XVIII. COGNITIVE INFORMATION PROCESSING*

Prof. S. J. Mason
Prof. W. F. Schreiber
Prof. M. Eden
Prof. T. S. Huang
Prof. D. E. Troxel
Prof. O. J. Tretiak
Dr. P. A. Kolers
R. M. Baecker
W. L. Black
J. K. Clemens
J. R. Disbrow
R. W. Donaldson
J. K. Dupress
D. N. Graham
N. Gramenopoulos
G. S. Harlem
M. C. Harrison
K. R. Ingham
E. E. Landsman
P. Mermelstein
N. D. Strahm
D. E. Thornhill

A. PICTURE TRANSMISSION BY PSEUDORANDOM SCANNING

Pseudorandom scanning has been ingeniously used by Deutsch[1] to reduce television frame rate and hence bandwidth. Now most practical picture transmission channels have some lowpass effects. In this note we shall investigate the effect of a lowpass filter on pseudorandomly scanned pictures. We shall consider only monochrome still pictures.

In the system of Fig. XVIII-1, the brightness of the original picture is represented by $f(x, y)$, where (x, y) are the spatial coordinates of the picture point on the raster. Assume that the raster contains $(N+1)^2$ points (j, k), $j, k = 0, 1, 2, \ldots, N$. The distance between two neighboring points has been taken as the unit length. We write

$$f(x, y) = \sum_{j=0}^{N} \sum_{k=0}^{N} f(j, k)\, \delta(x-j, y-k),$$

where $\delta(x, y)$ is the two-dimensional unit impulse.

Then scanner 1 converts $f(x, y)$ to

$$\hat{f}(t) = \sum_{m=0}^{M} \hat{f}(m)\, \delta(t-m),$$

where $\delta(t)$ is the one-dimensional unit impulse,

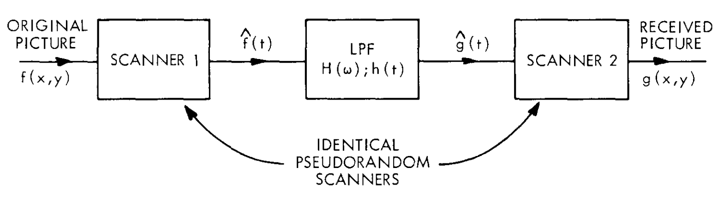


Fig. XVIII-1. A picture-transmission system.

*This work was supported in part by the National Science Foundation (Grant GP-2495), the National Institutes of Health (Grant MH-04737-04), and the National Aeronautics and Space Administration (Grant NsG-496).

$$M = (N+1)^2 - 1, \quad \text{and} \quad \{\hat{f}(m);\ m=0, 1, \ldots, M\}$$

is a random ordering of $\{f(j, k);\ j, k=0, 1, \ldots, N\}$. We have assumed that the scanning starts at $t = 0$. We have

$$\hat{g}(t) = \int_{-\infty}^{\infty} \hat{f}(t-u)\, h(u)\, du$$

$$= \sum_{\substack{k=-\infty \\ k=\text{integer}}}^{\infty} \hat{f}(t-k)\, h(k).$$

The values of $\hat{f}(k)$ for $k > M$ and $k < 0$ are taken to be zero.

Assume that the filter is realizable, then $h(t) = 0$ for $t < 0$, and

$$\hat{g}(m) = \sum_{k=0}^{m} \hat{f}(m-k)\, h(k).$$

From $\hat{g}(m)$, scanner 2 (which is identical to scanner 1) reconstructs $g(x, y)$.

$$g(x, y) = \sum_{j=0}^{N} \sum_{k=0}^{N} g(j, k)\, g(x-j, y-k).$$

Assuming that the width of $h(t)$ is small compared with the width of the raster, except for points near the edges, we have approximately

$$g(j, k) = h(0)\, f(j, k) + \sum_{u=0}^{\infty} h(u)\, f_{jk, u},$$

where $f_{jk, u}$ are selected in random order from the original picture $f(x, y)$. We then write

$$g(j, k) = s(j, k) + n(j, k),$$

where $s(j, k) = h(0)\, f(j, k)$ is the signal, and $n(j, k)$ is the noise. The noise is independent of the signal and is uncorrelated from point to point.

