SEQUENTIAL RESPONSE BIAS

A STUDY ON CHOICE AND CHANCE

W.A. Wagenaar

1972

INSTITUTE FOR PERCEPTION TNO

SOESTERBERG - THE NETHERLANDS

Printed in The Netherlands by Bronder-offset N.V., Rotterdam

CONTENTS

CHAPTER 1

INTRODUCTION

§1. SEQUENTIAL RESPONSE BIAS AND SUBJECTIVE RANDOMNESS

The concept of chance is man's tool to describe a world in which unpredictable events occur. The question that can be raised, however, is whether any event may really occur "just by chance" or that all events are connected by a chain of causal relations. Even if the latter view is correct, which implies that chance does not play a rôle in the physical world, man will be bound to use the concept of chance in those situations where he has no access to the "real" causes. As Brunswik (1955) stated: "while God may not gamble, animals and humans, ... cannot help but to gamble in an ecology that is of essence only partly accessible to their foresight", (p. 236).

Consequently, the use of the concept of chance may indicate nothing more than an inability to find causes, systematic relations, or order in the physical world.

The proper way to find systematic relations in sets of phenomena seemingly ruled by chance is indicated by scientific methods. In a sense the history of science describes how the frontiers of chance have been repelled by the numerous techniques and disciplines. Specifically designed for the discrimination between chance and not-chance are the disciplines of probability theory and statistics.

In normal every-day life it is as important to discriminate between chance and order as it is in science. Man may react to his environment in a more sensible way when the rules behind the phenomena are perceived.

It is useful to know when one may burn one's fingers and when not! In every-day situations, however, man can not apply statistical techniques, and hence will be forced to discriminate between chance and order on a more or less intuitive basis. In a classic review by Peterson & Beach (1967) entitled "Man as an intuitive statistician" a diversity of strategies is discussed by which man deals with an uncertain world.

The topics discussed include intuitive description of observed data, intuitive inferences about populations on the basis of observed data, and intuitive predictions concerning future data. The detection of systematic trends in observed data is only a small part of these "intuitive statistics". In tasks of this type people use the concept of chance intuitively, as far as they continuously judge whether a certain trend would or would not occur by chance.

What do people expect to happen by chance? At which point do they begin to reject chance and to accept the possibility of a causal factor? A large number of experimenters devoted their research to these and similar questions. Most studies were restricted to the perception of "chance" in sequences of events. Two major research techniques have been applied.

One technique is the randomization paradigm. Subjects are instructed to produce a random sequence of elements, as would be generated by a "fair" randomizator, like a coin or die. These randomization experiments generally show that man is a bad randomizer. Human subjects continuously show biases for or against certain alternatives , dependent on previous choices. This tendency is called *sequential response bias*. It is called "response bias" because it is the subject who introduces the effect into the response sequence, and it is called "sequential" because of its dependency on the sequential structure.

The other method uses a judgement paradigm. Sets of data with some order in it and sets of data without any order (random sets) are presented to a subject, with the request to indicate which sets look random. The point of *subjective randomness* is then expressed as the degree of order present in the sequences which looked random to the subjects.

The distinction between *sequential response bias* and *subjective randomness* should be kept in mind very clearly. The two notions are connected with different experimental paradigms: sequential response bias is measured in randomization experiments, and subjective randomness is measured in judgement experiments. Each paradigm has its own history.

Consequently the review of the experimental literature consists of two separate sections.

§2. SEQUENTIAL RESPONSE BIAS IN RANDOMIZATION EXPERIMENTS[1]

The interest in sequential response bias in various fields of psychological research was extensively reviewed by Tune (1964a). It appears that the interest was first raised in the early part of this century, when psychophysicists noticed that successive responses of a subject are mutually dependent (Fernberger, 1913). In the psychophysical setting the usual procedure is that the subject makes a binary choice. At the level of the absolute threshold, *i.e.* the point where half of the stimuli are perceived, the sequence of binary responses is supposed to be completely random. Careful analysis of response sequences, however, has shown that prediction of future responses benefits from the knowledge of previous responses. This implies that subjects are not responding exclusively to the actual stimulus, but also in part to the structure of the previous response series. Hence, the informational content of responses in psychophysical measurements is less than originally postulated.

More recent research on subjective probability, probability learning, and gambling behaviour, also revealed that successive responses of a subject are mutually dependent in experimental settings where independent responses were expected. In general, the subject's task in this type of research is to accumulate information contained in the previous stimulus sequence, in order to determine optimal present or future responses. Here sequential response bias implies that the subject is not only guided by the stimulus situation (including previous stimuli) but also by the sequence of previous responses, which complicates the nature of the response. An extreme example of this process is given in Fig. 1.1. The boy does not pay any attention to the stimuli (questions) at all, but only tries to produce a random sequence of responses.

Clinical psychologists became interested in sequential response bias in as far as they used a randomization test to diagnose neuroticism (*e.g.* Weiss, 1964).

Finally, randomization has been employed as a secondary task in the measurement of mental load, were it is assumed that the sequence becomes less random with an increased mental load imposed by the primary task.

Fig. 1.1. An extreme example of attending to previous responses[2].

The degree of sequential bias in the randomized sequence is then taken as a measure of mental load imposed by the primary task (*e.g.* Spence, Wolitzky & Pezenik, 1969).

Reichenbach (1949) was the first to claim that humans are unable to produce a random sequence of responses, even when explicitly instructed to do so: "If a person not trained in the theory of probability were asked to construct artificially a series of events that seems to him to be well shuffled, there would not be enough *runs* in it". Since then, 15 other studies on the randomization paradigm have been published. Tune (1964b) reviewed four such experiments. The present section is devoted to an analysis of all known publications to date.

Definition of the randomization experiment

The experiments to be discussed in this section are characterized by several restrictions:

1. The subject is explicitly instructed to produce a random series of events. Often the instruction refers to random processes like coin tossing, dice throwing, etc.
2. The series are long enough to prevent complete memorization.
3. No stimulus or feedback is given to the subject during the experiment, except for an possible pacing signal.
4. The subjects are normal human adults.

The restrictions stated above exclude the following types of related experiments:

1. Senders & Sowards (1952), Senders (1953) and Wagenaar (1968) used the "pseudo-psychophysical" paradigm. The subjects were told that they participated in the measurement of a sensory threshold, but actually no stimuli were presented. In this scheme subjects persisted in responding as if there really was a stimulus. Hence, non-randomness in the produced sequences is to be attributed to a response bias of the subject, and not to a variable sensitivity. A related paradigm was used by Zwaan (1964) who suggested to his subjects that successive outcomes of chance processes (like tossing a coin) were transmitted to them by means of telepathy. The response was to be written down as soon as they received the "telepathic message". The expectation was that subjects would try to react randomly. The latter technique was inspired by the famous Zenith radio experiments concerning telepathic transmission of short sequences (Goodfellow, 1938).
   Both paradigms, ingenious as they are, have the difficulty that random behaviour is not explicitly instructed. Hence, the exact nature of the task remains somewhat vague.
2. Experiments by Bendig (1951), Cohen & Hansel (1955a,b) and Liebermann (1963) were devoted to non-randomness in very short sequences (up to ten elements). In these sequences no high order sequential effects can be calculated.
3. Some experiments (*e.g.* Hake & Hyman, 1953) studied non-randomness in probability learning situations. Since information about the correctness of individual guesses was given, dependencies on both previous responses and previous stimuli occurred. In this respect the paradigm is very different from the randomization experiment.
4. A vast literature exists on alternation behaviour in children (Croll, 1966; Jeffrey & Cohen, 1965; Manley & Miller, 1968; Miller et al., 1969; Rabinowitz, 1969; Rieber, 1966) and in rats (Lester, 1968), while sequential response bias has also been studied in abnormals (Mittenecker, 1953; Weiss, 1964; Yavuz, 1963; Zlotowski & Bakan, 1963).

Comparison of experimental procedures and conditions

Experimental evidence on randomization is highly contradictory. One reason may be the striking divergence of experimental procedures used by the various experimenters. Some relevant factors that contribute to the disagreement among experimental results are presented in Table 1.1. These factors are:

1. Number of alternative events, ranging from 2 to 26.
2. The nature of the alternatives, varying from numbers and letters to circles on a paper and nonsense syllables.
3. Sequence length, ranging from 20 to 2520.
4. Presentation of the choice set: internally by a verbal definition of the set, or externally by a hardware display of the set.
5. Availability of previous responses, dependent on the mode of production; if the sequence is called out, the response can only be remembered whereas responses that are written down remain present.
6. Rate of production, ranging from 0.25 to 4.0 sec per response.
7. Number of subjects, ranging from 2 to 124.

Each of these factors may be relevant for the randomization paradigm. Among the 15 experiments under discussion, no two experiments can be found which differ in only one of the factors mentioned above. Hence, part of the contradiction in the results may be due to the fact that the experimental procedures are not completely comparable. An extensive analysis of the importance of the above mentioned factors will be an indispensable part of the next chapters.

Mathematical definition of non-randomness

With respect to the mathematical definition of non-randomness little standardization is evident concerning the criterion for calling a series random or non-random. Here a methodological problem arises as non-randomness is easier proved than disproved. For proving non-randomness it is sufficient to reveal one type of systematic trend in the series, whereas for the establishment of real randomness it is required to prove that *not any* serial regularity of the many possible ones is present. An endless repetition of the alphabet, for instance, is perfectly random regarding single letter frequencies, but extremely non-random with respect to frequencies of pairs. A similar difficulty occurs when an experimenter is interested in the increase or decrease of randomness: one series can be more random than another according to one criterion and, at the same time, less random in another respect. Recognition of this problem is crucial for the interpretation and comparison of experimental results. The measures of non-randomness most frequently used are presented in Table 1.2. If only frequencies of single responses are taken into account, analyses are said to be of zero order[3]. In zero order analyses no dependencies among responses can be established. For first order analyses frequencies of digrams (= pairs) are used, for second order analyses frequencies of trigrams etc. The general rule is that analyses of order n, which require a count of (n+1)-grams, can yield dependencies between responses that are maximally n places apart.

As shown in Table 1.2 the measures of non-randomness used by the various experimenters, differ with respect to both the order of analysis and the mathematical definition.

A detailed review of all advantages and disadvantages of the various measures of non-randomness will be presented in the next chapter. At this moment it will suffice to state that most measures of non-randomness are neither powerful enough for disproving all sequential regularities, nor adequate for establishing increases and decreases of sequential response bias.

Results and theories

Given the divergence of experimental procedures and methods of measurement, it is not surprising that results are quite contradictory. Actually there is no way of combining details of the results of the 15 publications into one coherent theory. Some major outcomes are presented in Table 1.3.

Firstly, almost all experimenters found systematic effects of sequential response bias. Only Ross (1955) claimed that his subjects were good randomizers.

Secondly, most experiments yielded negative recency, *i.e.* the occurrence of too many alternations or too many runs. Some authors used methods that do not distinguish between negative and positive recency. Positive recency was reported only for first order dependencies. Weiss' (1964) data seemed to point to second order positive recency, provided that his relative frequencies of trigrams were corrected to add up to 100%. Although Ross' (1955) experiments yielded real randomness, some objections can be raised. His subjects were requested to stamp symbols

(X or 0) on cards. This procedure may have favoured repetition (going on with the same stamp) over alternation (taking the other stamp). Thus, in Ross' experiment the frequently observed tendency towards alternation may have been counterbalanced by this unintended facilitation of repetition.

Thirdly, several other systematic deviations from randomness were found, such as preference for the natural order of the alternatives, and preference for as well as avoidance of symmetric patterns. In general, these systematic trends are related to the nature of the stimuli.

Finally, Table 1.3 presents some factors which are supposed to increase non-randomness. In view of difficulties in defining such an effect mathematically, these outcomes should be evaluated with caution.

Several theories have been proposed to account for sequential response bias. The main distinction is between theories that assume an incorrect concept of randomness and theories that attribute sequential response bias to the limiting effect of some psychological function. The concept hypothesis is implicitly or explicitly sustained by Chapanis (1953), Mittenecker (1953, 1958), Rath (1966), Ross & Levy (1958), Skinner (1942), Teraoka (1963), and Zwaan (1964). The reason why a subjective concept of randomness should deviate in the direction of negative recency always remains rather vague. Skinner (1942) suggested that in verbal behaviour repetition is a natural tendency which is extinguished during maturation by processes of conditioning. Rath (1966) referred to the Cournot-hypothesis in economics, according to which subjects expect repetition in the world instead of change. Randomness, which is the opposite of what the world looks like, should therefore contain many alternations.

In the other type of theory sequential response bias is attributed to some functional limitations of the subject. Weiss (1964, 1965) and Wolitzky & Spence (1968) proposed an attention hypothesis: the subject should maintain a set to be random and to inhibit other modes of responding. In a sense, this is paying attention in order not to pay attention to previous responses. Tune (1964b) argued that randomization is hampered by the limited capacity of human memory. Subjects who can tally frequencies of all n-grams, may be random up to the order (n-1). Baddeley (1962, 1966), however, claimed that the very existence of human memory is responsible for sequential bias, since a system without any type of memory could never produce sequential dependencies. Instead, he proposed the hypothesis that humans have a limited capacity for generating information. Rath (1966) made the same suggestion. Both authors failed to explain why and how such a mechanism should operate.

The experimental data presented thus far confirm and descredit the various theories to almost the same extent. Therefore, in the next chapters we shall enter into a detailed analysis of all theories in the light of new experimental evidence.

§3. SUBJECTIVE RANDOMNESS IN JUDGEMENT EXPERIMENTS

In the preceding section it was mentioned that several investigators attributed sequential response bias to an incorrect concept of randomness: subjects are bad randomizers because they do not know what "random" means. If this were true, the same error should be made in a judgement experiment: sequential response bias and subjective randomness should be highly related phenomena. If, on the other hand, subjects who show a sequential response bias are able to succesfully discriminate between random and non-random sequences, there is no reason to hypothesize an "incorrect concept of randomness".

Three of the seven authors suggesting the existence of such an incorrect concept mention the judgement experiment. Additionally, Cook reported an experiment of this type in 1967. Some relevant information concerning those four experiments is summarized in Table 1.4. The results are in mutual disagreement. The existence of subjective randomness was twice denied and twice confirmed. In Chapter 5 a more detailed analysis of the experimenters will be presented.

Judgement of non-randomness in two-dimentional displays was studied by Ross & Weiner (1963) and Gemelli & Alberoni (1961). This topic is beyond the scope of the present survey.

Author(s)	Number of alternatives in the choice set	Nature of the alternatives	Number of series per subject per condition	Number of responses per series	Internal or external represented choice set	Mode of production	Number of seconds per response	Paced or unpaced	Number of previous responses visible for the subject	Number of subjects
Baddeley 1962	26	letters	2	100	i	calling out	0.5,1, 2,4	p	0	24
	2,4,8, 26	letters	1	360	i	calling out	1	p	0	50
	16	letters	1	360	i	calling out	1	p	0	10
	2,4,8, 16,26	digits	1	360	i	calling out	1	p	0	38
	26	letters	8	100	i	calling out	2	p	0	12
Baddeley 1966	26	letters	2	100	i	calling out	0.5,1 2,4	p	0	12
	26	letters	16	100	i	calling out	2	p	0	12
	2,4,8, 16,26	letters	3	120?	i	writing down	1	p	all	124
	2,4,8, 16,26	letters	3	?	i	writing down	?	u	all	120
	2,4,8, 16,26	numbers	3	120?	i	writing down	1	p	all	101
	2,4,8, 16,26	numbers	3	?	i	writing down	?	u	all	92
Bakan 1960	2	heads-tails	2	150	i	writing down	?	u	all	70
Chapanis 1953	10	digits	1	2520	i	writing down	±2	u	all?	13
Lincoln & Alexander 1955	8	discs	4	700	e	touching	1.2	p	0	16
	8	discs	4	700	e	calling names	1.2	p	0	16
Mittenecker 1953	10	digits	1	100	i	calling out	$\pm0.6^4$	u	0	56
	10	digits	1	?	i	calling out	$\pm0.6^4$	u	0	30
	10	digits	1	?	i	calling out	? (slow)	u	0	30
Mittenecker 1958	9	circles on a paper	1	180	e	pointing	1	p	0	20

Author(s)	Number of alter-natives in the choice set	Nature of the alternatives	Number of series per subject per condition	Number of re-sponses per series	Internal or exter-nal represented choice set	Mode of production	Number of seconds per response	Paced or unpaced	Number of previous responses visible for the subject	Number of subjects
Rath 1966	2	digits	10	250	i	writing down	0.5	p?	all	20
	10	digits	10	250	i	writing down	0.75	p?	all	20
	26	letters	10	250	i	writing down	1.0	p?	all	20
Ross 1955	2	digits	1	100	e	stamp-ing	?	?	1?	60
Ross & Levy 1958	2	heads-tails	8	20	i	writing down	?	?	all	15
	2	heads-tails	4	20	i	writing down	?	?	all	15
Teraoka 1963	5	digits	1	1252	i	calling out	±1	u	0	4
	5	letters	1	1252	i	calling out	±1	u	0	4
	5	nonsense syll.	1	251	i	calling out	±1	u	0	2
	5	digits	1	751	i	calling out	1	p	0	3
	5	digits	1	751	i	calling out	±0.5[4]	u	0	3
Warren & Morin 1965	2,4,8	digits	6	500	i	calling out	0.25, 0.50, 0.75	p	0	2
Weiss 1964	2	push-buttons	1	600	e	pushing	1	p	0	28
Wolitzky & Spence 1968	10	digits	1	45	i	writing down	2.45	p	1	20
Zwaan 1964	2	heads-tails	1	32	i	writing down	?	u	all	48
	4	suits of a pack of cards	1	32	i	writing down	?	u	all	48
	6	sides of a die	1	32	i	writing down	?	u	all	48

Table 1.1. Experimental conditions of 15 experiments on randomization by humans.

Author(s)	Order of analysis	Description of the measure for non-randomness
Baddeley 1962	1	repetition of digrams
	0	redundancy
	1	stereotyped response
	1	information per response
Baddeley 1966	0	redundancy
	1	stereotyped and repeated digrams
Bakan 1960	1	number of runs
	2	alternation and symmetry in trigrams
	0	frequency of alternatives
Chapanis 1953	0	frequency of alternatives
	1,2	frequency of digrams and trigrams
	1-?	autocorrelation function
Lincoln & Alexander 1955	0-2	redundancy
	0	frequency of alternatives
	1	spatial distance between two alternatives in the digrams
	2	frequency of trigrams
Mittenecker 1953	mixed	distance of repetition (recurrence)
	0	frequency of alternatives
Mittenecker 1958	0,1	redundancy
	1-11	autocorrelation function
Rath 1966	0	frequency of alternatives
	1	frequency of digrams corrected for frequency of alternatives
	2	frequency of trigrams corrected for frequency of digrams
	1	frequency of digrams as a function of the distance between the two elements in the natural ordering

Author(s)	Order of analysis	Description of the measure for non-randomness
Ross 1955	1	number of alternations
	0	frequency of alternatives
Ross & Levy 1958	1	number of alternations
	1-?	occurrence of runs
Teraoka 1963	0	frequency of alternatives
	1	conditional probabilities
	1	frequency of digrams as a function of the distance between the two elements in the natural ordering
	1-4	occurrence of runs
Warren & Morin 1965	0-3	redundancy
Weiss 1964	1-9	frequency of (n)-grams corrected for lower order dependencies
	2	frequency of trigrams
Wolitzky & Spence 1968	0	frequency of alternatives
Zwaan 1964	1	number of runs
	mixed	distance of repetition (recurrence)

Table 1.2. Measures of non-randomness used by various experimenters.

