A Thesis submitted to the Lakehead
University Faculty of Arts in partial fulfillment of the requirements
for the Master Degree in Psychology

THE CAPACITY FOR VISUAL PRODUCTION
AND
DISCRIMINATION OF RANDOM ARRANGEMENTS

Dr. N. Ginsburg (Supervisor)

Dr. S. Goldstein (Second Reader)

Submitted by: Garth R. Donald ©

October, 1987

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Abstract

Humans are reportedly poor randomizers. One reason suggested for this is that they may have difficulties in discriminating random from non-random stimulus arrangements. If this is so then there should be a correlation between production and discrimination of stimulus arrangements. Ginsburg and Goldstein (1987) have advanced a method of measurement that allows discrete quantification of subject responses along a continuum of arrangement from regularity through randomness to contagiousness: the cluster continuum. The present research has incorporated the cluster method of measurement in two randomization and two discrimination tasks. Participants were required to (1) arrange a number of dots on each of a square and hexagonal field so as to represent their subjective view of randomness and (2) discriminate between stimuli of varying degrees of randomness with separate tasks requiring this discrimination on each of the grid fields used in randomization.

Results indicate an inability by subjects to produce random arrangements with a tendency for subjects to err on the side of regularity in production. There was a significant correlation between the number of correct choices in the discrimination task and the measure of cluster in the production task. Results are

discussed within the context of previous investigations. Models for interpretation of the production/discrimination relationship are discussed.

Acknowledgement

Many thanks to Norman Ginsburg for his patience and support during the preparation of this manuscript. As well, I wish to express gratitude to my parents, Robbie and Marlyn Donald. Without their support and nurturance, this manuscript could never have seen fruition. Judie, for you, my appreciation is ineffable.

Garth Donald

TABLE OF CONTENTS

LIST OF TABLES, FIGURES AND APPENDICES

TABLES

FIGURES

APPENDICES

APPENDICES (cont'd)

A variety of research has sought to investigate the randomization abilities of humans. Experimental evidence has generally indicated that humans are poor randomizers (Wagenaar, 1972). This result, though striking in nature, is not fully understood. Two interpretations have been offered for the failure to randomize properly: (1) factors specific to the randomizating situation, such as the concept of randomness a subject holds, or strategies they may use in tasks, and (2) non-specific factors, such as memory, attention, and boredom. Wagenaar (1972) has labeled these non-specific factors "functional limitations." While the investigation of non-specific factors is important and will be reviewed here, the present study is more concerned with factors specific to randomness. Two approaches to randomness will be examined, it's production and it's recognition. Then, the relationship between the two and the light this might shed on the concept the subject has of randomness will be considered. This concept may underlie the kinds of results found in both tasks. Finally, the difficult topic of the measurement of randomness will be dealt with .

Production Tasks

A number of experiments have been conducted where subjects were instructed to produce a random series of events from a finite set of responses. Generally, subjects have shown an inability to randomize. Some investigators have attributed this to functional limitations. For example, Tune (1964) suggests that non-randomness in production is due to limitations of short term memory. Other theorists suggest limitations based on one's limited capacity for generating information, boredom on the part of the subject and attention/distraction difficulties of tasks (Baddeley, 1966; Weiss, 1964). All of these suggestions find some support; however, none finds complete agreement across the literature.

Rath (1966) conducted an experiment in which paced randomization was investigated on three levels. Levels consisted of paced randomization of either binary, decimal or alphabetical characters. Subjects were provided with booklets in which to record responses and were instructed to randomize using the urn model as an example. Pacing consisted of the experimenter writing booklet page numbers on a blackboard every five minutes, a cue that subjects were instructed to keep abreast of. Results indicated that subjects were poor randomizers. Biases in this respect included a culturally bound preference for specific characters and a preference for symbols found in higher frequency in the English language (i.e., e,t,a,o,n,r,i and s). Further, it was noted that avoidance of repetition and preference for alternation occurred across all symbol sets. A negative correlation between rate of character generation and number of symbols in set was found.

The results of this investigation point to a number of factors that mediate experimental results. First, it has been shown by Rath (1966) that as the number of alternatives in a set increases, a subject's ability to randomize decreases. Similar results have been reported by Warren and Morin (1965), Baddeley (1966) and Newman and Collier (1952). This finding suggests a functional limitation which may involve memory, where as load increases, ability to randomize decreases (Tune, 1964).

Another element suggested in Rath's (1966) investigation that is related to memory load is pacing. Baddeley (1966) required subjects to call out a sequence of 100 letters. Responses were required (paced) at one of four rates; one letter per .5, 1, 2 or 4 seconds. The randomization task was explained on the basis of the urn model. Analysis revealed increasing sequential redundancy as rate of generation increased. This relationship between rate and ability has been supported by others, however, rate has been shown to both increase and decrease non-randomness leaving the effects of this factor unresolved (Warren and Morin, 1965; Baddeley, 1966; Teraoka, 1963, Evans, 1978).

As had Rath (1966), Chapanis (1953) reported a tendency for subjects to arrange alternatives in a natural order when producing a random sequence. In Chapanis's investigation (1953), a comparison between subjects with (sophisticated) and without (unsophisticated) training in mathematical and probability theory was made. Analysis of subject's self paced randomization of the digits 0 to 9 indicated a subjective preference for particular numbers. Repetitive pairs and triplets were generally avoided by unsophisticated subjects. Sequences in which all digits differed were preferred. The sophisticated group more closely approximated randomness though neither group produced truly random sequences. Similar results reflecting response preferences and a natural ordering have been reported (Teraoka 1963, Lincoln and Alexander, 1955, Goodfellow,1940). However, this factor is not active when alternatives with no natural ordering are used as stimuli (Teraoka, 1963).

Mode of production within experiments has also been seen as influencing functional limits (Wagenaar, 1972). For example when previous responses are unavailable to subjects they may have difficulty in achieving randomness because of the limitations of short term memory (Tune, 1964). As well, production tasks are boring. Boredom has been found to reduce randomization abilities (Weiss, 1964). More specifically, response sterotpy increases with the amount of time spent on task (Weiss, 1964).

Another procedure which may reflect functional difficulties is one of time sharing (Wolitzky and Spence, 1968). When subjects are required to generate random sequences while simultaneously engaging in a secondary task, they are less likely to approximate randomness. Such a procedure would reduce attention and increase memory load. Baddeley (1966) has shown non-randomness to increase with secondary task demands. Evans (1978) however, notes that such deterioration of randomization abilities from baseline show gradual improvement as learning of a secondary task occurs.

The results of these studies point to the conclusion that humans are generally unable to produce random sequences. Instead what we find is a pattern of response with non-symmetrical and alternating patterns being preferred to long runs of homogeneous responses and/or statistically random patterns (Wagenaar, 1972).

Discrimination Tasks

Few authors have conducted experiments involving both production and discrimination though this would seem desirable since it is possible that the same processes underly both tasks (Kahneman and Tversky, 1972).

Cook (1967), for example, investigated subject abilities in recognition of different degrees of bias in binary strings. In this investigation, subjects were instructed to make comparisons of string list pairs and report which of the pair was more patterned. Results indicated that subjects were able to discriminate, however; stimuli were so obviously non-random that results are not seen as decisive (Wagenaar, 1972). No production task was included. Similiarly, Baddeley (1966), indicated that subjects were able to discriminate random from non-random series, though no data were presented.

In a more thorough investigation Wagenaar (1970) had subjects judge binary sequences of black and white dots presented visually via slides. Sequences of slides were constructed so the probability of repetition was .2, .3, .4, .5, .6, .7 and .8. Subjects chose around .4, indicating they were unable to select random sequences (i.e. random equals .5 in this experiment). Wagenaar interpreted this as indicating bias against long runs. Similiar results are reported by Mittenecker (cited in Wagenaar, 1972) and Zwann (cited in Wagenaar, 1972).

Wiegersma (1982) conducted a comparative study of production and discrimination of randomization. Discrimination involved subjects selecting one of seven sequences of black and white dots which appeared to be a chance sequence. Probability of repetition was identical to that used by Wagenaar (1970). Production tasks involved randomization of numbers (1-8) and a tone sequence. In general, no significant correlations between randomization and judgements were found. Wiegersma (1982) concluded that the subjective concept of randomness may be of little use as correlations did not support this notion. That is, Wiegersma (1982) predicted that if a subject used a subjective concept of randomness as an internal model of randomness then he would both produce and discriminate on the basis of this model. Wiegersma (1982) concludes that this is not the case as his results do not support this prediction. In turn, Wiegersma (1982) suggests that the subjective concept of randomness is of little use as a theoretical explanation for the diverse results in this field of research.

It should be noted, however, that Wiegersma's investigation (1982) compared production in one modality with discrimination in another. The discrimination task involved visual stimuli, while the production task used the vocal-auditory domain. Any relation between production and discrimination might have been obscured by this modality difference. One of the purposes of the present study was to compare the performance of subjects on production and discrimination tasks within the same sense, namely vision.

Judgements of Representativeness

An alternate and perhaps more fundamental approach to the problem of failures in randomization tasks has been forwarded by Kahneman and Tversky (1972). In contrast with the authors outlined so far, Kahneman and Tversky (1972,1982) suggest that failures in randomization result not from a failure to apply implicit knowledge because of functional limitations but because of the use of heuristics. More specifically, Kahneman and Tversky (1972) suggest that when making judgements, subjects " replace the laws of chance by heuristics, which sometimes yield reasonable estimates and quite often do not" (Kahneman and Tversky, 1972, p.430).

One such heuristic that Kahneman and Tversky (1972) discuss is called representativeness. According to these authors, subjects who follow this heuristic when making judgements, evaluate " a sample , by the degree to which it is: (i) similiar in essential properties to its parent population; and (ii) reflects the salient features of the process by which it is generated" (Kahneman and Tversky, 1972, p. 430). A subject will therefore select or produce a given event if it appears more representative of it's parent distribution than a given alternative. The notion of representativeness is in stark contrast to the notion of humans as intuitive statisticans and forms another side in the randomization paradigm. If such a representative heuristic is operative then it may affect both production and discrimination judgements of randomness.

This notion of an internal model and it's effect on a subject's performance is viewed from a different perspective by Diener and Thompson (1985). As opposed to the notion of a random representative heuristic Diener and Thompson (1985) propose that, "subject's judge a sequence to be random only after eliminating all tenable alternative hypotheses for the production of the series" (Diener and Thompson, 1985, p. 433). That is, "subjects decide that a sequence is random by eliminating alternative nonrandom hypotheses, rather than by directly recognizing the series as representative of a random process" (Diener and Thompson, 1985, p. 433). What Diener and Thompson (1985) are suggesting is quite different from Kahneman and Tversky (1972) and leads to a very different set of predictions for the relationship between production and discrimination.

Relationship Between Production and Discrimination

As we see, it is possible to view both production and recognition of randomness in relation to cognitive structures, or what Kahneman and Tversky have refered to as heuristics (1972). Cognitive structure was defined by Neisser as " a nonspecific but organized representation of prior experience" (1967, p. 287). It is suggested that we carry within us cognitive structures that represent parts of the cluster continuum (see Figure 1). When we are asked to judge whether a stimulus is random, we compare it with these structures or schemata. But what kind of schemata are these? Two possibilities have been introduced. On the one hand, they could be models that represent randomness itself, as Kahneman and Tversky (1972) have suggested. But there is another possibility, that was put forth by Diener and Thompson (1985). It is that the schema represent nonrandomness, rather than randomness itself. According to this view, when asked to judge whether a stimulus is random, a subject will not try to match it to an internal random template. Instead, the attempted match is to regular or clustered models. If the stimulus fails to match either of these, then it will be classified as random. In other words it amounts to classification by default.

Each of these models leads to it's own prediction in regard to the relationship between production and discrimination tasks;

(1) Random Model (Kahneman and Tversky, 1972).

This model assumes that the basic process involves a comparison with an internal model of randomness. That is, when a subject is asked to make a judgement as to whether a sequence is random, he compares that sequence to his internal random model and bases a decision on how closely the two match. The more realistic the internal model, the closer the subject's production would be to physical randomness and similiarily the more accurate his discrimination ability. Therefore this model predicts a negative correlation between discrepancy from randomness in production and success in discrimination.

(2) Nonrandom Model (Diener and Thompson, 1985).

This model assumes that the basic process involves comparison with nonrandom schemas. Two cases need to be considered:

(A) Clustered Schema: If using a clustered schema, a subject will decide on randomness by contrasting stimuli with it. The more successful he is in doing this, the further away from the cluster end of the continuum his productions will be. Within this context we would predict a negative correlation between a subject's production score and his success in discriminating clustered stimuli from random stimuli.

(B) Regular Schema: If using a regular schema, a subject will decide on randomness by contrasting stimuli with it. The more successful he is in doing this, the further away from regular his production will be. We can therefore predict a positive correlation between a subject"s production score and his success in discriminating regular stimuli from random stimuli.

For the cluster continuum at least two kinds of discrimination are possible:

(1) Discrimination between random and regular. Within this range, there could be a random standard combined with regular variables, or visa versa.

(2) Discrimination between random and clustered. Within this range, there could be a random standard combined with clustered variables, or visa versa

Thus the production-discrimination relation could be studied in four ways. Ideally one would use all four procedures, but for practical reasons it is desirable to limit this experiment to one of them. While the choice is arbitrary, the random-cluster part of the continuum was chosen since this is an area about which little is known.

Measurement of Randomization

Perhaps the most difficult problem in this area of research has been the criterion necessary for calling a series random (Wagenaar, 1972). Sufficient conditions for non-randomness exist to the extent that one trend in a series can be shown (Wagenaar, 1972). On the other hand, conditions for randomness require that no single trend be present. In essence, randomness is disproved with relative ease while proving it is difficult. A related problem is that of degrees of randomness. In this respect, we find a difficulty in quantifying increases and decreases in randomness such that a series may or may not be random, depending upon criterion used. To date, no parsimonious method of analysis has been available for the discreet quantification necessary in this respect. Most measures available cannot disprove all serial regularities nor can they be used to quantify increases/decreases in randomness (Wagenaar, 1972).

