Psychological Review
1969, Vol. 76, No. 2, 216–220

THEORETICAL NOTES

QUANTITATIVE SPECIFICATION OF INFORMATION IN SEQUENTIAL PATTERNS

E. L. L. LEEUWENBERG

University of Nijmegen, The Netherlands

A theory is presented dealing with the specification of the units of information that are independently processed in the human perception of sequential patterns. This theory implies that sequential patterns, such as visual patterns and also Letter Series Completion Test items can be encoded in a formal language. The amount of information that can subsequently be reduced from these formal representations predicts rather successfully the degree of complexity of patterns such as measured in various tasks recording complexity judgments, the number of tachistoscopic presentations required for correct reproduction, fixation times, duration of appearance, stabilized retinal images, and errors in the extrapolation of letter series.

The question as to which aspects of objects in human perception are assimilated as independent units of information is of crucial importance in psychology. Where knowledge of these units is nonexistent, it is impossible to determine the level of difficulty of performance tasks dealing with any kind of complex stimulation. Also, for this reason it is extremely difficult to trace general relationships between, for example, the short-term memory and the long-term memory information transmission capacity, between metrical and structural information and, in fact, between all psychological functions concerned with the processing of information.

Without the pretension of providing the final answer to the above question, an analysis is proposed which primarily deals with visual patterns. This analysis is restricted to a description of patterns in terms of units of structural information. The structural information of a pattern corresponds to the dimensionality of that pattern. The concept of "structural information" has been introduced by MacKay (1950), who distinguishes it from "metrical information," which refers to the number of relevant categories on each dimension (selective information theory). In the present approach this metrical information content is left out of consideration.

In the recent past, several attempts have been made to assess the amount of information contained in sequential patterns (Attneave, 1954; Chomsky, 1957; Oldfield, 1954; Payne, 1966; Shannon, 1948; Simon & Kotovsky, 1963). The reason for presenting yet another analysis, which moreover builds upon these earlier attempts, is threefold: (*a*) it is in better agreement with the Gestalt principles; (*b*) it requires fewer symbols for the description of patterns; (*c*) it provides a better quantitative prediction of the data obtained under various kinds of experimental conditions. The correlation between this new information measure and the judged complexity of visual patterns is .97; the correlation with the number of presentations for correct representation is also .97. Moreover, there is a very clear relationship of the above information measure of patterns with the spontaneous fixation time and also with the time in which these patterns remain visible when presented as stabilized retinal images. These experimental results, as well as the rationale for the information assessment procedure to be described here, are fully reported elsewhere (Leeuwenberg, 1968). In the present context we shall restrict ourselves to a brief treatment of the coding procedure, illustrated by a few examples of visual patterns. Subsequently, we shall apply the proposed procedure to Letter Series Completion Test items.

Coding Procedure

The following formal agreements will serve as preliminaries to the coding procedure:

$$\begin{aligned} a,\, b &= a,\, b \\ (a,\, b) &= a,\, b,\, a,\, b,\, a,\, b,\, a,\, b, \text{ etc.} \\ 2(a) &= a,\, a \\ 2(a,\, b) &= a,\, b,\, a,\, b \\ 2\{a,\, b\} &= a,\, a,\, b,\, b, \end{aligned}$$

In general, brackets ( ) refer to the whole group of terms, while a brace { } operates upon the terms individually; this will become clear as we proceed.

$((a, b), (c, d)) = a, b, c, d, a, b, c, d,$ etc.
$((a, b), \{c, d\}) = a, b, c, a, b, d, a, b, c,$ etc.
$(\{a, b\}, \{c, d\})\quad a, c, b, d, a, c, b, d,$ etc.
$2\{(a, b),\ c, d\} = a, b, a, b, c, c, d, d,$

If no comma is placed between the groups of elements enclosed by brackets or a brace, the series formed by the group on the left determines the structural relationships between the elements of the group on the right, for example,

$$(2)\ (a, b)$$

This means that the numbers of b's interposed between the a's make up the series (2), that is 2, 2, 2, 2, 2, etc. Thus the whole series becomes $a, b, b, a, b, b, a, b, b, a,$ etc. Another example,

$(2, 1)(\{a, b\}, (c, d,)) = a, c, d, c, d, b, c, d, a, c, d, c, d, b, c, d,$ etc.