Author(s)	Are subjects good randomizers?	Positive or negative recency	Other systematic deviations from randomness	Factors increasing non-randomness
Baddeley 1962 and 1966	no	?	unbalanced 1- and 2-gram frequencies, stereotyped digrams	increase of rate of production and number of alternatives, introduction of secondary task
Bakan 1960	no	neg.	avoidance of symmetric response patterns	
Chapanis 1953	no	neg.	unbalanced 1-, 2-, 3-gram frequencies preference to decreasing series, avoidance of increasing series	naivety of subjects
Lincoln & Alexander 1955	no	neg.	preference to the easy motor responses, to alternatives with a large spatial distance to the previous alternative, and to clockwise or counterclockwise ordered sequences	giving verbal response instead of motor response
Mittenecker 1953 and 1958	no	neg.	balancing of frequencies within small samples	neuroticism decrease of rate of production
Rath 1966	no	neg.	preference to symbols adjacent in the natural sequence	increase of number of alternatives

Author(s)	Are subjects good randomizers?	Positive or negative recency	Other systematic deviations from randomness	Factors increasing non-randomness
Ross 1955	yes	pos.		
Ross & Levy 1958	no	pos. and neg.[5]	overuse of runlength with expected frequency of at least 1	naivety with respect to the expected frequency of runs
Teraoka 1963	no	neg.	response-chaining related to the natural order of the alternatives, dependencies over at least 5 places, periodicity with period of 5 responses	presence of a natural order of alternatives, decrease of rate of production
Warren & Morin 1965	no	?		increase of rate of production and of number of alternatives
Weiss 1964	no	pos?	preference for symmetric trigrams	boredom
Wolitzky & Spence 1968	no	?		increase of the informational load of a secondary task
Zwaan 1964	no	neg.		

Table 1.3. Some results of experiments on randomization by humans

§4. WORKING PROGRAM

Thus far some general information on sequential response bias and subjective randomness is presented. It appears that a multitude of problems are waiting for elaboration and solution. The program, then, is as follows. In Chapter 2 the definition of "randomness" is treated and a solution to the measurement problem will be proposed. Chapter 3 reports an introductory experiment on sequential response bias in randomization tasks. The topics discussed include the nature and the extent of sequential response bias. In Chapter 4 more specific details of sequential response bias are studied. The focus is mainly on the effect of various experimental conditions, as listed in Table 1.1. The results are discussed in relation to the supposed contributions of psychological functions like memory and attention.

Chapter 5 is devoted to subjective randomness in judgement experiments and to the relation between sequential response bias and subjective randomness. Finally, in Chapter 6 some elements of a theory of sequential response bias are provided.

Author	Stimulus material	Results
Baddeley 1966	digit sequences generated by subjects and random sequences	subjects selected the true random sequences
Cook 1967	12 binary sequences, one random and 11 with various types of bias	subjects recognized degrees of sequential bias
Mittenecker 1953	ten-digit sequences, real random, or subjectively random, according to a hypotheses of the experimenter	subjects selected the subjectively random sequences
Zwaan 1964	2-, 4- and 6-alternative sequences with non-randomness varied from negative to positive recency in five steps	subjects selected sequences with negative recency

Table 1.4. Summary of experiments on subjective randomness.

CHAPTER 2

MEASUREMENT OF HIGH-ORDER NON-RANDOMNESS IN SHORT SEQUENCES

§1. A PARADOX

Specifying a "degree of (non-)randomness" leads to a paradoxical situation. When a coin is tossed 100 times, 2^{100} different output-sequences are possible, each with the same probability of occurrence. In this sense all binary sequences containing 100 elements have an equal "degree of randomness", which makes it difficult to understand why some of those sequences should be labeled as "non-random". If certain sequential properties are defined, which distinguish "random" sequences from "non-random" ones, the possibility exists that, at a certain moment, the outcome of a toss *must* be "heads" and not "tails" in order to fulfil the requirement of randomness. Then the paradox is that an element of a random sequence, which is unpredictable by definition, becomes predictable by the very requirement of randomness.

The paradox is solved by two considerations. First, it should be stressed that although all binary sequences that are a result of tossing a coin 100 times are equiprobable, some relevant properties of the sequences have no uniform distribution. The frequency of "heads", for instance, follows a binomial distribution with a mode at f (heads) = 50. The probability of exactly 50 times "heads" is 0.80×10^{-1}, whereas the probability of 100 times "heads" is 0.79×10^{-30}. As a result a prediction can be made concerning the frequency of "heads" in a random sequence, notwithstanding the equiprobability of the 2^{100} possible sequences.

Secondly, part of the confusion is due to a careless usage of language. Specifying non-randomness of a sequence means, properly speaking, specifying the non-randomness of *the source that generated the sequence.* If the process of coin-tossing is a priori defined as random, the resulting sequences are also random by definition. In that case it is of no use to speak about a degree of randomness. If, on the other hand, the source that generated the sequence is unknown, it is possible to infer some properties of that source on the basis of the sequences it generated. The unknown source is connected with the sequences it generated by laws of probability. Therefore statements concerning the source can only be made with a limited degree of confidence when based on the observed sequences. As a special case one may determine the confidence with which an unknown source is described as a specified random generator. If the confidence falls below a certain criterion the hypothesis is rejected that the actual source can be identified with the random generator. The paradox is originated because in daily use it is then said that the sequence is rejected as being "non-random".

1. Related notions	2. Definition of the source, producing randomness		3. Descriptions of random behavior	4. Operational definitions
accidental	without:	aim	abnormal	unpredicted
by chance		attention	disorderly	unexpected
casual		care	incoherent	
desultory		cause	irregular	
haphazardly		concern	nonsymmetrical	
needless		design		
incidental		intention		
occasional		method		
		notice		
		purpose		
		strategy		
		system		
	ruled by:	destination		
		fate		
		fortune		
		fortuity		
		hap		
		luck		

Table 2.1. Definition of "random", according to The Oxford Dictionary and Websters Dictionary of the New World.

§2. THE DEFINITION OF RANDOMNESS

Consultation of a dictionary shows that definitions of the notion 'random' fall into four classes (see Table 2.1). The first class gives related notions, such as "haphazardly" or "by chance". The next class contains definitions that describe the sources that produce randomness. Thirdly, random is defined by a description of what random behaviour looks like. Finally, randomness is defined in terms of inadequate expectation and prediction.

The last of these classes contains definitions which are most directly related to the definition of randomness as given by Reichenbach (1949). After a detailed enumeration of mathematical definitions he presented a "psychological definition of randomness", which essentially shows few differences with our formulation: random sequences are characterized by the property that not any person predicting the sequence may enhance the predictions of future responses by use of the previous responses. Although a formulation of this kind "refers only to a limitation of the technical abilities of human observers", Reichenbach stated that "for all practical purposes the psychological definition of randomness is sufficient".

Two statements are added to the psychological definition for mathematical use in the next sections:

1. In a random sequence of choices among A alternatives, the probability of a certain alternative being selected on the n^{th} place equals 1/A, irrespective of the n-1 previous choices.
2. The combined probability of a (part of a) sequence is the product of the probabilities of the successive elements.

§3. FIRST DESIGN OF A MEASURE OF HIGH-ORDER NON-RANDOMNESS IN SHORT SEQUENCES[6]

Development of a suitable measure of sequential non-randomness in behalf of the study of sequential response bias should start from the following requirements:

1. The number of elements in the series should not be extremely large. A sequence of 100 responses should be sufficiently long.
2. Non-randomness is to be calculated at least to the sixth order and

independently for each order of analysis.

3. The measure should be able to distinguish between negative recency (too many alternations) and positive recency (too many repetitions).
4. The method should yield comparable outcomes for series of different length and with different numbers of alternatives.
5. The values of non-randomness are to be defined at least on an interval scale, such as to permit the calculation of mean scores.

In the next section the existing techniques for the measurement of non-randomness are reviewed, especially with respect to the forementioned requirements.

Review of existing techniques

The techniques used in the experiments reviewed in the previous chapter vary with respect to both the order of analysis and the mathematical origin of the measures.

One class of measures bears relation to occurrence of runs, which are strings of responses of one and the same alternative, and bordered on both sides by responses of a different alternative. The number of runs, used by Bakan (1960) and Zwaan (1964), is essentially a first order measure since it equals the number of digrams with unequal elements. The frequency distribution of runs with length i, as used by Ross & Levy (1958) and Teraoka (1963) is a measure with all orders mixed in a mathematically complex way. Recurrence distributions (Mittenecker, 1953, 1958; Zwaan, 1964) give the frequency distribution of gaps with length i between two identical responses, which are actually runs of non-occurrence of that alternative. This again is a measure with all orders mixed.

A second class contains measures from information theory, like information per response (Baddeley, 1962) and relative redundancy in the series (Baddeley, 1962, 1966; Lincoln & Alexander, 1955; Mittenecker, 1958; Warren & Morin, 1965). Measures of this type require very long series for higher order analyses. The very requirement of long series, especially if there ar more than two alternatives in the choice set, is a basic obstacle for the application of these measures of non-randomness. A sixth order analysis of series with four alternatives should be based on an average of five observations in 16.384 cells, which amounts to over 80.000 responses (Attneave, 1959). Such experiments would deal with boredom or aggressive feelings against the experimenter, rather than with sequential response bias.

Furthermore, for analyses above order 4, auto-correlation functions are often used, which have again the disadvantage that estimates of dependencies are not given separately for each order (Chapanis, 1953; Mittenecker, 1958). A series with an endless repetition of the digram 0-1 will yield an endless auto-correlation function with values +1, -1, +1, -1, etc. Yet the simplest description of the dependencies is a first order alternation model.

One way to overcome this difficulty is to calculate a power spectrum on the basis of the auto-correlations (Pöppel, 1967). For the computation of a power spectrum with six terms, however, at least 72 auto-correlations are needed in order to secure a reliable outcome, whereas the computation will be successful only if the auto-correlation function is fairly periodic over this interval. Unfortunately, this is not a priori true for attempted random sequences.

A promising approach is discussed by Vitz & Todd (1969) who code the sequence into a hierarchy of larger elements until the sequence itself is the final coded element. They define the complexity of the sequence by using Garner's multi-variate uncertainty analysis (Garner, 1962). For the time being, this method neither gives independent estimates of non-randomness of successive orders of analysis, nor is it well enough developed to allow for comparison of series with unequal length or an unequal number of alternatives.

Finally, Weiss (1964) proposed a technique which is an outgrowth of information theory techniques, but with fewer requirements as to sequence length. Briefly, his technique consists of calculating expected frequencies for all n-grams on the basis of observed frequencies of the (n-1)-grams. These expected frequencies are compared with observed frequencies of n-grams by a squared deviation measure. This method cannot distinguish between positive and negative recency and does not allow a comparison of series with unequal length or an unequal number of alternatives.

In general, it can be concluded that most measures of non-randomness are neither powerful enough for disproving a satisfactory range of serial

Measurement technique	Series not longer than 100	Independent scores up to the sixth order	Distinction between neg. and pos. recency	Independence from N and A	Definition on interval scale
Total number of runs	+	-	+	-	+
Frequency distribution of runs	+	-	+	-	?
Recurrence distributions	+	-	+	-	?
Information per response	-	+	-	-	+
Auto-correlation functions	+	-	+	?	?
Power spectrum of auto-correlation functions	-	+	+	?	?
Coding principle of Vitz & Todd (1969)	+	-	?	-	?
S^2-score of Weiss (1964)	?	+	-	-	?

Table 2.2. Comparison of existing techniques for measuring sequential non-randomness with the five requirements stated in §3.

regularities, nor adequate for establishing increases and decreases of non-randomness. As shown in Table 2.2, no technique fulfils all requirements stated above. Therefore, it was necessary to develop a new technique for the measurement of sequential non-randomness in short sequences. The proposed method is related mostly to Weiss' approach, which also took its start from expected and observed frequencies of n-grams.

The ϕ-coefficient

The main idea behind the proposed method is that it is possible to predict frequencies of (n+1)-grams on the basis of observed frequencies of n-grams if the n^{th} order non-randomness is zero. For example, frequencies of pairs may be predicted on the basis of the frequencies of the single elements if there are no first order dependencies.

The general formula for predicting the frequency of the n-gram

$e_1 \ldots e_n$, is

$$f(e_1 e_2 \ldots e_{n-1} e_n) = \frac{f(e_1 e_2 \ldots e_{n-1}) \times f(e_2 \ldots e_{n-1} e_n)}{f(e_2 \ldots e_{n-1})}, \qquad (2.1)$$

which reduces for the first order case to:

$$f(e_1 e_2) = \frac{f(e_1) \times f(e_2)}{N}. \qquad (2.2)$$

In this way frequencies of digrams are predicted starting from single-element frequencies. Comparison of predicted and observed frequencies of digrams renders ϕ_1. The frequencies of trigrams are predicted starting from observed frequencies of digrams, etc. This method is schematically presented in Fig. 2.1.

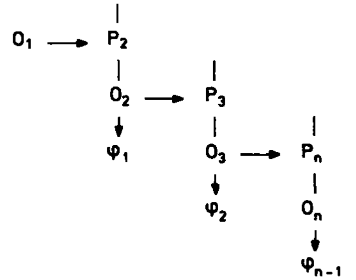


Fig. 2.1. Schematic representation of computation of ϕ-coefficients.O(n) and P(n) are observed and predicted frequencies of n-grams. P(n) is calculated on the basis of O(n-1). Comparison of O(n) and P(n) delivers ϕ(n-1).

The correspondence between predicted and observed frequencies of n-grams is a measure for the n^{th} order non-randomness. The next section deals first with the two-alternative case (A = 2) and then with the general case of many alternatives (A > 2).

A = 2

The n-grams contain n events $(e_1 e_2 \ldots e_n)$ each event with a value of zero or one. Such n-grams may be classified into two groups:

1. Repetitive n-grams: $e_1 = e_n$ (first and last event are both zero, or both one);
2. Alternating n-grams: $e_1 \neq e_n$ (one event zero, the other one).

By addition a sum of observed frequencies of repetitive n-grams and a sum of observed frequencies of alternating n-grams is obtained. The same is done for predicted frequencies.

The predicted values are compared with the observed values by a χ^2. A coefficient for non-randomness is given by $\phi = \sqrt{\chi^2/N}$, N being the sequence length. The range of ϕ is from zero to one. The value receives a sign which is arbitrarily minus for more alternating n-grams observed than predicted (negative recency), and plus for too many repetitive n-grams observed (positive recency).

<u>A > 2</u>

The n-grams contain n events ($e_1 e_2 \ldots e_n$) each event assuming one of the alternative values $a_1 a_2 \ldots a_x$. The sequence is translated into a binary sequence by replacing a zero for a_1 and a one for all other alternatives. Thus the distribution of the alternative a_1 over the sequence is obtained. In this way x different sequences can be constructed by singling out successively the elements $a_1 a_2 \ldots a_x$.

From the binary sequences thus obtained, frequencies of n-grams are counted and added again separately for repetitive and alternating n-grams. Then predicted values χ^2 and ϕ are calculated as in the case $A = 2$.

The method described above shows two important differences from redundancy measures used by information theory. First, sequences with more than two alternatives are translated into binary sequences which reduces the number of n-grams drastically; second, the n-grams are grouped as alternating and repetitive. In this way some of the original information contained in the frequencies of the n-grams is lost, but for sequences shorter than five times the number of n-grams ($N < 5\ A^n$), part of this original information had only a spurious value.

The reduction of the number of cells, obtained in this fashion, may mask some of the effects of non-randomness. Therefore, the ϕ-coefficient has a type II error like most of the other measures. The efficiency depends on the nature of the expected deviations from randomness. The type I error of ϕ is treated extensively in §5. Whether the ϕ-coefficient has the required properties of a measure for non-randomness stated above, is a question that can only be answered after some properties of ϕ have been reviewed.

§4. SOME PROPERTIES OF THE ϕ-COEFFICIENT

Extreme values of ϕ

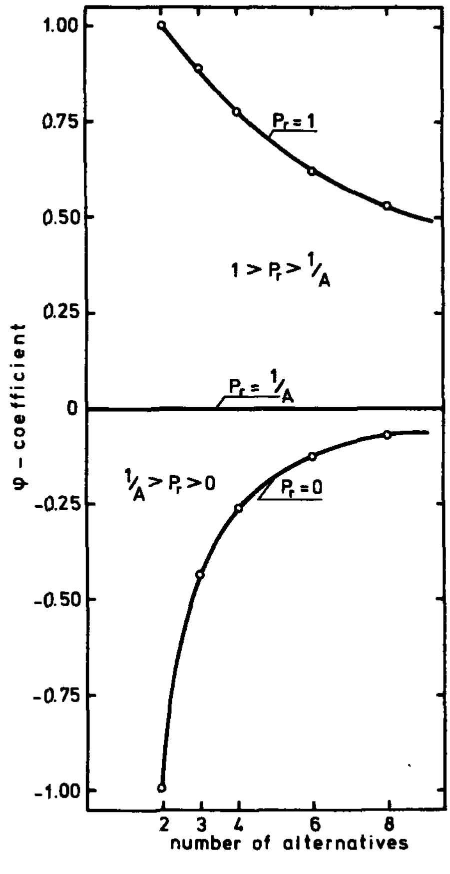


Fig. 2.2. Extreme values of ϕ, in a sequence with $N = \infty$, as a function of A. P_r = probability of a repetition.

Fig. 2.2 shows the extreme values of the first order ϕ-coefficient for $2 \leq A \leq 8$ in a sequence of infinite length. It appears that the extremes are -1.0 and +1.0 only for A = 2, whereas they approach zero when A increases. Moreover, an effect of skewness is apparent from the asymmetry of the two curves. The approach to zero is due to the fact that a certain difference is always expected between the number of repetitive and alternating n-grams if the frequencies of the single elements are smaller than N/2 (*i.e.* A > 2), whereas X^2 has a maximal value if the expected frequencies of alternations and repetitions are equal. The skewness is also an effect of single-element frequencies being less than N/2: the maximum effect of repetition makes N-1 digrams repetitive ones, whereas maximum alternation can produce only a number of alternating digrams equal to 2(N-1)/A, provided that the sequence has been translated into a binary code. This means that for A > 2 the maximum effect of repetition is larger than the maximum effect of alternation. Both effects, related to the extreme values of ϕ, and which occur for higher order ϕ-coefficients as well, are due to the translation into binary sequences in the case A > 2.