Pielou (1977), through investigation of dispersion of natural phenomenon (i.e. trees), has suggested that randomness within a distribution of elements be quantified on a continuum. According to Pielou (1977) natural phenomenon are distributed contagiously, appearing clumped or aggregated. In this respect, a set of elements can be quantified within space and time, their distribution varying from regular through random to contagious. Quantification along this continuum is calculated via variance estimates. Variance of elements across space, in relation to mean number of elements per unit space, is the defining characteristic of distributions. According to Poisson, when items are distributed randomly their variance is equal to their mean. Such an arrangement has come to be called a Poisson distribution (Lewis, 1960). Divergence above (V> mean) and below (V< mean) this point represents a distribution's degree of contagiousness or regularity respectively (see Figure 1).

Use of this system of spatial analysis in specifying the randomness of a stimulus has been forwarded by Ginsburg and Goldstein (1987). These authors advocate the use of the cluster score (C - score) which is equal to the variance of a distribution divided by it's mean (see Analysis section of Method). This will be used within the context of the present randomization/discrimination experiment.

This system of classification has certain advantages. As a method of measurement, it allows the discreet quantification of subject responses, overcoming past methodological difficulties in this respect. Further, the continuum of quantification allows a precise analysis with respect to increases and decreases in subject's responses and thereby degrees of randomness. This research seeks as one of it's goals, to test the C-score as an alternative to methods such as runs, autocorrelation, information and n-grams. It is proposed that it will be a more exacting method of analysis in the randomization experiment.

A second advantage of this system is that the measure can be applied to any kind of task in any modality. In the present context, this facilitates the administration of both production (output) and discrimination (input) tasks within the same modality. In this study, the use of visual arrays for both production and discrimination will be employed. This approach will afford task comparision as well as minimize factors which have been seen as functional limitations. For example, as visual tasks, events will be simultaneously available to subjects, therefore eliminating the necessity of short term memory. It is proposed that this system of classification and the use of visual arrays will allow more detailed understanding of subjects' performance on tasks. Further, should the method be found reliable, it will be useful as a tool in future research.

In sum, this study has the following purposes:

(1) To measure individual differences in the production of randomness

(2) To measure individual differences in the discrimination of randomness

(3) To examine the relationship between individual

differences in the ability to produce and to discriminate random

(4) To determine the reliability of scores on production across different stimulus numbers and field shapes

(5) To determine the reliability of scores on discrimination across different stimulus numbers and field shapes.

METHOD

Subjects. A total of 40 subjects participated in the study. Twenty two were introductory psychology students with the remaining eighteen taken from a psychology statistics course. All subjects were attending summer sessions at Lakehead University. Participation was on a voluntary basis with students from the introductory psychology class receiving course credit for participation. Of the forty subjects, ten were male while thirty were female. Mean age across the group was 29.4 with mean ages 29.8 and 29.2 years for males and females, respectively. All subjects participated in every task with data being collected on an individual basis.

Task 1: Production using the square grid.

Apparatus. A thirty by thirty grid, containing nine hundred internal cells, served as the stimulus field (see Appendix 1). Subjects were provided with a box containing fifty dots with which to complete the experimental task. Dots were 3/16 of an inch in diameter and punched out of black bristol board.

Procedure. Upon entering the experimental room subjects were seated and given the following instructions:

> In front of you lies a grid and a number of dots. Your task is to arrange the dots randomly on the grid. In doing this, place only one dot on the grid at a time. You may place no more than one dot per square, however, you may rearrange the dots until you have what you want as a final product. Remember, try to place the dots into what you consider to be a chance or random arrangement. Any questions? Begin.

At this point, questions were addressed and instructions repeated until subjects were believed to have a thorough understanding of task requirements. No further explanation of random was given.

Analysis. Variance calculations were computed for completed grids in the following manner: A grid overlay, consisting of a ten-by-ten matrix was placed over the task grid (see Appendix 2). The overlay had one hundred internal cells, each enclosing nine of the task grid cells. Frequency of dot occurrence within the overlay grid was recorded. These figures were then applied to a standard variance formula yielding a variance score for each subject. Subsequent variance scores were then transformed to cluster scores (c score = variance/mean).

Task 2: Discrimination using the square grid.

Stimuli. A total of sixteen stimuli, eight random and eight contagious, were constructed in the following manner: Dots, (N=50), having a diameter of 3/16 of an inch, were punched out of black bristol board and spread over a square field containing one hundred cells (see Appendix 3).

Each cell constituted a sample of space and enclosed a three-by three matrix used for dot location. Patterns were determined by the Poisson distribution: $P(n) = c(m^N/n!)$, where n = number of events (dots) per sample of space, m = mean number of events per sample and $c = 1/e^m$ (e = 2.7183). Given one hundred samples of space, $100P(n) = F(n)$ = frequency of n to the nearest whole number. A probability table was constructed for each level of variance and adjusted so that $F(n) = N$, where N = Grand Total. This was necessary as N was a whole number (i.e. whole dots). For each of the eight random patterns, the mean number of dots per sample of space was equal to the variance of dots across samples, thus fulfilling the criterion for randomness. For the eight contagious patterns, variance of dots across samples was greater than the mean number of dots per sample of space, this to the extent needed to product required C's (clusters).

To ensure all displays were approximately equal in area, four peripheral cells were occupied (1,10,91,100;

see Appendix 3). Once peripheral assignment was completed, the remaining cells, along with corresponding internal dot position, were determined by a random number table. Once dot position was determined, dots were glued to their specified locations on a 21.6 by 27.9 cm. blank sheet of typing paper. Photocopies of these served as the experimental stimuli.

Procedure. Upon completion of Task 1, subjects were informed they would be completing a discrimination task and were instructed in the following manner:

> I will be presenting you with pairs of stimuli.In each case, I want you to make a decision. You must decide which of the pair is the most random arrangement of dots. If the one on the left looks random to you, then mark the appropriate space on your answer sheet as such. If you choose the right, then record this. If both stimuli look like equally random arrangements then record your answer as equal. Any questions?

At this point, any questions were addressed and instructions repeated until subjects were believed to have a thorough understanding of task requirements. No further explanation of random was given.

Eight set of stimulus pairs (contagious and random stimuli), glued to 35 by 50 cm. sheets of black bristol board, were presented. Positioning of the random pattern (left or right) (see Appendix 4) in the pair was determined using Gellerman's table (Gellerman, 1933). Stimulus pairs were arranged in an ascending manner such that contagious stimuli appeared in order from lowest (.75) to highest (2.5) variance. Using a metronome, pairs were presented to subjects for approximately five seconds with an interstimulus interval of equal duration. Subjects recorded responses on answer sheets provided (See Appendix 5).

Information regarding subject sex and age was also recorded at this time. Each subject's score on the discrimination task was computed by giving one point for each correct decision. Incorrect responses or those recorded as equal were not considered in calculations.

Task 3: Production using the hexagonal grid.

Apparatus. A stimulus field consisting of 259 loci was used (see Appendix 6). This field was based on a field of thirty-seven contiguous hexagons, each containing seven possible locations (six vertices and one central point, see Appendix 7). Subjects were provided with a box containing seventy-four dots with which to complete the experimental task. Dots used were identical to those used in Task 1.

Procedure. Upon completion of Task 2, subjects were informed they would be completing another production task and were instructed in the following manner:

> In front of you lies a number of dots and a sheet of paper on which there is a number of points. Your task is to arrange the dots randomly on the paper using the points as dot locations. In doing this, you may place only one dot on the paper at a time. You may place only one dot per point, however, you may rearrange the dots until you have what you want as a final product. Remember try to place the dots into what you consider to be a chance or random arrangement. Any questions? Begin.

At this point, questions were addressed and instructions repeated until subjects were believed to have a thorough understanding of the task requirements. No further explanation of random was given.

Variance calculations were completed as in Task 1. In the present context, a hexagonal grid overlay consisting of thirty-seven hexagons was used (see Appendix 8). Each hexagon, comprising one sample of space, enclosed seven possible dot locations.

Task 4: Discrimination using the hexagonal grid.

Stimuli. A total of twenty-four stimuli, twelve random and twelve contagious were constructed. Here, stimuli were based on the hexagonal field format (see Appendix

7). Details regarding construction, mounting, positioning and presentation of stimuli were as in Task 2. In the present context, seventy-four dots were used, with probability tables for variance constructed and adjusted on this basis. Again area was controlled for with six peripheral hexagons being occupied (1,4,16,22,34,37; see Appendix 7).

Twelve contagious stimuli, ascending in steps of .5, with variances ranging from 2.5 to 8, were constructed. Random stimuli were constructed from three distributions thus yielding three random patterns. Four duplicates of each random pattern were produced giving the twelve necessary to complete the pairs of stimuli to be presented. Pair members from the random patterns were selected on a alternating basis with duplicate pattern stimuli being rotated 180 degrees in orientation prior to mounting.

Procedure. Upon completion of Task 3 subjects were informed they would be completing a final discrimination task and were instructed as in Task 2. Stimulus pairs were presented in an ascending manner such that contagious stimuli appeared in order from lowest (2.5) to highest (8) variance. Subject's score on discrimination was calculated as in Task 2.

RESULTS

For each stimulus generated by a subject, a production score was obtained as follows. For the square grid, the numbers of dots in each of the 100 cells were tabulated, and the variance calculated. This variance was then divided by the mean(0.5) to yield the cluster score (C). These C scores (production scores) are given in Table 1. In the present context, the limits of attainable C scores range from 0.5 to 8.0. However, the greater majority fall between 0 and 1. In relation to the chance expectation of 1, there is a significant tendency for scores to fall below this value (x^2= 14.4, df = 1, p<.005).

Table 1. C - scores for production using the square field (CSQU)

C-Score Range	Frequency	Cumulative Frequency (%)
.5 - .65	17	42.5
.66- .85	12	72.5
.86- 1.05	3	80.0
1.06- 1.25	4	90.00
1.26- 1.45	1	92.5
1.46- 1.65	2	97.5
1.66- 1.85	1	100.0

Total--------N=40

Production scores for the hex grid (CHEX), are shown in Table 2 (see Appendix 9 for examples of these). Inspection of Table 2 indicates that the majority of scores on this task (95%) also fall between 0 and 1 with this again highly significant (x^2= 32.4, df=1, p< .005). In the hex grid, the limits of attainable C-scores range from 0 to 4.84. Measures of central tendency for production tasks are given in Table 3.

Table 2. C- score for production using the hex field (CHEX)

C-Score Range	Frequency	Cumulative Frequency (%)
.1 - .25	15	37.5
.26- .45	14	72.5
.46- .65	5	85.0
.66- .85	3	92.5
.86- 1.05	1	95.0
1.06- 1.25	1	97.0
1.26- 1.45	1	100.0

Total--------------N=40

Table 3. Measures of central tendency for production/ discrimination tasks

	CSQU	CHEX	DSQU	DHEX
Mode	.500 & .66	.243	7	8
Median	.660	.298	7	9
Mean	.7795	.4078	6.675	8.025

Total--------------N=40

Discrimination scores for the square (DSQU) and hexagonal (DHEX) grids are given in Tables 4 and 5, respectively. Maximum obtainable scores are 8 for DSQU and 12 DHEX, with no subject reaching this on the latter task. Measures of central tendency for discrimination tasks are given in Table 3. Individual results on all tasks are presented in Appendix 10.

Table 4. Discrimination scores for discrimination using square grid (DSQU)

Correct Decisions	Frequency	Cumulative Frequency (%)
3	1	2.5
4	1	5.0
5	3	12.5
6	10	37.5
7	15	75.0
8	10	100.00

Total-------------N=40

Table 5. Discrimination scores for discrimination using hex grid (DHEX)

Correct Decisions	Frequency	Cumulative Frequency(%)
3	2	5.0
6	6	20.0
7	7	37.5
8	5	50.0
9	13	82.5
10	4	92.5
11	3	100.0

Total-------------N=40

When collapsed across tasks, correlation between overall production and discrimination was significant ($r = -.3923$, $p = .012$). This relationship indicates an increased discrimination performance as produced arrangements become more regular. Individual correlations among tasks are given in Table 6. Inspection of Table 6 reveals the following significant relationships:

CSQU with CHEX ($r = .6321$, $p < .01$), CSQU with DSQU ($r = .3482$, $p < .05$), CSQU with DHEX ($r = .3983$, $p = .01$), CHEX with DCQU ($r = .332$, $p < .05$), DSQU with DHEX ($r = .4155$, $p < .01$).

The first and last of these correlations give estimates of reliability for tasks across field types. Significant correlation for CSQU with both DSQU and DHEX and CHEX with DSQU indicates that as subjects deviate either below or above random on production they discriminate better and worse, respectively. This relationship, however, does not completely generalize in the case of CHEX and DHEX.

Table 6. Correlation between tasks. (* represents significance greater than .05/ ** represents significance greater than or equal to .01)

	CHEX	DSQU	DHEX
CSQU	.6231 (p=.0002) **	-.3482 (p=.020) *	-.3983 (p=.01) **
CHEX		-.3320 (p=.036) *	-.1374 (p=.398)
DSQU			.4155 (p=.008) **

Discussion

These experiments have found, (1) an inability by subjects to produce random arrangements using either the square or hexagonal field, (2) a tendency to err on the side of regularity in production, (3) a negative correlation between production and discrimination (4) a strong reliability relationship between production on the square and hexagonal fields and, (5) a strong reliability relationship between discrimination on the square and hexagonal field.

Production

In a critical review of the literature, Wagenaar, (1972) concluded, that generally subjects have shown an inability to randomize. The present findings are consistent with Wagenaar's (1972) conclusions and demonstrate this with two levels of available dots using two matrices.

Though Tune (1964) argues that production of randomness is limited by short term memory, the present study cannot support this. Production, in this case did not require short term memory as all responses were equally available at all times, a variable which has not been overlooked as in previous research. Similarly, boredom would not seem to be a function in the present case as neither task repetition nor time on task were of great duration, variables which have been viewed as increasing boredom (Weiss, 1964).

