FOLDEDNESS FACTOR IN LETTER ARRAYS

JAMES PUDER
Saratoga, California

An idle diversion with pencil and graph paper involves writing some ordered sequence of letters in the cells of a grid and searching the horizontal, vertical and diagonal rows of the resulting array for accidentally-formed new words. Since there is usually a multiplicity of possible ways to inscribe a given letter string into a grid, the question arises as to whether some inscription patterns are likely to be more productive of new words than others. Logic suggests that the answer is yes. Because any new words in the array must form along straight lines, and because no new words can form entirely within the line of the original letter sequence, grid space that is occupied by straight-line segments of that sequence is partly unavailable for new-word formation. Ergo, those inscription patterns that break or fold the original letter string into the shortest straight-line segments presumably offer the greatest opportunity for new words to form.

But is this **foldedness factor**, as it might be called, a potent enough force to have any noticeable effect at the modest dimensions of ordinary letterplay? In an effort to find out, a letter sequence was inscribed into a series of identical small grids with varying degrees of foldedness and the resulting arrays compared for rates of new-word productivity.

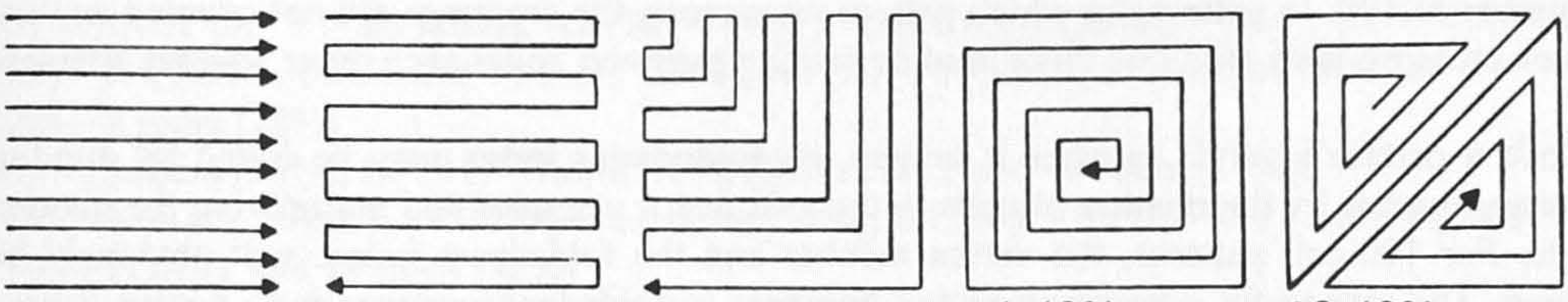
The Alpha-Omega Sequence

The letter sequence chosen for the trial was the familiar sequence of the spelled-out Greek letter alphabet: **alpha beta gamma delta epsilon zeta eta theta iota kappa lambda mu nu xi omicron pi rho sigma tau upsilon phi chi psi omega**. In addition to being well-known and alphabet-related, this sequence commends itself by having a favorable (45 to 55) ratio of vowels to consonants and by happening to consist of exactly 100 letters, which may be neatly arrayed as a 10x10 square. This is advantageous because a square is the optimal rectangular shape in which to array a letter sequence in order to maximize the opportunities for new-word formation.

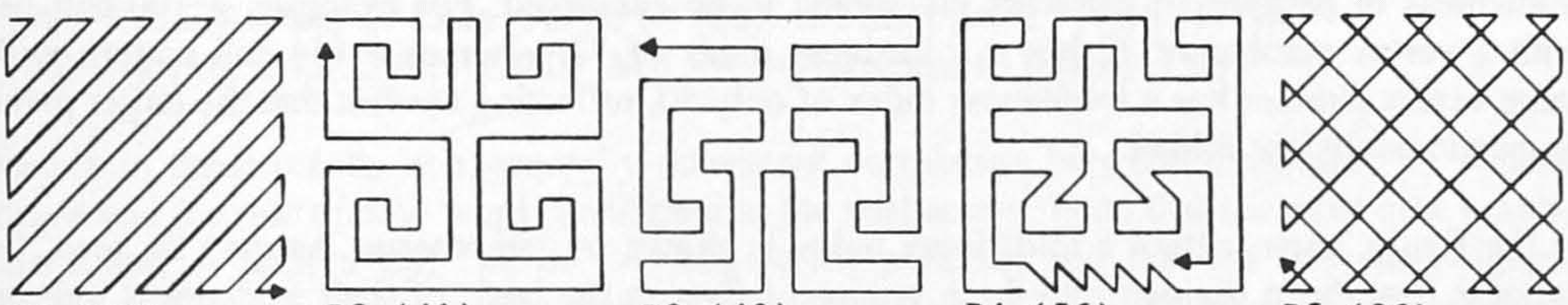
The Inscription Pattern

Diagrams of the fifteen 100-cell inscription patterns selected for comparison are shown in the figure on the next page. Note that in every pattern but two (A1 and C5), the letter sequence is a single unbroken string. In pattern A1 the sequence is divided into ten segments whose serial order is from top to bottom, and in pattern C5, a knight's tour, the sequence is completely disjoined, with no two consecutive letters in it being adjacent in the grid.