The zero point

In principle, real randomness should be represented by $\phi = 0$ for all values of A and N, and for all orders of analysis. Actually, ϕ is slightly biased depending on A and N as is shown by a Monte Carlo analysis (Table 2.3).

N \ A	2	3	4	6	8
50	-0.015	-0.004	-0.001	+0.001	-0.000
100	-0.010	-0.001	-0.003	+0.001	+0.001
200	-0.003	-0.002	+0.001	-0.001	-0.001
400	-0.006	-0.004	+0.002	+0.001	-0.000

Table 2.3. Median values of first order ϕ-coefficients in samples of 200 random sequences for different combinations of A and N.

The first order bias is explained by the fact that single element frequencies seldom equal N/A exactly. The unequal frequencies result generally in an expected frequency of repetitions which is too high[7]. Because a true random sequence has no systematic bias against alternation, the number of repetitions seems to be too low. A similar reasoning holds for higher order effects. This bias of the zero point is due to the small value of N, rather than to the specific definition of ϕ. Weiss' (1964) method appears to suffer from the same bias.

Standard deviation of ϕ

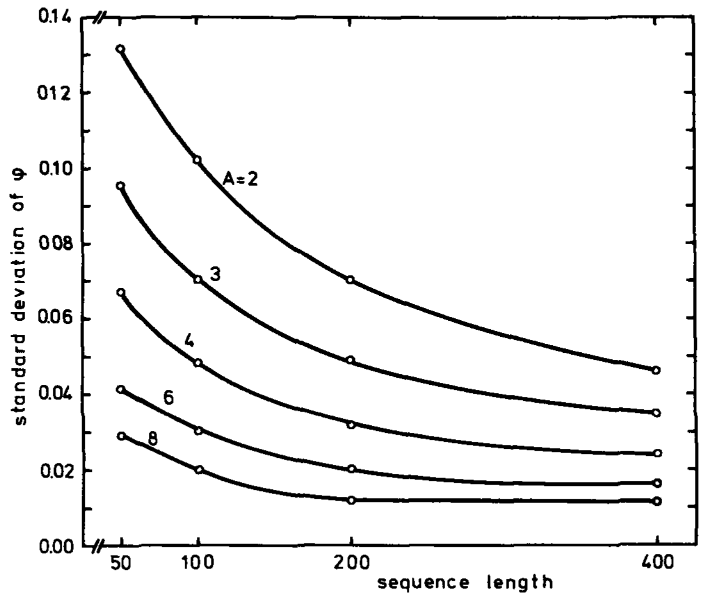


Fig. 2.3. Standard deviation of ϕ in 200 random sequences for each combination of A and N.

The sequence length, N, has also an effect on the standard deviation of ϕ. Fig. 2.3 shows σ_ϕ for first order ϕ-coefficients as a function of N and A. Once again these data are based on samples of 200 random sequences for all combinations of A and N. The standard deviations for higher order ϕ-coefficients do not differ markedly. The graph simply illustrates that a given difference between expected and observed proportions of repetitive and alternating n-grams becomes less probable (or more significant) with an increase of N.

Independence of successive orders of analysis

The values of ϕ of successive orders of analysis are independent, or uncorrelated, as is shown in Table 2.4 for the case A = 2 and N = 100.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
ϕ_2	-0.059				
ϕ_3	0.012	0.035			
ϕ_4	0.062	0.027	-0.061		
ϕ_5	-0.008	0.056	-0.037	0.064	
ϕ_6	0.069	-0.112	0.022	-0.016	0.037

Table 2.4. Correlations between ϕ-coefficients of six orders of analysis, based on 225 sequences with A = 2 and N = 100. Only correlations exceeding ± 0.131 are significant at 5%.

§5. NORMALIZATION OF ϕ

In the previous section it is shown that ϕ fulfils the first three requirements stated in section 2. The fourth and fifth requirement, however, are not satisfied. These requirements state that the measure of non-randomness should permit comparison of series of different length and with different numbers of alternatives, and that the measure should be defined as an interval or ratio scale.

Apart from the very small effect of sequence length on the bias of the zero point (§4) it can be stated that ϕ is independent of N: combining x identical sequences of 100 elements (which means increasing N without changing the proportions of observed and predicted repetitions and alternations), will not result in a change of ϕ. The data in Fig. 2.3 only imply that the significance of the non-randomness will increase with N. Therefore it is justified to compare ϕ-coefficients based on different sequence length.

On the other hand, the dependency of ϕ on the number of alternatives (A) is a problem that is difficult to overcome. The solution proposed is to express non-randomness as a percentile. For this purpose the distribution of ϕ in random sequences with N = 100 and A = 2, 3, 4, 6 and 8 were approximated with a Monte Carlo procedure. Each distribution was based on 500 to 1000 random sequences[8]. By relating an experimentally obtained ϕ-coefficient (based on any value of N, $N \geq 100$) to the accessory distribution, a percentile value P_ϕ can be found where $0 \leq P_\phi \leq 100$. The normalized value of ϕ, ϕ_n, is calculated by

$$\phi_n = 2P_\phi - 100 \qquad (2.3)$$

Thus, a measure is obtained which ranges from -100 for negative recency to +100 for positive recency, while absence of recency is represented by a zero score. The value of ϕ_n can be considered as the supplement of the one-tailed significance: if $\phi_n = -95$, the one-tailed significance of the non-randomness (when occurring in a sequence with N = 100) is 5%.

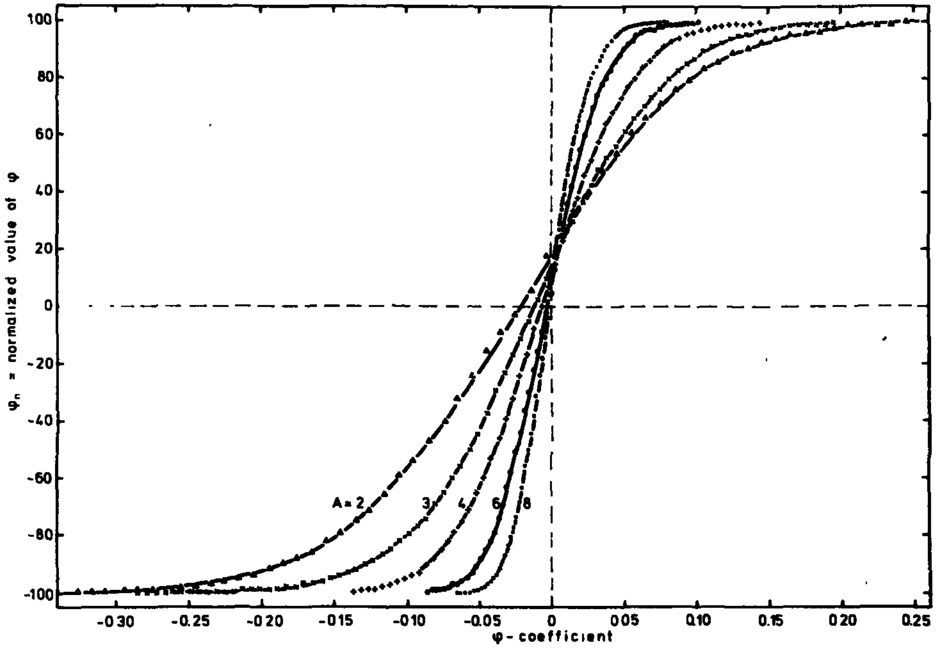


Fig. 2.4. Relation between ϕ and ϕ_n, for different values of A, based on the cumulative distribution of ϕ in 500 - 1000 random sequences with N = 100, averaged over six orders of analysis.

The relation between ϕ and ϕ_n is presented in Fig. 2.4. The curves are based on cumulative distributions averaged over six orders of analysis. Although the differences among orders of analysis are rather small, in practice separate curves for each order may be used. When ϕ is averaged over six orders of analysis its distribution is approximately normal. Therefore, the conversion of ϕ into ϕ_n can be done by expressing ϕ as a standard score by

$$z = m\phi + k \tag{2.4}$$

where m and k depend on A as shown in Table 2.5.

A	m	k
2	10.400	0.270
3	14.094	0.177
4	19.850	0.135
6	32.929	0.084
8	45.250	0.065

Table 2.5. Values of the coefficients m and k in formula (2.4), for different values of A.

The relation between A, ϕ and z is approximated by

$$z = \phi(57.27\ e^{0.072\ A} - 56.27) + 0.61\ e^{-0.522\ A} + 0.06 \tag{2.5}$$

With the use of the table of the normal distribution, z is converted into an area under the normal distribution, which is equal to $P_\phi/100$. Normalization with the formulas presented yields only small errors (< 5%).

An example will illustrate the different possibilities for converting ϕ into ϕ_n. In a sequence with N = 100 and A = 4 the third order ϕ is found to be -0.04.

1. Comparison with the distribution of third-order ϕ-coefficients in the case A = 4 (not presented in this study) gives $\phi_n = -52$.
2. Relating ϕ to the distribution averaged over six orders of analysis (Fig. 2.4) gives $\phi_n = -50$.
3. Using formula (2.4) with the coefficients of Table 2.5 one finds z =

-0.66, $P_\phi = 25.46$, $\phi_n = -49.08$.

4. Using formula (2.5) one finds $z = -.67$, $P_\phi = 25.14$, $\phi_n = -49.72$.

ϕ_n lends itself to comparing sequences with different values of A. At the same time normalization of ϕ solves the problem of defining the scale: ϕ_n, being essentially expressed as a probability, is defined on a ratio scale. This permits the operations of addition, subtraction, multiplication and division.

It is appropriate to mention that the normalization of ϕ with the use of experimental, non-analytic distributions is not a very elegant procedure. An attempt to find the analytic solution, however, would overshoot the mark since the accuracy of the approximated curves is high enough for experimental use, and the theoretical distributions (which are only weakly related to χ^2) have no wider fundamental importance.

§6. CONCLUSION

To the knowledge of the present author ϕ_n is the only measure that fulfils all requirements stated in section 2. The advantages of ϕ_n centre mainly on the ability to detect high order effects in *short* sequences. A disadvantage lies in the possibility of overlooking some effects of non-randomness. Another shortcoming is the not too elegant conversion of ϕ into ϕ_n.

CHAPTER 3

NATURE AND EXTENT OF SEQUENTIAL RESPONSE BIAS

§1. INTRODUCTION

In the present chapter an introductory experiment on randomization by humans is reported. The experiment was designed for three purposes. First, a comparison of several methods for the measurement of sequential response bias was felt to be profitable. Second, the experiment was intended as a general orientation towards the nature and extent of sequential response bias. In this context the relevant topics are the identification of various systematic sequential trends, the maximal order of sequential dependencies and the degree of predictability of attempted random sequences. Finally, the experiment was more specifically designed for the study of a very evident controversy: the influence of rate of production.

§2. RATE OF PRODUCTION

Mittenecker (1953) was the first to establish the effect of rate of production on sequential response bias. He demonstrated the existence of a deviating concept of randomness in a judgement experiment and showed that in randomization experiments subjects come closer to their individual concept when they are allowed to spend more time at the task. Mittenecker explained the result that sequential response bias is inversely related to rate of production by asserting that the actualization of a concept needs deliberation, and deliberation takes time.

Baddeley (1962, 1966) worked with a different theory. He hypothesized that man has a limited capacity for generating information: the total amount of information generated by randomization in a certain period of time is constant. If a pacing signal forces the subject into a higher rate of production he should generate sequences that are more redundant, as he is not able to produce more information per time unit. This investigator found indeed that increase of rate of production goes together with increase of sequential response bias, which is completely contradictory to Mittenecker's results!

When Teraoka (1963) and Warren & Morin (1965) also conducted experiments with a variable rate of production, their results failed to confirm Baddeley's hypothesis. Teraoka found, like Mittenecker, decrease of non-randomness with increasing response rate, whereas Warren & Morin found again an increase of non-randomness in the same condition, but the increment was not large enough to keep the amount of information generated per time unit at a constant level. The confusion seems to be at a maximum!

Part of the contradiction in the results may be attributed to the important differences among the various experimental procedures. Another part may be due to the different measures of sequential response bias

Author(s)	Alternatives	Sequence length	Rate of production (sec/response)	Order of analysis	Sequential response bias, as a function of increasing rate of production
Baddeley 1962, 1966	26 letters	2 series of 100 elements	0.5, 1, 2, 4 (paced)	0; 1	increases
Mittenecker 1953	10 digits	?	±0.6 (as fast as possible) ?(slow, unpaced)	mixed, recurrence distribution	decreases
Teraoka 1963	5 digits	1 series of 751 elements	0.5 (unpaced) 1.0 (paced)	0; 1; 1 - 4	decreases
Warren & Morin 1965	2, 4, 8 digits	6 series of 500 elements	0.25; 0.5; 0.75 (paced)	0 - 3	increases

Table 3.1. Some relevant data about four experiments on the influence of rate of production on sequential response bias.

used. Some relevant data about the experiments are presented in Table 3.1 (for detailed information see Tables 1.1, 1.2 and 1.3, Chapter 1). The hope was that our own experiment would yield more conclusive results.

§3. EXPERIMENT 1[9]

Procedure

The subjects were eight paid undergraduate students. The subject was seated in front of the appropriate part of a panel with 2, 3, 4, 6 or 8 push buttons (see Fig. 3.1). The task was to press the buttons in random

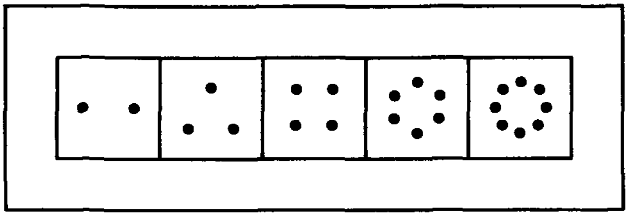


Fig. 3.1. Experiment 1.
Keyboard with 2, 3, 4, 6 and 8 push buttons.

fashion at a metronomic rate of 0.5, 1, 2 or 4 sec. per response. Only one finger of one hand was used. In each of 20 conditions (5 sets of alternatives x 4 rates of response) each subject produced 106 responses. The total number of responses was 8 (subjects) x 5 (number of alternatives) x 4 (rate of production) x 106 (length of sequence) = 16.960.

All conditions were presented randomly over three consecutive days. The responses were recorded on punch tape to facilitate analysis by a PDP-7 computer. In order to compare the various methods described in Chapter 1, sequential response bias will be expressed by five different measures:

1. ϕ_n-scores.
2. Information per response and relative redundancy.
3. Run-distributions.
4. Recurrence distributions.
5. Sampling distributions.

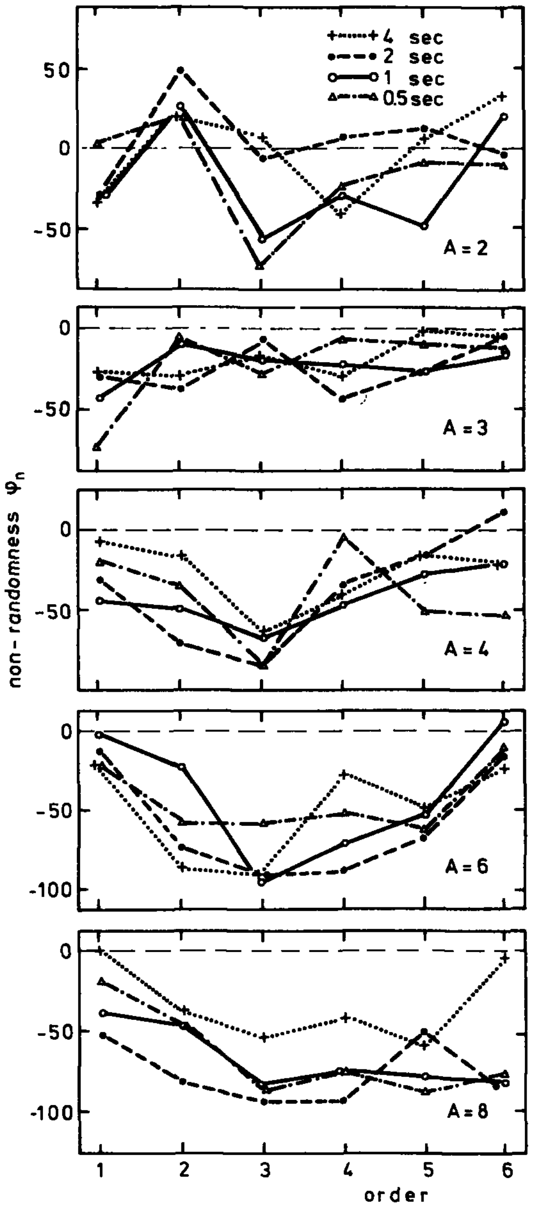


Fig. 3.2. Experiment 1. Graphical representation of ϕ_n-scores, as a function of order of analysis, separately for A = 2, 3, 4, 6 and 8 alternatives and for rate of production is 4, 2, 1 and 0.5 sec. per response.

Results

1. A graphic representation of ϕ_n scores is given in Fig. 3.2. An analysis of variance of these data revealed that negative recency in-

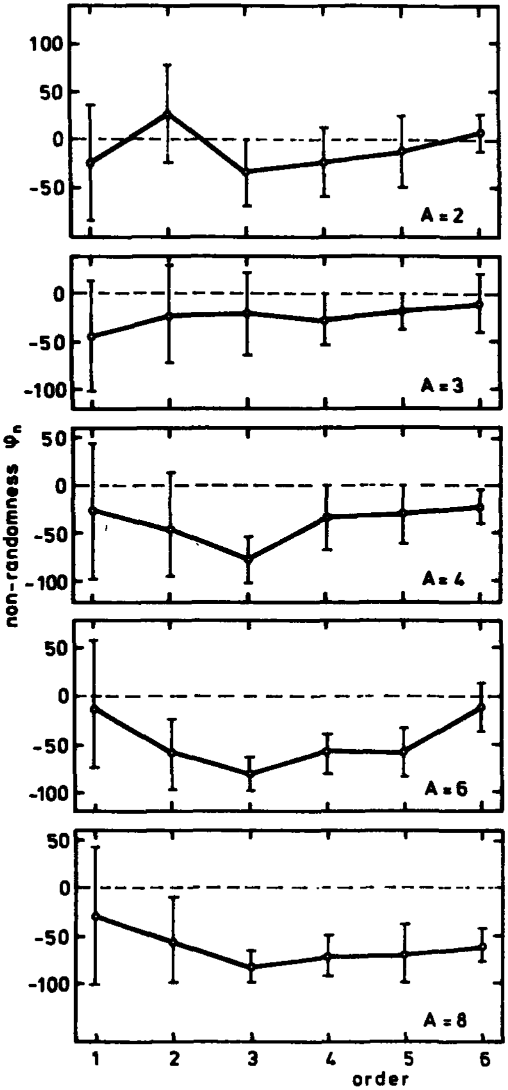


Fig. 3.3. Experiment 1. ϕ_n-scores as a function of order of analysis, separately for A = 2, 3, 4, 6 and 8 alternatives, averaged over various rates of production and subjects. The vertical lines indicate standard deviations among subjects

creased with the number of alternatives ($F^{***}_{4,28}$ = 21.49). The effect of rate of production was not significant ($F_{3,21}$ = 1.89). In Fig. 3.3 the data are plotted with standard deviations between subjects. The differences among these five curves suggest an interaction between orders and alternatives. The analysis of variance supported this observation ($F^{**}_{20,140}$ = 3.26). This interaction is mainly due to the fact that conditions with many alternatives show stronger high order effects. The differences among subjects are large ($F^{***}_{7,420}$ = 18.11) but there is a strong interaction with order of analysis ($F^{***}_{35,420}$ = 6.30): the individual differences are large only for first and second order effects.

