acoustic imagination. Letter names possess acoustic properties that can be used as objective indicators of acoustic imagination. Thus, some letter names have a long e sound, /i/ (b, c, d, e, ..., z), and other letter names do not (a, f, h, ..., y). If S can correctly classify successive alphabetic letters according to their acoustic properties without actually saying them aloud, we can be reasonably sure that acoustic imagination is at work. When doing this task, there is again the strong impression of saying each letter prior to extracting its acoustic property. This may be necessary only for producing and abstracting the acoustic property of the imagined letter, but it may also have a sequential control function, as previously conjectured for visual imagery.

In some preliminary work, we have found the actual rate of letter classification to be very similar for visual and acoustic properties. But, subjectively, visual and acoustic imagination are very different. Accordingly, we have sought to find at least one objective variable that clearly separates them. The variable selected was imagined vs perceptual representation of the alphabetic sequence. It was reasoned that letters visually present (percepts) would have their visual properties immediately available for abstraction, whereas their acoustic properties would be no more available than in imagination. In addition, when the letters are visually present, as percepts, the sequencing problem could be handled by eye movements rather than by implicit verbalization. Hence, the differences in rate between written and verbal response modes should diminish.

Method

Subjects.--Twenty undergraduate volunteers were paid for their participation. One additional S was discarded for inability to do the task. Ten Ss were assigned at random to each of two between-Ss conditions.

Design and Procedure.--The design consisted of a factorial arrangement of two methods of representation (image, percept) by two properties (visual, acoustic) by two response modes (spoken, written). Response mode was between Ss, and the other two factors were within Ss. Representation could either be imaginal, meaning that the Ss somehow implicitly generated the successive alphabetic letters, or perceptual, meaning that a printed lowercase serial list of the alphabet was visually present before the S while he made his responses. The properties of the letters were either visual (in which case S had to distinguish between vertically large and vertically small letters of the lowercase alphabet) or acoustic (in which case S had to distinguish between letters with the long e sound, /i/, in their names from those that did not). Vertically large letters and long e letters were to be classified as YESs, and the other letters were to be classified as NOs. For the spoken response mode, S said YES or NO for each successive letter of the alphabet, as he classified it for its properties. Thus, in the visual property-image condition, S would emit a string of the following sort: NO, YES, NO, YES, ..., NO, because the successive letters of the alphabet (a, b, c, d, ..., z) have that pattern of YES/NO classifications of vertical height. If the letters were to be processed for their acoustic properties, the corresponding string of YES/NOs would be: NO, YES, YES, YES, ..., YES. When all letters are so considered, there are, for visual properties, 12 YES and 14 NO letters.[1] For acoustic properties, there are 9 YES and 17 NO letters. Hence, there is somewhat less response uncertainty for acoustic than for visual properties. For the written response conditions, the same classification was employed, except that a vertical line and dot were chosen to correspond to YES and NO, respectively. The S wrote his response on a blank 8½ x 11 in. sheet, but he did not visually monitor his writing.

The nature of the acoustic and visual properties of letters was pointed out to S, as well as the particular response mode he was to use. He was then given four practice trials, one for each condition. He was told not to make any errors. Later, during the experiment proper, if an error was made, it was pointed out to S at the end of the trial. This procedure leads to low error rates (Weber & Castleman, 1970), and errors are not further considered in this study.

The order of presentation of the image-percept by visual-acoustic property conditions was random within each of six blocks. There were four different random lists of the six blocks. The beginning of a trial began with the E's designation of the trial as percept/image and the presentation of a card indicating whether visual or acoustic properties were to be processed. Response time for processing the 26 letters of the alphabet was the dependent variable. The response time interval began with presentation of the card and ended with S saying, "Stop," after completing the last letter of the alphabet. Times were recorded on a Standard Electric clock to the nearest .01 sec. The S was not given temporal feedback, but he was urged to go as rapidly as possible. There was about a 30-sec interval between successive trials while the E recorded the response time and reset the clock.

Results and Discussion

Table 1 shows mean times to process the 26 letters of the alphabet as a function of conditions. Each mean is based on 10 Ss and six trials. The SE_M is a between-Ss measure of variation. Significance tests for main effects indicate the following: Response time is greater for image than for percept representation, $F(1,18) = 121.84$, $p < .01$ Response time is greater in the spoken than in the written mode, $F(1,18) = 5.53$, $p < .05$. And response time is greater in the acoustic than in the visual property conditions, $F(1,18) = 15.85$, $p < .01$.

The findings for main effects must be qualified, however. Perhaps of most interest are the significant two-way interactions: Response Mode by Image-Percept Representation, $F(1,18) = 56.57$, $p < .01$; and Visual-Acoustic Property by Representation, $F(1,18) = 30.77$, $p < .01$. The significant Response Mode by Representation interaction indicates that the differences between spoken and written responses were much greater in the image than in the percept conditions. This is to be expected if verbal sequential control is involved in the image conditions. In that case, the verbal control process would compete for processing space with the spoken YES/NO response. The significant Property by Representation interaction indicates that the difference between visual and acoustic property processing time is greater for percept than for image representation. Again, this is to be expected, since visual properties are directly available in the percept representation, whereas acoustic properties are not directly available. The Property by Representation interaction illustrates clearly that visual and acoustic imagination are separate processes in accord with our subjective impressions. Neither the Response Mode by Property nor the triple interaction approached significance at the .05 level.