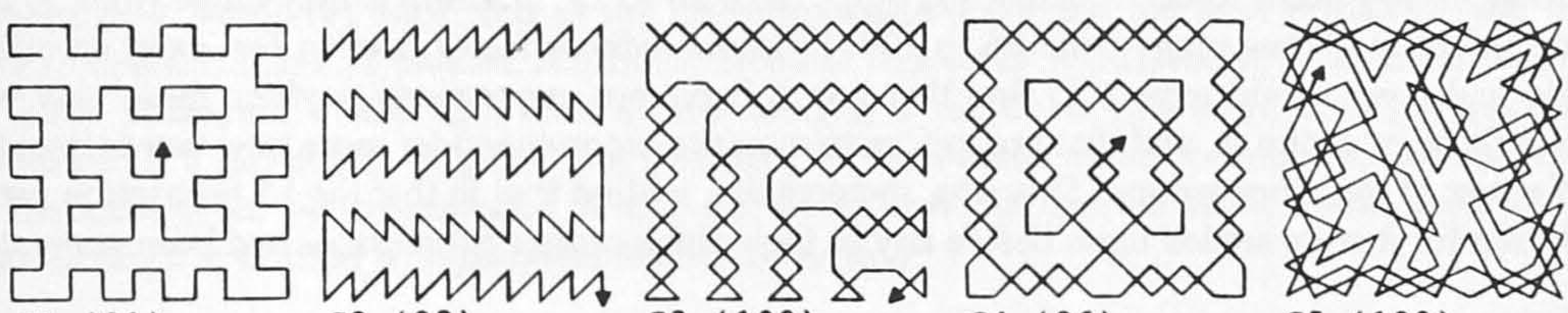
A way to quantify the foldedness of an inscription pattern is to count the number of cells in the grid that it fills in which a pattern line does not enter and exit the cell on opposite sides or corners. In practice, this amounts to simply counting the number of vertices in a pattern, with each bend and each line end counting as one vertex. Thus, pattern A1, with 20 line ends and no bends, has a **vertex number** of 20, while pattern A2, with 18 bends and two line ends, also has a vertex number of 20. Patterns that jump between non-adjacent cells are considered to create a vertex in each of those cells, and so pattern C5, which consists of 99 such jumps, has a vertex



A1 (20) A2 (20) A3 (28) A4 (20) A5 (20)



B1 (36) B2 (41) B3 (40) B4 (56) B5 (36)



C1 (91) C2 (92) C3 (100) C4 (86) C5 (100)