The (spatial) mean of $s(j, k)$ is

$$\langle s\rangle \equiv \frac{1}{(N+1)^2} \sum_{j=0}^{N} \sum_{k=0}^{N} s(j, k) = h(0) \langle f\rangle,$$

while that of the $n(j, k)$ is

$$\langle n\rangle = \left[\sum_{u=1}^{\infty} h(u)\right]\langle f\rangle.$$

We have

$$\frac{\langle n\rangle}{\langle s\rangle} = \frac{\sum_{u=1}^{\infty} h(u)}{h(0)}.$$

Note that the average of the noise is not zero. This situation will probably reduce the contrast of the received picture.

The variance of s is

$$\sigma_s^2 \equiv \frac{1}{(N+1)^2} \sum_{j=0}^{N} \sum_{k=0}^{N} [s(j,k)-\langle s\rangle]^2$$

$$= h^2(0)\, \sigma_f^2,$$

and the variance of n is

$$\sigma_n^2 = \left[\sum_{u=1}^{\infty} h^2(u)\right] \sigma_f^2.$$

Hence,

$$\frac{\sigma_n^2}{\sigma_s^2} = \frac{\sum_{u=1}^{\infty} h^2(u)}{h^2(0)}.$$

Assuming $H(w) = 0$, for $w \notin [-\pi,\pi]$, we have

$$\sum_{u=0}^{\infty} h^2(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(w)|^2\, dw.$$

Also recall that

$$h^2(0) = \frac{1}{4\pi^2}\left[\int_{-\infty}^{\infty} H(w)\, dw\right]^2.$$

Then we have

$$\frac{\sigma_n^2}{\sigma_s^2} = \frac{2\pi}{W_e} - 1,$$

where

$$W_e = \frac{\left(\int_{-\infty}^{\infty} H\right)^2}{\int_{-\infty}^{\infty} |H|^2} = \text{two-sided equivalent bandwidth of } H(w) \text{ in rad/unit length.}$$

The following table gives some numerical values.

$\frac{W_e}{2\pi}$	1	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$\frac{\sigma_n}{\sigma_s}$	0	.6	1	1.7

If $W_e = 2\pi$, h(t) is an impulse since we have taken the distance between two neighboring points as the unit length; hence, there is no noise in the received picture. When W_e decreases, the signal-to-noise ratio decreases quite rapidly, reaching 1 when $W_e = \pi$.

The preliminary analysis given here indicates that the effect of a lowpass filter on pseudorandomly scanned pictures will be adding some independent white noise to the received pictures. The noise power is quite large even if the bandwidth of the lowpass filter is reasonably wide.

T. S. Huang

References

1. S. Deutsch, Narrow-band TV uses pseudorandom scan, Electronics, Vol. 35, pp. 49-51, April 27, 1962.

B. PICTURE TRANSMISSION BY Δ-MODULATION

The transmission of speech by Δ-modulation has been quite successful. The application of ordinary Δ-modulation to picture transmission, however, gives very poor received pictures because picture signals vary much more rapidly than speech signals. A modified Δ-modulation scheme, which we call exponential Δ-modulation, has been conceived.[1] This method gives much sharper pictures than ordinary Δ-modulation.