It has been suggested that subjects cannot produce random sequences because they " replace the laws of chance by heuristics, which sometimes yield reasonable estimates and quite often do not" (Kahneman and Tversky, 1972, p. 430). Further, this use of individual heuristics results in preferences for certain patterns of responses, where responses are locally representative of their parent distribution/heuristic (Kahneman and Tversky, 1972). Subjects in the present study have shown a preference to arrange dots in a preferred pattern, that of a regular nature. If Kahneman and Tversky's (1972) notion is correct then regular patterns (low C-scores) reflect local representations of the subjects' random heuristic. For example, subject number one has produced a c-score of .540 on the square production grid (see Appendix 10). Here, the produced score of .540 is a reflection of the internal model the subject holds and thus is a local representation.

In looking to past investigations, it is possible to directly compare this pattern type. For example, Bakan (1960) reports that subjects exhibit consistent patterns of responses in their randomization production task. In measuring this, Bakan (1960) has quantified these patterns via runs. Generated patterns in Bakan's (1960) investigation have deviated from randomness by showing too many runs in a series as opposed to chance (i.e. chance = 151 runs, $\bar{x}$ runs = 176). In the present experiment we may convert subject data to represent runs whereby the number of runs is given by the number of cells which are occupied, as specified by Ginsburg and Goldstein, (1987). Within this context we find that the mean number of runs for the square field is 43.9 while the figure for the hexagonal field is 33.3. The number of runs by chance, as given by the Poisson distribution, for the square and hexagonal field is 39 and 32 respectively. Thus we see, the current subjects, like Bakan's (1960) have deviated from randomness by having too many runs in their generated displays. Related to the cluster continuum, we may say that Bakan's (1960) subjects erred on the side of regularity in their productions.

On the other, the subjects of Evan's (1978) experiment erred on the clustered side in their productions. They were asked to generate random series using the numbers 1 to 10. It is possible to calculate approximate values of C from the data reported by Evans (1978). When this is done, the mean is around 1.4, which shows clustered responses. From these comparisons, we do not see the concurrence needed to establish a universally operating heuristic.

Discrimination

Brunswik's (1951) notion of ecological validity suggested that experimental stimuli be more representative of the ecological relationships in the natural environment. In turn, Gibson (1960) suggests we consider and incorporate the laws of stimulus information in the organism's natural environment into experimental stimuli. This trend was carried a step further by Ginsburg & Goldstein (1987), who showed how stimuli could be assigned to positions on the cluster continuum .

The present study has used the measure suggested by Ginsburg & Goldstein (1987), the C-score, to generate a set of ordered stimuli in order to investigate the ability of subjects to discriminate stimuli differing in organization. It was found that subjects were able to discriminate between random and clustered patterns, and individual differences in this ability were measured. Though it is unclear on what basis subjects have made their judgements, Diener and Thompson (1985) have suggested that, "subjects decide that a sequence is random by eliminating nonrandom hypothesis, rather than by directly recognizing the series as representative of a random process." It is possible that some of the nonrandom hypotheses subjects have eliminated here may involve cluster.

Cook (1967), presented results on a discrimination task. Subjects in Cook's (1967) study were presented pairs of string lists of binary digits and asked to, "decide whether the first was more or less patterned than the second" (Cook, 1967,p. 1005). Though Cook's

(1967) results indicated that string lists which were compiled by random number tables were judged to be less patterned than others, he was unable to quantify the differences between distinguishably different strings. In attempting this quantification, Cook (1967) was only able to use descriptive techniques. One way to quantify Cook's (1967) stimuli is by use of a method related to C-scores and runs, that of the range of run sizes. In the case of Cook's (1967) stimuli, a string's range is given by the maximum number of digits in a run minus one (i.e.: string 1's range =8, string 2's range = 6, etc). Once calculated, we may rank order the strings with respect to the number of times subjects preferred one string list to another and perform a correlation between a string's rank and its range.

Performance of a rank order correlation on these data reveals a significant correlation which directly predicts Cook's (1967) results ($r = -.7$, $p<.01$). Further, we now have a means of specifying stimulus organization.

The Relationship Between Production and Discrimination

In considering this relationship let us look to the models which may be tested with the data at hand:

(1) Random Model (Kahneman and Tversky, 1972)

We assume that the basic process a subject uses while producing and discriminating is one in which he compares stimuli to an internal model of randomness. The more realistic this model is, the closer a subject's production would be to physical randomness and the more accurate his discrimination ability. This model predicts a negative correlation between descrepancy from randomness and success in the discrimination task. In order to test this model production scores will have to be transformed to represent distance from randomness. With these data transformed in this manner [i.e. absolute valve(c-score minus 1)], an overall correlation of .2985 is obtained (p. = .061). Though this relationship is not significant, it's direction violates what the Kahneman and Tversky (1972) model predicts.

(2) Nonrandom Model: Clustered Schema (Diener and Thompson, 1985)

We assume that, subject's decide that a sequence is random by eliminating alternative nonrandom hypotheses, rather than by directly recognizing the series as representative of a random process " (Diener and Thompson, p. 433, 1985). In this instance we are able to test one of the possible alternative hypotheses that the subject is eliminating, namely, clustered schema or models. If a subject decides on randomness by contrasting with clustered schema then the more successful he is in doing this, the further away from the cluster end of the continuum his productions will be. Simply put, we predict a negative correlation between c-scores and success in the discrimination task.

When collasped across tasks, the correlation between overall production and discrimination is significant and supports the prediction (r= -.3923, p = .012).

Although the data of the present study do provide evidence which supports the Diener and Thompson (1985) interpretation of nonrandom models, it is possible to suggest further studies that would be more exhaustive. Consider the whole cluster continuum, from stimuli with c-scores of 0 to stimuli with a maximum value for the items used. It is possible to measure difference thresholds with three kinds of standards: (a) random, as was done in the present study, (b) regular, or (c) clustered. If the essential comparisons that subjects make are with models that represent their concept of randomness, as claimed by Kahneman and Tversky (1972), then we would expect random standards to yield the lowest difference thresholds. If, on the other hand, random standards give difference thresholds greater than those that are regular or clustered, this would lend convergent support to the Diener and Thompson (1985) interpretation of nonrandom models.

Measurement

"The first problem to be solved is the problem of measurement" (Wagenaar, W.A. 1972, p. 71). As a methodological problem randomness is much more difficult to prove than disprove.

In previous investigations, a variety of approaches to the quantification of randomness have been used. Among these are runs, autocorrelation, information and n-grams. Wagenaar, (p. 69, 1972), in reviewing these methods has concluded that, "most measures of randomness are neither powerful enough for disproving all serial regularities nor adequate for establishing increases and decreases of non-randomness". Clearly, the use of the cluster ratio has provided this investigation with a discrete and precise tool for the measurement of randomness. As a method of analysis it carries with it a predicted value for randomly occurring events, that is C=1.0, which does not vary by subject response. The cluster ratio has proven to be a reliable measure across both the combination of field type and number of dots. Further, it's use has allowed discrimination tasks whose results are directly comparable to production tasks. In sum, the cluster ratio, is a decisive tool which may be used to investigate any factor seen as contributing to non-randomness.

REFERENCES

Baddeley, A.D. (1966) The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology*, 18, 119-129.

Bakan, P.,(1960) Response tendencies in attempts to generate random binary series. *American Journal of Psychology*, 73, 127-131.

Brunswik, E., and Herma, H. (1951) Probability learning of perceptual cues in the establishment of a weight illusion. *Journal of Experimental Psychology*, 41, 281-290.

Chapanis, A., (1953) Random number guessing behavior. *American Psychologist*, 8,332.

Cook, A. (1967) Recognition of bias in strings of binary digits. *Perceptual and Motor Skills*, 24, 1003-1006.

Diener, D., and Thompson, W.(1985) Recognizing randomness. *American Journal of Psychology*, 98, No. 3. 433-447.

Evans, F. J. (1978) Monitoring attention deployment by random number generation: An index to measure subjective randomness. *Bulletin of the Psychonomic Society*, 12(1), 35-38.

Gellerman, L.W. (1933) Chance orders of alternating stimuli in visual discrimination experiments. *Journal of General Psychology*, 42, 207-208.

Gibson, J. (1960) The concept of the stimulus in psychology. *American Psychologist*, 15, 694-703.

Ginsburg, N., and Goldstein, S. R. (1987) The measurement of visual cluster. *American Journal of Psychology*, 100, 193-203.

Goodfellow, L. D. (1940) The human element in probability. *Journal of General Psychology*, 23, 201-205.

Kahneman, D. and Tversky, A. (1972) Subjective probability: A judgement of representativeness. *Cognitive Psychology*, 3, 430-454.

Kahneman, D. and Tversky, A. (1982) On the study of statistical intuitions. *Cognition*, 11, 123-14

Lewis, D. (1960) *Quantative Methods in Psychology*. New York: McGraw Hill.

Lincoln, R. S., and Alexander, L. T. (1955) Preferred patterns of motor and verbal responses. *Journal of Experimental Psychology*, 50, 106-112.

Neisser, U. (1967) Cognitive Psychology. New York: Appleton.

Newman, E. and Collier, N. (1952) Guessing sequences using two, four and eight alternatives. American Psychologist, 7, 272-273.

Pielou, E. (1977) Mathematical Ecology. New York: Wiley.

Rath, G. J. (1966) Randomization by humans. American Journal of Psychology, 79, 97-103.

Ross, B. M. (1955) Randomization of a binary series. American Journal of Psychology, 68, 136-138.

Teraoka, T. (1963) Some serial properties of "subjective randomness". Japanese Psychological Research, 5, 120-128.

Tune, G. S. (1964) A brief survey of variables that influence random generation. Perceptual and Motor Skills, 18, 705-710.

Wagenaar, W. A. (1970) Appreciation of conditional probabilities in binary sequences. Acta Psychologica, 34, 348-356

Wagenaar, W. A. (1972) Generation of random sequences by human subjects; a critical survey of literature. Psychological Bulletin, 77, No.1., 65-72.

Warren, P. A. and Morin, R. E. (1965) Random generation: Number of symbols to be randomized and time per response. Psychonomic Science, 3, 557-558.

Weiss, R. L. (1964) On producing random responses. Psychological Reports, 14, 931-941.

Wiegersma, S. (1982) A control theory of sequential response production. Psychological Research, 44, 175-188.

Wolitzky, D. L. and Spence, D. P. (1968) Individual consistencies in the random generation of choices. Perceptual and Motor Skills, 26,1211-1214