Diagrams of the trial inscription patterns (above) and their alpha-omega sequence expressions (below)

```
a l p H a b e t a g   a l P H a b e t a g   A L A G O N A M P S   a l p H A b e t a g   t e h T A t e a t u
A m m A d e l T a e   E a T L E d a M m a   h p t a l z L b u i   e T A i o T a k a a   A e d A M M a e m n
P s i L o n z E t a   P s i L O N z E t a   a b e M i E a d u L   h R O n p i R h p m   i L h P L g z a u u
E t A T h e T A i O   O i a t E H T A t e   E d A m S t p A A o   T c l O n P h O p m   o t A A A n d a u x
T A k A p P a L a M   T a k a p P A L a m   L T a E p A P M y n   a i i o M e i s A A   t a B T o b t s p i
B D a m U n u x I O   o i x u n u M a d b   E h T a T E A u a p   t m S I a g c i L D   a e E l M a P i s o
m I c R o N P i r H   m i c r o N P I r h   T A i O T a k n m h   e o P S p i h G a e   k P i a m I a O i m
o S I g m A T a u u   u u a t A m g i S o   R c I m O i x u g i   a I u U A t A m m l   A s l g h g E m L i
p S i l O n p h i c   p s i L o n p h i c   o n p i r H o s i c   T x u N u M a d b t   p a i c i h p N o C
h i p S i o m e g a   a g E m o i s p i h   a g e m o i S p i h   e z n O l i s p e a   p s o h r i p n o r
A1 [36%]              A2 [29%]              A3 [39%]              A4 [30%]              A5 [25%]

a l b e D E E T p p   a l p h l T E t a T   a L p h a B E t a g   M M A P S i l e T a   p L s i a p n p h i
p a t A L z a A a r   a a t a e A A z N h   a g E M O D A m M a   a E d E a t o A T t   a h L p p K i o c p
h a M t N E k L c o   g g e b D e T E o E   h i p S I E L T A e   g L T a m a N z E h   z O a A E R A H R n
G m A o t A A I n u   e a m M A P S I L T   c u a T A H T p T p   a O s I g U u P S e   n e L b H A i t O C
A E L A T m M P a u   m o I S P A T O i a   i u i g m E T A E S   T H r i P n O l i t   b a T O e P t l O i
P i T o b o i t p c   a T a u i k N u x i   h p s o N u O i a I   e u x i n P s i o A   m D S a S T I L M i
S h i D I R A S i h   m l I u h A u d b o   p S i h u M t e t L   B n o o R I h p M i   m I A i e S A O E a
E a a x h m i h i m   g o s P c p m a m m   n o L r x a a z n O   a U m I c C h I e o   g A O m p t i G T d
t m u o g l p p o e   i n p h I p A L A I   o n p I i d k a p P   h m A b a A P a g t   a M T u u x a E A A
u n s I O N S i g a   s o h r i P n o r c   r c i m O b m a l a   p L a D m l p a k a   e g u a u n h t M m
B1 [36%]              B2 [34%]              B3 [36%]              B4 [38%]              B5 [42%]

p H A t A M M E L T   a p a e a a M d l a   a A p L A D E M a A   i c t o T p e r a i   a G d h L A p x A a
l A b E g A a d e a   L h b T G M A E T e   e L t h E b A t M g   m e r A n h i t h o   a i L m A P T t N i
c R o i R s i g p s   I T H a E t Z o i P   p E t B i A T A k p   a o H I G H m m t o   i a H E I P u O U e
i M N P h O a m l i   O A E T T A E N L s   h S D A M O L a p a   i T P A C A i a s a   c m S P N k L u L m
i o i O m a T a o n   T K P A A b a U u i   i T A R O M P i r o   e x t N A A d p k i   g h E I h S a I p u
x u s p E g u U e z   a A P L M d M N x o   a l c M A N G I S h   u z O E G L S E G A   a o M m T o a p M a
U N H I P N p s T a   u T m I O r P O c m   o t u I T L O H p i   N N B L p E L I P m   a E A t P E A S c O
m A c i h o l i t e   u A a G S H I N r i   e n M n i A I N C H   U O A h I T m O a p   H t G e o a N t m I
b D a p k a i a a t   p I O p i H P i m g   z a u o u S P O M g   L M P D a S a u a t   a b Z I T S O D i R
m A l p a T O T E h   s L n h c i S o e a   t e i x p u i S a e   i S a E b m p l u a   E R T I N p a o l B
C1 [35%]              C2 [52%]              C3 [50%]              C4 [43%]              C5 [51%]
```

number of 100. In patterns in which pattern lines cross, the crossings are not counted as vertices; such crossing lines might be thought of as passing over and under each other without intersecting.

Once a pattern's vertex number is known, its **foldedness index** may be found by dividing its vertex number by the number of cells in the grid that it occupies and multiplying the quotient by 100. For 100-cell patterns, the vertex number and the foldedness index. will obviously be the same. The reason for converting vertex numbers to foldedness indexes is to permit degrees of foldedness in patterns of different magnitude to be compared. For example, a 100-cell pattern with a vertex number of 72 has a foldedness index of 72, whereas a 144-cell pattern with the same vertex number has a foldedness index of only 50, reflecting the fact that the larger pattern is proportionately less folded.

In the figure, each pattern's foldedness index is shown in parentheses. As may be seen, the 15 patterns have been divided into three groups of five on the basis of their foldedness indexes. In group A the pattern foldedness indexes range from 20 to 28, in group B they range from 36 to 56, and in group C they range from 86 to 100. If the foldedness factor does in fact exert an effect at this scale, we would expect to find that group B pattern expressions produce more new words than those of group A, and that group C pattern expressions engender more new words than those of either of the other groups. This was, incidentally, a blind trial in that the 15 inscription patterns chosen for it were settled upon before any of their alpha-omega expressions had been generated.