An ordinary digital Δ-modulation system is shown in Fig. XVIII-2. Q is a two-level quantizer whose output is "1" when the input is positive, or "0" when the input is negative. G is a two-level generator that puts out A or –A when the input is "1" or "0",

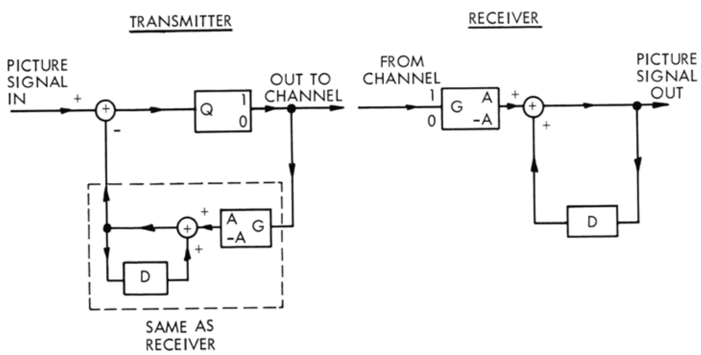
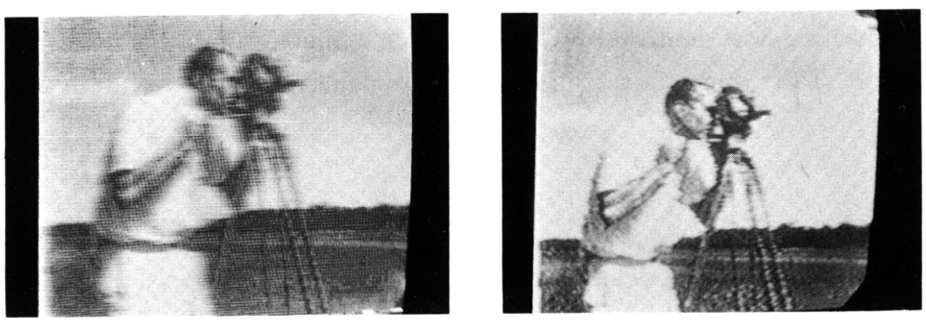


Fig. XVIII-2. Δ-modulation system.



(a) (b)

Fig. XVIII-3. (a) An ordinary Δ-modulated picture. A = 30. The picture brightness varies from 0 to 1023. (b) An exponentially Δ-modulated picture. $a_i = 10 \times 2^{i-1}$, limited at 80.

respectively. D is a one-unit (time between successive samples) delay.

In an exponential Δ-modulation system, the two-level generator G is replaced by a multilevel generator G_1. If in the input to G_1 we have a string of n 1's after one or more 0's, the output of G_1 corresponding to the string of 1's will be $a_1, a_2, \ldots, a_n$, where a_i are positive numbers. Similarly, if in the input we have a string of n 0's after one or more 1's, the output corresponding to the string of 0's will be $-a_1, -a_2, \ldots, -a_n$. To increase rise time, we usually choose $a_1, a_2, \ldots, a_n$ to be a more or less exponentially increasing sequence, hence the name exponential Δ-modulation. Notice that G is a special case of G_1 with $a_i = A$.

Exponential Δ-modulation will give sharp pictures, but will introduce overshooting and ringing. To avoid excessive overshooting, some kind of limiting has to be used; and by properly adjusting the output levels of G_1 a compromise might be reached between ringing and loss of edge sharpness. Computer simulation of the exponential

Δ-modulation has been carried out[2] and some preliminary results obtained. The purpose of the computer simulation is to find an optimum combination of system parameters. Figure XVIII-3 shows an exponentially Δ-modulated picture and an ordinary Δ-modulated one. By changing parameters, we expect to get better pictures than Fig. XVIII-3b.

In our discussion, we have assumed that the channel is noiseless. If the channel is noisy, the ordinary Δ-modulation system illustrated in Fig. XVIII-2 is not practical. An error occurring in a scanning line ruins the remaining samples in the line. A single error gives rise to a white or black streak along the scanning direction. Exponential Δ-modulation suffers the same defect. Hence, in practice we have to find ways of combatting noise. One possible method[3] is to separate the picture signal into two parts: a low-frequency part and a high-frequency part. The former is coarsely sampled and sent by PCM, while the latter is sent by exponential Δ-modulation in which the transmitter and receiver are modified by inserting an amplifier of gain less than one right after the delay. This and other noise-combatting methods will be studied.