The letter R stands for the reversal operation:

$3R\ (a, b, c,) = a, b, c, c, b, a, a, b, c,$

$\int 2,\ 1, 3, 2, = 0, 2, 2+1, 2+1+3, 2+1+3+2$
$= 0, 2, 3, 6, 8$
$\int 2,, 1, 3, 2 = 2, 3, 6, 8$

Example: The figure

can be represented as follows

$$\int\{(\int 5(1))(\{0, 2\}, 1)(90°, -90°)\}$$

Reconstruction of the figure from the formula proceeds as follows. The subformula 5(1), occurring in the formula represents the series: 1, 1, 1, 1, 1

$$\int 5(1) = 0, 1, 2, 3, 4, 5$$

The values in this series correspond to the numbers of 1s interposed between the elements of $\{0, 2\}$

$(\{0, 2\}, 1) = 0, 1, 2, 1, 0, 1, 2, 1, 0, 1, 2, 1, 0, 1, 2, 1,$ etc. $(\int 5(1))(\{0, 2\}, 1) = 0, 2, 1, 0, 1, 1, 2, 1, 1, 1, 0, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 0$

These values refer to the numbers of $-90°$s that are interposed between the 90°s.

$(\int 5(1))(\{0, 2\}, 1)(90°, -90°) = 90°, 90, -90, -90, 90, -90, 90, 90, -90, 90, -90, 90, -90, -90, 90, -90, 90, -90, 90, -90, 90, 90, -90, 90, -90, 90, -90, 90, -90, 90, -90, 90, -90, -90, 90, -90, 90, -90, 90, -90, 90, -90, 90, -90, 90, 90$

$\int\{(\int 5(1))(\{0, 2\}, 1)(90°, -90°)\} = 0, 90, 180, 90, 0, 90, 0, 90, 180, 90, 180, 90, 180, 90, 0, 90, 0, 90, 0, 90, 0, 90, 180, 90, 180, 90, 180, 90, 180, 90, 180, 90, 0, 90, 0, 90, 0, 90, 0, 90, 0, 90, 0, 90, 180$

Step-by-step reconstruction of these angles with respect to a fixed base-axis results in the figure given above. The lengths of line forming the connections between the angles are here taken to be equal.[1]

Information Assessment

One single unit of information is conveyed by each of the operations R and $\int$ occurring in the formulas. One information unit is also assigned to each numeric value except in the series (0, 1) and (1, 0), each of which series carries only one unit of information. The justification of this procedure may be found elsewhere (Leeuwenberg, 1968, p. 22).

A matter of discussion is raised by the structure as given by the bracket and the brace signs. The question is whether or not this should be regarded as an increment in the information assessment. In order to come to a solution of this problem, we shall consider the formula

$$3(b)$$

more closely. Clearly both 3 and b should be regarded as information units, since 3 as well as b are mutually independent and indivisible data. However, if 3 has no bearing on the

[1] Of a more general nature is the coding procedure in which patterns are represented by distances instead of by angles. These distances are defined by the heights of the successive points of a pattern from a fixed base axis; these successive points are separated by a minimal discriminable standard distance. This representation is accomplished by letting the series of operations,

$$\int h,, \sin n\{\int r,,$$

precede the formulas for the patterns give above (omitting the first $\int$).