2. Information per response and relative redundancy can be calculted only at the zero and first order levels of analysis, provided that the sequences are combined irrespective of rate of production (see Table 3.2). The neglect of differences among rates of production seems to be justified, according to the ϕ_n-scores.

	A = 2	3	4	6	8
Information per response (bits)					
zero order of dep.	1.00	1.58	2.00	2.56	2.97
first order of dep.	0.98	1.50	1.93	2.47	2.86
Relative redundancy (%)					
zero order of dep.	0.0	0.0	0.1	1.0	1.0
first order of dep.	1.8	5.0	3.6	4.6	4.6

Table 3.2. Information per response and relative redundancy in the response sequences of Experiment I for zero and first order of dependency.

Runlength	A = 2	3	4	6	8
1	1049 (864)	2002 (1522)	2105 (1920)	2490 (2364)	2776 (2604)
2	379 (428)	376 (503)	367 (475)	294 (390)	205 (322)
3	200 (212)	112 (166)	109 (118)	81 (64)	46 (40)
4	121 (105)	45 (55)	35 (29)	11 (11)	9 (5)
5	50 (52)	18 (18)	12 (7)	4 (2)	4 (1)

Table 3.3. Runlength distributions in the response sequences of Experiment I. Bracketed figures are theoretical values.

3. The distributions of runs in the data averaged over subjects and the various rates of production, are presented in Table 3.3. (For a definition of runs see §3, Chapter 2.) For N $\geq$ 100 theoretical run distributions for the individual sequences with length N and consisting of A alternatives are very well approximated by the formula

$$f(x) = N.\left[\frac{A-1}{A}\right]^2 \left[\frac{1}{A}\right]^{x-1}, \tag{3.1}$$

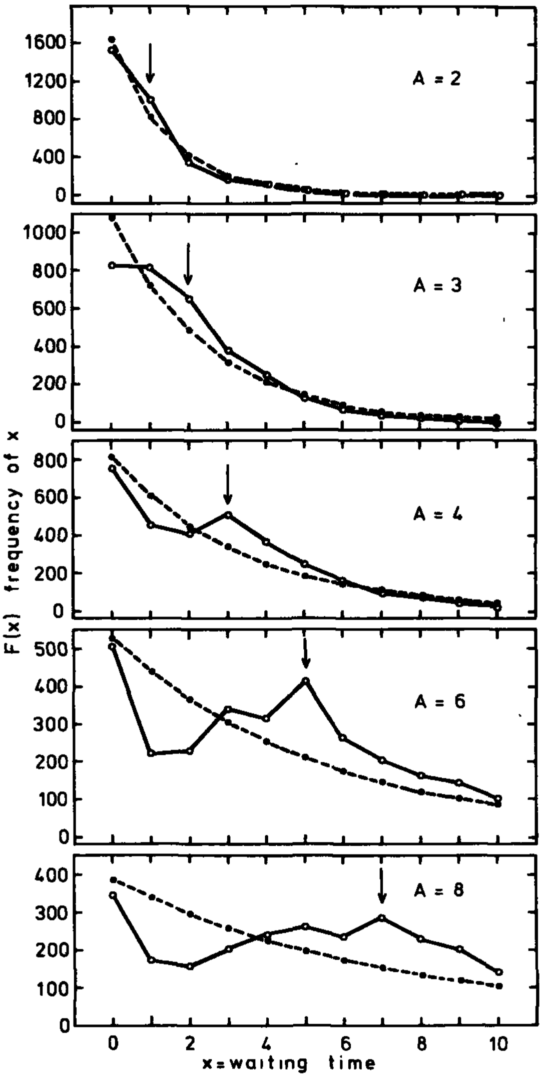


Fig. 3.4. Experiment 1. Recurrence distributions for A = 2, 3, 4, 6 and 8 alternatives, averaged over various rates of production and subjects. Arrows indicate the maximum difference between theoretical (dotted lines) and observed distributions.

where f(x) is the frequency with which a runlength x occurs. For the exact formula see Mood (1940) and Wilks (1947).

4. Recurrence distributions for the results averaged over subjects and the various rates of production are presented in Fig. 3.4. For $N \geq 100$ theoretical curves for the individual sequences with length N and consisting of A alternatives are very well approximated by the formula

$$f(x) = \left[N - A\right] \left[\frac{A-1}{A}\right]^x \left[\frac{1}{A}\right], \qquad (3.2)$$

where f(x) is the frequency of a gap with length x between two identical responses (shortest gap = 0).

5. Sampling distributions for the results averaged over subjects and the various rates of production are presented in Fig. 3.5. A sampling distribution is obtained by dividing the sequence into equal parts, in the present case each containing ten responses. The frequencies

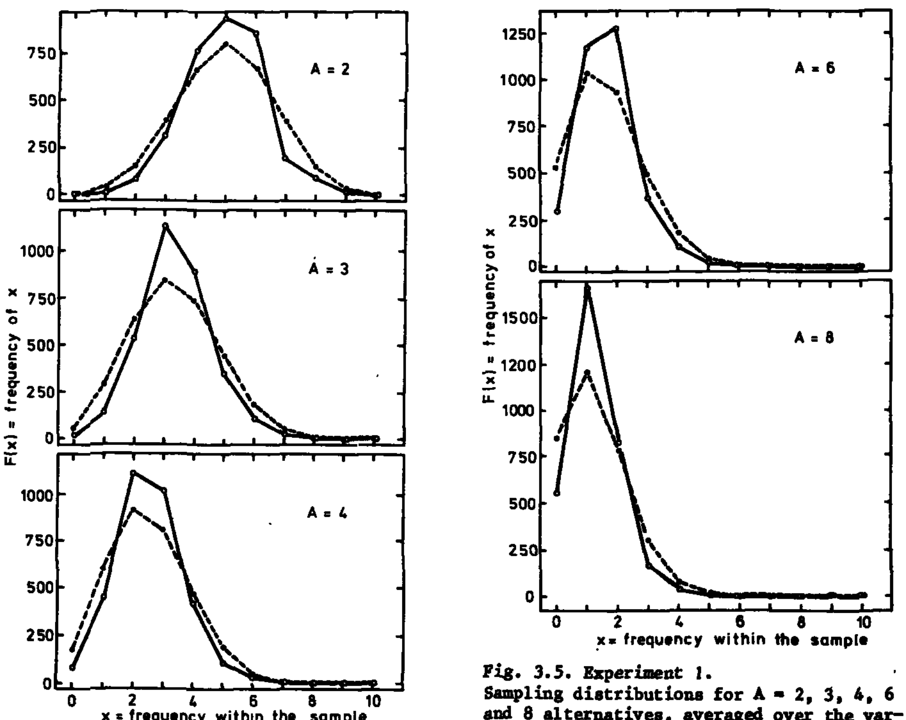


Fig. 3.5. Experiment 1.
Sampling distributions for A = 2, 3, 4, 6 and 8 alternatives, averaged over the various rates of production and subjects. Dotted lines: theoretical distributions.

with which the alternatives occur within these samples are tabulated. The tabulated distributions are called "sampling distributions". The theoretical distributions for sequences with A alternatives and divided into m samples each containing n responses are binomial distributions given by the formula

$$f(x) = \left[m.n.\right] \binom{n}{x} \left[\frac{1}{A}\right]^{x} \left[\frac{A-1}{A}\right]^{n-x} \qquad (3.3)$$

where f(x) is the frequency with which an alternative occurs x times in a sample ($o \leq x \leq n$)

§4. THE NATURE OF SEQUENTIAL RESPONSE BIAS

The results of Experiment 1 clearly point to negative recency, or too many short runs. The number of negative ϕ_n-scores was by far in the majority, for all values of A (see Table 3.4, p. 42).

The same tendency is demonstrated by the runlength distributions: runs with length = 1 occurred too often, and longer runs were relatively scarce (Table 3.3). In the following parts of this study the effect of negative recency will not be stressed again, only exceptions to the rule will be notified.

A supplementary datum is provided by the recurrence distributions and the sampling distributions. The recurrence distributions show that subjects tended to produce a gap of A - 1 places between two identical responses. Formulated in other words, they tended to follow a cycle with length A. This effect was reported before by Mittenecker (1953). He suggested that a "subjective random sequence" would consist of successive parts with length A, in which each alternative occurs just once. In terms of an urn model this would mean that successive responses are drawn from an urn which contains one specimen of each alternative, without replacement. When, after A selections, the urn is empty, it is filled again, and drawing continues. In this formulation the model is too rigorous, but it describes a tendency which is undeniably present.

A similar tendency is shown by the sampling distributions (Fig. 3.5) which are too much peaked for all values of A. Small samples are too representative and, conversely, frequencies are balanced in too small samples. This finding is in contradiction with data of Peterson, Du Charme & Edwards (1968) and Wheeler & Beach (1968). These investigators asked their subjects to express by a number the subjective probability that x red chips would occur in n draws, provided that the overall probability of red chips is p. When x goes from zero to n, a complete subjective sampling distribution is obtained. These distributions were too *flat* in all instances. Intuitively it is felt, however, that the present approach gives more direct information on man's ideas about chance processes, whereas the technique of Peterson *et. al.* contains the indirect and questionable step of expressing subjective probabilities by numbers. An extensive discussion on subjective sampling distributions is presented by Wagenaar (1971b).

More specifically the results showed no difference among the various rates of response. Hence it is difficult to support the idea that there should exist a limited capacity for generating information: the amount of information generated was easily increased with a factor 8. The discussion on the influence of rate of production is continued in the next chapter. The finding that sequential response bias increases with an increasing number of alternatives was reported before by Baddeley (1962, 1966), Rath (1966) and Warren & Morin (1965). The new element in the present results is the finding that the functional relation between response bias and order of analysis is so different for various values of A. Another important new piece of evidence, namely the large individual differences for first and second order sequential response bias, will be treated extensively in Chapter 6.

§5. THE EXTENT OF SEQUENTIAL RESPONSE BIAS

Since dependencies are necessarily based on partial or complete recall of previous responses, a psychologically relevant question concerns the longest distance over which sequential dependencies occur. The rôle of memory will be discussed later on into more detail. At the moment we will confine ourselves to the determination of the longest distance of dependency.

The ϕ_n-scores showed significant deviations up to, and including sixth order effects (see Table 3.4) even when tested with the rather conservative sign test. The urn model of Mittenecker predicts that the maximal distance of dependency is a linear function of A. Indeed Table

				Order of analysis				
A	1	2	3	4	5	6	Total	z
2	23*	7**	22	23*	19	16	110	1.98
3	23*	19	21	22	22	19	126	4.24
4	21	24**	31**	23*	23*	21	143	6.65
6	19	26**	31**	30**	29**	19	154	8.20
8	21	27**	30**	30**	30**	27*	165	9.76

Table 3.4. Negative ϕ_n-scores in Experiment 1. Maximum value per cell in column 2-7: 32. Maximum value in column 8: 192.* p (sign test) $\leq$ 0.05;** p (sign test) $\leq$ 0.01.

3.4 suggests such an effect, but possibly this tendency is caused by the overall influence of the number of alternatives. Therefore a more sensitive method for the determination of the maximal distance of dependency is needed.

The method proposed here uses sampling distributions, as shown in Fig. 3.5, as a starting point. If the sample size (n) is equal to or

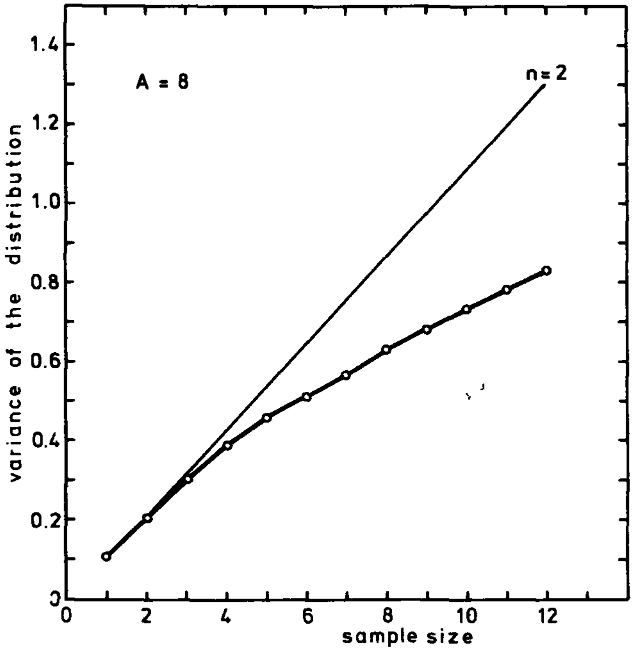


Fig. 3.6. Experiment 1. Demonstration of the extrapolation technique (see text).

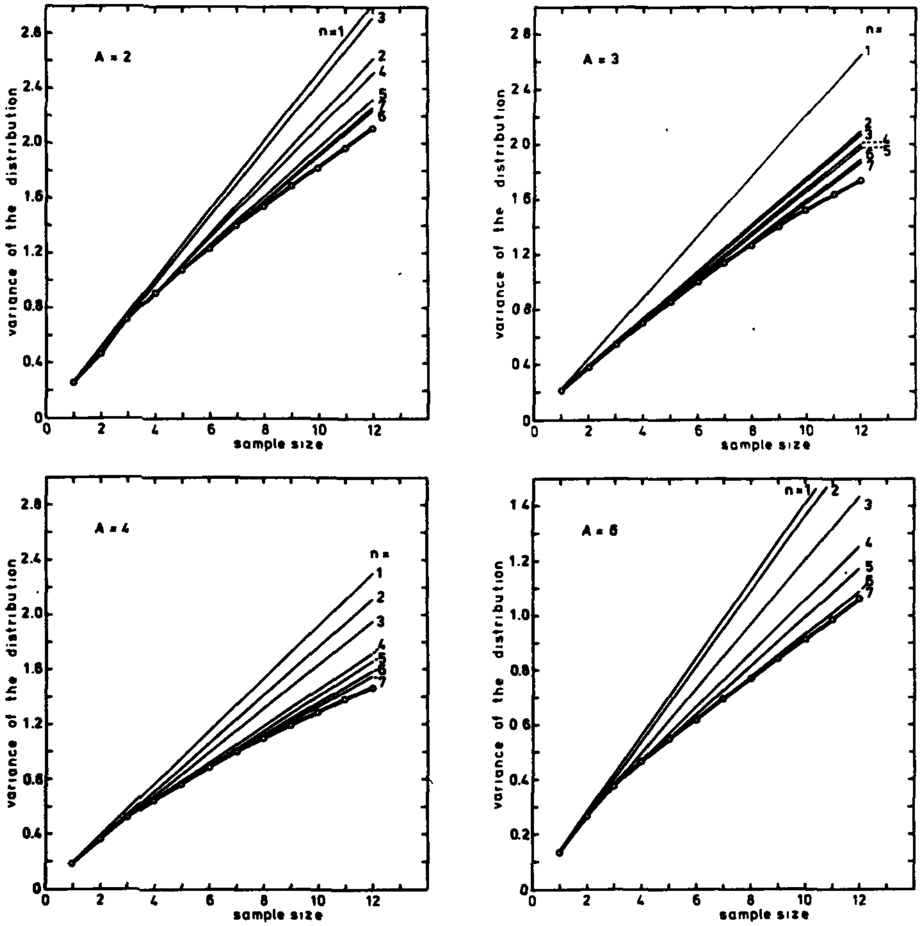
larger than the maximal distance of dependency, it is possible to predict the sampling distribution that will be obtained if sample size is increased to n+1. When the prediction is not confirmed, it should be concluded that the maximal distance of dependency was at least n+1 instead of n.

The computational procedure is based on the prediction of frequencies of (n+1)-grams starting from frequencies of n-grams, with use of formula 2.1. A simple example may clarify the procedure. If a sequence with A=8 is divided into samples with two elements, and there is only a first order effect of non-randomness, it is possible to predict the sampling distribution that might be found when the sequence was sampled with sample size three, four, five and so on. To test the correctness of the predictions, variances of the observed and predicted sampling distribution are compared. In Fig. 3.6 an example is presented graphically for the eight-alternative sequences of Experiment 1. Sampling distributions have been calculated, with sample size n varying from 1 to 12. With n=2 the variance was 0.21. Extrapolation with formula 2.1 delivers distributions with larger variances than the observed distributions actually had. Hence the maximal distance of dependency was longer than 1.

Complete data of this kind are presented in Fig. 3.7. The effect that frequencies were balanced in too small samples (see the preceding section) can be specified by asserting that the sample size was seven or slightly above seven responses, relatively independent of the value of A.

Another important quantitative aspect of sequential response bias concerns the question what proportion of the responses is predictable when, for instance, six previous responses are taken into account. The seventh order redundancy, which can be reliably calculated only for the case A=2, gives some indication in this respect. When all two-alternative sequences (3392 responses) are taken together the seventh order redundancy amounts to 3.5%, which places the odds at 61 to 39 for optimal prediction using information of six previous responses.

Warren & Morin (1965) calculated fourth order redundancy for A=4 and A=8. It is possible to estimate predictability on the basis of their data, corrected with the Miller-Madow technique (Miller, 1955). For two subjects the predictability is 38% and 38% with A=4, 22% and 26% with A=8.