T. S. Huang

References

1. This modification originated with Richard Witt and others of Raytheon Company's Communications and Data Processing Division. They have called this modified method "Δ^2-modulation."

2. The computer simulation was done by Robert Dick as part of his Senior Thesis research (1964).

3. W. F. Schreiber, Picture Coding, a paper presented at the Picture Data Processing Seminar, Yeshiva University, August 1963.

C. HIERARCHICAL CONCEPT FORMATION

One of the most common properties of the naturally occurring pattern-recognition devices is the ability to break a pattern down into components in such a way as to preserve only the essential information about the pattern. The components are usually hierarchical or recursive in structure, and will be referred to as "properties." These properties denote the presence or absence of particular relations between particular elements or other properties. More general relations between types of elements or properties are referred to as "concepts." The problem of pattern recognition is intimately concerned with the selection and recognition of such properties and concepts.[1] This report represents an attempt to state a simplified version of this problem in mathematical form, and suggests a tentative heuristic solution.

We consider a pattern to be an ordered sequence of n binary variables, and suppose that it is assigned to a class according to the values of m other binary variables. Our problem is to deduce the class to which an unknown pattern belongs from a series of patterns of known class. That is, we must select a Boolian function of n variables for each of the m class variables, and we must use our experience to select the function that has the highest chance of predicting the class of the unknown functions. In order to allow hierarchical or recursive structure, we must allow these variables to correspond to properties also. Thus we must imagine each of our m Boolian functions to be written as a function of functions of functions, etc. We now have a convenient method of using our previous experience, which can take the form of a number of functions that were found to be of value in the past. We shall assume that the patterns that are being investigated are such that the m Boolian functions are likely to be expressible simply in terms of these previously useful property functions, or functions 'similar' to them. In order to define the meaning of 'similar,' we can imagine that experience is stored not as a long list of property functions, but is abbreviated by using expressions to denote types of property functions which have been found useful. These expressions can be called concepts, and it is seen that the formation of concepts from property functions is essentially the same as the formation of property functions from patterns. In view of this recursive structure, we will restrict our consideration to the lowest level.

Suppose, therefore, that we have no difficulty in choosing between two possible functions both of which fit the data and are decomposable. First, how can we derive a function that is simply expressible in terms of established property functions; second, if our experience is small, how can we establish property functions that will prove to be effective? It would seem that we have two ways of treating the first question, synthetic and analytic. That is, we can attempt to build up a function from the established property functions to fit the data, or we can use the data to select likely property functions. The second method would have the advantage of finding property functions that are not well-established, and hence solve the second problem.

We shall therefore consider the following problem: Given a set of well-established property functions and a set of patterns, we wish to find a Boolian function that gives the correct pattern discrimination, and is simply decomposable, preferably into functions that are well-established.

Since Shannon's work[2] on relay circuits, Boolian functions have received much attention. In particular, much work has been done on the simplification of such functions, most of which has been concentrated on functions of a particular form, i.e., conjunction-of-disjunctions or disjunction-of-conjunctions. In the terms of the switching theorist, our problem can be stated: Given a function with many don't-care conditions, find a simple form that is decomposable, preferably disjunctively, into preferred subfunctions. Disjunctive decomposition has been treated by Ashenhurst[3] and more general decomposition by Curtis,[4] but their results are mostly of theoretical value if applied to more than approximately 6 variables. We wish to investigate a heuristic solution to the problem of decomposition.

We consider here a process that is repeated for each new pattern presented. At each stage, we shall suppose that we have derived a function that is acceptable as a theory of the discrimination process. If the next pattern is in accord with this theory, then it will be retained, and we shall be more confident of its ultimate success. If we get a disagreement, we shall wish to alter the theory to fit the new data. We consider the following process:

1. Test a pattern.
2. Select a number of functions that simplify the form of a function which fits the data. Henceforth these will be considered as auxiliary variables, indistinguishable from the original variables.
3. Drop any variables that do not appear to play a significant part in the function.
4. Test the next pattern, and go to 2.