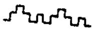
The formula for the presented figure then becomes:

$$\int h,, \sin n\{\int r,, \{(\int 5(1))(\{0, 2\}, 1)(90°, -90°)\}\}$$

where h stands for the height of the starting point of the pattern; while n and r stand for the length and direction of the first contour element of the pattern, respectively. Seeing the patterns under discussion do not differ with regard to these units of "frame information," we shall not consider them further.

TABLE 1

CODING PROCEDURE

Pattern	Figure	Code	Information
a		$\int(\alpha^\circ)$	2
b		$\int n(\alpha^\circ)$	3
c		$\int\{(2)(90^\circ, -90^\circ)\}$	4
d		$\int\{(1, 2)(90^\circ, -90^\circ)\}$	5
e		$\int\{4(0, 1, 2)(90^\circ, -90^\circ)\}$	7
f		$6(-90^\circ, 90^\circ)$	3
g		$\int\{6R(90^\circ, -90^\circ)\}$	5
h		$\int\{2(\{-90^\circ, 90^\circ\}, (7R(90^\circ, -90^\circ)))\}$ $\int\{2(7R(90^\circ, -90^\circ)), \{90^\circ, -90^\circ\})\}$	8
i		$\int\{4R(2(90^\circ, -90^\circ))\}$	6
j		$\int 3R(3(90^\circ, -90^\circ))$	6
k		$\int\{2(R(0, 1), 2)(90^\circ, -90^\circ)\}$	7
l		$\int\{2(2\{(0, 1, 2), 1\}(90^\circ, -90^\circ))\}$	9
m		$\int\{2(3)(3, 1)(1, \{2, 0\})(90^\circ, -90^\circ)\}$	10
n		$\int 10R((1, 2, 0, 1, \frac{1}{2})(90^\circ, -90^\circ))$	10
o		$\int 2((2, 0, 2), R(0, 1))(90^\circ, -90^\circ)$ $\int 2((2, 0, 2), \{0, 1\}, \{1, 0\})(90^\circ, -90^\circ)$ $\int 2(3, 5)(\{2, 0\}, 1)(90^\circ, -90^\circ)$	9
p		$\int\{3(3)(0, 1)(\{2, 0\}, 1)(90^\circ, -90^\circ)\}$ $\int\{3((2(0, 1, 2)), 1)(90^\circ, -90^\circ)\}$	9
q		$\int\{(0, \{2, 1\})(90^\circ, -90^\circ)\}$	6
r		$\int\{2R((4)((2, 2), \{0, 1\})(90^\circ, -90^\circ)))\}$	9

number of *b*'s, its value has no content. In such a case there is no unit of information that corresponds to 3. The fact that 3 implies a quantity and the fact that it refers to the number of *b*'s are intrinsically associated in each specific case. Therefore the two aspects of the numeric value correspond for each together to one single information unit. This is also the reason for not adding the brace and bracket structure to the information assessment. In effect this structure serves to show which values are referred to by numbers or operations such as ∫ and *R*. And, as we have seen, these relations are inherent in numbers and operations.

Examples

In Table 1 are given some examples of the coding procedure and of the assessment of information. An important characteristic of the coding procedure is apparent in the formula representing Pattern h (see Table 1). The first part of this formula corresponds to that of Pattern g and the second part to that of Pattern f (or vice versa). This implies that in this case the main configuration f (superstructure) of Pattern h is integrally represented in the formula of Pattern h, irrespective of the elaboration (substructure) corresponding to Pattern g. This is in agreement with the Gestalt principle concerning the invariance of structures.

From another point of view, in the proposed coding procedure, different aspects of a pattern which bear an hierarchic interrelationship are represented by coordinate units, the total of which defines the information content of the pattern. It must be realized that, consequently, the perceptual process—if it is to correspond to the present coding procedure—will not subtract units of information from a pattern in random fashion but in a specific order. This order is defined such that each newly subtracted unit of information changes the result of the units already stored towards a closer representation of the original pattern.