TABLE 1

SECONDS TO PROCESS ALPHABET (N = 10 for Each Condition)

Representation and Property		Response Mode			
		Spoken		Written	
		Mean	SE_M	Mean	SE_M
Image	AP	22.41	4.16	15.60	5.66
	VP	21.42	3.21	14.86	3.51
Percept	AP	15.78	2.29	15.16	4.54
	VP	12.53	1.57	12.35	2.90

With regard to sequential control, the objective results are consistent with the subjective impression that S implicitly says each successive alphabetic letter before visualizing it and abstracting its spatial height property or before "hearing" it and abstracting its acoustic sound property. For the image conditions, when a spoken YES/NO response is required, it competes for verbal processing capacity with the implicit speech generation of the alphabetic sequence. The written response actually leads to more rapid processing because it does not compete for processing capacity with the implicit verbal generation of the seque e. In a long serial list there is, then, evidence for an important interrelation of visual and implicit verbal codes. But this relation does not hold for the percept conditions, where sequencing between successive letters is no longer under verbal control. The sequencing problem is solved simply by moving the eyes from one successive letter form to another. Thus, the difference between spoken and written response time disappears in the percept conditions. Of course, this argument applies only to a long ordered string like the alphabet. Undoubtedly, visual imagination for some kinds of materials is lacking in a verbal sequencing component.

Several final comments are in order. First, when the means of Table 1 are converted to a rate measure, the most rapid image condition at 1.75 letters per second is for visual properties with a written response. This is very comparable to previous subjective measures of visual imagery (Weber & Bach, 1969; Weber & Castleman, 1970, Experiment 1), which neither required overt letter-by-letter responding nor had the additional within S acoustic property task. On the basis of this comparison, we can be reasonably sure that the written response does not appreciably interfere with and slow down visual image representation. Second, the most rapid imagined acoustic property condition also occurred with the written response mode and was 1.67 letters per second. This is in marked contrast to the approximately 6/sec rate of saying the letters implicitly (Weber & Castleman, 1970). Thus, there is something quite different in "pure" visualization (without letter-by-letter overt responding) as compared to the "pure" verbalization (saying the successive letters silently to one's self). But this difference tends to disappear when each individual letter must be overtly responded to, suggesting a complex set of relationships. Third, the unequal distribution of YES/NO letters for the visual and acoustic property tasks probably gives rise to a conservative difference. Thus, if YES/NO uncertainty were equal for the two tasks, the acoustic property task might have been even slower.

References

Paivio, A. Imagery and verbal processes. New York: Holt, Rinehart & Winston, 1971.

Weber, R. J., & Bach, M. Visual and speech imagery. British Journal of Psychology, 1969, 60, 199-202.

Weber, R. J., & Castleman, J. The time it takes to imagine. Perception & Psychophysics, 1970, 8, 165-168.

Note

1. A few letters, such as "i," are ambiguous, but once specified by E, there is no further difficulty.

EXPERIMENTS 4 AND 4': VISUAL IMAGERY FOR WORDS

(These experiments have now been submitted for publication. Experiment 4' was not part of the original proposal but has been added for completeness.)

Abstract

Two experiments were conducted to examine Hebb's test of the "picture theory" of visually imagining words. Experiment 4 examined retriveal from visual image representations of words by using a method of probing for the spatial properties of the i-th letter in an imagined word. Evidence was obtained for a limited capacity visual image system that can represent, in a spatially parallel manner, three-letter words at least as efficiently as they can be represented in a visual percept system. Experiment 4' examined the growth of visual image representations of words. All parts of a word did not grow simultaneously, and image growth did not in general proceed from left to right. The results of the two experiments were interpreted as indicating the existence of a visual image operating memory of between three and five letters capacity, the contents of which can be examined in much the same way as visual percepts.

Hebb (1966, 1968, 1972) and Woodworth and Schlossberg (1954) describe at length what we might term the "picture theory" of visual imagery. The picture theory consists of at least three facets. (1) Subjective Experience. Many people report visual images of whole lines or stanzas of verse, or at least of single words. The presumption would also seem to be that these images are spatially parallel, that is, all words or letters in the composite image are simultaneously available. (2) Function. The use of visual image representations of lines of text aids one's memory of the text. In the case of words, it may aid in learning to spell a word or in recalling the correct spelling of a word. (3) Description/Explanation. It seems that a mental picture or image can appear in one part of the mind at which another part of the mind can look.

Hebb proposes a test of the picture theory (1966, 1968, 1972). Hebb's test seems to consist of at least the following propositions. (1) It is a fact that when words are perceptually (visually) available they can be spelled almost as fast backwards as they can forwards. (2) To demonstrate the picture theory it would be necessary to show the following. If there is actually a spatial visual image at which one can look, the S ought to be able to spell the imagined word in a backward direction nearly as fast as in a forward direction. Thus after forming a clear image of a long word such as "university," he should be able to spell it backward nearly as rapidly as forward. (3) Hebb finds however that the backward spelling of the imagined word is much slower than the forward spelling. (4) He concludes that whatever the nature of the visual memory image, it is not like having a picture in the mind at which another part of the mind can look: "...the subjective impression that one can 'look at' one's image freely is shown by objective test to be wrong" (Hebb, 1966, pp. 43-44). Since it is possible to spell much more rapidly in a forward than in a backward direction "...the fact that the person with visual imagery can only 'see' the letters of a word in left-to-right order shows clearly that the memory image... is a <u>series</u> of events in a <u>particular order</u>, not a picture whose parts could be looked at in any order" [or simultaneously] (Hebb, 1966, p. 46; italics added). Thus Hebb's test for the picture theory of visual imagery consists of a general requirement that there exist analogs between perception and imagery for the processing of spatial information and a particular requirement that these analogs reveal themselves in the spelling test.