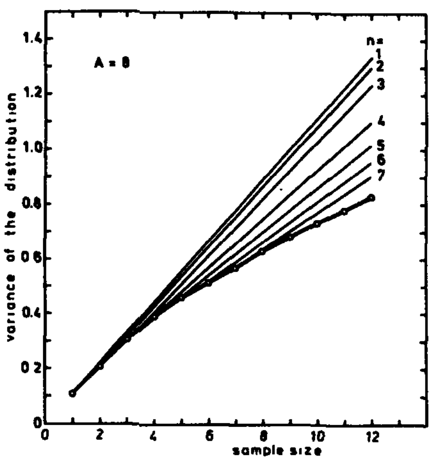


Fig. 3.7. Experiment 1.
Results of extrapolation of the sampling distributions.

§6. CONCLUSIONS

The general conclusions of the first experiment are:

1. In a randomization experiment subjects show sequential response bias. The bias is in the direction of negative recency.
2. Subjects tend to follow a cycle with a length equal to the number of alternatives.
3. Subjects tend to balance frequencies within too small a sample.
4. The longest distance of dependency is about seven, more or less irrespective of the number of alternatives.

The more specific conclusions are:

1. Sequential response bias neither increases or decreases with an increasing rate of production. Hence, the hypothesis that man acts like a generator with a limited capacity for generating information is not confirmed.
2. Sequential response bias increases with the number of alternatives. There is a strong interaction between number of alternatives and the order of analysis.
3. Individual differences are limited to first and second order effects.

CHAPTER 4

SEQUENTIAL RESPONSE BIAS AS A FUNCTION OF VARIOUS EXPERIMENTAL CONDITIONS

§1. INTRODUCTION

Following the first exploration of sequential response bias as presented in the previous chapter, it will be tried to deal with the subject matter in a systematic way. The strategy of the present chapter is to present experimental verification of the predictions made by the various theories mentioned in Chapter 1.

The main distinction among theories on sequential response bias is between theories hypothesizing a subjective concept of randomness on the one hand, and theories that attribute the effect to limitations of a psychological function on the other hand. Baddeley (1966) stated that it is highly unlikely that a subject "should vary his *concept* of randomness with the rate at which he is required to randomize". Actually he suggested that the concept theory would predict sequential response bias to be completely independent of external experimental conditions. Hence, the manipulation of these conditions might yield useful data to decide upon the two lines of thought. However, Mittenecker (1953) argued that the actualization of a "concept of randomness" may be inhibited or facilitated by the experimental conditions, which implies that a possible relation between the extent of sequential response bias and external conditions cannot be taken as lack of support for the concept theory. The discussion does not appear to be very fruitful unless a method for the measurement of a "concept of randomness" is defined. This problem will be discussed in Chapter 5. In the present chapter only those theories are evaluated that attribute sequential response bias to a functional limitation.

§2. FIVE THEORIES

Tune (1964b) attributed the inability to randomize to the limited capacity of human memory. In order to produce a random sequence, the subject "must be able to remember, or at least have available for use and reference, a tally of the times that each of the n-grams has been used"....."It is likely, therefore, that conditions which militate against these abilities will tend to increase the non-randomness in selections". More specifically Tune related non-randomness to the limitations of short-term memory. It is tempting, indeed, to relate the longest distance of sequential dependency (about seven places) to the traditional size of the short-term memory span of seven items, plus or minus two (Miller, 1956). The general idea that randomization should be facilitated by a good memorization of previous responses will be called *positive memory theory.*

The positive memory theory was contested by Baddeley (1966) who argued that precisely the fact that the subject uses his memory explains why he is inferior to a random number table. Sequential dependencies can occur only when previous responses have been stored in some way, and all factors enhancing the functioning of memory would deteriorate randomization. This point of view will be called *negative memory theory.*

At this point it should be stressed that both memory theories are rather vague.

1. Tune seems to refer to long-term, rather than short-term memory, when he speaks about a tally of frequencies, whereas the inability to memorize more than the seven last responses is connected to short-term memory.
2. Positive memory theory predicts that sequential dependencies reach a maximum at distances longer than six places, since especially the responses that are farther away have not been stored.
3. Even when both memory theories suppose a relation between *short-term* memory and sequential response bias, it will be difficult to use experimental data about this function, since the memory span seems to be highly dependent on the experimental paradigm by which it is meas-

ured. The "magical number seven" holds for a span paradigm in which a limited string of symbols is presented and recalled. It holds also for a probe paradigm in which after presentation of the stimuli it is indicated by a probe which symbol should be recalled. But in a continuous memory paradigm the span is largely decreased to about three items. In this paradigm a continuing sequence of items is presented in stead of a relatively short block of known length. Each item is reproduced after a number of new items have passed (Mackworth , 1959; Pollack & Johnson, 1963; Sanders & van Borselen, 1966). Finally in a missing-item paradigm the span seems to be increased to ten items or more. This result was obtained by Buschke (1963) who presented strings of digits with the request to indicate which symbols out of a limited set were missing.
At this moment it is unclear which of these or other paradigms is most related to the randomization experiment.

4. Finally, it should be noted that in randomization experiments subjects are never instructed to memorize the response sequence

The four considerations mentioned above make predictions about human randomization on the basis of experimental evidence on short-term memory very questionable. As a consequence, the predictions stated in the present chapter are weak.

As mentioned before, Baddeley (1962, 1966) proposed a third theory in which man has a limited capacity for the generation of information. When information per second, to be produced by randomization, surpasses a certain level, man is forced to introduce some redundancy into the sequence. Therefore, any experimental condition increasing the demanded information per unit time increases sequential response bias. This theory will be called *limited capacity theory*.

Obviously, Baddeley's theory is also rather vague. It remains unclear why such a limitation would exist, why a certain type of redundancy occurs when the system is overloaded, and what the level of the capacity limit actually is. Taken strictly the theory would lead to the absurd conclusion that the limit is below 0.25 bits/sec., since subjects cannot even generate a binary random sequence at a rate of one response in four seconds.

Another contestor of positive memory theory was Weiss (1965). He argued that the problems of storing frequency tallies are so overwhelming that "the use of memory devices may lead to even less success than is demonstrably true for most subjects". Instead he proposed the so-called "Attention-Distraction Hypothesis". On the one hand in order to effectively maintain a random sequence, "one must suppress, or otherwise inhibit paying attention to each response after it occurs". Hence ... "a certain level of distraction is required to ensure maximal uncertainty between successive responses". This theory will be called *negative attention theory*. In addition, Weiss stated that the task requires the "ability to maintain a set for randomness in the face of competition from almost every other source, since any other source of stimulation is likely to be more diverting". This idea is essentially identical with Weiss' previous conjecture (1964) that randomization performance is related to the subjects' arousal level. As mentioned in Chapter 1, the combination of this *positive attention theory* with a negative attention theory creates a peculiar kind of circularity. In a sense the subjects are required to pay attention in order not to pay attention. When they pay more attention (positive) they will pay less attention (negative) and the reverse. Any result may be explained by this combination since any factor favouring distraction (Weiss listed sleep deprivation, drugs and schizophrenia) would decrease both attention for previous responses and attention to the task. It is unpredictable what the result would be. In order to evade this difficulty, the two attention theories will be treated separately.

The attention theories are again not very strict as they refer to the notion of attention without any indication about how to measure attention in this situation. The general confusion created by the notion of attention in the past century originated partly because of the lack of an adequate operational definition (*e.g.* Sanders, 1967). Consequently, it will be difficult to predict the effect of any experimental condition on the basis of either attention theory.

The short analysis of the five theories relating sequential response bias to a functional limitation leads to the conclusion that the theories are vague and their predictions weak. Nevertheless, because these theories are the best there are, the present chapter attempts to compare the predictions with experimental results anyhow. This is done mainly to provide a framework in which the next series of experiments can be placed.

Theory \ Experimental conditions	Increase of production rate	Increase of number of alternatives	Increase of artificial memory	Vocalization	Increase of task duration and monotony	Addition of secondary task
Positive memory theory	letters and digits: ? spatial positions: 0	letters and digits: 0 spatial positions: -	+	+	-	-
Negative memory theory	letters and digits: ? spatial positions: 0	letters and digits: 0 spatial positions: +	-	-	+	+
Limited capacity theory	-	-	0	0	0 or -	-
Positive attention theory	+	+	?	?	-	-
Negative attention theory	-	-	-	-	+	+

Table 4.1. Predictions concerning the influence of several experimental conditions made by five theories on randomization.
+ = better randomization
− = worse randomization
0 = no effect
? = no prediction possible

In Table 4.1, the effects of several experimental conditions as presumably predicted by the five theories are presented schematically. A detailed explanation will be given in the next sections.

§3. GENERAL EXPERIMENTAL PROCEDURE

For experiment 2 to 7 the following general procedures were followed, unless mentioned otherwise.

Subjects were paid undergraduate students. The task was to press push-buttons in random fashion at a pace set by a metronome. Only one finger of one hand was used[10]. The order of experimental conditions was varied systematically over subjects and the responses were recorded on punch tape. The design was always complete factorial which made the analyses of variance quite uniform. Therefore, only the final F-values of the analyses are presented, accomapnied with the signs *,**, or ***, respectively for $p < 0.05$, $p < 0.01$ and $p \ll 0.01$. F-values without asterisks are not significant.

§4. AGAIN: RATE OF PRODUCTION

In the previous chapter it was shown that the experimental evidence concerning the influence of rate of production on randomization is contradictory. So are the theories, witness Table 4.1.

Supporters of either memory theory would argue that the experiments listed in Table 3.1 are related to auditory short-term memory, since all experimenters asked their subjects to call out the response sequence. Unfortunately, the literature is not entirely conclusive about the influence of presentation rate on auditory short-term memory. In general, it is reported that more items are recalled when presentation rate is increased and rehearsal is prevented (Posner, 1963) except at very high rates (Yntema, Wozencraft & Klem, 1964). However, when the items are not recalled in temporal order, the effect disappears or even reverses (Posner, 1964). In the randomization situation it is not clear how recall of previous items is supposed to occur: ordered temporally, ordered reversely, or not ordered at all. Hence, it is impossible to predict, both for positive and negative memory theory, what influence the increase of production rate could have. For the situation of Experiment 1 the most relevant data are those gathered by Sanders (1968) about short-term memory for spatial positions. This author showed that the number of recalled spatial positions is independent of presentation rate. Hence, both positive and negative memory theory would predict that rate of production does not change randomization performance when the alternatives are spatial positions which in fact was confirmed by Experiment 1.

Increase of production rate is liable to increase the general level of arousal (Baddeley, 1966). Therefore, negative attention theory might predict a deterioration of randomization at higher rates, as arousal would intensify the attention for previous responses. Positive attention theory might predict just the reverse, since a high level of arousal would increase the attention for the randomization task.

Finally, limited capacity theory predicts an adverse effect of rate of production, as indeed was found by Baddeley (1962, 1966). Warren & Morin (1965), however, showed that the amount of information generated per time unit increases with rate of production, implying that a capacity theory is not applicable at the slower rates. Moreover, the latter authors reported that the effect of production rate depended on the number of alternatives to be randomized: for two alternatives the effect of production rate was small or absent, whereas for eight alternatives a considerable effect was found. In the light of this finding it is fitting that Baddeley found a large effect with 26 alternatives.

A possible explanation of the interaction between rate of production and number of alternatives is suggested by the fact that Baddeley and Warren & Morin used only an internal representation of the alternatives: the set was defined only verbally, by instruction. Hence, the subject's task was two-fold:

1. Activation of an internal "mental picture" of the set.
2. Random selection out of the activated (part of the) set.

In case of small choice-sets the difference between an internally or externally represented set may be negligible, but it is at least doubtful whether one may have 26 letters equally available during a whole session. In the latter case it is plausible that subjects used one small subset at a time, that they tried to make random selections only within the subset, and changed subsets occasionally. This hypothesis may be easily tested by repeating Baddeley's experiment with both internal and external represented sets of alternatives.

Experiment 2

Procedure

Nine subjects were required to call out random sequences of 100 letters paced by a timer clicking at one of two rates: one click per 1 or per 4 sec. The set of alternatives contained the letters of the alphabet, except the Z. The choice-set was presented either internally by instruction, or externally by visual presentation of the letters on a card. The total number of responses was

9 (subjects) x 2 (production rates) x 2 (presentation modes) x 4 (sequences) x 100 (responses) = 14.400

In a control condition five other subjects repeated the experiment with a set of four alternatives (the letters ABCD). Here the number of responses was

5 (subjects) x 2 (production rates) x 2 (presentation modes) x 4 (sequences) x 100 (responses) = 8000.

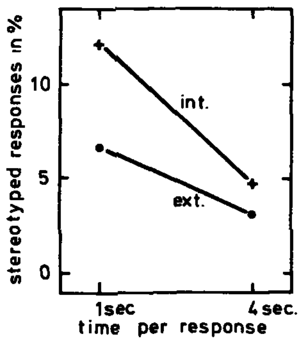


Fig. 4.1. Stereotype scores in the four conditions of Experiment 2.

Results and Discussion

Apart from the ϕ_n-scores one measure for non-randomness used by Baddeley, *viz.* the number of stereotyped responses (digrams containing letters in the alphabetical order) was calculated. Fig. 4.1 shows, first, that Baddeley's effect was reproduced nicely with internal representation of the choice-set. The effects of both presentation mode and production rate were significant ($F^{***}_{1,16} = 38.05$; $F^{***}_{1,16} = 41.44$). At the same time, however, the interaction between presentation mode and production rate indicated that the influence of production rate was much smaller when the set of alternatives was presented visually ($F^{**}_{1,8} = 11.45$). A *post hoc* Newman-Keuls analysis revealed the effect of production rate to be significant for the "internal" conditions only. The frequency of stereotyped responses was not far above the chance level (4%) in the external conditions. The composition of the digrams is pictured in Fig. 4.2. Almost all distances in the alphabetical order occurred with their expected frequencies, except the very short and the

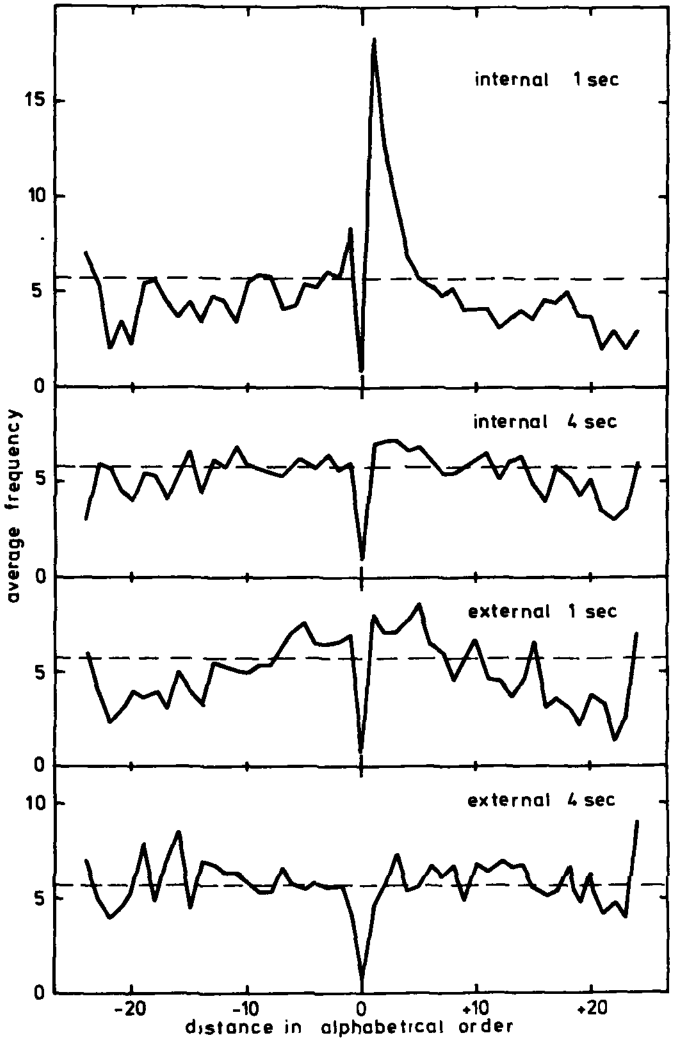


Fig. 4.2. Compositions of the digrams in the four conditions of Experiment 2.

very long ones. The distance +1 corresponds with Baddeley's stereotyped measure. High frequency of stereotyped responses went at the cost of the very long distances. More interesting, however, is the distance zero which occurred with too small a frequency, independently of presentation mode or production rate. The same result is revealed by the ϕ_n-scores (Fig. 4.3) which showed no effect of production rate ($F_{1,8} = 1.06$) and an effect of presentation mode ($F^{**}_{1,8} = 28.00$) which interacts signifi-

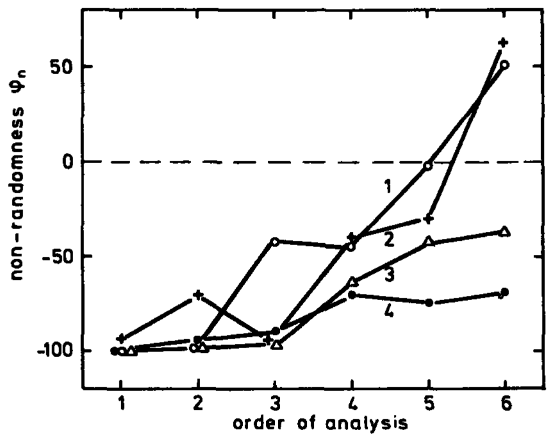


Fig. 4.3. ϕ_n-scores in the four conditions of Experiment 2.
1 = internal, 4 sec.; 2 = internal, 1 sec.;
3 = external, 4 sec.; 4 = external, 1 sec.

cantly with order of analysis ($F^{**}_{5,80} = 11.44$). The interactions between presentation mode and production rate, and between production rate and order of analysis were not significant ($F_{1,48} < 1$; $F_{5,40} = 1.16$). There was a tendency to repeat the items every six responses when the set of alternatives was presented internally. This suggests that the subset of the alphabet, activated at a time, had a size of six letters approximately.

In the control condition with four alternatives no effects of presentation mode and production rate were found (stereotype scores: $F_{1,8} < 1$, $F_{1,8} = 3.35$; ϕ_n-scores: $F_{1,8} < 1$, $F_{1,8} < 1$) nor an interaction between these two factors (stereotype scores: $F_{1,4} < 1$; ϕ_n-scores: $F_{1,4} < 1$). For this condition also first and second order redundancy scores were computed, which showed again no differences for either factor.

Conclusion

The effects of rate of production, reported by Baddeley (1962, 1966) and Warren & Morin (1965) are to be attributed to a deteriorating internal representation of large sets of alternatives with higher production rates. Though these effects are sound, they are not related to the randomization task proper. The remaining evidence points to no effect of increasing production rate, or very small effects in the direction of either increase or decrease of sequential response bias, even when the alternatives are digits or letters. This result conflicts with any of the five theories presented in §2 except the memory theories which did not permit a prediction at all for digits and letters, whereas they did predict the observed lack of effect for spatial positions.