It is seen that as patterns are examined, the variables are constantly changing. A new pattern will first be used to calculate the values of the auxiliary variables, and then the unwanted variables will be ignored. If the selection procedure in steps 2 and 3 is very good, the number of variables considered at each stage will diminish, until there is only one variable, which is a decomposable function of the original variables. If the selection is not so good, we may find that we cannot reduce the number of variables, or we cannot fit the next pattern with the remaining variables. In the first case, we probably have to give up because the computation will become too large; in the second case we have thrown away necessary information, and have to go back a few steps to recover it. This procedure seems intuitively to be satisfactory, but we cannot examine it in detail until we further define the process described in steps 2 and 3. We consider this now.

The procedure that we suggest for the selection of useful auxiliary variables is derived from the procedure devised originally by Quine[5] for the simplification of Boolian functions. He defines a prime implicant of a Boolian function as a conjunction of variables which has the property that it implies the function (i.e., if it has the value 1, then the function has the value 1), but that the conjunction of a subset of the variables is not a prime implicant. Using the notation of the switching theorist, who writes "." for "and", "+" for "or", primes to denote negation, "1" for "true", and "0" for "false", a prime implicant is of the form $v_1 . v_2 . v_3 . \dots . v_k$. Quine also shows that all irredundant expressions for a Boolian function written as a sum of products (sp) will be written in terms of prime implicants. He also gave automatic methods of deriving the prime implicants, and later work[6] has given automatic methods for deriving the irredundant expressions from the prime implicants. In particular, some of these methods can be used when the function is incompletely specified.[7] By considering the dual procedure, irredundant expressions in the form of a product of sums (ps) can be obtained in a similar way. What we suggest here is that both the prime implicants and the irredundant expressions can be used with advantage as the auxiliary variables.

To take a simple example, suppose the patterns a'b'cd, ab'cd', and ab'c'd' give result 1, and the patterns a'b'c'd, a'bcd and ab'c'd give result 0. Then the prime implicants are d', ac, b'c, b', (a+c), (a'+d'), (c+d'), and the irredundant expressions are (d'+b'c), b'(c+d'), and b'(a'+d')(a+c). If it is found that the pattern a'b'cd' gives result 0, we find that the new prime implicants are ac, ad', c'd', ab'c, b'cd, b', (a+c), (a+d), (a'+d'), (c+d'), and the irredundant expressions are (ad'+b'cd), (ac+c'd'+b'cd), (c'd'+ab'c+b'cd), b'(a+d)(c+d'), and b'(a+c)(a+d)(a'+d'). If, however, the expression (d'+b'c) had been the variable e, our procedure would have found that the expression (a+d)e would also fit the data. We might note in passing that the suggested procedure is more powerful than the standard ones used for simplifying switching circuits, and may well be useful in that area.

The reasons for suggesting prime implicants and irredundant ps or sp expressions are only intuitive. It is reasonable to suppose that with the presentation of each new pattern an attempt is made to modify the "theories" that were represented by previous functions that fitted the data. The simplest calculable functions provided for us by switching theory are these two-level irredundant expressions, which themselves are made from prime implicants. The effectiveness of the method depends directly on its ability to decompose a function, and work is in progress to determine the extent of this ability.

If the function can be simply decomposed in this way, there remains the less straightforward problem of which functions to retain as variables, and which to drop. One of the primary conditions will be that the function be in accord with previous experience. We also have to consider the case of starting from scratch, with no previous experience, so we have to fall back on the criterion of the utility of the variable in finding a simple expression for the function. Again, 'simple' is not definable, but we can probably assume that simplicity of expression in the normal algebraic sense will be a good approximation. If we accept this definition, then we shall wish to keep available those variables that have occurred recently in the simpler irredundant expressions, and drop the remainder.