It is the contention of this paper that, with qualification, the picture theory of visual imagery is at least partly correct and also that Hebb's general test is a perfectly reasonable requirement. However, we feel that Hebb's particular test, a comparison of forward and backward spelling times, is not appropriate. In fact there are several problems associated with the spelling test. (1) The spelling test is not necessarily <u>visual</u>. If a <u>S</u> is asked to form a visual image of a word and then spell it as rapidly as possible in a forward direction, his performance may not be based on a visual representation at all. He may simply draw on a highly practiced verbal/speech representation of the word. Hence if Hebb's spelling test is essentially verbal in its requirements, then we should not be surprised that for a long word it is faster to spell in a forward than in a backward direction. After all, we have years of learning and repeated practice in the forward verbal spelling of long words and very little comparable training in backward spelling. Unwinding a serial verbal habit is easier if we begin at the initial step than if we try to go backwards. It is of course possible that forward and backward spelling do not draw on the same processes. While forward spelling is probably a verbal/speech process, it is possible that backward spelling does indeed draw on the visually imagined representation. But if this is the case, we are simply comparing verbal sequencing with visual image sequencing and finding that the verbal process is more rapid. This is not surprising in view of the findings of Weber and Castleman (1970) that the letters of the alphabet can be spoken either aloud or silently at a rate of about six per second, while sequentially visualizing the same letters occurs at a rate of about two per second.

With visual percepts, where the word is visually available, it also should not be surprising that the letters of the word are spelled almost as rapidly in a backward as in a forward direction. Any small difference that occurs in favor of forward spelling may again mean that the <u>S</u> is partly drawing on a verbal representation rather than on an exclusively visual representation in perception. Thus there is nothing to prevent him from looking at the word but basing its spelling on a stored verbal sequence rather than the visually available letters. However, if the perceptually displayed word is to be spelled backward, there would not be a stored verbal representation. The <u>S</u> would then rely on the visual string; and since all the visual information would be simultaneously available, he could simply visually scan the word from right to left. He could presumably do this at about the same rate no matter how long the word was because he could simply move his eyes from right to left. The capacity of the visual perception system would be limited only to the number of letters in the perceptually presented line.

(2) The visual image system may be very limited in capacity. Why indeed does Hebb insist on a long word? If a short word (e.g., "toy") is visually imagined, and there is uncertainty on the part of S as to which direction he will be cued to spell the word, then it can readily be spelled backwards about as rapidly as forwards. This could be because the word is within the capacity of the visual image system for spatially parallel (simultaneous) representation, whereas longer words are not. That is, a longer word might have to be generated a syllable, or a few letters, at a time in order not to exceed a limited capacity visual image system. It might also be possible, however, to spell "toy" as rapidly backwards as forwards on the basis of purely verbal processes. For example, S might first spell the word in a forward direction and then pull out the backward spelling from a short-term operating memory of limited capacity. It is in fact difficult to decide between the alternatives of visual versus verbal representation for either short or long words. When S reports the letter names he could be getting them from a visual image representation or from a verbally stored code of the word that he has just recovered from saying it first in a forward direction and then recovering the backward direction from operating memory. Of course, the fact that S claims to have a visual image of a word is not a guarantee that when he spells it he is basing his spelling on the visual representation rather than a verbal representation. Even a short word yields ambiguity when we try to decide if it draws on verbal rather than visual processes. What is needed is a way of knowing that S is drawing on a visual representation of a word.

If S had to respond to spatial features of a word, then we could be more certain that the instructions to imagine the word and then retrieve information from the imagined representation were being followed. One solution would be to have S report on the visual properties of the letters in a word (Weber & Castleman, 1970; Weber & Kelley, 1972; Weber, Kelley & Little, 1972). In particular, we might ask S to imagine lower-case printed letters and to classify them for the spatial property of vertical height. Those letters that are vertically large (b, d, f, g, ...,y) would fall in the YES category and all other letters would fall in the NO category. This procedure allows for a set of converging factors pointing to the conclusion that a visual image system rather than a verbal system is involved. The converging factors are: (a) instructional set, asking S to visualize and classify the letters in a word for their spatial properties; (b) the ability of Ss to correctly classify the imagined letters for vertical height; (c) frequent subjective reports from Ss that they do visualize during the task; and (d) the finding (Weber and Castleman, 1970) of highly distinctive rates of processing for visual image versus implicit speech instructional sets. Weber, Kelley and Little (1972, Experiment 2) have already compared image representations for word and alphabet strings in which the height property of each letter in the string was reported from left-to-right. The results indicated different representations for the two types of strings, with alphabetic representation sequential and word representation at least partly parallel. However, even for words, the letter-by-letter report may well have biased toward sequential representation of letters. Thus, in the present experiments, a probe is presented indicating a letter position; and the height of the letter occupying the corresponding position in a test word is classified as YES/NO by the Ss. The one word/one probe technique avoids building in unnecessary sequential processing, minimizes response requirements, and may thereby allow for maximum parallel representation of a visually imagined word.

Experiment 4: Word-Probe

This experiment is concerned with comparing retrieval times from an available image versus from an available perceptual representation. Specifically, comparisons are made for image versus percept representations of three and five letter words. If visual imagery is more limited in capacity than visual perception, then we would expect percept-image differences in response time to be smaller for three than for five letter words; this should reveal itself as an interaction between representation mode and word length. To the extent that processing three and five letter words is within the capacity of both the visual image and visual percept systems, and to the extent that these processes overlap in shared components (Hebb, 1966, 1968, 1972; Neisser, 1972), then the following should also be the case. Response times (RT) should be similar in the sense that there would be no main effects for percept-image representation or word length; and also there should be no interaction between serial probe position and representation mode. In line with this reasoning, it is possible to specify several possible alternative hypotheses regarding serial position curves relating response time and letter position probed.