§5. NUMBER OF ALTERNATIVES

According to either memory theory an increase of the number of alternatives should leave randomization performance unaffected as far as letters or digits are concerned, since Conrad & Hull (1964) found that the memory span is independent of the number of alternatives. For spatial positions, however, Sanders (1968) reported a clear relation between memory span and size of the set of alternatives. Therefore, in this case an effect of set size might be expected. It should be stressed, however, that the prediction is very weak, since Mackworth (1962) obtained an effect of response mode in span experiments: when the response is dictated the span is somewhat larger than in conditions where the response is given by means of push-buttons. Conrad & Hull (1964) had their subjects dictate responses, whereas in our randomization experiments push-buttons are used. It is not certain whether there will be no effect of number of alternatives in a span experiment where subjects respond on push-buttons. The limited capacity theory predicts a deterioration of randomization when the number of alternatives increases, since the amount of demanded information per token increases linearly with $^2\log A$. In contrast, positive attention theory predicts that randomization is facilitated by an increase of the number of alternatives, because of the accompanying increased level of arousal (Baddeley, 1966). For the same reason negative attention theory will predict again an adverse effect of increasing the set of alternatives. The various predictions are summarized in Table 4.1.

It has been reported rather consistently that increase of the number of alternatives makes the task more difficult. Baddeley (1962, 1966) and Rath (1966) found, with unpaced production, that rate of generation slows down when the set of alternatives increases. It is questionable, however, that the rate of generation in an unpaced condition is in any way comparable to some measure for non-randomness in paced conditions. *Warren & Morin (1965) showed that the relative redundancy increases* with increase of the choice-set. Because the authors used digits, this result seems to contradict either memory theory and positive attention theory. It should be noted, however, that Warren & Morin used only two subjects and internal representation of the choice-set which is known to interact with set size (see Experiment 2).

The results of Experiment 1 revealed a significant increase of sequential response bias as a function of the number of alternatives. An inspection of Figs. 3.3 and 3.7 learns that it is not the maximum distance of dependency that increases with A, but only the degree of dependency. This result, obtained with spatial positions as alternatives, is not incompatible with positive memory theory. The effect should disappear, however, when the push-buttons are marked with digits, which is verified in the next experiment.

Experiment 3

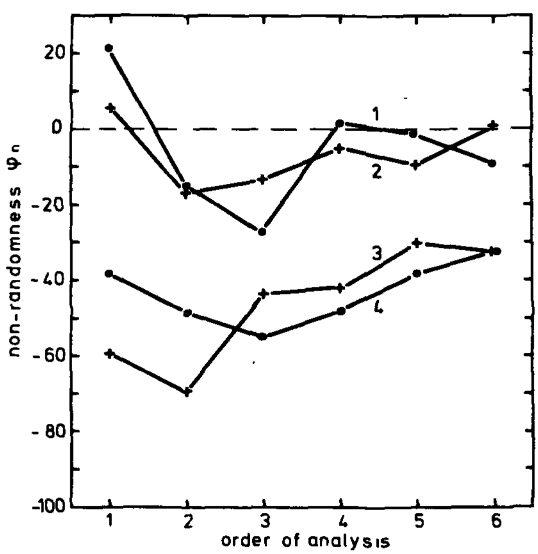


Fig. 4.4. ϕ_n-scores in the four conditions of Experiment 3.
1 = 2 alternatives, digits; 2 = 2 alternatives, spatial positions;
3 = 8 alternatives, digits; 4 = 8 alternatives, spatial positions.

Procedure

Eight subjects were required to push in random fashion on either two or eight buttons in a row, at a metronomic rate of 1 second per response. The buttons were white ("spatial positions") or marked with digits from 1 to 8 ("digits"). Each subject produced six sequences of 100 responses in each condition. The total number of responses was

8 (subjects) x 2 (sets of alternatives) x 2 (spatial positions or digits) x 6 (sequences) x 100 (responses) = 19.200

Results and Discussion

The ϕ_n-scores are displayed in Fig. 4.4. The effect of number of alternatives was confirmed again ($F^{***}_{1,42} = 75.81$) for all orders of analysis (interaction not significant: $F_{5,70} = 1.33$). The condition digits-spatial positions did not result in any significant difference ($F_{1,42} < 1$) nor in a significant interaction with number of alternatives ($F_{1,7} < 1$) or with order of analysis ($F_{5,35} = 1.20$).

Obviously, the effect of number of alternatives is independent of the alternatives being digits or spatial positions. There exists, however, the small possibility that the kinaesthetic factor inthe task was much more important for recall of previous responses, than the lay-out of the push-buttons. Actually this is not the case, as will be shown in §7.

Conclusion

Increasing the set of alternatives causes an increase of sequential response bias, irrespective of the nature of the alternatives. This finding is compatible only with limited capacity theory and negative attention theory.

§6. NUMBER OF PREVIOUS ELEMENTS VISIBLE TO THE SUBJECT

Positive memory theory is the only theory to predict a favourable effect of presenting the subject with some part of the produced sequence, because in this way the memory span would be artificially extended. For the same reason negative memory theory predicts the reverse effect. More specifically, positive memory theory predicts that sequential response bias *decreases* as a function of artificial span, especially in conditions with *many* alternatives, since non-randomness - which should be due to poor recall - is at the extreme in those conditions. Negative memory theory is bound to explain the effect of set size in terms of poor recall with few alternatives. According to this theory the artificial memory should *increase* sequential response bias, especially in conditions with few alternatives. At any rate either memory theory predicts that the effect of set size decreases when more previous responses are presented. The predictions of both memory theories are presented schematically in Fig. 4.5.

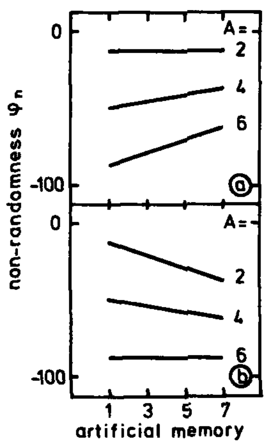


Fig. 4.5. Predictions on the basis of the two memory theories concerning the effect of the presentation of previous responses.
a = prediction for positive memory theory;
b = prediction for negative memory theory.

Limited capacity theory predicts that presentation of previous responses does not change randomization performance at all, since the demanded amount of information generated per time unit remains constant. Negative attention theory would reason that presentation of previous responses increases the attention for them, and hence increases sequential response bias. Finally, it is not so obvious what result positive attention theory would predict. On one hand the display of previous responses may distract the attention from the task and thus favour non-randomness. On the other hand, the level of arousal may increase which should go together with better randomization.

Thus far no experimenter manipulated the presentation of previous responses systematically. Table 1.1 (Chapter 1) shows that most frequently either *no* previous responses or *all* previous responses were visible for the subject, conditional on whether the production mode was calling out or writing down the responses. Baddeley (1966) used both methods, apparently without noticing the difference. In the next experiment this factor will be studied in a more formal way.

Experiment 4

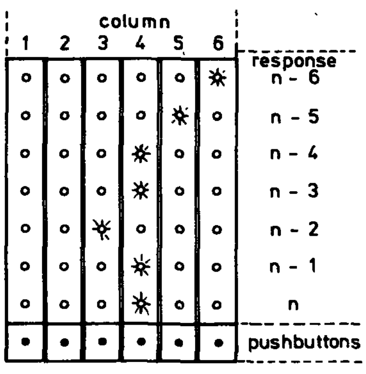


Fig. 4.6. Apparatus for Experiment 4. The lights indicate in which column previous responses occurred. The two last responses were in Column 4, before the third button was pressed, and so on.

Apparatus

The experimental apparatus is pictured in Fig. 4.6. A rectangular display is divided into six columns. At the bottom of each column a push-button is mounted, and over it seven lights. When at the beginning of the experiment one button is pressed, the light directly over it will go on. When the second response is made, the first light will move one line up, and in the first line a light will burn in the column of the last response. After seven responses a light will burn in each line, indicating in which column the seven previous responses occurred. In this way an artificial continuous memory is presented to the subject. The length of this memory, *i.e.*, the number of previous responses visible, can be varied by covering some lines of the display. The number of alternatives to be randomized is varied by covering some columns.

Procedure

The number of alternatives to be randomized was two, four or six. The number of previous responses visible was one, three, five, or seven. Each of eight subjects was required to produce four series of 100 responses in each condition, at a metronomic rate of 2 sec. per response. The total number of responses was

8 (subjects) x 3 (sets of alternatives) x 4 (length of artificial memory) x 4 (sequences) x 100 (responses) = 38.400

Results and Discussion

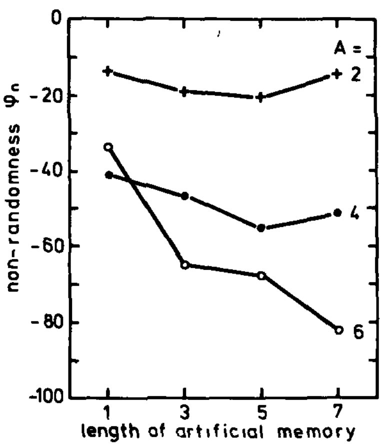


Fig. 4.7. ϕ_n-scores as a function of the length of the artificially increased memory (Experiment 4).

The ϕ_n-scores (Fig. 4.7) showed a significant adverse effect of the artificial memory ($F^{***}_{3,231}$ = 8.04) and a significant interaction of this factor with number of alternatives ($F^{**}_{6,252}$ = 3.25). A Newman-Keuls test revealed that the artificial memory increased sequential response bias only in the conditions with six alternatives. The interaction between artificial memory and order of analysis was not significant ($F_{15,315}$ = = 1.36). The main effect of number of alternatives was confirmed again ($F^{***}_{2,224}$ = 67.86).

The results disagree with the predictions of limited capacity theory and of both memory theories. Another problem for both memory theories is the fact that the effect of artificial memory was equal for all orders of analysis. Any memory theory would predict that low order effects will already change with a small artificial span, whereas only a large artificial memory will influence the high order sequential dependencies. Of the remaining theories negative attention theory seems to be compatible with the present findings but it is not clear how to explain the interaction with set size.

Conclusion

Presentation of previous responses increases sequential response bias, but only when the number of alternatives is large. This finding was not predicted by any of the five theories.

§7. VOCALIZATION AND KINAESTHESIS

Calling out the responses reduces the artificial memory to zero, but adds the factor of vocalization, which is known to be important for recall accuracy of the recency part of the serial position curve in short-term memory (Murray, 1966; Conrad & Hull, 1968). Crowder & Morton (1969) attributed this effect to a precategorical acoustic storage which may last several seconds. Moss (1971) found that the effect of vocalization tends to be reversed for the earlier items of the stimulus sequence. Probably this phenomenon is due to the fact that the time spent for vocalization cannot be used for rehearsal. Vocalization should have an effect rather similar to the extension of artificial memory with three items. Hence the predictions of all theories are, at least for lower order effects, the same as in the preceding section.

The use of push-buttons excludes both artificial memory and vocalization but introduces a kinaesthetic factor. Although little is known about this factor, Adams (1967) stated that, in general, short-term retention of motor responses is rather poor. If there is any effect at all, it aids recall of previous responses and has about the same influence as vocalization.

The next experiment was designed as a test of the predictions concerning both factors, vocalization and kinaesthesis.

Procedure

Sixteen subjects were requested to produce a random sequence of eight alternatives, at a metronomic rate of 1 sec. per response. It was decided to use eight alternatives because of the finding that factors enhancing recall of previous items are more effective with the larger sets of alternatives. There were four modes of production: calling out digits from 1 to 8, pressing buttons with digits from 1 to 8 on them, pressing on the same buttons plus calling out the digits, and pressing on buttons without digits. The last condition (spatial positions) was added as a replication of Experiment 3. Each subject produced six sequences of 100 responses in each condition. The total number of responses was

16 (subjects) x 4 (modes of production) x 6 (sequences) x 100 (responses) = 38.400

Results and Discussion

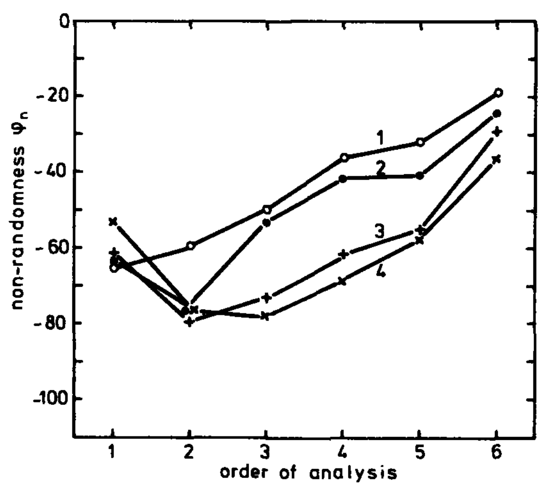


Fig. 4.8. ϕ_n-scores in the four conditions of Experiment 5.
1 = push-buttons labeled with digits;
2 = unlabeled push-buttons;
3 = pressing buttons + vocalizing;
4 = vocalizing only.

The results, pictured in Fig. 4.8, show first of all a considerable difference between vocalized and non-vocalized conditions. The effect was very significant ($F^{**}_{3,45} = 8.67$) as was the interaction with order of analysis ($F^{**}_{15,225} = 2.82$). A separate analysis of variance on vocalized conditions showed no effect of kinaesthesis ($F_{1,15} = 1.52$) and no interaction between kinaesthesis and order of analysis ($F_{5,75} < 1$).

The finding that vocalization impedes randomization seems to fit negative memory theory. However, the effect should be large for low order, and small or absent for high order dependencies. Actually quite the reverse was found, which cannot be explained by any of the five theories.

Conclusion

The kineasthetic factor is not important when subjects try to randomly press push-buttons, whereas vocalization increases the higher order effects of sequential response bias. None of the five theories predicted these results.

§8. TASK DURATION AND MONOTONY[11]

The predictions about the influence of sequence length show again an extreme diversity. The positive memory theory argues that it will become increasingly difficult to store the growing amount of frequency tallies as the task goes on. As a consequence randomization should become poorer as a function of sequence length. On the contrary, negative memory theory would state that the increasing amount of information to be stored would impede memorization and thus favour sequential randomness. The limited capacity theory predicts "no effect", unless it is hypothesized that the amount of information generated in say half an hour is also limited. Since randomization is rather a boring task, it is not inconceivable that task duration will be inversely related to the general level of arousal of the subject. As a consequence, positive attention theory predicts that sequential response bias increases with time spent at the task. On the other hand, Weiss (1965) argued that attention for previous responses will decrease when the subjects get bored. Therefore, negative attention theory predicts less sequential response bias in the later parts of the sequences.

Strangely enough the effect of task duration in randomization experiments has never been investigated seriously, although the length of generated sequences varied from 20 to 2520 responses, among the various publications (see Table 1.1, Chapter 1). Goodfellow (1940) suggested that longer series "would merely conceal the tendencies rather than eliminate them". In the context of psychophysics Day (1956) stated that factors like fatigue and boredom "undoubtedly introduce a certain amount of organisation into the series". These speculations may be contrasted by the observation of Weiss (1964) that some psychiatric patients, when generating a binary random sequence, actually improved in this ability from the first to the last quarter of the task. The purpose of the next experiment is to provide more conclusive data on this issue.

Some ambiguity exists with respect to the possible origins of an effect of task duration. On one hand the amount of produced information (sequence length) is considered to be the effective factor while, on the other hand, the low level of arousal or monotony is held responsible. These two factors may be correlated, but are theoretically more or less independent. With a fixed duration the task may possess a variable degree of monotony. One way of manipulating monotony is to introduce two production rates. Slow and fast series may be grouped separately (massed condition) or ordered alternately (distributed condition). The alternation of production rate in the last condition may break the monotony of the task.

Experiment 6A

Procedure

Six subjects were requested to press on four push-buttons in random fashion, at a metronomic rate of either 2 or 0.5 sec. per response. All subjects served in two sessions a day, one massed session, and the other distributed, during four consecutive days. All sessions contained eight slow series followed by eight fast series (or the reverse) whereas slow and fast series alternated in the distributed sessions. The total number of responses was

6 (subjects) x 4 (days) x 2 (massed-distributed) x 16 (series) x 100 (responses) = 76.800

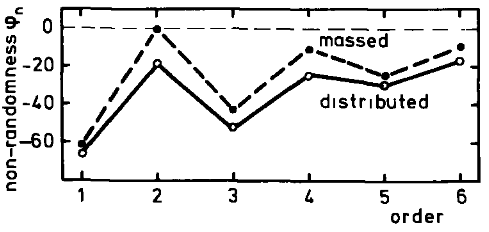


Fig. 4.9. ϕ_n-scores in Experiment 6A, separately for massed and distributed conditions.

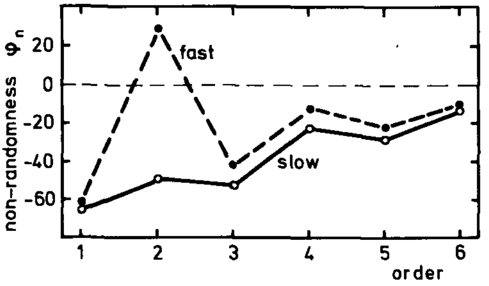


Fig. 4.10. ϕ_n-scores in Experiment 6A, separately for slow and fast conditions.

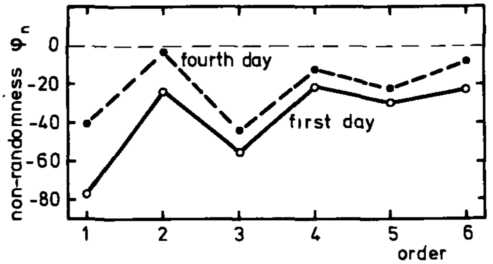


Fig. 4.11. ϕ_n-scores in Experiment 6A, separately for the first and the fourth day of the experiment.

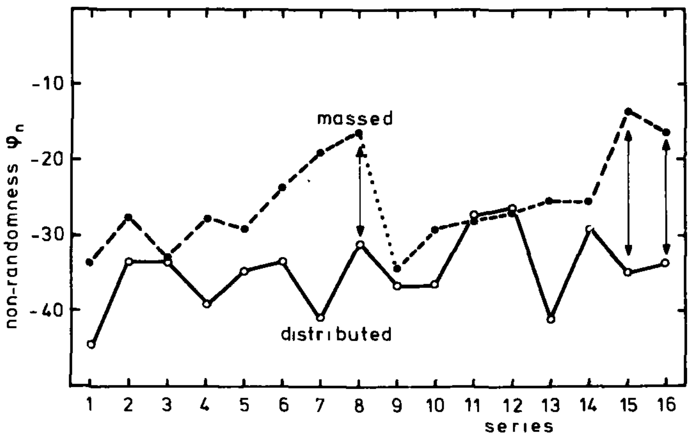


Fig. 4.12. The course of ϕ_n during 16 series in Experiment 6A, separately for massed and distributed conditions. Arrows indicate significant differences.