Alternatively, if we observe that the determination of prime implicants requires extensive calculation, we may suppose that any successful procedure will probably generate these by a more heuristic procedure, which will use a process for generating probable prime implicants. This generator will probably be such that it tends to generate functions in accordance with previous experience, and will contain parameters for each variable denoting their probable usefulness. A simple procedure for dropping variables will be to arrange some sort of decay for these parameters, with possible resetting if they are used in a simple expression.

M. C. Harrison

References

1. M. Minsky, Proc. IRE 49, 8 (1961).
2. C. E. Shannon, Trans. A.I.E.E. 57, 713 (1938).
3. R. L. Ashenhurst, Proc. Symposium on Switching, Harvard University, 1957.
4. A. Curtis, A New Approach to the Design of Switching Circuits (D. Van Nostrand Company, Inc., New York, 1962).
5. W. V. Quine, Am. Math. Monthly 59, 521 (1952); 62, 627 (1955).
6. See, for example, S. R. Petrick, Proc. of the I.C.I.P., 1959.
7. J. T. Chu, J. ACM, 8, 497 (1961).

D. READING AS A PERCEPTUAL SKILL

Extracting information from the printed word is a uniquely psychological process inasmuch as the black marks on the page are entirely arbitrary visual targets. There are many thousands of alphabets and each is virtually as good as any other as the medium of information exchange; only conventions and rules limit their form. Reading therefore is uniquely psychological in that the target configurations, the printed words, have much less immediate and direct significance than is true for the other kinds of visual forms and objects that make up the perceptible world. Furthermore, in reading, the reader actually looks at relatively little of the printed material from which he gets his information.

We have begun to analyze some of the constituent processes of reading by varying the temporal and spatial properties of material which normal adult subjects are required to read. In experiments described earlier[1] college students were required to identify words of various kinds spelled temporally only. By means of a motion picture projector, the constituent letters of a word were made to fall on the same part of a viewing screen, presented for and separated by various temporal intervals. The principal relevance of that work here is the finding that subjects were often able to recognize and identify the constituent letters of a word without being able to identify the word itself. This phenomenon, which we encountered even more often in the work reported below, identifies a difference between two high-level perceptual skills: organizing materials for identification, and perceiving their organization. Perceiving visual targets as words would require this second stage of recognition.

Normal silent reading rates are of the order of 300 to 600 words per minute, considerably higher than those now known for auditory or tactile inputs. At 600 words per minute and five to seven letters per word, the reader is scanning fifty to seventy letters per second. If reading were phonetic, the reader would be processing 30-50 phonemes per second. But from a large number of psychophysical experiments we know that such rates exceed the sequential capacities of the nervous system. The latter is practically incapable of segregating serial events that occur within 25 milliseconds of each other. Yet even 600 words per minute is a low rate for the accomplished reader. Speed reading courses aim for 1000 to 1500 words per minute and the records, while few, are real of people capable of reading at rates greater than 10,000 words per minute. These rates of course are associated with high levels of comprehension; reading rates, however high, accompanied by lack of understanding of the printed material would be of limited interest.

This background defines a problem of compelling interest for the investigator interested in perception and in the "meaning" the observer imposes on his perceiving. Clearly, input rates handling words at 600 to 1000 words per minute define the fact that the reader is not looking at every visual element, and is certainly not sounding words to himself, but in a fashion not yet clear, short-circuits and leap-frogs about the printed page at the same time that he extracts a very high degree of its informational content. We are concerned with these processes and, ultimately with the implicit language literate educated adults use to translate the visual input into information. This, the more abstract concern, is accompanied by practical considerations for the psychology of reading. At the present time, approximately 30 per cent of all youngsters in grade school read so poorly that they require supplementary instruction, and of these, something again like 30 per cent seem to be deficient in specific perceptual mechanisms.