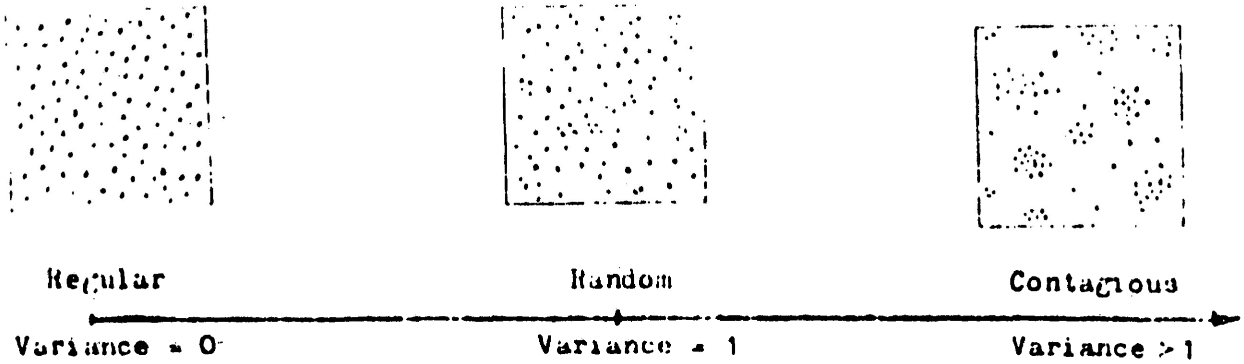
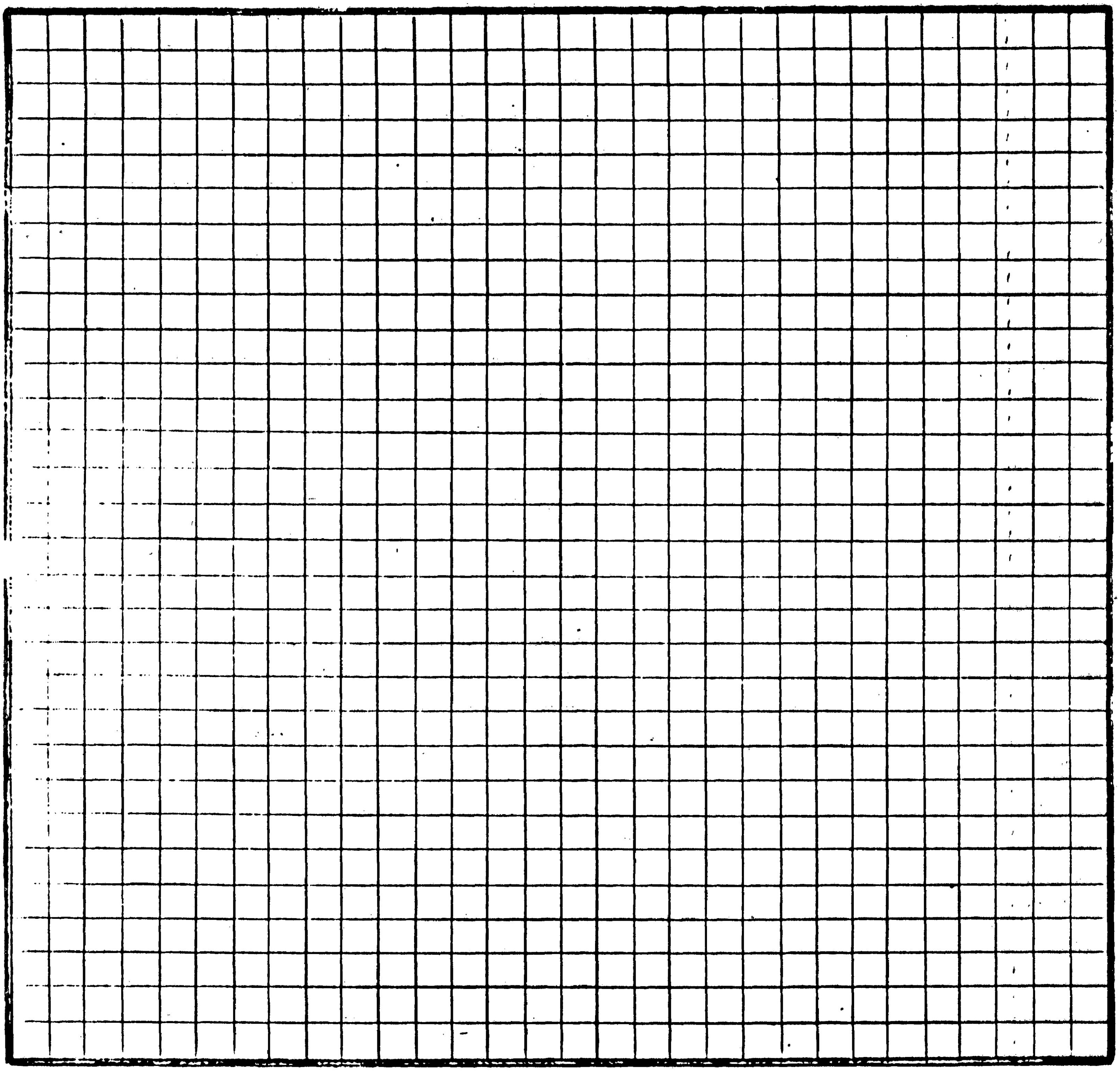
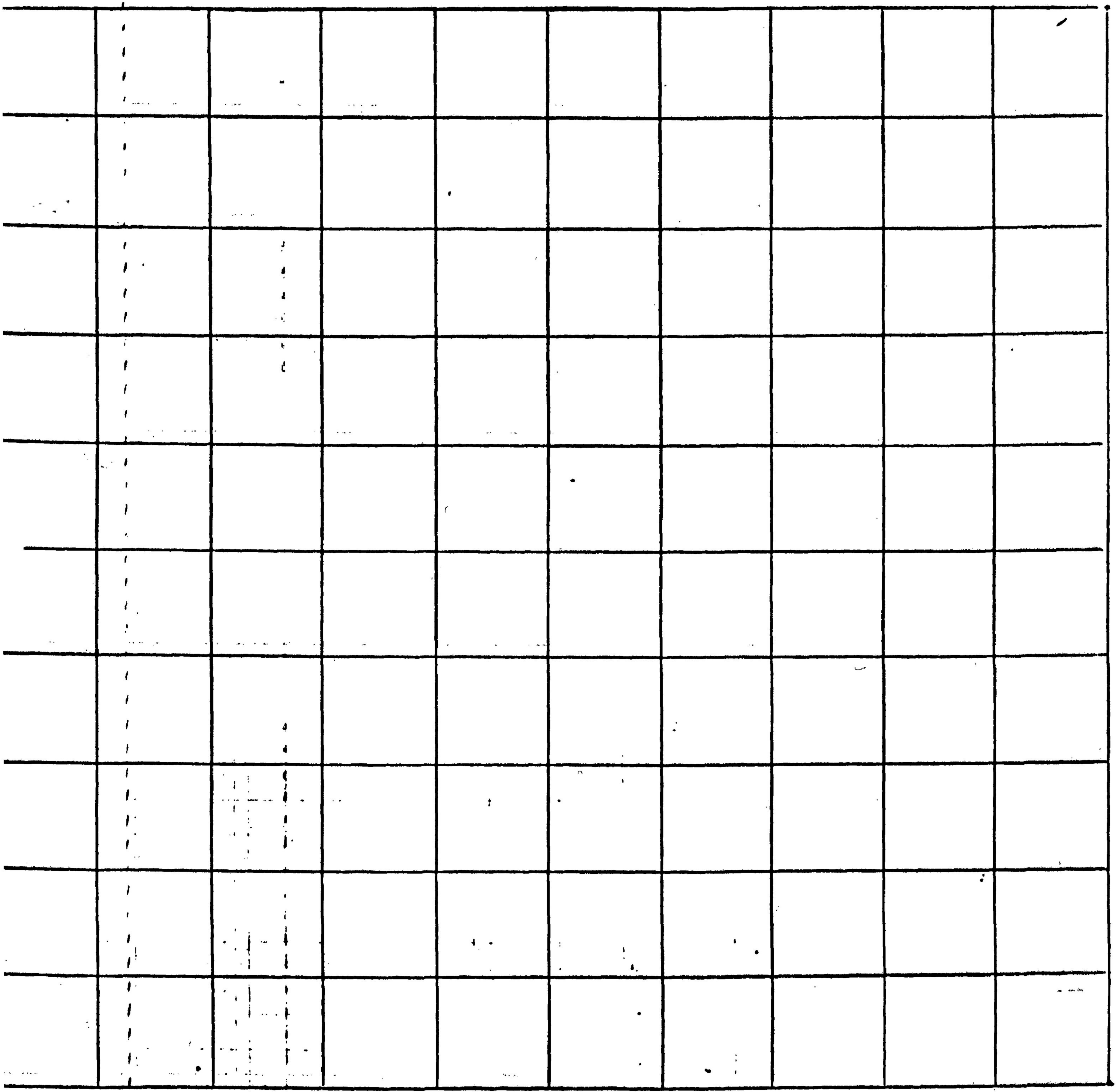


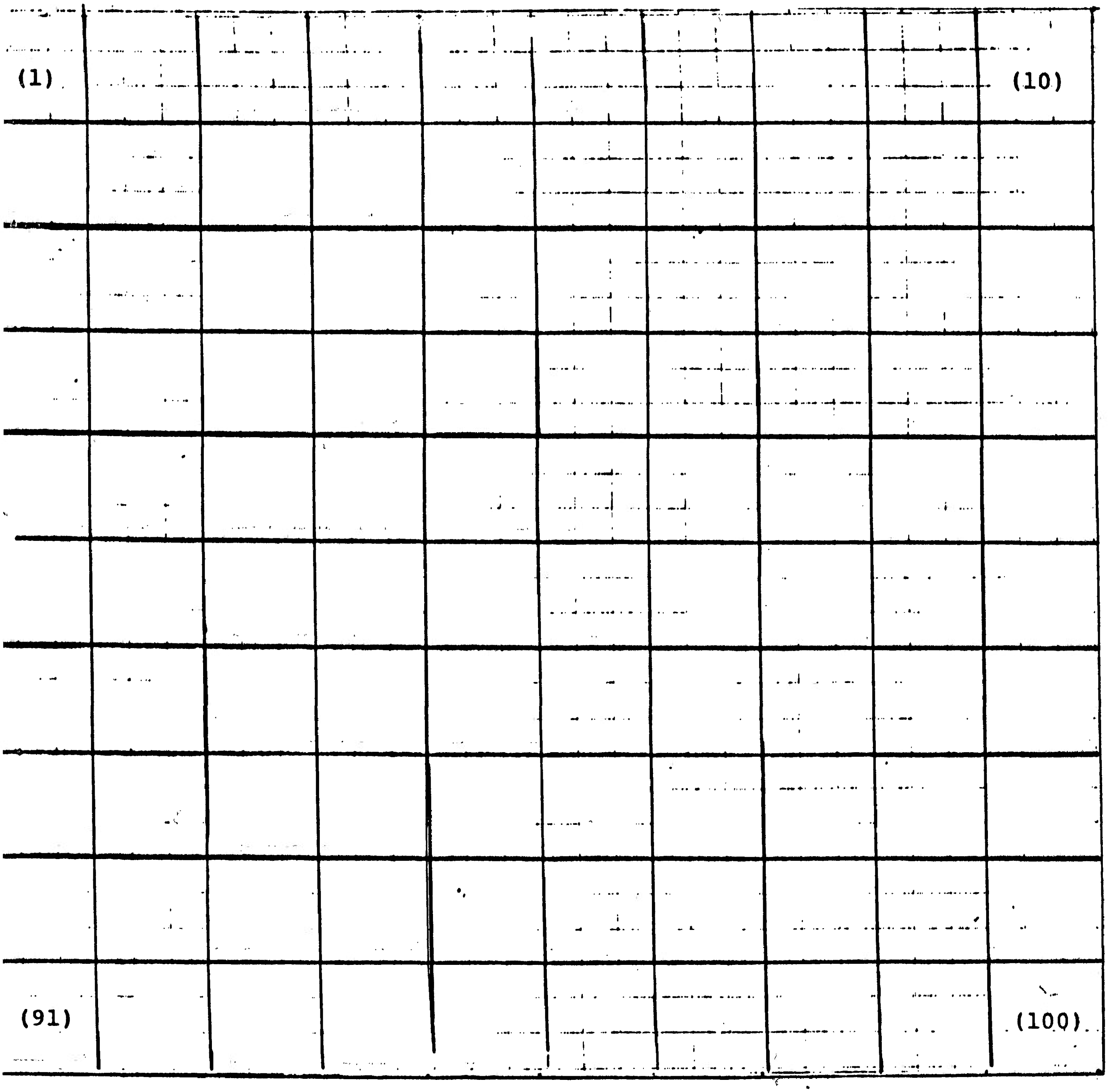
Figure 1. The change in variance along the continuum from regularity to contagiousness for a mean of one event per sample of space.



APPENDIX 1 : Stimulus field for production using the square grid.



APPENDIX 2 : Grid overlay for production using square item field.



APPENDIX 3 : Square field used to construct all experimental stimuli. To equate overall area, squares 1, 10, 91, 100 were always occupied.

(A) For Grid

* random	test
left	right
right	left
right	left
left	right
right	left
right	left
left	right
left	right

(B) For Hex

* random	test
left	right
right	left
right	left
left	right
left	right
right	left
left	right
right	left
right	left
left	right
left	right
right	left

APPENDIX 4: Gellerman's series for determining stimulus placement (left/right) in pair presentation.