Results and discussion

Fig. 4.9 shows ϕ_n-scores for massed and distributed conditions. Sequential response bias was slightly but significantly less in the massed conditions ($F^{*}_{1,5} = 9.63$). There was no interaction with order of analysis ($F_{5,2060} = 2.08$). The effect of rate of production, as displayed in Fig. 4.10 was significant ($F^{***}_{1,2040} = 141.71$), but the interaction with order of analysis ($F^{**}_{5,25} = 13.24$) showed, after a Newman-Keuls test, that the difference was caused by the second order scores only. Sequential response bias was significantly decreasing during the four days of the experiment ($F^{*}_{3,15} = 4.67$). The interaction with order of analysis was not significant ($F_{15,75} = 1.08$). Fig. 4.11 shows the total decrease during four days. The course of sequential response bias in 16 series of one session, irrespective of production rate, is shown in Fig. 4.12. In the distributed condition non-randomness did not significantly change during 16 series (a Friedman-test on the mean scores of six subjects in the 16 series yielded $\chi^2 = 10.83$; df = 15; $0.80 > p > 0.70$). On the other hand a significant change of non-randomness in the course of 16 series was found in the massed condition (Friedman-test: $\chi^2 = 27.18$; df = 15; $0.05 > p > 0.02$). Separate t-tests for each series showed that non-randomness is significantly less in the massed condition only in series 8, 15 and 16 (with 5 df, $t = 2.60$, $p < 0.05$; $t = 4.28$, $p < 0.01$; $t = 2.61$, $p < 0.05$ for series 8, 15, and 16, respectively).

The results are clearly in contradiction with positive memory theory, positive attention theory and limited capacity theory. It is remarkable that sequential response bias is not so much decreased by task duration proper, but rather by task duration accompanied by monotony. This seems to disqualify negative memory theory in favour of negative attention theory, since it is boredom instead of the amount of response interference that decreases sequential response bias. For the effect of production rate on the second-order ϕ_n-scores no clear explanation is available. Possibly some motor factor was introduced by the relative high speed in the fast condition, although the comparable conditions of Experiment 1 did not reveal such an effect.

Conclusion

Task duration, when accompanied with monotony, decreases sequential response bias. This finding is compatible only with negative attention theory.

The data presented in Fig. 4.12 suggest that sequential response bias could disappear when the sequence becomes long enough. The possible application of this idea, for instance in the field of psychophysics, makes it worthwhile to study sequential response bias in still longer sequences. For this purpose the next experiment was conducted.

Experiment 6B

Procedure

Ten subjects were requested to press four push-buttons in random fashion, at a metronomic rate of 2 sec. per response. The one-hour sessions contained 16 sequences of 100 responses each. The total number of responses was:

10 (subjects) x 16 (sequences) x 100 (responses) = 16.000

Results

The results in Fig. 4.13 showed that sequential response bias decreases up to one half during one hour of production ($F^*_{15,810} = 1.18$).

Conclusion

There is no reason to hope that sequential response bias will disappear during the course of an experiment when the sessions do not take more than one hour.

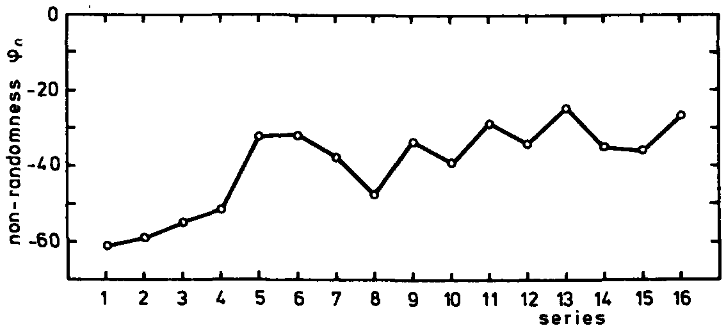


Fig. 4.13. The course of ϕ_n during 16 series in Experiment 6B.

§9. ADDITION OF A SECONDARY TASK

The limited capacity theory, as formulated by Baddeley (1966), predicts that there should be a linear relationship between the load imposed by a second task and the redundancy of a generated random sequence. Baddeley used card sorting as a secondary task, because of Crossman's finding (1953) that the load of this task is a linear function of log number of sorting categories.

It was shown in Table 4.1 that positive memory theory and positive attention theory predict the same effect as Baddeley's limited capacity theory, as recall of previous responses and attention for the randomization task are both supposed to suffer from the secondary task (*e.g.* Murdock, 1965; Routh, 1970). Negative memory theory and negative attention theory predict a beneficial effect of the secondary task for the same reason: recall of previous responses and attention for previous responses decrease when the load imposed by the secondary task increases.

The only experimental evidence was provided by Baddeley (1966) who found a nice monotone relation between redundancy of the randomized sequence and the number of sorting categories, although the relation was positively accelerated instead of linear. In the latter experiment again 26 alternatives were used for the randomization task, without an external representation. Therefore, by similar argument as production rate (§3), the effect may be due to increasing deterioration of the internal representation of the large set of alternatives. This consideration is reason enough to repeat the experiment with less alternatives which are visually represented.

Experiment 7

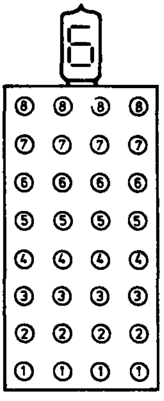


Fig. 4.14. Apparatus for Experiment 7. 32 Push-buttons are mounted in an array of 4 columns x 8 rows. On top a nixie tube is mounted, which may display the numbers 1-8.

Apparatus

The experimental apparatus is pictured in Fig. 4.14. The keyboard contained four columns with eight push-buttons each. The buttons are all marked with the number of the row in which it is located. On top of the keyboard a nixie tube is mounted, on which the digits from 1-8 may appear. The nixie tube is operated by the PSARP equipment (Van Doorne & Sanders, 1968).

Procedure

Eight subjects are requested to randomly choose among the four buttons in the row indicated by the number on the nixie tube. Thus selecting the appropriate row was the "sorting task", selecting the appropriate button was the randomization task, and the outcomes of both tasks were combined into one single response. The number of sorting categories was 1, 2, 4 or 8. The subjects were paced at a rate of 2 sec. per response by the appearance of the numbers on the nixie tube. Sequence-length was 100 responses. The total number of responses was

8 (subjects) x 4 (sorting categories) x 100 (responses) = 3200

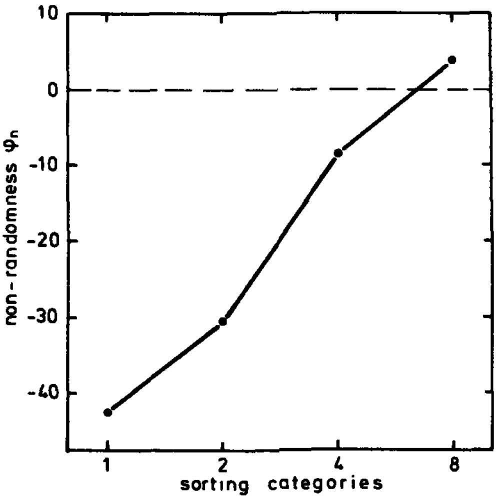


Fig. 4.15. ϕ_n-scores in Experiment 7.

Results

The outcome was surprising because it showed an effect opposite to that reported by Baddeley: sequential response bias decreased when the load of the secondary task increased (see Fig. 4.15). The effect was very significant ($F^{**}_{3,126} = 6.08$) and had no interaction with the order of analysis ($F_{15,105} = 1.43$). The relation between sequential response bias and log number of sorting categories appears to be linear within the range of the experimental conditions. It is questionable, however, whether the linear trend will remain when the number of sorting categories is further increased.

Obviously, Baddeley's results are to be attributed to the influence of the secondary task on the internal representation of the very large choice-set, as was the case in his experiment on production rate. The present finding is in contradiction with the limited capacity theory, as well as with positive memory and positive attention theories.

Conclusion

Introduction of a secondary task reduces sequential response bias, which was predicted only by negative memory and negative attention theory.

§10. SYNOPSIS

Each of the experiments discussed so far could have been executed on a larger scale, with more variables and with more subtility. Moreover, it is possible to add a score of other experiments with equally or more interesting paradigms. For the limited scope of this chapter, however, the newly obtained information is sufficient. The present set of experiments covers all experimental factors in which earlier studies differed. In addition it was possible to verify six times the predictions of the five theories that assume a functional limitation on the part of the randomizing subject.

Theory \ Experimental conditions	Increase of production rate	Increase of number of alternatives	Increase of artificial memory	Vocalization	Increase of task duration and monotony	Addition of secondary task
Positive memory theory	letters and digits: ? spatial positions: +	letters and digits: - spatial positions: +	-	-	-	-
Negative memory theory	letters and digits: ? spatial positions +	letters and digits: - spatial positions: -	-	±	±	+
Limited capacity theory	-	+	-	-	-	-
Positive attention theory	-	-	?	?	-	-
Negative attention theory	-	+	±	±	+	+

Table 4.2. How theories predicted the results of Experiment 1-7.
+ = prediction confirmed
± = prediction partly confirmed, partly falsified
- = prediction falsified
? = no prediction available

Table 4.2 lists the successes and failures of the predictions based on the various theories. Although the predictions were weak and sometimes rather hypothetical it is not encouraging for any theory that each of them was falsified at least three times. The results of the individual experiments may seem confusing at first but one aspect is very clear, *viz. that a change in external experimental conditions does influence sequential response bias*. This finding does not point to any theory specifically but explains why experimental evidence in the literature has been so inconclusive. A more thorough analysis of the aggregate results will be presented in Chapter 6.

The alternative to the hypothesis that a psychological function limits randomization is the idea that subjects have a faulty concept of randomness. Now that theories of the first type failed to prove tenable in their strict formulation, it is appropriate to go into the concept theory more deeply. Since *subjective randomness* in judgement experiments is traditionally considered a measure for the concept of randomness, this topic is next to be treated.

CHAPTER 5

SUBJECTIVE RANDOMNESS

§1. INTRODUCTION

Subjective randomness is most frequently measured in order to prove that the subjective concept of randomness does not differ from "objective randomness" (Baddeley, 1966), or as a demonstration that this concept shows the same tendency towards negative recency as encountered in sequential response bias (Mittenecker, 1953; Zwaan, 1964). In either case, subjective randomness is considered to be related to the concept of randomness rather than to a functional limitation. Yet the experimental situation of the judgement paradigm might call for a number of psychological functions apart from judgement proper, connected for instance with the scanning of the stimulus material. Therefore the claim that subjective randomness measures a subjective concept of randomness is as questionable as Baddeley's contention that sequential response bias is uniquely related to functional limitations.

In the light of this study the problem of measuring a pure concept of randomness is not too relevant. The purpose of the present chapter is to verify the hypothesis that such "concept-like" effects as measured in a judgement experiment are responsible for sequential response bias.

The conflicting results reported by Baddeley (1966), Mittenecker (1953) and Zwaan (1964) (see Table 1.4, Chapter 1) are difficult to in-interpret since none of these authors presented their stimulus material. In a fourth study by Cook (1967) the stimulus sequences are defined in a more formal way. Table 5.1 shows the first 30 elements of the 12 binary strings that were used.

String no.	
1	1 1 0 0 0 1 0 1 0 1 0 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 1
2	0 1 0 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 1 1
3	0 1 0 0 1 0 1 1 1 1 0 1 1 1 0 1 1 0 0 1 1 1 1 0 1 0 1 1 1 1
4	1 1 1 1 1 1 1 0 0 1 0 1 0 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0
5	0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1
6	0 1 1 0 0 1 1 1 0 1 1 0 0 1 1 0 0 1 1 1 0 1 1 0 0 1 1 0 0 1
7	0 1 1 0 0 1 1 0 1 1 1 0 0 1 1 0 0 1 1 0 1 1 1 1 0 1 1 0 0 1
8	0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 0 1 1 1 0 1
9	0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1
10	0 0 1 0 0 1 0 0 1 0 0 0 0 0 1 1 0 1 0 1 1 0 0 0 0 0 1 1 0 1
11	0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1
12	0 1 1 1 0 1 0 1 1 0 0 0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1

Table 5.1. First 30 elements of each string of digits used by Cook (1967).

Cook gave the following description:
"Twelve lists of 100 binary digits each were drawn up and printed on 12 separate sheets of paper, each sheet consisting of three rows of digits. List 1 was truly random, *i.e.*, half the digits were zeros, half were ones, the order being determined by a table of random numbers. Another three of the lists were random in the sense that the order was derived from a table of random numbers; but in them the proportion of zeros to ones was altered, these proportions being 55-45, 75-25 and 90-10 for list 2, 3 and 4 respectively[12].

List 5 consisted of a pattern 0110 repeated and list 9 consisted of 001 repeated. Lists 6, 7 and 8 were derived from list 5 by obliterating, respectively every fourth, every third, and every second digit and replacing these with digits based on a random number table (50-50 proportion). Similarly, lists 10, 11 and 12 were derived from list 9".

Subjects compared the lists two at a time, and decided which of the two lists was most patterned. The results showed that strings 1, 3 and 2 were judged as the most random ones, whereas 5 and 9 looked most patterned to the subjects. From these data it was concluded that subjects are able to recognize different degrees of bias.

For a description of Cook's stimulus sequences in terms of ϕ_n-scores some new material was generated following the rules stated above. In each of the 12 conditions ten sequences of 100 items were analyzed. The average results, shown in Table 5.2 give rise to the following comments.
1. Obviously sequential bias was confounded with marginal probability

Sequence No.	Marginal probability of zeros	ϕ_n-scores of order 1	2	3	4	5	6
1	0.50	0	22	31	35	23	- 17
2	0.45	8	- 24	21	- 23	43	- 4
3	0.25	- 17	- 3	24	11	35	23
4	0.10	- 22	86	1	- 38	- 88	- 19
5	0.50	0	-100	0	0	0	0
6	0.37	- 51	-100	- 97	100	17	18
7	0.50	19	-100	2	98	23	- 98
8	0.50	20	-100	16	98	21	- 45
9	0.67	-100	-100	0	0	0	0
10	0.62	-100	-100	+100	-100	- 93	- 24
11	0.83	-100	-100	+100	- 97	- 99	100
12	0.58	33	- 95	- 83	- 99	37	92

Table 5.2. Description of the 12 stimulus sequences used by Cook (1967) in terms of ϕ_n.

even in sequences 5 to 12. As it is not clear whether a subject will judge marginal probabilities, conditional probabilities, or both, it will be wise to deal with these effects separately.

2. Sequential bias in strings 5 to 12 was varied in a rather unsystematic way. Most of the sequences are extremely structured, with large differences between effects of successive orders. Not any of the sequences show resemblance to the sequences generated by subjects in a randomization experiment. This is even true for the sequences with a fifty-fifty ratio of zeros and ones.

Taking the arguments together it must be concluded that Cook studied the detection of extreme degrees of bias, with bias in the marginal probabilities and sequential biases of various orders confounded. Hence, the possibility is not excluded that subjects will mistake biased sequences for random ones *if the appropriate sequences are presented.*

The outline of the present chapter will be first to study subjective randomness in a more systematic way, in order to establish whether subjective randomness coincides with objective randomness, or not. The experiments will be limited to binary sequences only. The use of more alternatives would increasingly emphasize the perceptual problems involved in the scanning part of the task, thus complicating the paradigm too much for the limited scope of this chapter. In two further experiments the relation between subjective randomness and sequential response bias will be investigated. A more detailed interpretation of the aggregate data will be presented in Chapter 6.

§2. MEASURING THE POINT OF SUBJECTIVE RANDOMNESS[13]

Experiment 8 A

Stimulus material

It was decided to vary sequential bias for the first three orders of dependency separately. When the binary elements are described by zeros and ones, the experimental variable is the conditional probability of zero after zero (or one after one). This conditional probability was, after some piloting, given the values 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 or 0.8. The corresponding values of ϕ range from -0.6 with steps of 0.2 to +0.6. For the so-called "first order" stimuli, the probability that zero is directly repeated was varied. For "second order" and "third order" stimuli, the conditional probability that zero is repeated two or three positions later was manipulated. Altogether there were 3 x 7 = 21 different types of stimuli. Of each type 16 different examples were constructed by a PDP-7 computer. Care was taken to vary only one order of dependency at a time, while dependencies of all other orders up to the sixth were kept zero. As an example Fig. 5.1 shows non-randomness of six orders of dependency for the first order sequences. Values of ϕ were

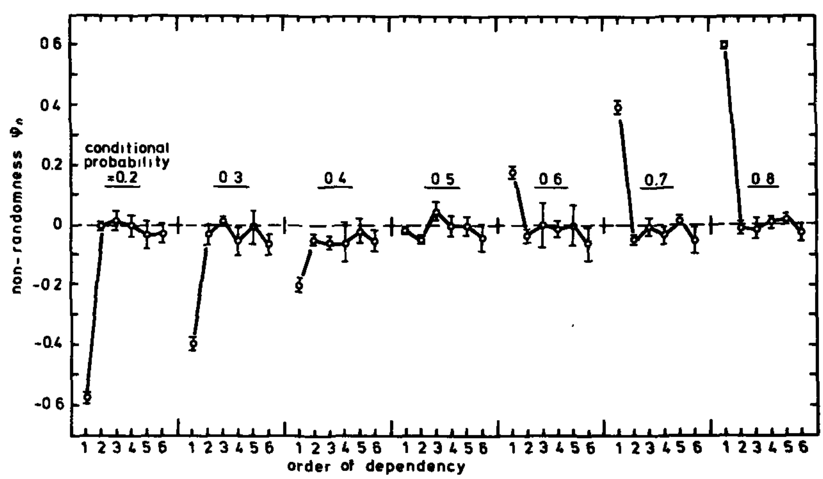
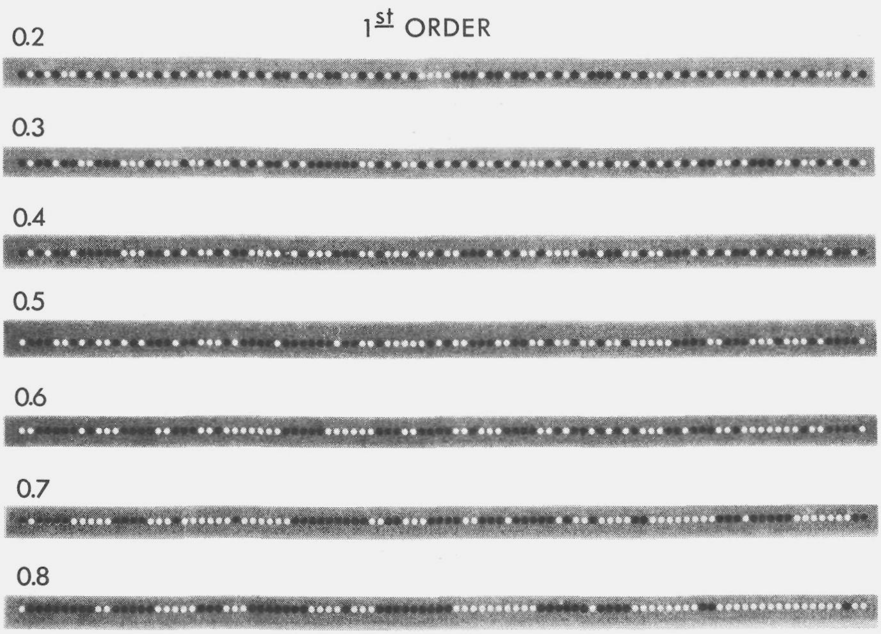


Fig. 5.1. Non-randomness in the "first order" pictures used in the experiments of the present chapter.

used, rather than ϕ_n, because of the extreme conditions which fall all in the hundredth percentile. It is justified to use ϕ as the present chapter is limited to the study of binary sequences only. In Chapter 2 it was shown that for binary sequences ϕ is distributed symmetrically, with the extremes being -1.0 and +1.0. The conditional probabilities may be converted into values of ϕ (for the first order values exactly and for the other orders approximately) by substituting:

$$\phi = 2\ (\text{conditional probability}) - 1 \qquad (5.1)$$

From the sample of 16 x 21 = 336 binary sequences, 240 different sets of seven sequences were composed (*i.e.* all sequences were used five times). Each set contained sequences of one and the same order of dependency (first, second or third) but all different with respect to the degree of dependency. For the actual presentation the binary sequences were translated into series of white and black dots on a neutral grey background (see Fig. 5.2). The seven sequences of a set were put on one slide in a randomized order.