(1) Forward Serial Hypothesis. Reaction time would be a linear increasing function of the serial position probed. This would be consistent with a letter-by-letter, left-to-right, self-terminating scan of the visual image representation of the word. Whenever the probed position was reached S would terminate the scan, abstract the height property of the letter in the probed position, and respond YES or NO. Since decision time and response time should not vary with the letter position probed, the slope of the function would represent the time per letter required to scan the image.

(2) Backward Serial Scan Hypothesis. Here RT would be a linear decreasing function of the letter position probed. This would indicate that S scans the image from right to left, i.e., in a backward direction.

(3) Random Access Hypothesis. Response time would be a flat function of the letter position probed. This would occur because S can "see" in a spatially parallel image all the letters at once. Insofar as

this image approximates a perceptually availabel word, the image and percept functions should be comparable in form.

(4) Centerpoint Scan Hypothesis. Here S would form an image of the word, and fix his attention on the centerpoint (middle letter) of the word. Then if an initial letter position is probed S would scan his image from the center to the left. If a terminal letter is probed, S would scan his image from center to right. Under these conditions the serial position function should dip in the middle and rise toward the end positions, with an increasing number of letters scanned reflecting increased RT.

(5) Centerpoint Interference Hypothesis. Here also S would generate an image of the word and hold it in focal attention. However the center of the word would yield a longer response time than the ends. This could be because of (a) an information-seeking strategy in which S habitually extracts information from the most informative parts of a word first, or (b) a "perceptual" effect in which the end letters suffer less visual interference--and therefore information is easier to extract, since the end letters are bounded on one side only by other letters (Woodworth & Schlosberg, 1954), and thus do not suffer the same interference that would occur if letters were present on both sides.

(6) Partial Word Hypothesis. In hypotheses (1) to (5) it has been assumed that the whole image of the word is present at once. However, if the size of the word exceeded the spatially parallel capacity of the visual image system, S would necessarily produce the word a segment at a time until he produced the segment in which the probed letter appeared. The shape of the RT function might differ among individual Ss, depending on what segment of the word was produced first and which of the preceding strategies was used.

Method

Subjects.--Sixteen experimentally naive volunteer Ss with normal or corrected vision were tested individually. Each S was previously screened for the ability to respond to a visualized alphabet in a manner similar to Weber and Castleman (1970).

Stimuli and Procedure.--The test stimuli were nouns and adjectives with a Thorndike-Lorge count of five or more per 100,000 occurrences.[3] A total of 42 different words were used. Twelve were test words of five letters in length and twelve were test words of three letters in length. In addition, there were eight filler words three letters in length (for statistical balancing purposes); and, finally, there were five practice words for each word length. The twelve five-letter test words were selected such that lowercase vertically large letters (b, d, f, g, h, j, ..., y) and vertically small letters (a, c, e, ..., z) appeared equally in all five letter positions (first through fifth). The twelve three-letter test words and eight three-letter filler words were also selected so that vertically large and small letters appeared in all three-letter positions.

Each S was given a preliminary spelling test for the words that were to be used. He was required to "spell" the words both with letters and also with YES and NO indicating the vertical height of the letters. For example, the word "cat" would be spelled "c, a, t," and "NO, NO, YES," respectively.

In the experiment the stimulus words were presented as images or as percepts. For the image conditions the E spoke the word and S was instructed to form a visual image of it with his eyes closed. Imagining something visually was explained to be like "picturing it in the mind." For the percept condition each word was rear projected with a Kodak Carosel projector equipped with a solenoid operated shutter. Letters appeared in elite type, lowercase form, on a 30.48 cm x 20.32 cm rear projection screen. The vertically small letters appeared 0.63 cm in height. Since vertically large letters were slightly irregular in height, the "j" is taken as the standard; and it was 1.27 cm in projected height. The S viewed the letters from a distance of 53.34 cm.

Four seconds after presentation of a stimulus word the E orally presented a probe digit ("one," "two," ..., "five"). The probe digit indicated a letter position in the word the S was imagining or perceiving. In a word like "cat" the digit "one" referred to the "c," "two," referred to the "a," and "three" referred to the "t," and similarly for the five-letter words. The spoken digit activated a Lafayette voice relay and started a Standard clock. The S then responded YES or NO to indicate whether the letter denoted by the probe digit was vertically large or not. The S's response stopped the clock, and RT from onset of the probe to onset of the response was recorded. For the percept conditions, after S responded the slide terminated and a blank adaptation field was displayed from another projector with a solenoid shutter. All Ss were instructed to respond as rapidly as possible, with 100 percent accuracy. Any words which resulted in error were repeated at the end of the run. Hence all time scores were for correct choices only.

Design.--There were two levels of word length (three-letter and five-letter words) and two modes of representation (percept and image). Word length was a between-Ss variable and representation mode was within-Ss, with half of the Ss proceeding in the order image-percept and the other half in the order percept-image. Within each of the between Ss conditions there were 60 trials (each test word appeared five times in the five-letter treatment and three times in the three-letter treatment). This allowed each letter position in each word to be tested once and only once; the additional three-letter filler words made it possible to equate practice on three- and five-letter words. Probe position was also treated as a separate within Ss variable for each word length. Two different random orders of words were used.

Practiced S.--In addition to the foregoing group conditions, the identical procedures were replicated with a single S highly practiced in visual image tasks dealing with alphabet and word strings similar to those described in Weber and Castleman (1970) and Weber, Kelley, and Little (1972). This practice was of an informal sort and cannot be quantitatively described, but it occurred over a period of several months on a fairly intensive basis.