The method that we have chosen for investigating some of these processes studies the acquisition of skill in reading various kinds of text subjected to geometric transformations. Material of this kind, illustrated in Figs. XVIII-4 and XVIII-5, is unfamiliar enough to most college students that their acquiring skill with it magnifies the normal reading process. When slowed down in this way, these normal students reveal phenomena of the kind found characteristically among children with reading problems of a certain type (dyslexia), thereby providing us with an insight into that dysfunction. The transformations shown in Fig. XVIII-4 are simple ones with respect to the plane of the paper: rotation, reflection, and inversion. In Fig. XVIII-5 the four samples of Fig. XVIII-4 have been subjected to a second transformation, reversal of letters.

Faced with a page of material in one of the transformations shown, which they are required to read aloud as rapidly as they can, normal college students begin by attempting to solve the problem of the nature of the transformation. They verbalize explicitly the relation that the transformation bears to normal text, thus: "This looks like a p and the text is rotated so it must be a b (or q or d, the choice depending upon the transformation); Oh, that makes this word bound." But after some practice, they forego this rationalistic strategy, for they become able to read the material directly. It is this latter, programlike stage of the reading process in which we are more interested; and we find that subjects given equal amounts of practice on each of the various transformations cannot then read them equally well. In fact, the material shown in

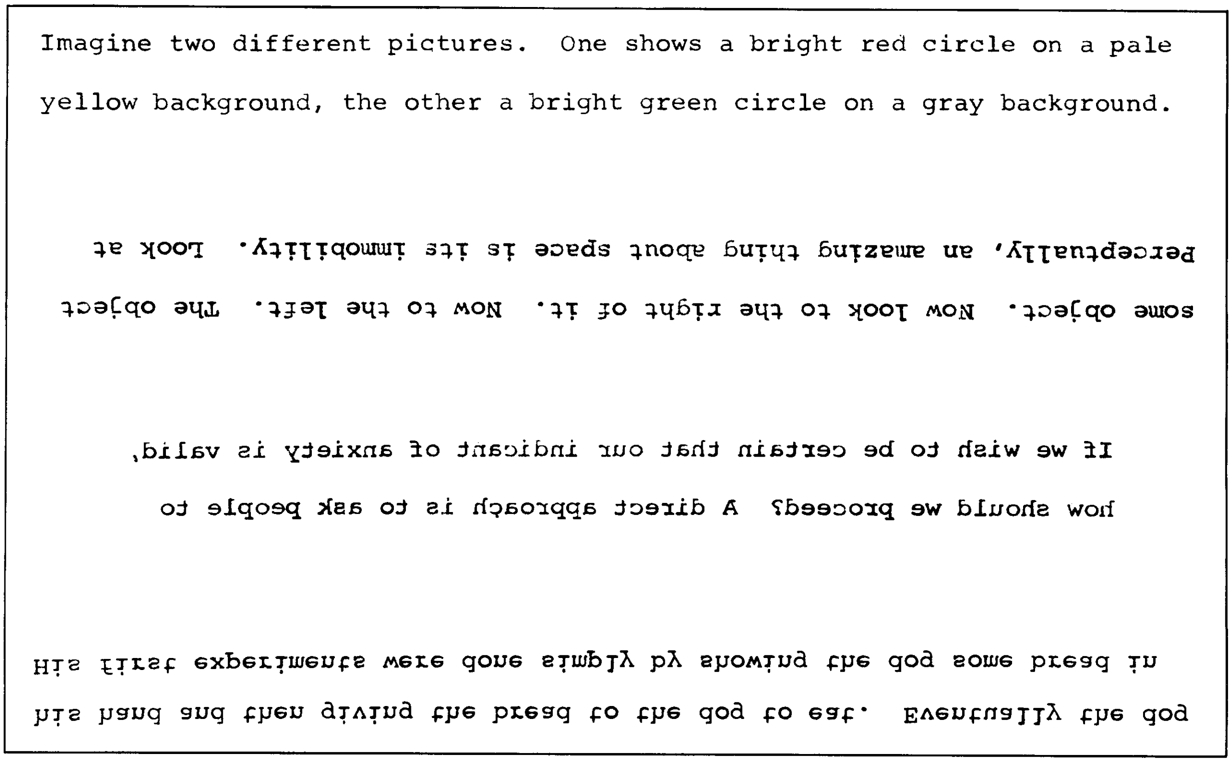


Fig. XVIII-4.