Word Search Results

The alpha-omega sequence expressions of the trial patterns are displayed in the bottom half of the figure. All countable new words found are shown in capitals. Counted were complete uncapitalized words of four or more letters that are printed in boldface type in the main section of Merriam-Webster's New Collegiate Dictionary, Tenth Edition, and their inferred plurals. Not counted were words already embedded in the alpha-omega sequence such as ABET, AGATE, CHIPS, LAMB, LISP, MADE, MICRON, PEAT, SILO, TAKA, TEAT and various other. However, these words were deemed countable if their formation in an array was at least in part the product of chance. Reversals were counted as two words.

Three methods of measuring the amount of new-word formation in a letter array suggest themselves. One is by a **simple word count,** which is the total of all countable new words in an array; another is by a **weighted word count,** which arbitrarily assigns higher point values to longer words; and the third is by **word density,** which tallies the number of letters in an array that contribute to countable new words. In each case, dividing the raw count by the number of letters in the array yields a ratio that is comparable between arrays of different size. It is not readily apparent which of these standards provides the fairest measure of new-word formation, but word density—i.e., the percentage of an array that is occupied by countable new words—may be the simplest of the three to work with. In the figure, each array's word density is shown in brackets.

Twenty-three five-letter and 188 four-letter countable new words were found in the trial arrays. No countable new words of six or more letters were found, although the touring knight does, in a seven-letter phrase, bid us with BRIO to EAT PEAS. The 23 five-letter words are listed below. Among the 188 four-letter words (not listed) were such possibly unfamiliar terms as agon, baal, deet, geta, holp, lota, luna, mana, mano, mora, nett, puna, sett, sori, taka, tala, tali, tapa, tele, tepa, tipi and toom, the phrase fragments alai, alma, lese, noms and pima and the hyphenated words no-no and non-U. In the following summary of results, each pattern's foldedness index is shown in parentheses and its alpha-omega expression's word density is shown in brackets:

A1 (20)—SIRUP and 9 fours [36%]; A2 (20)—IMPEL and 8 fours [29%];
A3 (28)—RATES, SHOOT and 14 fours [39%]; A4 (20)—AORTA and 9 fours [30%];
A5 (20)—8 fours [25%]
B1 (36)—12 fours [36%]; B2 (41)—ADMIT, ANOLE, PIPIT and 10 fours [34%];
B3 (40)—BATTS, EASEL, POLIS and 12 fours [36%]; B4 (56)—TOTES and 13 fours [38%];
B5 (36)—TOAST and 14 fours [42%]
C1 (91)—LEMMA, TUTOR, UNHIP and 8 fours [35%]; C2 (92)—20 fours [52%];
C3 (100)—MEDAL and 23 fours [50%]; C4 (86)—STELA and 13 fours [43%];
C5 (100)—GAMES, GLENS, HEMIN, MILPA, NITRE and 15 fours [51%]

Discernible in these results is a generally consistent correlation between the degree of pattern foldedness and the rate of new-word production in the trial arrays. Note that the most new words, (24) are found in pattern C3, one of two 100-vertex patterns, and that the two next-highest scorers, (20 new words each) are also group C patterns, C2 and C5. At the other end of the scale, the fewest new words (8) are found in the 20-vertex pattern A5 and the next fewest (9) are in another foldedness index 20 pattern, A2. In the arena of five-letter words, the 100-vertexed knight is the clear champion of the lists with a total of five, a muscular two more than his nearest competitors. (It's the power of peas!" he trumpets, as his Peasco sponsors beam. Once upon a more innocent time, he championed peace, not peas; note his storied name, now somewhat transposed but still infused with *pax,* leading off his array.)

Analysis

Average folded indexes and average new-word counts of the patterns in each of the three inscription pattern groups are collected in the following table. For the weighted word counts, four-letter words were assigned a value of one and five-letter words a value of six.

	Foldedness Index	Simple Word Count	Weighted Word Count	Word Density (%)
Group A	21.6	10.6	15.6	31.8
Group B	41.8	13.8	21.8	37.2
Group C	93,8	17.8	27.8	46.2

Even more strikingly than the individual pattern results, these group results imply a definite connection between the degree of inscription pattern foldedness and the rate of new-word productivity in the trial arrays. Where there is an increase in pattern foldedness index from group A to group B, there is a corresponding increase in group B's new-word productivity over group A's, and where there is an increase in pattern foldedness index from group B to group C, there is a corresponding increase in group C's new-word productivity over group B's. These consistent correspondences would appear to leave little doubt that the foldedness factor does indeed exert a significant influence in arrays of this size.

A less-expected finding was that new-word production in the arrays does not increase in direct proportion to the increase in pattern foldedness, but rather increases at a decreasing rate as foldedness increases. In retrospect that relationship might have been anticipated, given that nothing hinders pattern foldedness from rising to 100%, whereas additional new-word formation becomes increasingly unlikely as the word density of an array approaches 100%