NAME:________________

AGE:________________

SEX:________________

1.________

2.________

3.________

4.________

5.________

6.________

7.________

8.________

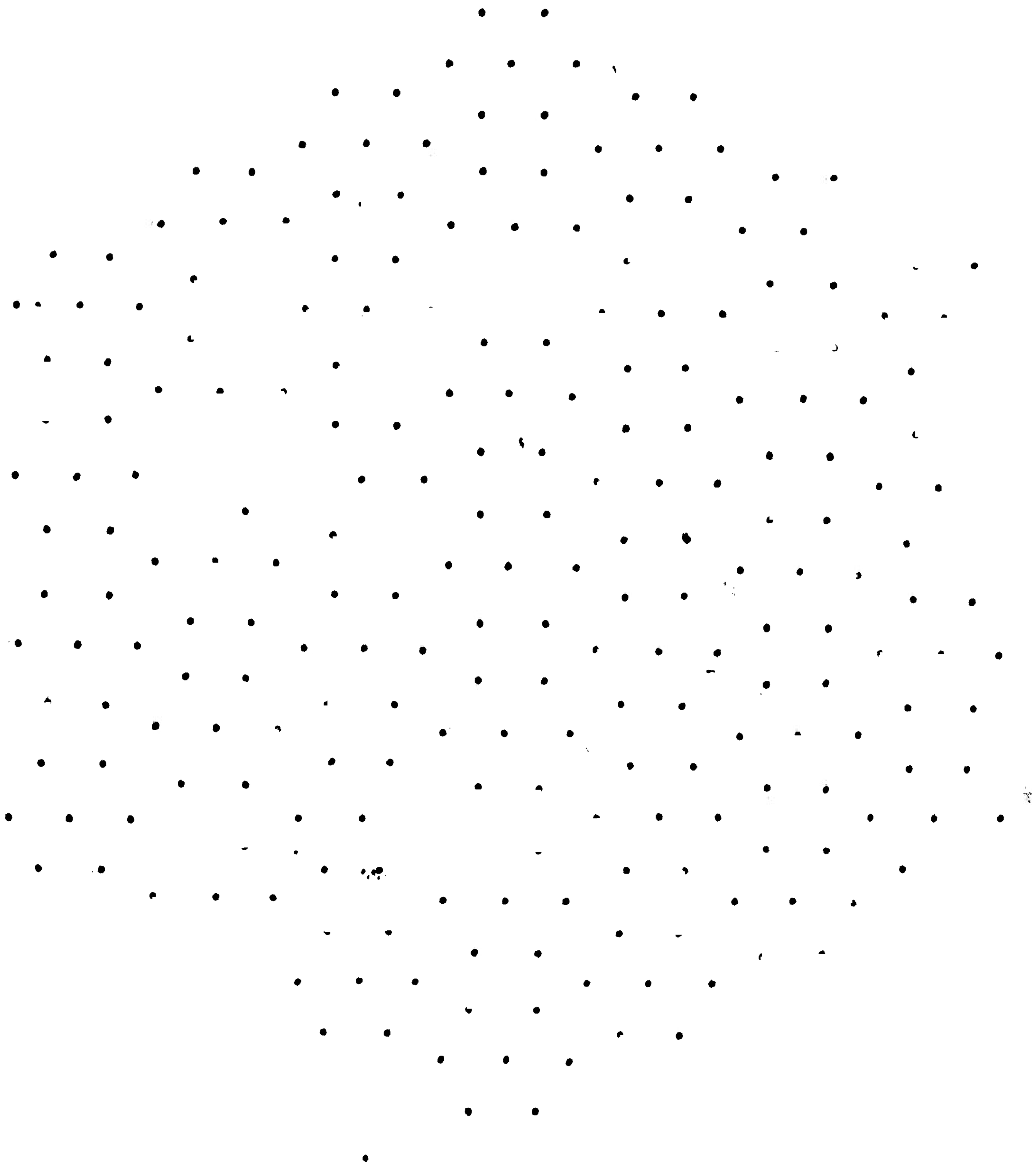
9.________

10.________

11.________

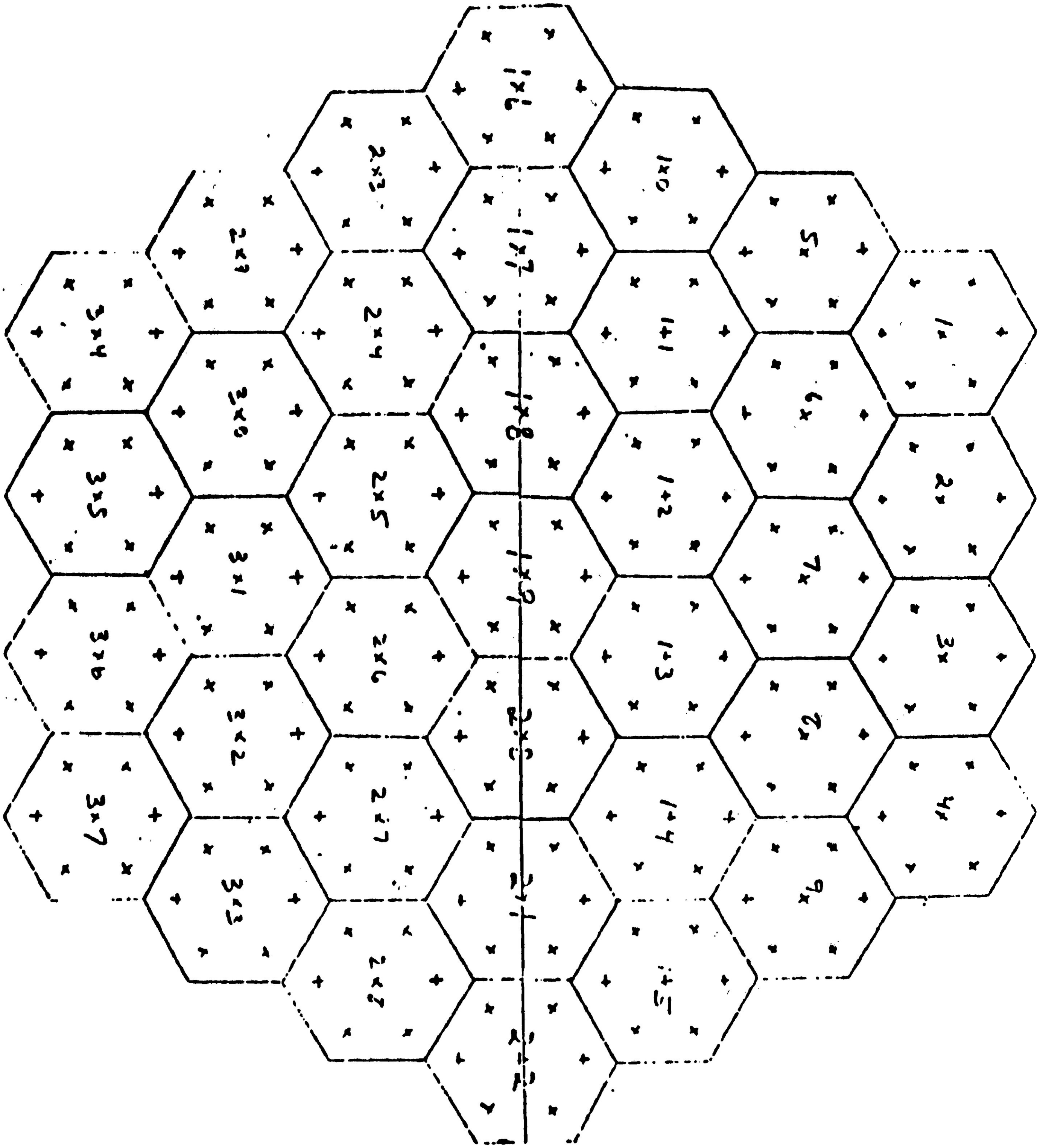
12.________

APPENDIX 5: Response sheets for discrimination tasks.

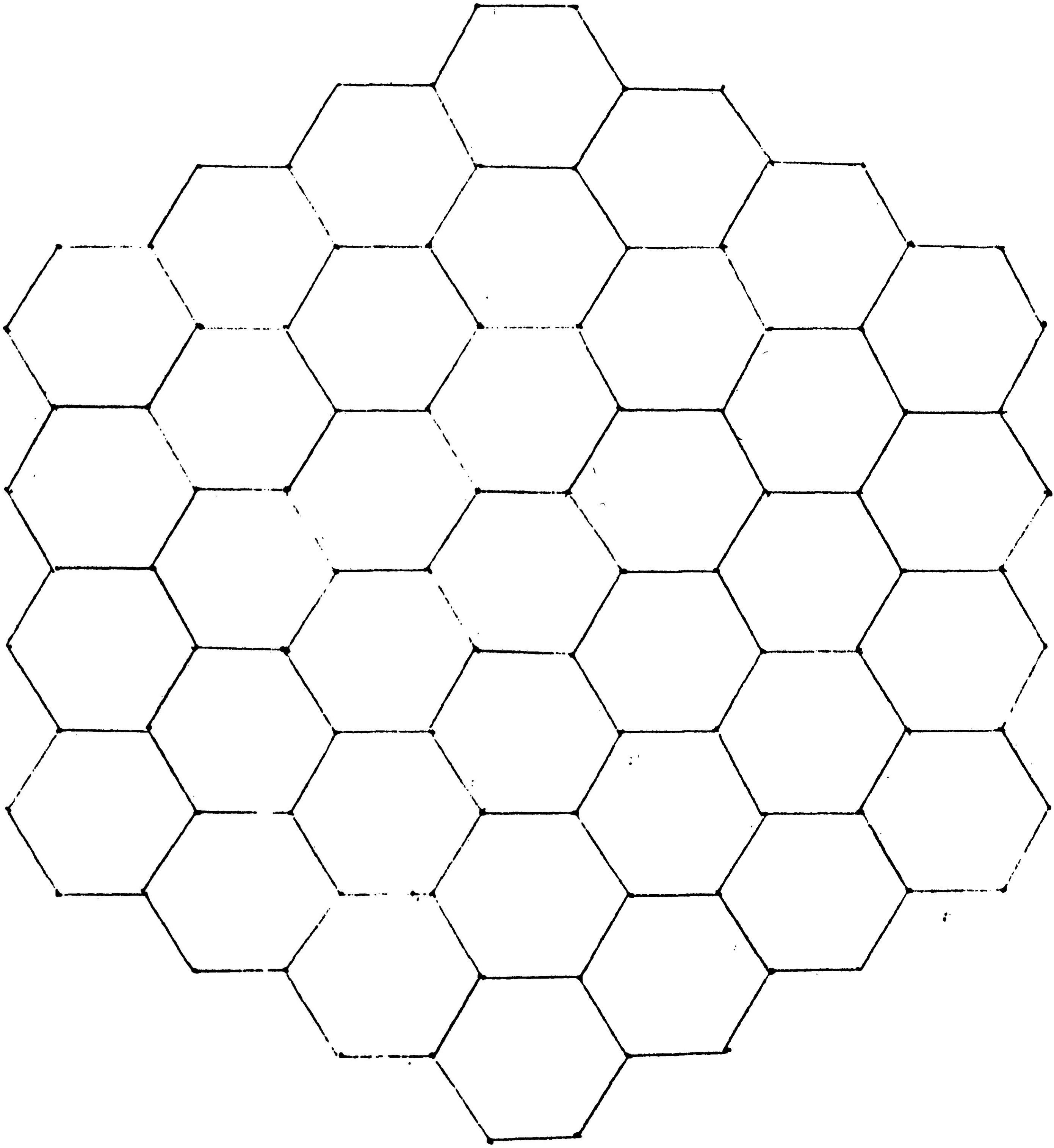


APPENDIX 6 : Hexagonal grid used in production task.

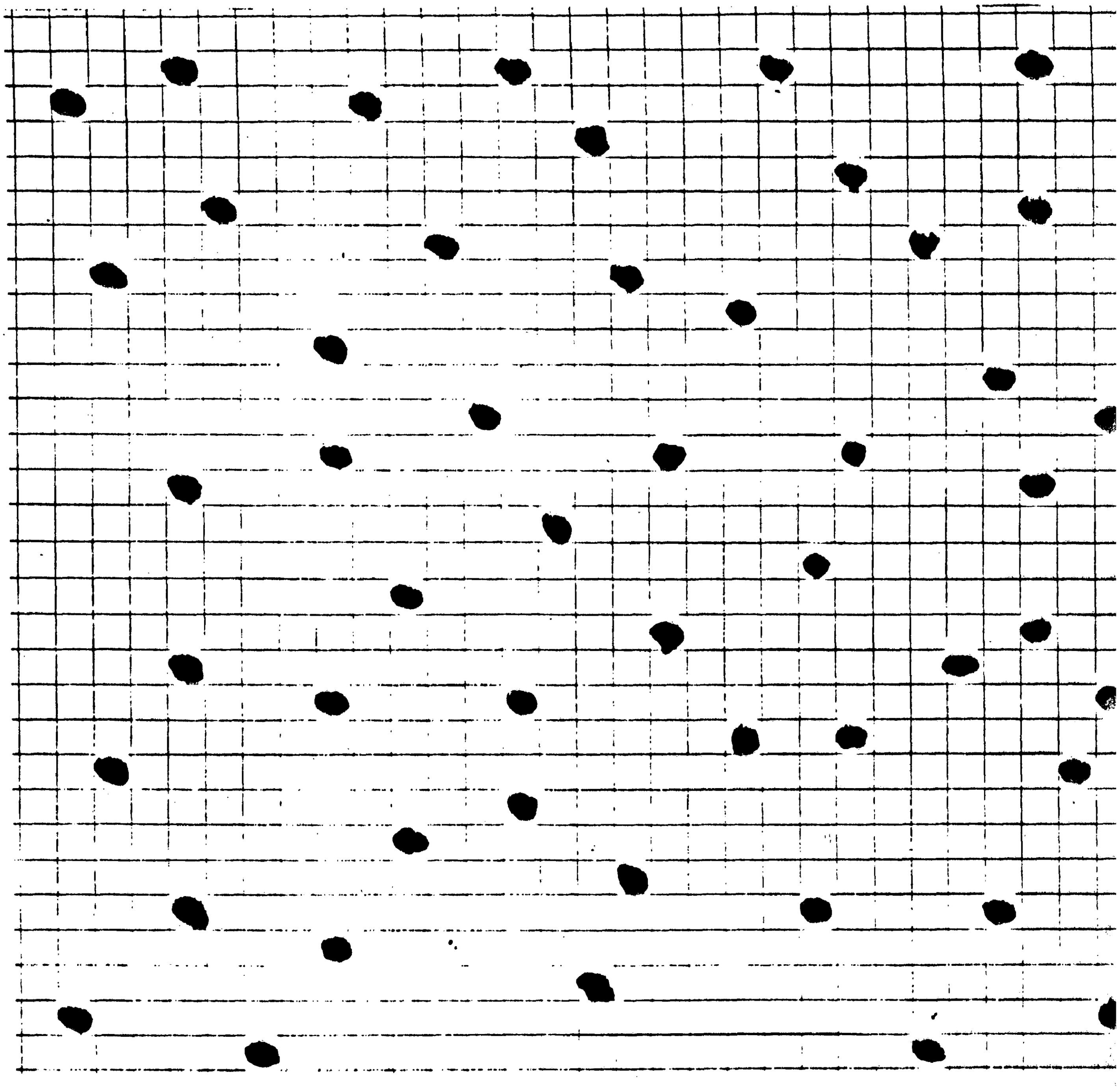
Appendix 7



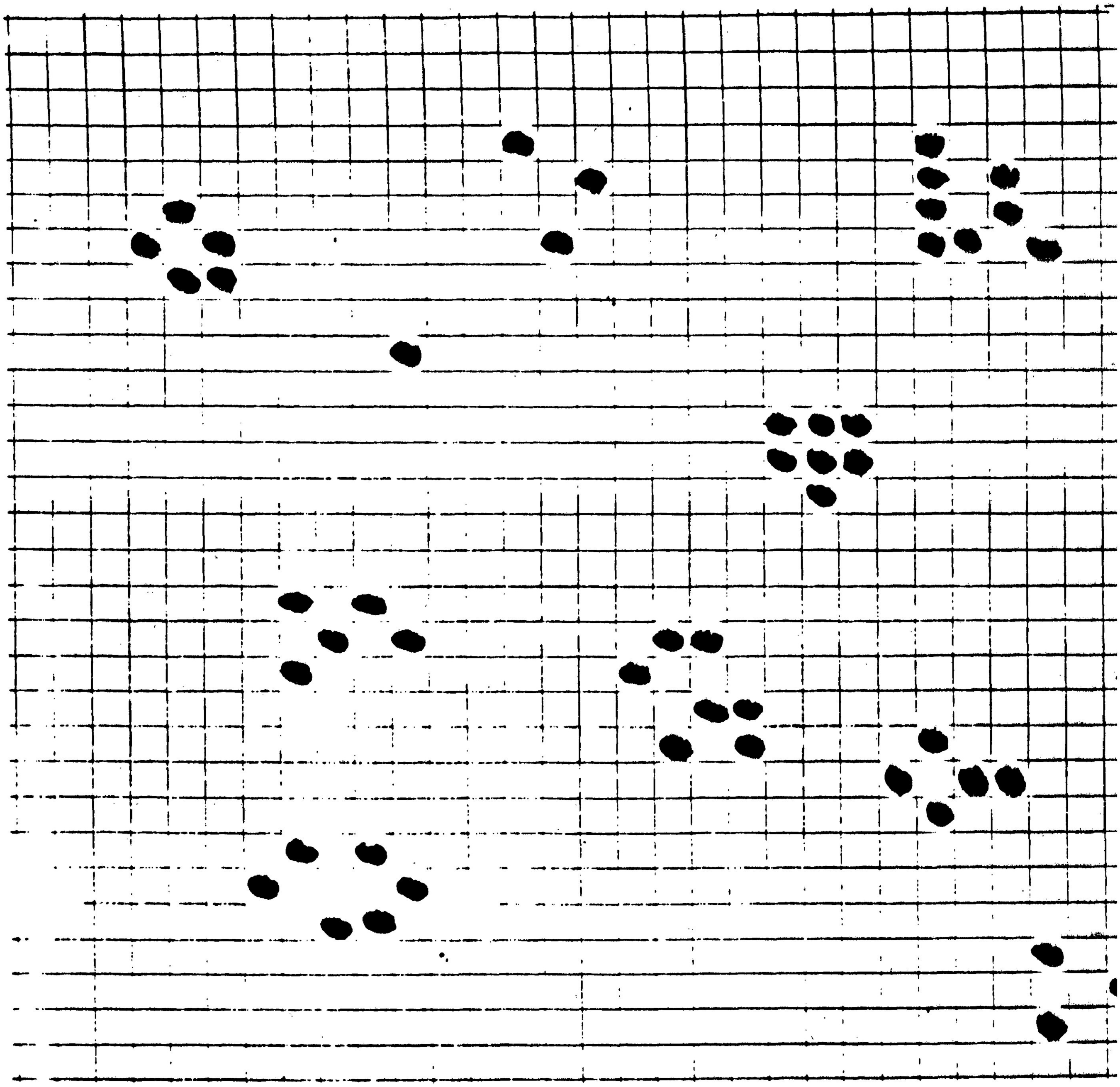
Appendix 7: Hexagon field used to construct all experimental stimuli. To equate overall area, hexagons 1, 4, 16, 22, 34 and 37 always occupied.