Results and Discussion

Figure 1 illustrates the principal group results in the left panel. Each S's median response time for a given condition was computed and then the mean of the medians obtained for Fig. 1. Thus each point represents 96 events (12 words x 8 Ss). Medians were used because of the substantial variability encountered. The number to the right of each function is the mean RT obtained by collapsing across probe positions. The right panel is the same data for the practiced S but it is based on medians only, 12 events (words) per point.

Several statistical analyses were conducted. In each case RT was the dependent variable. In the first analysis of variance, data was collapsed across probe positions. That is, each S's score was the mean of his separate median probe position RTs. The resulting two-way classification of two-word lengths (three, five) by two representation modes (image, percept) was analyzed with the first factor between-Ss and the second within-Ss. For word length, $F\ (1,\ 14) = 5.20$, $p < .05$; for representation mode, $F\ (1,\ 14) = .48$, $p > .05$; and for the Word Length by Representation interaction, $F\ (1,\ 14) = 2.82$, $p > .05$. Thus RT significantly increases with word length, but images and percepts are not reliably different in how rapidly spatial information may be extracted from them. Indeed retrieval time from the three-letter image condition is at least as fast as from the three-letter percept. The data for the practiced S collapsed across probe positions yields results somewhat comparable. The five-letter image condition requires the longest mean retrieval time, and the three-letter image and percept conditions are very comparable. The absolute times are generally less for the practiced S. No statistical analysis was possible, since only one S was involved.

The second series of analyses examined probe position effects. Consider the three-letter words. Here the analysis was based on a two-way classification of three-probe positions (1, 2, 3) by two representations (image, percept), with both factors within-Ss. For probe position, $F\ (2,\ 14) = 7.64$, $p < .05$; for representation mode, $F\ (1,\ 7) = .94$, $p > .05$; and for the probe position by representation interaction, $F\ (2,\ 14) = .09$, $p > .05$. Thus the dip in RT for the middle letter is significant, and this would be consistent with the centerpoint scan hypothesis. To examine more closely the probe position effect, separate Newman-Keuls tests were performed for the three-letter percept and three-letter image conditions. The results are indicated in Fig. 1 by the letters adjacent to each point. Those points with a letter in common do not differ significantly, while those points without a letter in common do differ significantly ($p < .05$). Hence the

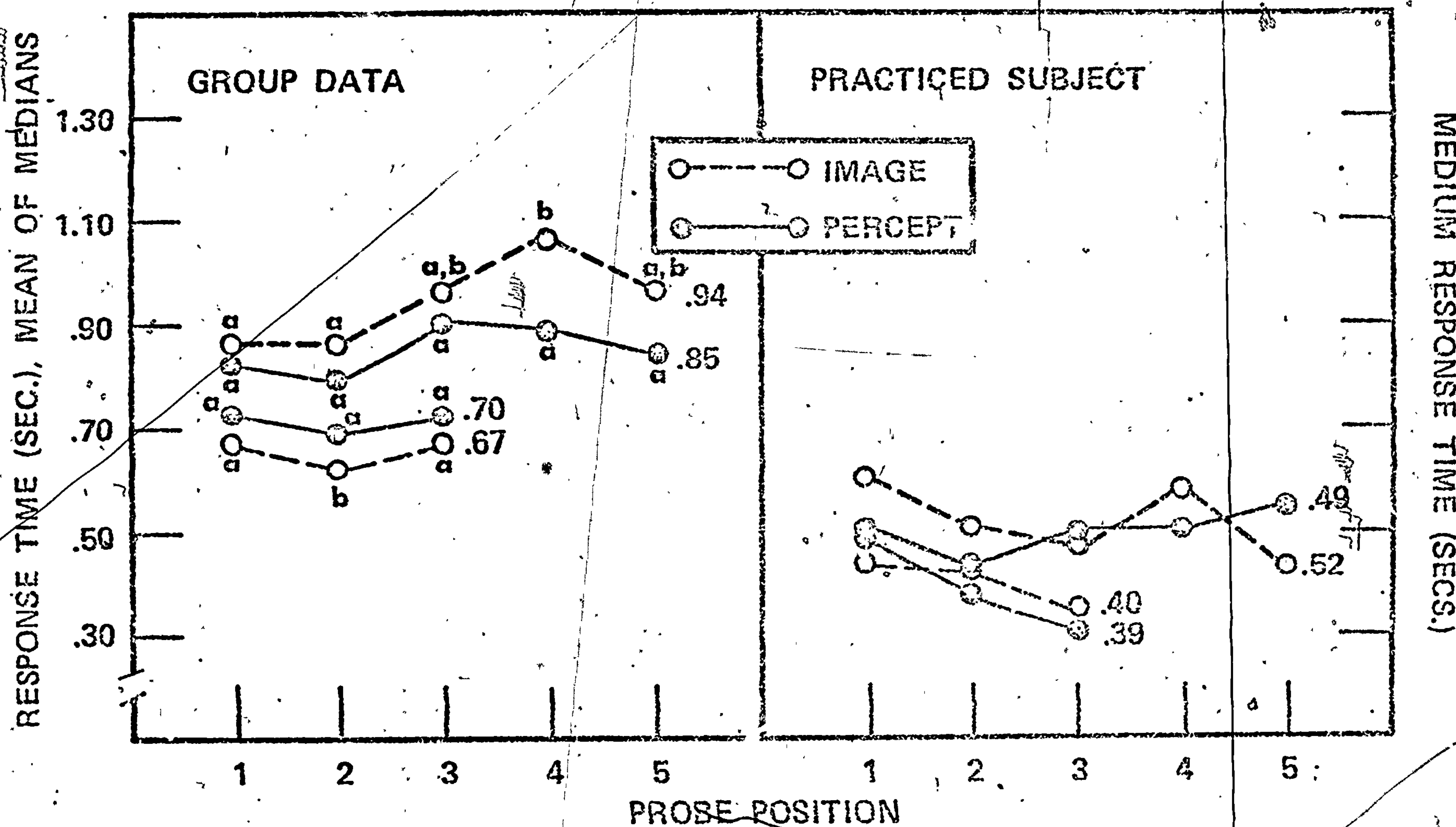

Figure 1. Experiment 4, Word-Probe presentation. Response time as a function of Image-Percent representation, and letter position probed. The left panel is for Group Data and the right panel is for a Practiced Subject