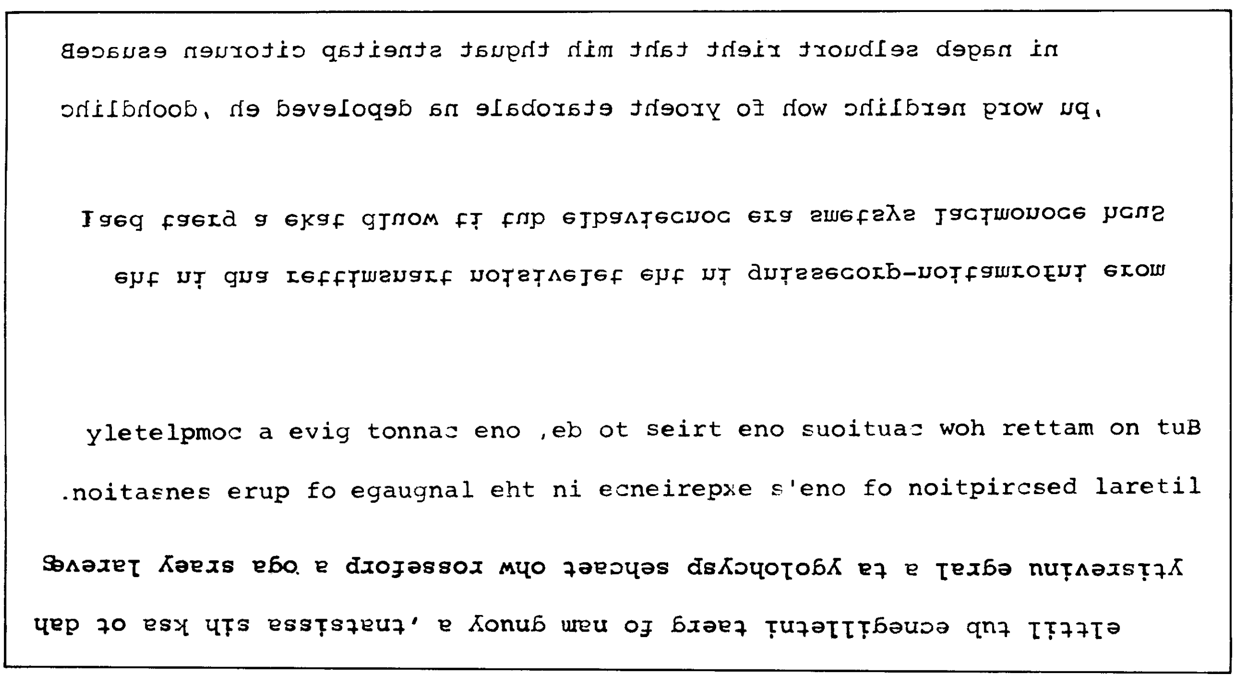


Fig. XVIII-5.

Fig. XVIII-4 is ordered there with respect to its difficulty, normal text being the easiest to read, and inverted text the most difficult of the three transformations. The fact that letter reversal is not a mere summative transformation – adding its own contribution to that of the rational transformations of Fig. XVIII-4 – is demonstrated by the finding that the order of difficulty shown in the first figure is not maintained in Fig. XVIII-5. The transformation shown second in Fig. XVIII-5 characteristically is found to be the hardest to read.

The programlike character of these acquired skills is demonstrated in turn by the fact that subjects, given a page of normal text to read after working with several pages of a single transformation, are usually unable to read the normal text, or even recognize it for what it is. But once the form of the text is identified, the page is read without impairment.

P. A. Kolers, M. Eden, Ann Boyer

References

1. P. A. Kolers and M. Katzman, Naming and Reading Sequentially Presented Letters, paper read to The Psychonomics Society, September 1963.