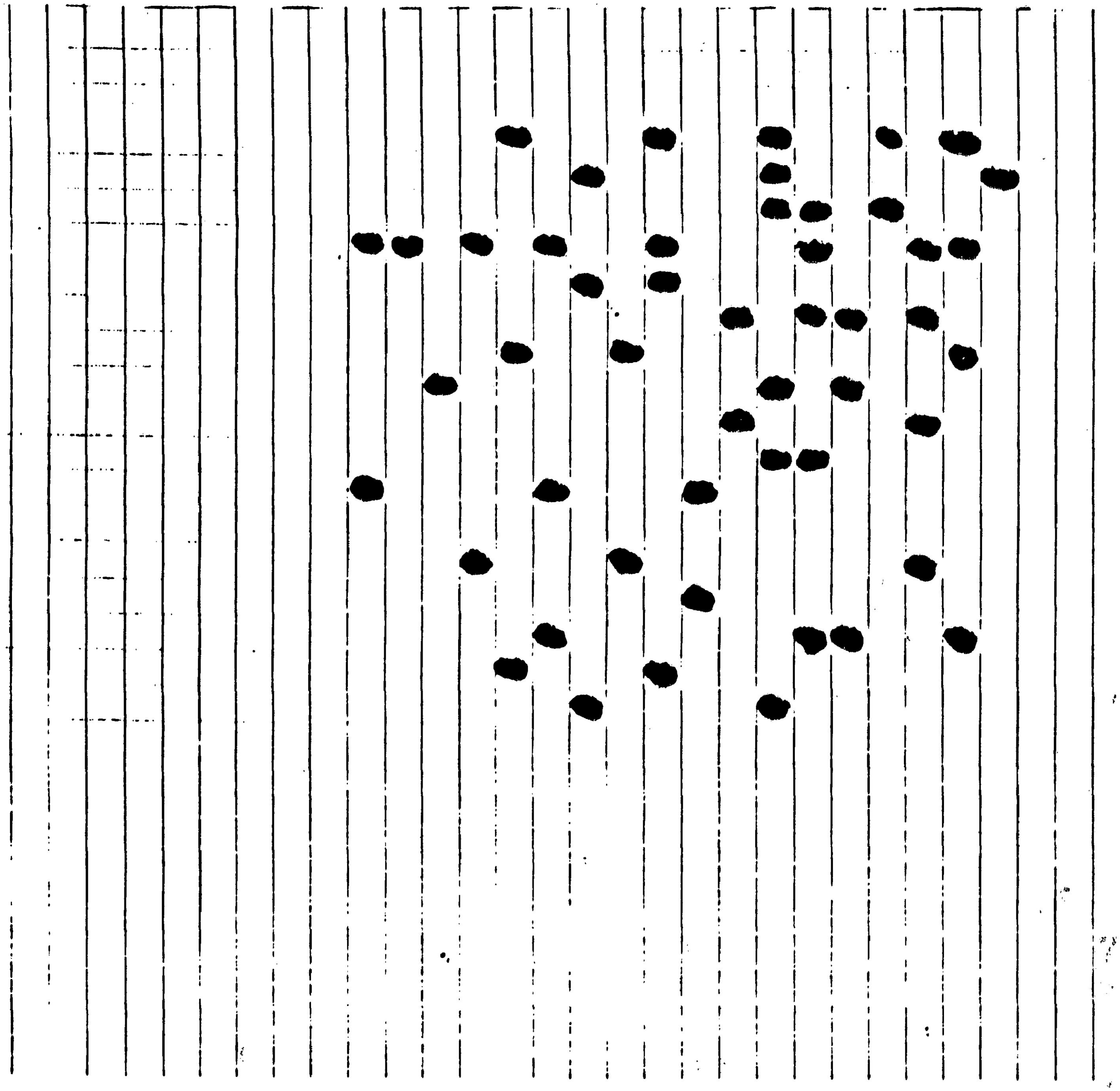
APPENDIX 8 : Hexagonal overlay



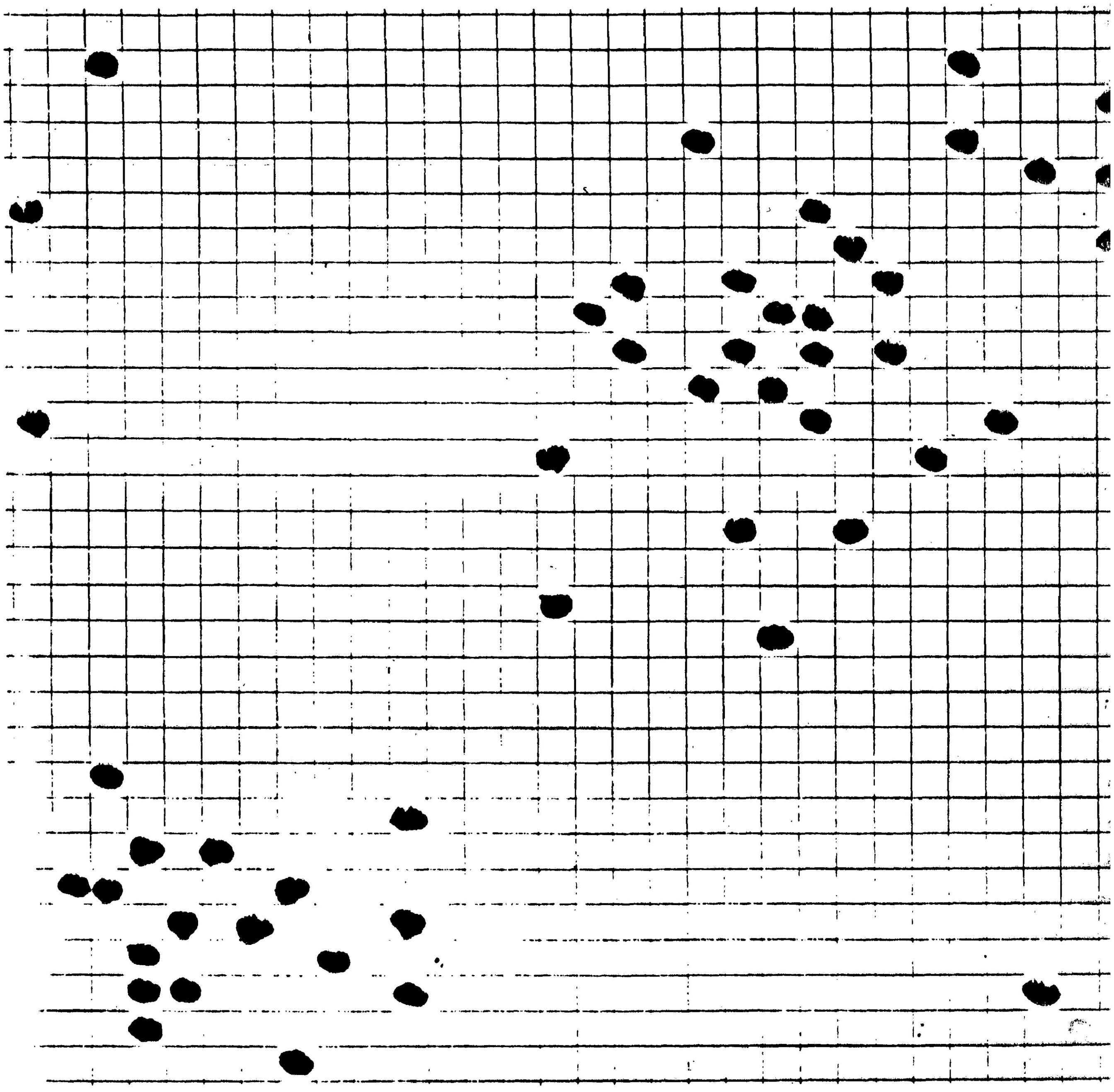
APPENDIX 9 : Examples of subject productions. (CSQU = .5)



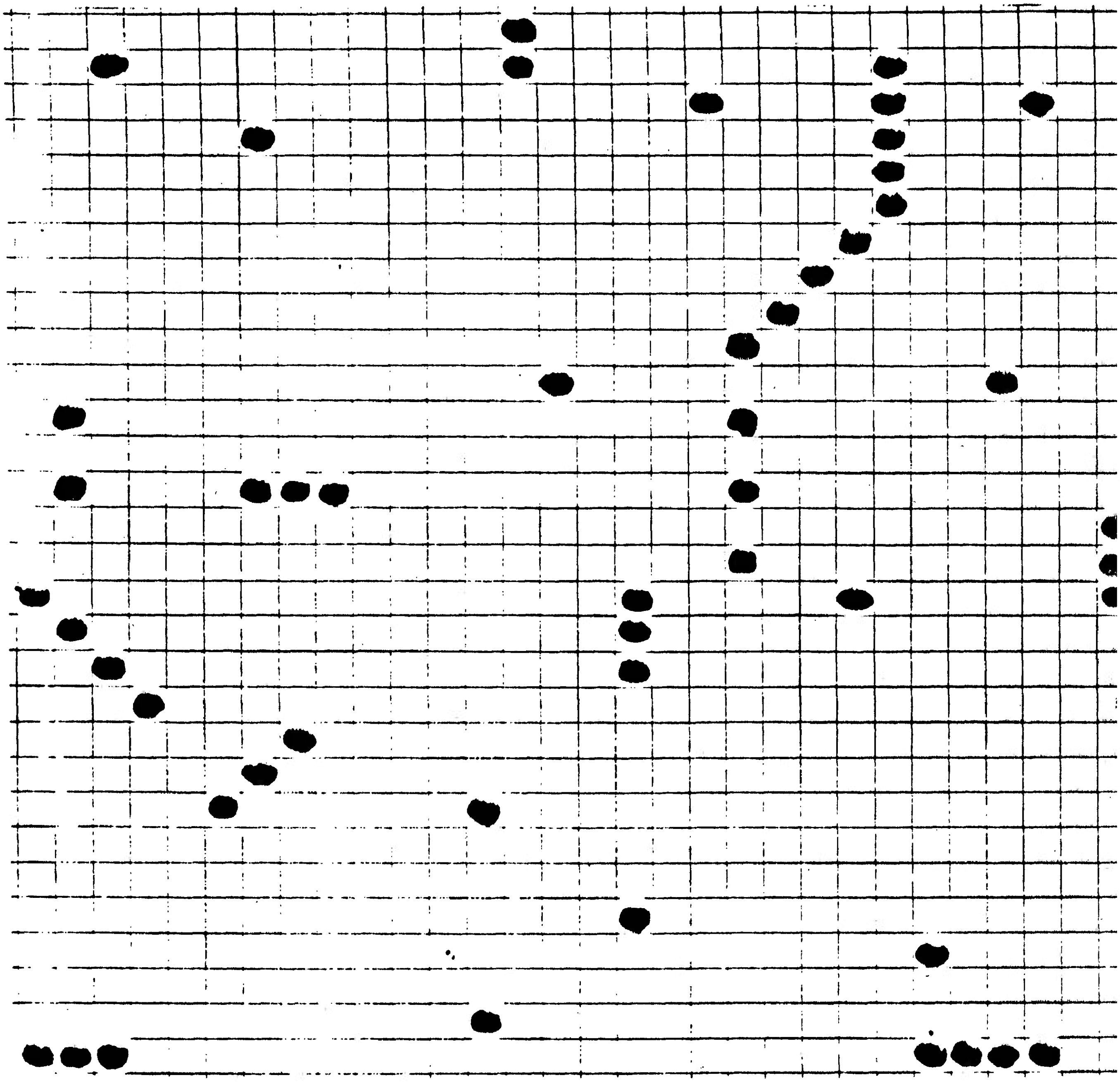
APPENDIX 9 : Continued (CSQU = 1.82)