three-letter percept condition may be regarded as statistically flat, a result consistent with the random access hypothesis. But the three-letter image condition indicates a significantly faster retrieval time when the second-letter position of a word is probed. The probe position effect for three-letter words from the practiced S suggests a right to left backward scan of the word for both percept and image conditions. Evidently the strategies used are not obligatory, and as a result vary among Ss.

In a comparable series of analyses for the five-letter word conditions, the following results were obtained. For probe position, $F(4, 28) = 4.97$; $p < .05$; for representation, $F(1, 7) = 1.88$, $p > .05$; and for the Position by Representation interaction, $F(4, 28) = 2.28$, $p > .05$. Not all parts of a word are equally available as shown by the significant probe position effect. And representation is not significant, possibly because of the substantial variability. The Newman-Keuls test results are again summarized in Fig. 1. The five-letter percept function is statistically flat, a finding consistent with the random access hypothesis. However the five-letter image function shows a maximum RT for the fourth probe position, and that fourth position RT is significantly greater than for the first and second positions. The first, second, third, and fifth probe positions are statistically the same, since they are all covered by the common letter a. Positions three, four, and five are also statistically the same, since they are covered by the common letter b. These results seem most consistent with the partial word hypothesis in which a fragment at a time of the word is generated into visual operating memory. The practiced S data for five-letter words follows a somewhat different pattern for the image conditions but seems reasonably flat for the percept conditions.

It is appropriate to note that preliminary analyses were also performed in which YES/NO responses were treated as a separate factor. No consistent effects distinguishable from the present results were apparent.

To summarize, there are differences between three- and five-letter words both in absolute RT and in the pattern of probe position effects. For image versus percept representation there do not seem to be significant differences in absolute RT, but there are differences in probe position effects. Finally, none of the hypothesized scan modes seem to account for very much of the data from the image conditions. However, for the percept conditions a random access hypothesis is supported.

Experiment 4': Probe-Word

The hypotheses of Experiment 4 have bearing on the retrieval of spatial information from an image that is all or partly available at the time the probe is presented. However it is possible that the growth of generation of an image is distinguishable from retrieval from an available image. For example, Experiment 4 indicated that Ss do not retrieve the visual property of the i-th letter by scanning the