Series Completion Test

Finally we shall consider the applicability of the coding procedure to Series Completion Test items. When dealing with sequences of letters of the alphabet, we shall let them be represented by series of digits, such as *a*, *b*, *c*, *d*, *e*, *f*, *g*, by 0, 1, 2, 3, 4, 5, 6. Then Series Completion Test items may be coded as given in Table 2.

Using these letter series with 67 subjects, Simon and Kotovsky (1963) observed the number of correct extrapolations (C) in a given time and also the time needed for the correct extrapolation of each of the series (T). These results are correlated as follows with the proposed information measure (I):

$$r(C, I) = -.84$$
$$r(T, I) = .66$$

Simon and Kotovsky propose a coding procedure that, from a qualitative point of view, is better suited to letter sequences. On the other hand it is difficult to deduce a measure of complexity from their procedure. They them-

TABLE 2

CODING OF SERIES COMPLETION TEST ITEMS

Items	Code	Information
1. *c d c d c d c d*, etc.	(*c*, *d*)	2
2. *a a a b b b c c c d d d*, etc.	3{∫(1)}	3
3. *a t b a t a a t b a t*, etc.	((*a*, *t*), {*b*, *a*})	4
4. *a b m c d m e f m g h m*, etc.	((0, 1)({∫(1)}, *m*))	4
5. *d e f g e f g h f g h i*, etc.	∫4,, (3(1), − 2)	5
6. *q x a p x b q x a*, etc.	(2)(*x*, ({*a*, *b*}{*p*, *q*}))	6
7. *u a d u a c u a e u a b u a f*, etc.	(*u*, *a*), {{∫3,, (1)}, {∫2,, (− 2)}}	8
8. *m a b m b c m c d m*	(0, 1){{2{∫(1)}}, (*m*)}	5
9. *u r t u s t u t t u*, etc.	({*u*, *t*}, {∫17,, (1)})	5
10. *a b y a b x a b w a b*, etc.	((*a*, *b*), {24,, (− 1)})	5
11. *r s c d s t d e t u e f*, etc.	∫17,, ((1), {− 16, 15})	5
12. *n p a o q a p r a q s a*, etc.	(0, 1)({∫12,, (2, − 1)}, (*a*))	6
13. *w x a x y b y z c z a d a b*, etc.	∫22,, (1, − 21, 12)	5
14. *j k q r k l r s l m s t*, etc.	∫9,, ((1), {6, − 7})	5
15. *p o n o n m n m l m l k*, etc.	∫15,, (1, 1, −1)	5

selves suggest serveral methods. When a measure of complexity is obtained in accordance with the present procedure, that is, by totaling the number of symbols (S) occurring in each of the pattern descriptions given by Simon and Kotovsky, the correlations are as follows,

$$r(C, S) = -.76$$
$$r(T, S) = .66$$

An objection to the proposed coding procedure —in contrast to that of Simon and Kotovsky— is that for the time being it will hardly lend itself to computer programming.

REFERENCES

ATTNEAVE, F. Some informational aspects of visual perception. *Psychological Review*, 1954, **61**, 183–193.

CHOMSKY, N. *Syntactic structures.* The Hague: Mouton Press, 1957.

LEEUWENBERG, E. *Structural information of visual patterns.* The Hague: Mouton Press, 1968.

MACKAY, D. M. Quantal aspects of scientific information. *Philosophical Magazine*, 1950, **41**, 289–301.

OLDFIELD, R. C. Memory mechanisms and the theory of schemata. *British Journal of Psychology*, 1954, **45**, 14–23.

PAYNE, B. The relationship between a measure of organisation for visual patterns and their judged complexity. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 338–343.

SHANNON, C. E., & Weaver, W. *The mathematical theory of communication.* Urbana: University Press of Illinois, 1964.

SIMON, H. A., & KOTOVSKY, K. Human aquisition of concepts for sequential patterns. *Psychological Review*, 1963, **70**, 534–546.

(Early publication received July 15, 1968)