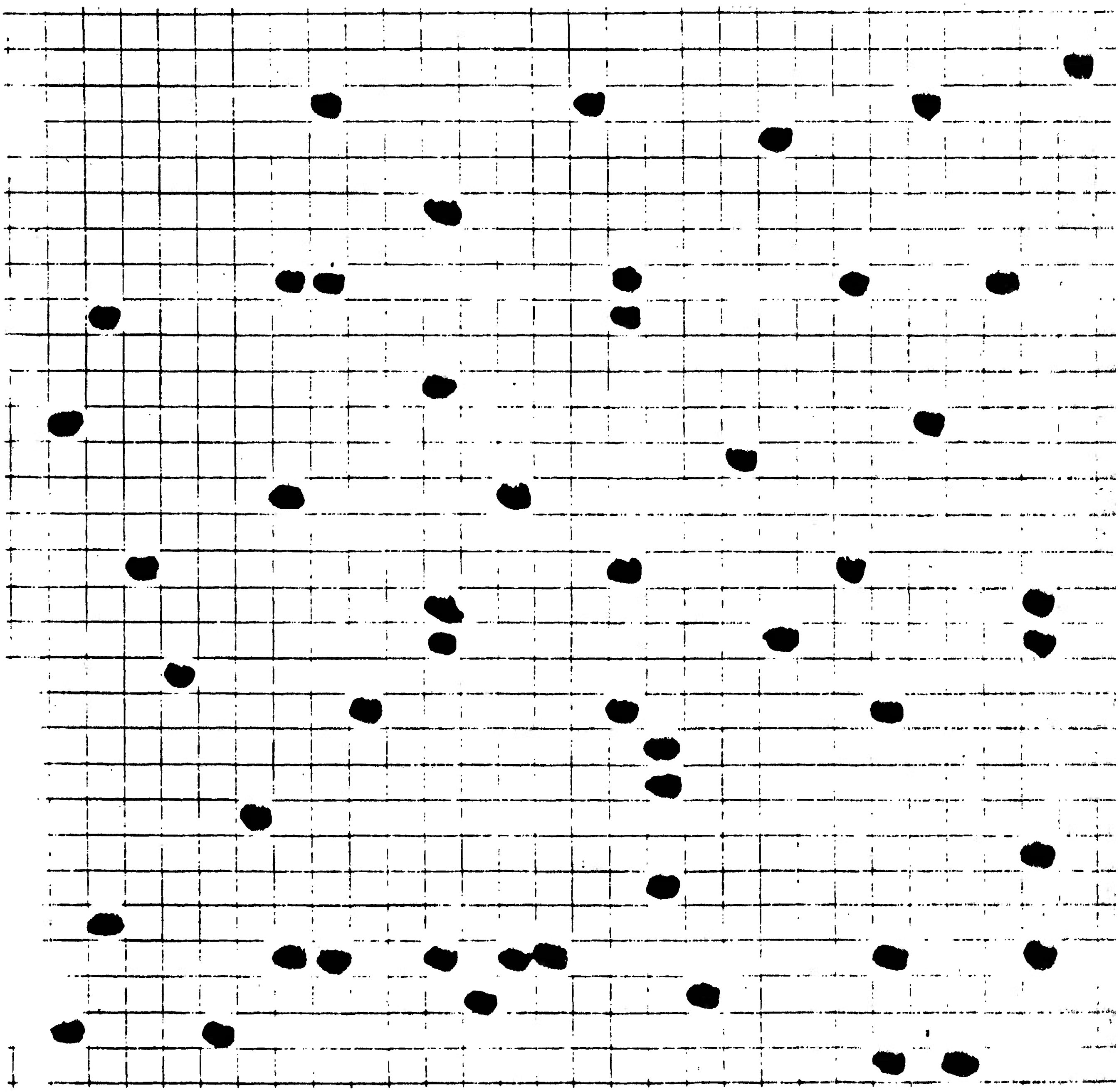
APPENDIX 9 : . Continued (CSQU = 1.54)



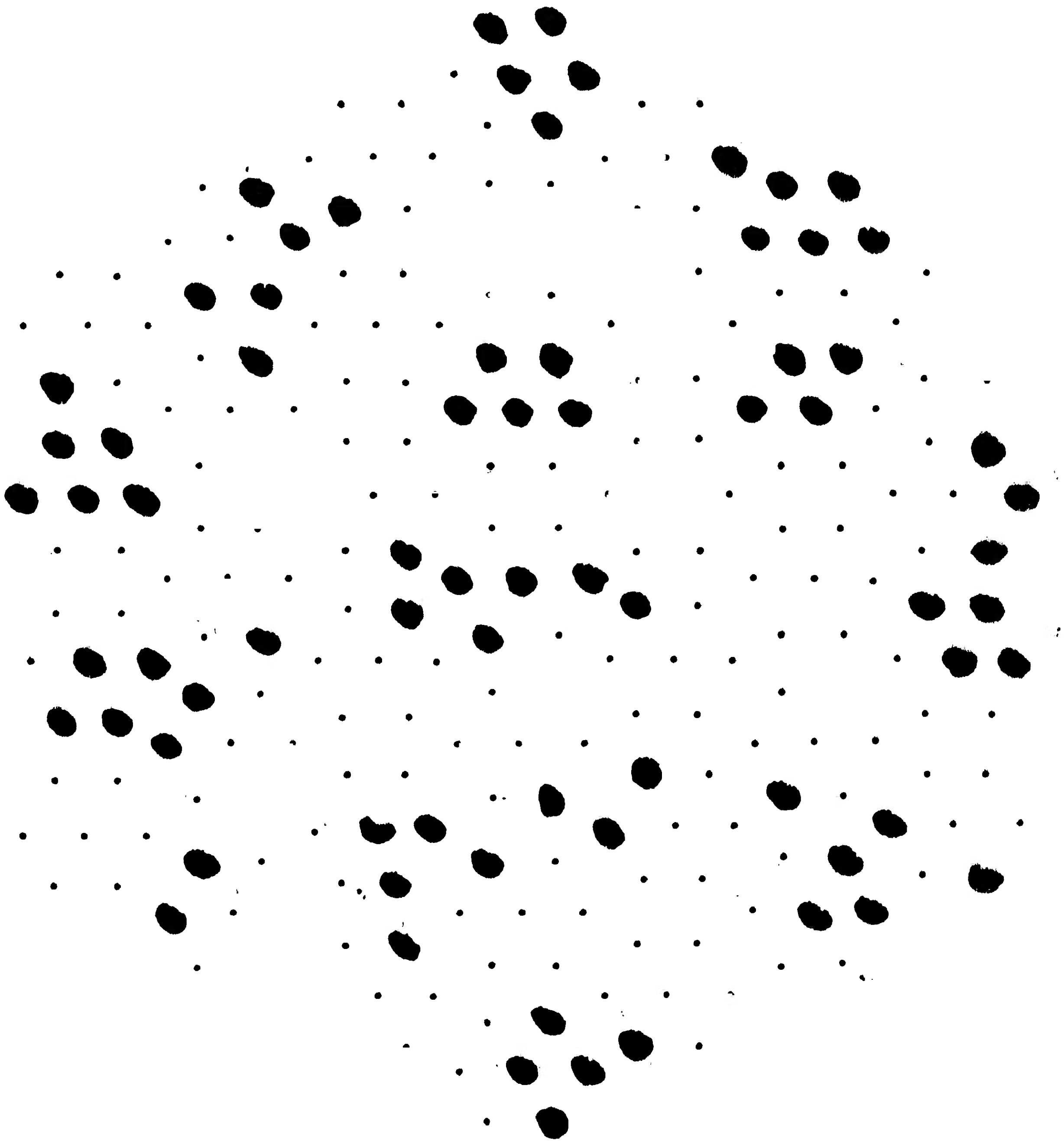
APPENDIX 9 : Continued (CSQU = 1.3)



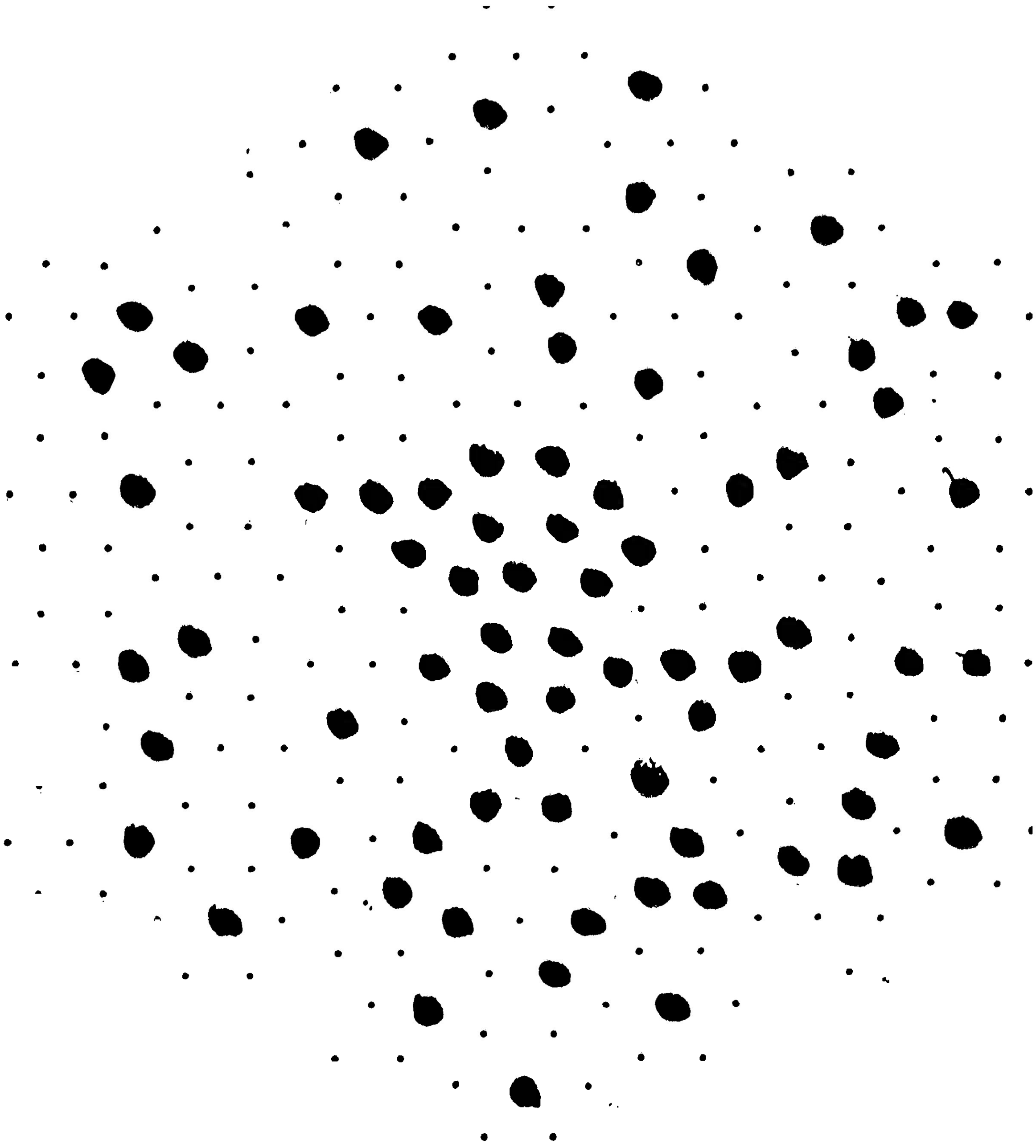
APPENDIX 9 : Continued (CSQU = 1.06)



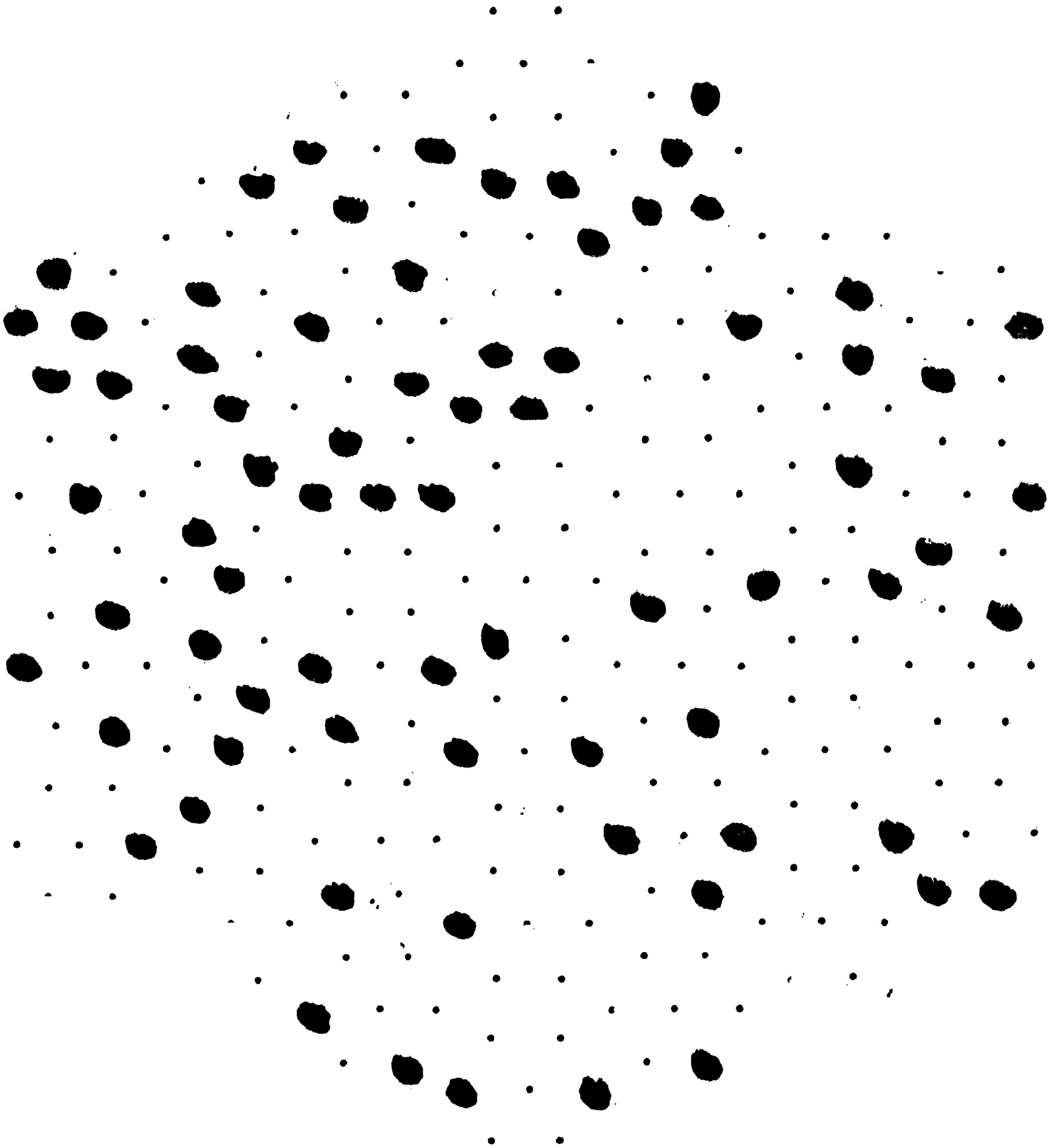
APPENDIX 9 : Continued (CSQU = .78)