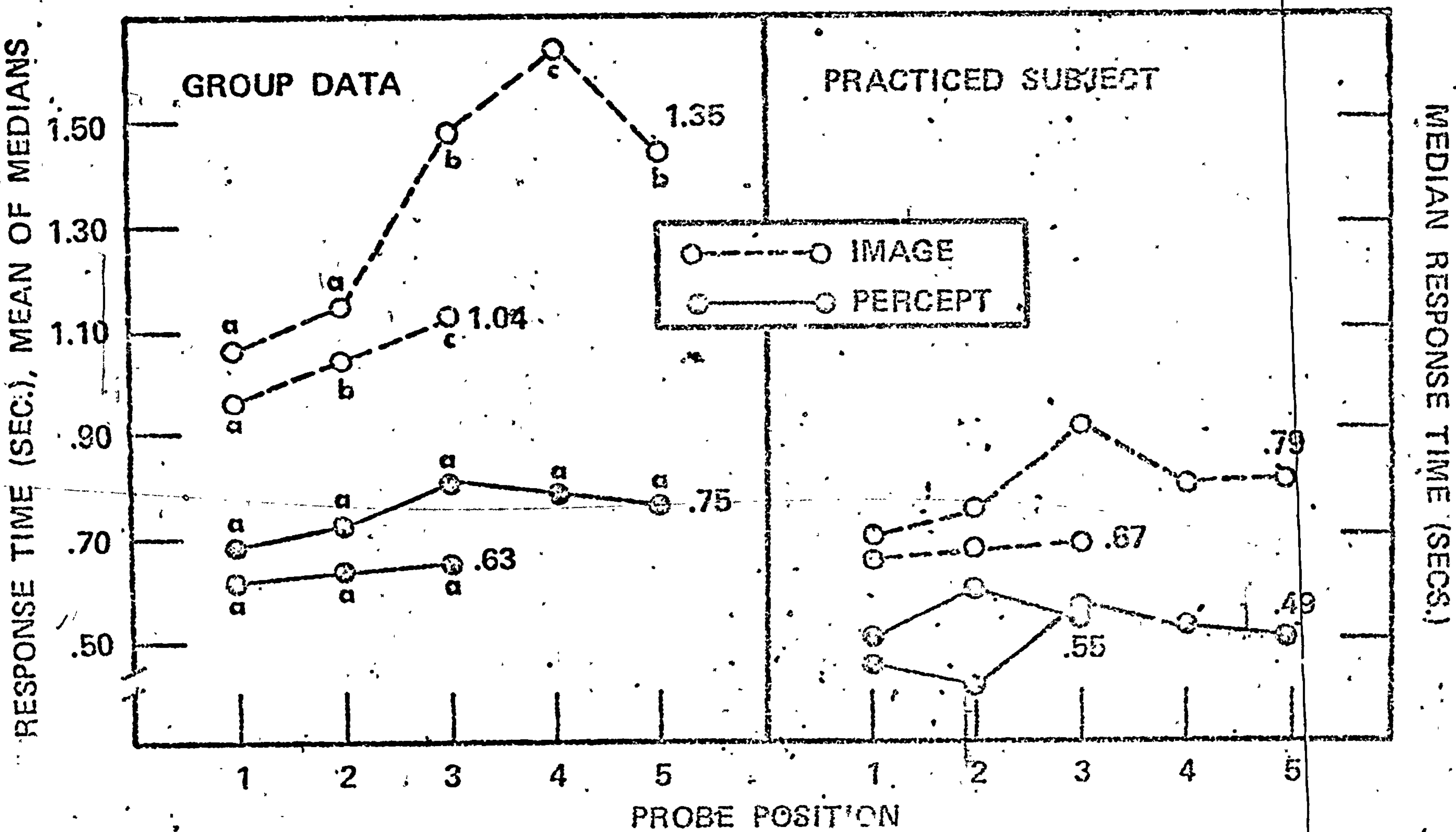

Figure 2. Experiment 4', Probe-Word presentation. Response time as a function of Image-Percept representation and letter position probed. The left panel is for Group Data and the right panel is for a Practiced Subject

completed image from left to right. But, it is still possible that the image is generated left to right in a sequential fashion. This is clearly the case with alphabetic sequences which require verbal sequencing between letters (Weber, Kelley, & Little, 1972). Thus with an alphabetic list S seems to say each letter before generating its image. In contrast, with words it is not the case that S says the successive letters in order to generate a visual representation of its letters (Weber, et al., 1972). Yet if the word is presented orally, it might still be processed as it comes, that is, from left to right. In Experiment 4' we emphasize the image generation or growth question by first presenting the probe and then the word that is to be imagined. Instead of uncertainty of probe position (Experiment 4) we now have word uncertainty. The word uncertainty is, of course, greater than the probe position uncertainty.

Method

Sixteen experimentally naive Ss with normal or corrected vision were paid for their participation and were tested individually. All stimulus materials and procedures were identical to those of Experiment 4 except that in this case the probe digit was presented first and then was followed by the stimulus word. The highly practiced S from Experiment 4 was also tested, but on different days than for Experiment 4.

Results and Discussion

The principal results are summarized in Fig. 2. The treatment of the results is exactly parallel to that of Experiment 4. When the probe position variable is collapsed, the resulting two-way analysis of word length by representation mode indicates that word length is significant, $F(1, 14) = 7.69$, $p < .05$; that representation mode is significant, $F(1, 14) = 249.58$, $p < .05$; and that the Word Length by Representation interaction is significant, $F(1, 14) = 9.72$, $p < .05$. While the word length variable was also significant in Experiment 4, the representation mode variable shows a substantial effect here in contrast to the first experiment. This is probably not too surprising, since the image representations must be generated before information from the image can be retrieved. In contrast, the percept conditions require no comparable generation before the abstraction of information can begin. Since the present image versus percept comparison is most comparable to Hebb's (1966) spelling test, it is to be expected that it takes longer to generate and retrieve spatial properties of images than it does to simply retrieve the spatial properties of an already available image or percept. The significant Word Length by Representation interaction is indicative of the limited capacity of the visual image system. As the word to be generated becomes longer and exceeds processing capacity, RT becomes disproportionally longer for the image conditions in comparison to the percept conditions.

Separate serial position analyses were performed for each word length. For three-letter words an analysis of Probe Position by Representation mode yielded a significant position effect, $F(2, 14) = 14.56$, $p < .05$; a significant representation effect, $F(1, 7) = 336.78$, $p < .05$; and no significant interaction, $F(2, 14) = 2.93$, $p > .05$. The Newman-Keuls tests for three-letter word conditions indicates that the percept function is statistically flat and that for the image function each point is statistically different from the other two. The increasing three-letter image function is consistent with a left-to-right generation of the image. The practiced S's three-letter image data is, however, quite flat and is consistent with the hypothesis of random access to an image that is completely and simultaneously available.

The serial position analyses for five-letter words indicate the following. For the probe position by representation analysis, position is significant, $F(4, 28) = 26.76$, $p < .05$; representation is significant, $F(1, 7) = 99.06$, $p < .05$; and the Position by Representation interaction is significant, $F(4, 28) = 13.77$, $p < .05$. The Newman-Keuls test indicates a statistically flat percept function. The image function, however, has statistically comparable values for probe positions one and two, comparable values for probe positions three and five, and a maximum value at position four. This is not consistent with left-to-right image generation. As in Experiment 4 the pattern of generation is suggestive of the partial word hypothesis, with positions one and two generated first and concurrently, followed by generation of the remaining positions, when necessary. The data for the practiced S indicates a much flatter five-letter image function. His absolute RT is also much less than that of the group data.

In summary, as in Experiment 4 percept functions remain statistically flat even when the probe is presented prior to the word. Generation processes of the type st[illegible]ed here are minimal in perception, and random access to visually available letters seems to be the result. However, image functions show a sharp increase from left-to-right. But it would be wrong to conclude from this that images are generated from left-to-right, because the increase is not monotonic for five-letter words and the data from the practiced S indicates reasonably flat image functions. Finally, this experiment shows a substantial time difference between percept and image conditions that was not seen in Experiment 4. It takes time to generate images (Weber & Castleman, 1970).

Conclusion

A modified version of the Hebb test of the picture theory of imagery has revealed that when long (five-letter) words are "spelled" for the spatial properties of their letters, there may indeed be some differences between the image and percept systems. But if the amount of information in the visual image system is small (three letters), so that the visual image capacity is not exceeded, then spatial information can be retrieved about as rapidly as when the word is presented to the visual perceptual system.