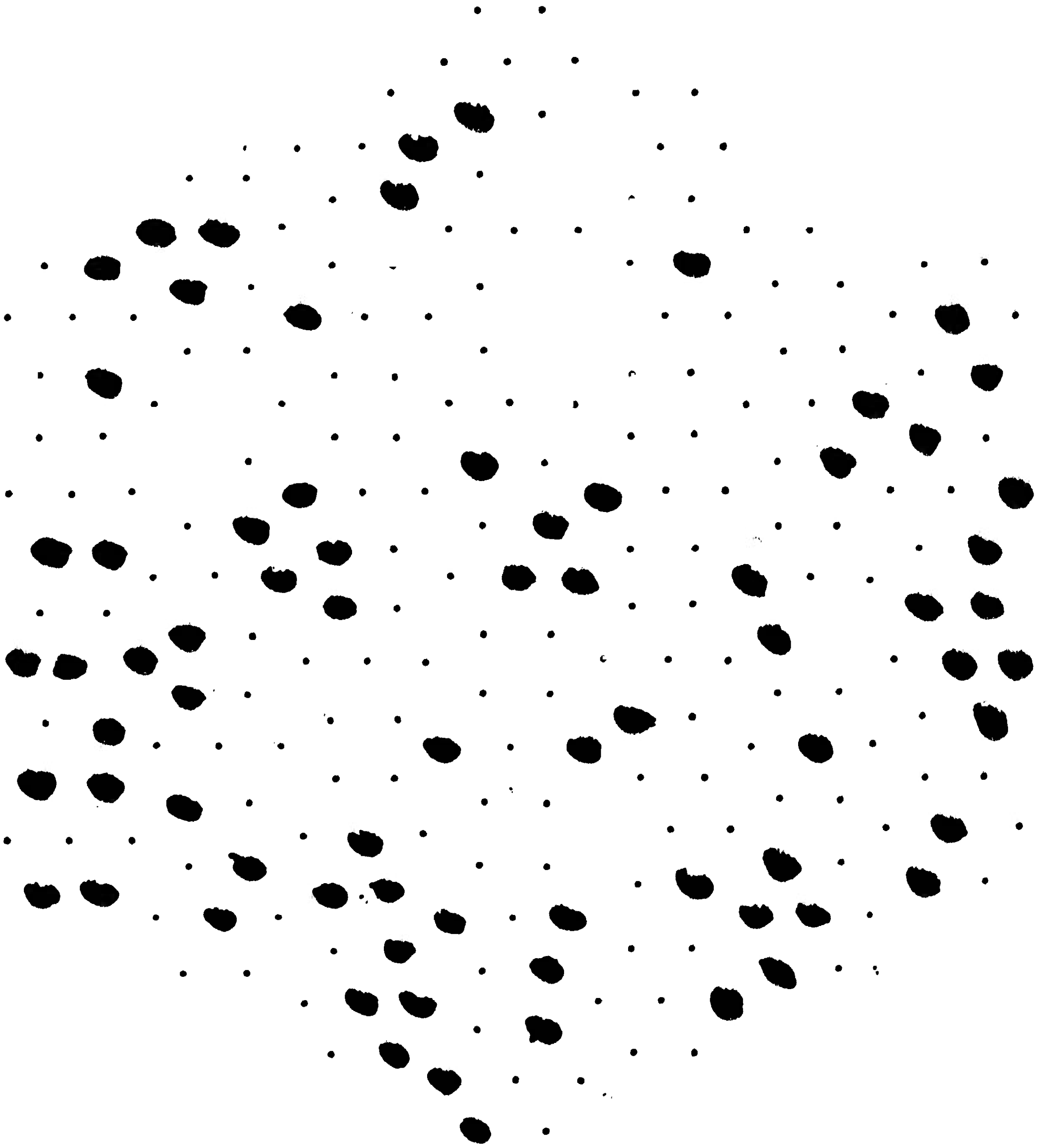
APPENDIX 9 : Continued (CHEX = 1.41)



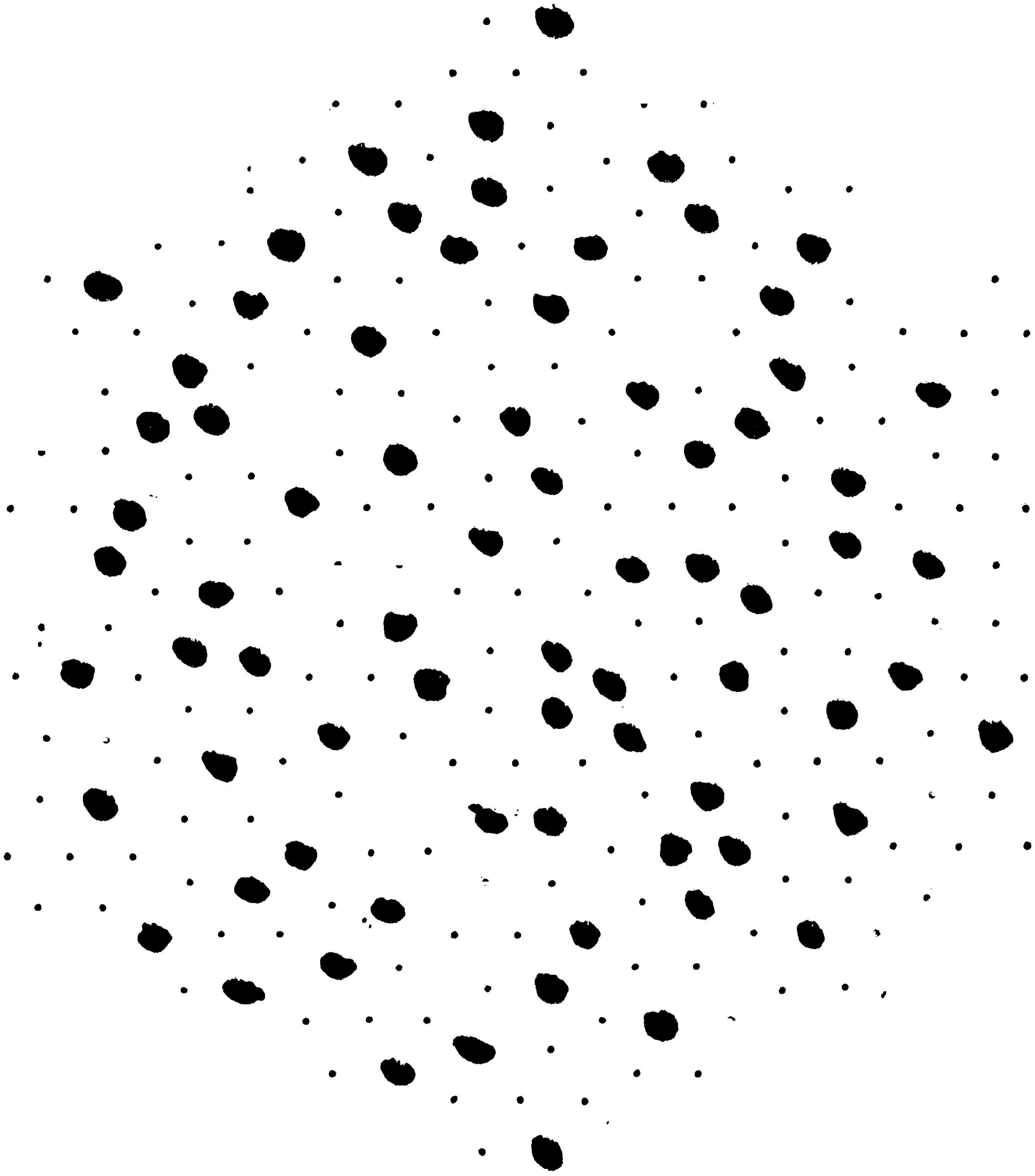
APPENDIX 9 : Continued (CHEX = .973)



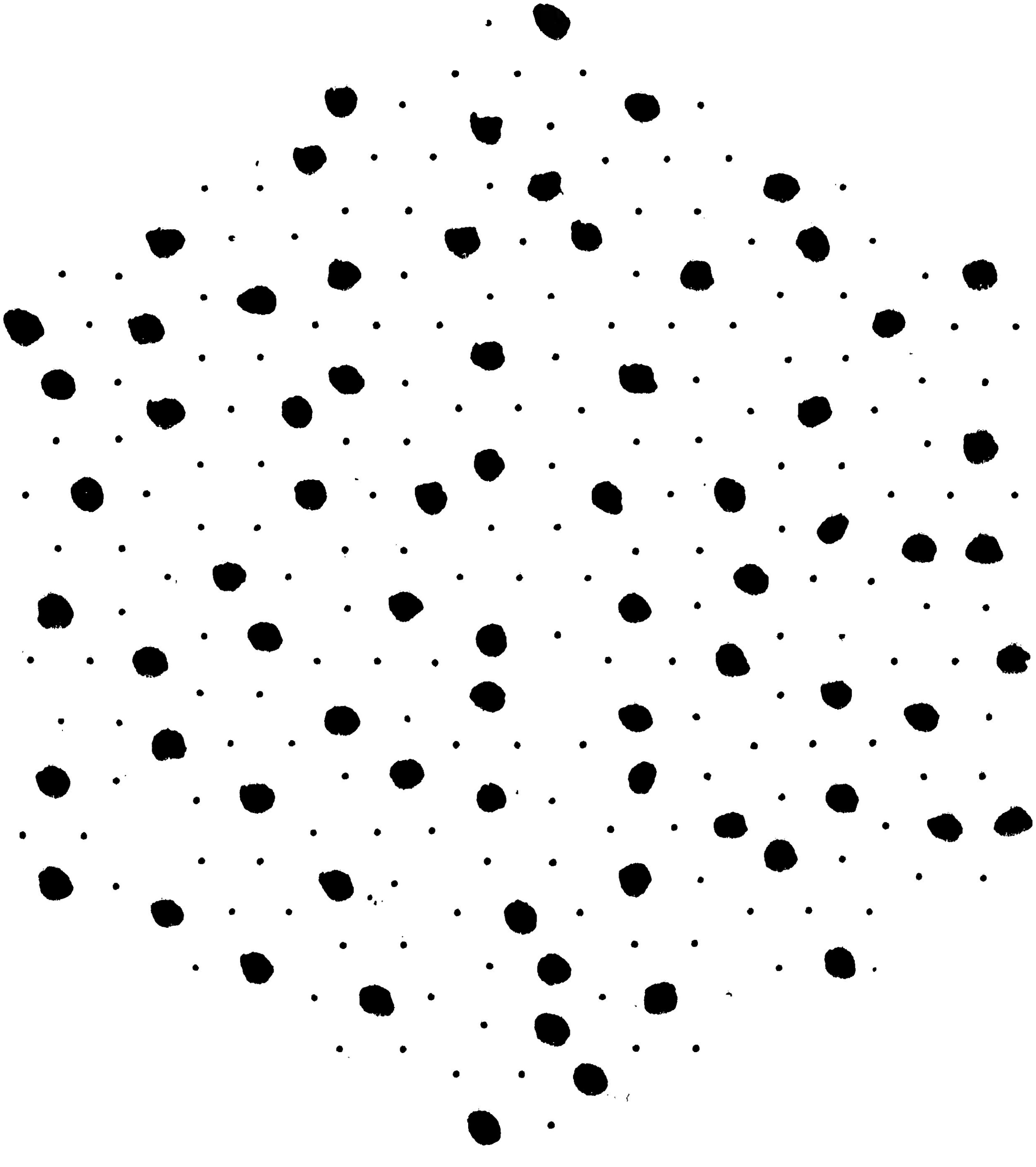
APPENDIX 9 : Continued (CHEX = .811)



APPENDIX 9 : Continued (CHEX = .784)



APPENDIX 9 : Continued (CHEX = .410)



APPENDIX 9 : Continued (CHEX = .162)

SUBJECT	CSQU	CHEX	DSQU	DHEX
1	.540	.162	7	9
2	.660	.378	8	9
3	.900	.271	7	6
4	.500	.243	7	6
5	1.060	.433	3	3
6	.580	.189	5	7
7	.740	.541	7	11
8	.500	.189	8	11
9	.580	.243	8	9
10	1.140	.298	7	9
11	.820	.352	6	6
12	.580	.162	8	7
13	.620	.410	7	7
14	.500	.216	8	10
15	1.540	.298	6	3
16	.780	.568	6	9
17	.500	.271	8	10
18	.540	.216	6	10
19	.860	.401	8	9
20	.540	.379	7	9
21	.780	.460	6	8
22	1.300	.676	6	8
23	.660	.162	8	8
24	1.820	1.410	4	9
25	.580	.243	6	10
26	1.140	1.082	7	7
27	.660	.298	5	7
28	.620	.189	7	9
29	.960	.595	7	9
30	.820	.784	6	6
31	.500	.352	5	9
32	1.060	.811	6	6
33	.500	.189	7	8
34	.740	.243	7	6
35	1.500	.515	8	7
36	.660	.973	7	7
37	.660	.243	7	9
38	.540	.271	8	11
39	.660	.216	6	9
40	.540	.379	7	8

TOTAL--------------N=40

APPENDIX 10

Individual results on all tasks