Our explanation for this is that the S is able to output from long term visual memory into a limited capacity operating memory (focal attention) a spatial representation of a word. Once the spatial features of a word are visually represented in operating memory, they can then be examined by other information processing routines. These routines are capable of examining the spatial properties of a letter in a given probe position, and an appropriate YES/NO decision can be made. In this sense, then, "one part of the mind can examine the activity of another part of the mind," contrary to Hebb's conclusion (1966). Thus Hebb's general criterion for visual imagery of words is perfectly appropriate, but his backward and forward spelling test for perceptually available versus imagined words is not appropriate for two reasons. First, as usually applied, the test advocates the use of long words, and any limited capacity system such as the visual imagery mode would be immediately overloaded. Second, visual imagery need not be involved at all for some spelling conditions, since S could draw on various verbal processing routines to spell the word rather than on any visual image representation he might have.

While the picture theory of visual imagery may be more correct than psychologists have been willing to grant, there are still some obvious differences between imagery and perception. These differences involve at least the following: capacity and sequencing considerations, serial position effects, practice effects, and subjective stability.

The present results suggest that the capacity of the visual image system for simultaneous letter representation is between three and five letters for high frequency words. Experiment 4 reveals that the image system does reasonably well with three-letter words, but it begins to show signs of strain (departures from flatness and percept-image differences) with five-letter words. Experiment 4' indicates that the manner in which the visual image capacity is filled may be consistent with left-to-right generation into visual operating memory for the three-letter image conditions but not for the five-letter conditions. The data for the practiced S is even less consistent with a left-to-right visual image growth.

It is of particular interest to know how the capacity limitations of the visual imagery system are handled. Weber, Kelley, and Little (1972) and Weber and Kelley (1972) have suggested that the limited capacity of the visual image system creates a sequencing problem. That is, when a spatial array of letters to be imagined exceeds the capacity of the visual image system for spatially parallel representation, it becomes necessary to generate the array into the visual imagery system a piece at a time. When this is done for alphabetic lists or long words ("supercalifragilisticexpialidocious"), it seems that a verbal sequencing operation is employed in the following way. The sequential ordering for such a long letter list is verbally encoded. To represent this order in the limited capacity visual image system, the S says each successive letter name or syllable to himself prior to visually imagining it (Weber, Kelley, & Little, 1972). This is in sharp contrast to perceptual presentation where the S need only move his eyes from one letter or syllable to another (Weber & Kelley, 1972) in order to represent a letter prior to abstracting its spatial properties. Thus the capacity of the visual perceptual system is unlimited as long as the time requirements are not so stringent as to limit visual scanning or eyemovement from one letter locus to another.

For both experiments, retrieval from percepts as a function of position probed looks like a random access process. But there does not seem to be any great consistency in retrieval from or generation of images that makes sense in line with the various hypotheses described in Experiment 4. Also, in both experiments the RTs of the practiced S are generally faster than those of the naive group, but differentially so for image and percept conditions. This could be due to individual differences, but we believe (without strong evidence) that it means that visual image generation and retrieval can profit from extensive practice. The limits of such improvement would be of interest in its own right but would require more systematic study of the same S over time.

Finally, there are undoubtedly image-percept differences in stability of representation. Images seem to fade in and out while percepts are relatively stable. The time course of image evanescence would also be of interest.

References

Hebb, D. O. Textbook of psychology. Philadelphia: Saunders, 1966, 1972.

Hebb, D. O. Concerning imagery. Psychological Review, 1968, 75, 466-477.

Neisser, U. Changing conceptions of imagery. In P. W. Sheehan (Ed.), The function and nature of imagery. New York: Academic Press, 1972, pp. 233-251.

Weber, R. J. & Castleman, J. The time it takes to imagine. Perception & Psychophysics, 1970, 8, 165-168.

Weber, R. J. & Kelley, J. Aspects of visual and acoustic imagery. Psychonomic Science, 1972, 27, 121-122.

Weber, R. J., Kelley, J., & Little, S. Is visual image sequencing under verbal control? Journal of Experimental Psychology, 1972, 96, 345-362.

Woodworth, R. S. & Schlosberg, H. Experimental psychology. New York: Holt, 1954.

CONCLUSIONS

A series of six experiments has been conducted which demonstrate a number of phenomena related to the visual image representation of letters and words. Those phenomena include the following:

(1) Sequencing between images of letters in an unpronounceable string like the alphabet is under verbal control: the subject encodes and decodes the sequence verbally; his generation of a visual image then depends on his immediately prior verbal activation of the name of the letter. Hence, this is a sequential, one-letter-at-a-time process.

(2) Sequencing between visually imagined letters in a pronounceable string, such as a short work, is not under verbal control. This is because the entire word can be represented simultaneously in visual imagination and there is no corresponding need to verbally generate it a letter at a time.

(3) The capacity of the visual image system for simultaneous representation of letters is very limited: approximately between 3-5 letters for short easily pronounceable letter strings (words).

(4) Visual percept and visual image representations of words are very comparable for three-letter words but with longer words the capacity of the imagery system for simultaneous letter representation begins to be exceeded. When this occurs, a sequencing problem arises: how to get from an initial letter (or set of letters) to the next letter(s). This is done differentially for the two systems, by implicit verbal control for the image system and by simple eye movement or scanning for the percept system.

The capabilities of the visual image system for representing letters and words suggest that the teaching of spelling might profitably employ visual as well as verbal encoding. For example, the word "child" when "spelled" in terms of its visual properties becomes "NO, YES, NO, YES, YES." If a student represents a word with the appropriate sequence of YES/NO, we can be reasonably sure that he has a visual encoding of the word. The extent to which visual encoding and rehearsal might aid spelling warrants follow up.