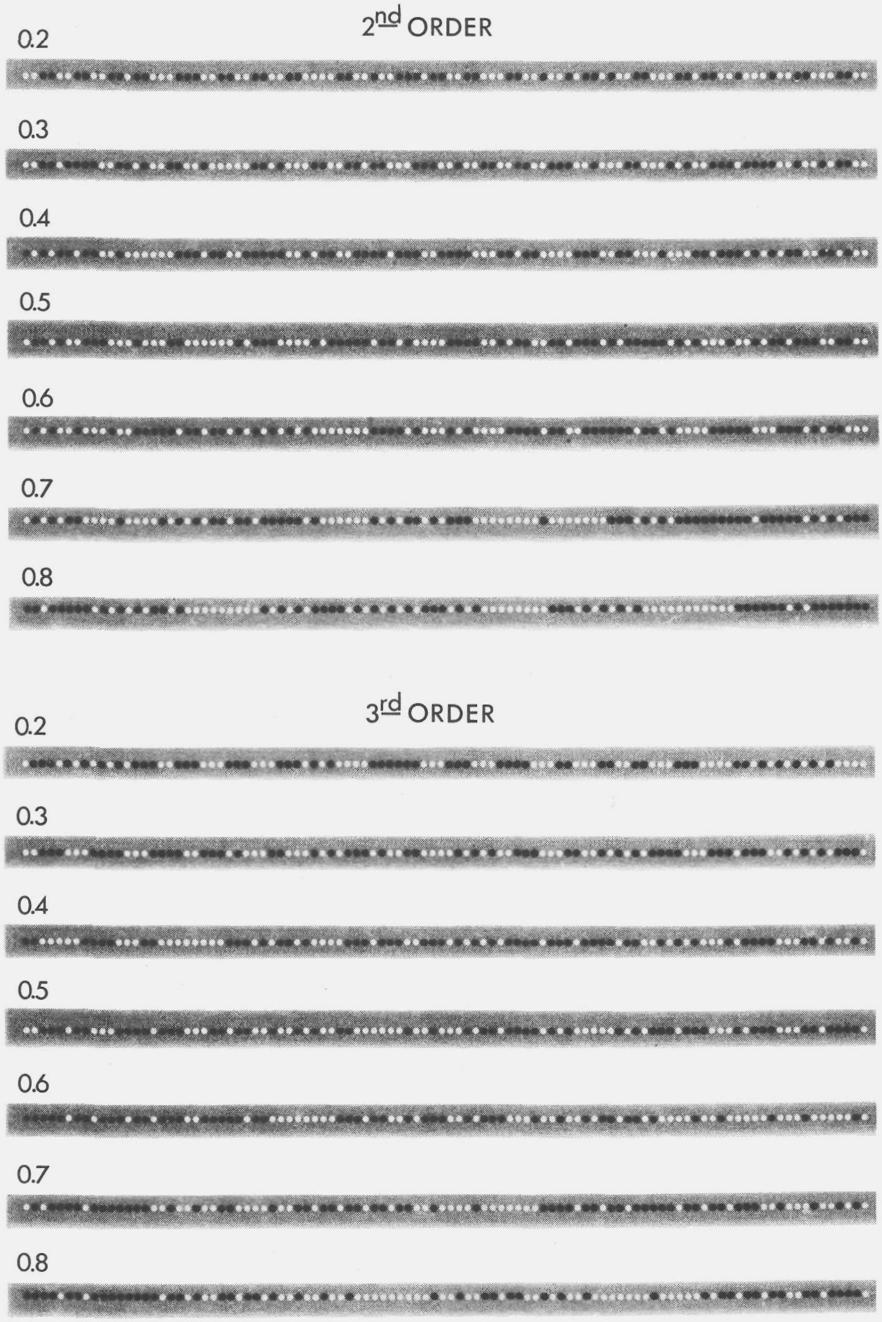


Fig. 5.2. Examples of the binary sequences presented for judgement.

Procedure

The slides were presented to the subjects with the instruction to write down the number of the one sequence that looked most random to them. The notion of randomness was made clear by mentioning random sequences produced by the flipping of a coin[14]. All subjects judged 48 slides, 16 for each order, starting with the first order sequences and finishing with the third order material. This means that the order of dependency was confounded with order of presentation. The reason for this was that some pilot studies showed that difficulty increases with order of dependency. It was decided to start every time with the easy slides in order to give subjects the best chance to perform optimally. Time of presentation was 30 sec. per slide. The total session lasted about 30 min.

Scoring

For each order of dependency subjects rendered 16 judgements. For each set of 16 judgements mean and standard deviation were calculated. In this way three means and three standard deviations per subject were obtained.

Subjects

Subjects were 93 undergraduate students of the University of Utrecht (32 females and 61 males) and 110 officer candidates of the Dutch Army[15]. Subjects were run in groups. The students were paid Dfl. 2.50 and the candidates received a free drink (afterwards).

Results

The average points of subjective randomness show a clear bias in the direction of negative recency (conditional probabilities < 0.5) for all three orders of dependency (see Fig. 5.3). These means are to be interpreted in relation to the standard deviations, since any subject may yield an average at a conditional probability level of 0.5 if he responds with complete neglect of the stimulus slides. In this case the

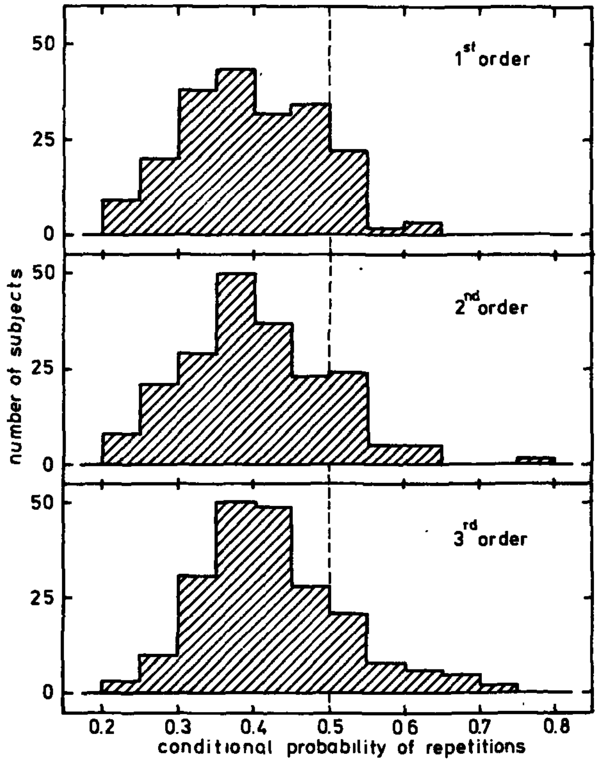


Fig. 5.3. Results of Experiment 8A, showing the ideal points of randomness of 203 subjects.

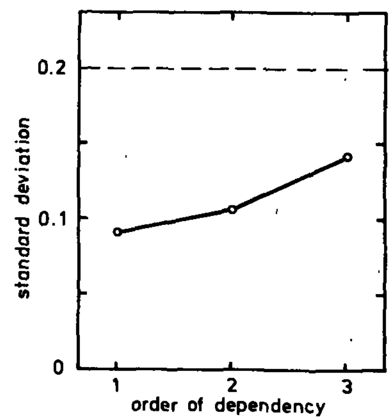


Fig. 5.4. Average standard deviation among successive judgements are well below 0.2, the level that would be obtained when subjects respond with complete neglect of the stimulus slides.

standard deviation among the 16 successive judgements would be 0.2. Fig. 5.4 shows that the actual standard deviations were less. At the same time it is shown that the discriminal dispersion for conditional probabilities in this material increases with the order of dependency. The female students showed a slightly better performance with the first order pictures; their average was 0.43, which differs significantly from 0.39, the average of the male students ($t = 2.58$, $df = 56$, $p < 0.02$). For second and third order stimuli no differences between male and female were found. The officer candidates showed the same performance as the male students.

Discussion

The effect of discriminal dispersion increasing with order of dependency calls for some further analysis, since the three orders are completely equivalent with respect to conditional probabilities or informational content. The present findings suggest that subjects did not process such a mathematical quantity at all. An alternative parameter, frequently encountered in the literature on human processing of sequential information is the *run structure*.

Derks & House (1970) concluded, after an analysis of studies by Bruner, Wallach & Galanter (1959), Derks & House (1965), Galanter & Smith (1958), Green (1958), Ludvigson (1966), Restle (1967), Vitz & Todd (1967) and Wolin, Weichel, Terebenski & Hansford (1965) that in the processing of a sequence of events "a repeated series of identical events is acquired, utilized and extinguished as a unit"...."In other words, if the subject's description of a sequential pattern is a series of numbers that gives the lengths of the runs, then the length of that number series will be critical and not the magnitude of the numbers in it". Additionally this idea was put forward by Goodnow & Pettigrew (1955), Keller (1961), Nicks (1959) and Restle (1961).

The average run structures of the stimulus sequences are compared with the theoretical run structures of a random sequence in Table 5.3. It appears that the non-randomness in terms of runlength distributions decreases with increasing order of dependency. Hence a subject, acting as a "statistical analyzer" of run structure would indeed have a greater discriminal dispersion at the higher orders. A more careful analysis

	Cond. prob.	Run-length 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Level of signif.
First order	0.2	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	<<0.001
	0.3	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	<0.01
	0.4	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	n.s.
	0.6	-	+	+	+	+	+	+	+	-	-	-	-	-	-	-	n.s.
	0.7	-	-	+	+	+	+	+	+	+	-	-	-	-	-	-	<0.05
	0.8	-	-	+	+	+	+	+	+	+	+	-	-	-	-	+	<0.005
Second order	0.2	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	<0.001
	0.3	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	<0.05
	0.4	-	+	+	-	-	+	-	-	-	-	-	-	-	-	-	n.s.
	0.6	+	-	-	-	+	+	+	+	-	-	-	-	-	-	-	n.s.
	0.7	+	-	-	-	-	+	+	+	+	+	-	-	-	-	-	<0.05
	0.8	+	-	-	-	-	+	+	+	+	+	+	-	-	-	-	<0.001
Third order	0.2	0	-	+	+	-	-	-	-	-	-	-	-	-	-	-	n.s.
	0.3	0	-	+	+	-	-	-	-	-	-	-	-	-	-	-	n.s.
	0.4	0	-	+	+	+	-	-	-	-	-	-	-	-	-	-	n.s.
	0.6	0	+	-	-	+	-	+	+	-	+	-	-	-	-	-	n.s.
	0.7	0	+	-	-	-	-	+	+	+	-	-	-	-	-	-	n.s.
	0.8	0	+	-	-	-	+	+	+	+	-	+	-	-	-	-	n.s.

Table 5.3. Comparison of the run-structures of biased sequences with the run structure of a real random sequence. + means frequency of run length too high; - means frequency of run length too low; 0 means frequency of run length correct. Significances of the differences with the theoretical random run structure are calculated by the Kolmogrov-Smirnov test.

reveals which aspect of the run structure determines the "judged ranomness" of a sequence. Apparently it is not the frequency of runs with length 1, since then positive recency for the second order and no recency for the third order would have been obtained. What the preferred sequences of the various orders, *i.e.* the sequences with conditional probabilities near 0.3 - 0.4, have in common is the lack of runs longer than six in favour of a high frequency of shorter runs. It is not unlikely that subjects simply selected sequences with few long runs for each order of dependency. This idea of run structure will be elaborated upon in the next chapter.

§3. CORRELATIONS BETWEEN SUBJECTIVE RANDOMNESS AND SEQUENTIAL RESPONSE BIAS

If the effect of subjective randomness acts upon the randomization task it should be found that subjects with extreme effects of subjective randomness display also extreme sequential response bias, and *vice versa*. This hypothesis was tested in the next experiment.

Experiment 8B

Procedure

The 110 officer candidates who served in Experiment 8A were requested to write down a binary random sequence of 100 responses, at a rate of approximately one response per sec. before the start of the experiment proper. Through mediation of the selection centre it was possible to correlate these data also with general intelligence and technical insight of the subjects. For this purpose the Otis mental ability test (Otis, 1954) and Bennet's test of mechanical comprehension (Bennet, 1947) were used. For general description and reference see Cronbach (1960). Care was taken not to impair the anonymity of the subjects.

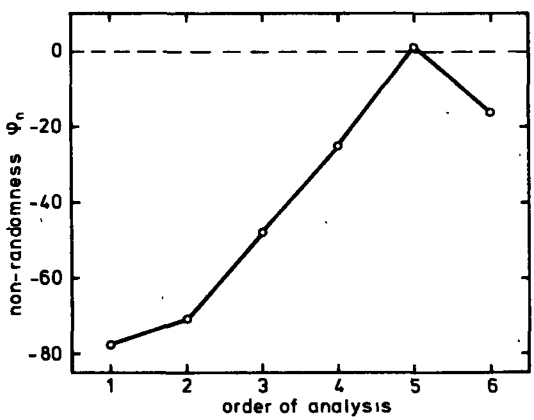


Fig. 5.5. Sequential response bias in the randomization condition of Experiment 8B.

Results and Discussion

The average sequential response bias in the sequences generated by the 110 subjects is shown in Fig. 5.5. The first-order effect corresponds with a 0.43 conditional probability for repetition, which is not too different from the average point of subjective randomness (conditional probability = 0.39). The correlations among the points of subjective randomness (three orders) and sequential response bias (six or-

		Subjective randomness of order			Sequential response bias of order					
		1	2	3	1	2	3	4	5	6
Subjective randomness of order	1	0.92**								
	2	0.73**	0.86**							
	3	0.59**	0.79**	0.80**						
Sequential response bias of order	1	0.27**	0.27**	0.28**	0.69**					
	2	0.10	0.23*	0.23*	0.29**	0.38**				
	3	0.03	-0.01	0.06	-0.11	0.04	0.18			
	4	0.10	0.17	0.20	0.05	0.03	-0.08	0.13		
	5	-0.05	-0.03	-0.06	-0.17	-0.07	0.17	-0.22**	-0.16	
	6	-0.12	0.00	-0.05	-0.09	0.12	-0.02	0.04	-0.04	-0.04

Table 5.4. Correlations among points of subjective randomness (three orders) and sequential response bias (six orders). Diagonal cells contain test-retest reliabilities. * means $p < 0.05$; ** means $p < 0.01$.

ders) are presented in Table 5.4. The diagonal cells contain the test-retest reliabilities of each quantity, calculated on the basis of split-half reliabilities. The data give rise to the following remarks.

1. The points of subjective randomness of all three orders correlate mutually extremely high. This is further support for the hypothesis that subjects judge some common feature, like the occurrence of long runs.
2. The effects of sequential response bias of six orders show no significant mutual correlations, except the correlation between first and second order effects.
3. The points of subjective randomness correlate with the first and second order effects of sequential response bias. These correlations are highly significant but not high, which suggests that subjective randomness contributes to sequential response bias only to a limited extent, and maximally up to the second order effects.
4. The test-retest reliabilities of the points of subjective randomness

are high.

5. The test-retest reliabilities of the effects of sequential response bias are only high and significant for the first order of dependency. Individual subjects can be described by their first order response biases, but not by the effects of higher orders.

On the matrix of correlations a factor analysis was performed resulting in one important and meaningful factor, which might be interpreted as a scale for subjective randomness (see Table 5.5). This factor accounted for 67% of the total variance. The second factor explained a further 17% of the variance.

	Effect	Factor loading
Subjective randomness	1st order	0.80
	2nd order	0.91
	3rd order	0.88
Sequential response bias	1st order	0.38
	2nd order	0.29
	3rd order	0.00
	4th order	0.21
	5th order	-0.11
	6th order	-0.06

Table 5.5. First factor resulting from a factor analysis on the matrix in Table 5.4. The factor accounts for 67% of the total variance.

For 78 out of the 110 subjects it was possible to correlate subjective randomness and sequential response bias with general intelligence and technical insight. The only significant correlations that emerged were *negative* correlations ($r = -0.30$ and $r = -0.26$) between technical insight on one hand and first order subjective randomness and first order sequential response bias on the other hand.

Conclusion

The so-called "concept theory" which claims that sequential response bias can be explained by the biases measured in a judgement experiment, is only partially right for first and second order sequential response bias. The higher order effects of non-randomness in a randomization task are not correlated with any aspect of subjective randomness.

§4. THE EFFECT OF UNLEARNING SUBJECTIVE RANDOMNESS

A further test of the hypothesis that subjective randomness causes sequential response bias is provided by a training experiment. When subjects are trained to recognize the real random series in the judgement situation, this should, according to the "concept theory", decrease sequential response bias in the randomization task. The next experiment is concerned with this paradigm.

Experiment 9

Procedure

Ten subjects were trained to recognize the real random series in the sets of seven binary sequences described in §2. This was accomplished for all three orders of dependency by giving the correct answer after each selection. For this purpose the subject had two push-buttons: one for the next slide and one for the correct answer. The order of actions taken by the subject was as follows:

1. Call for next slide.
2. Write down a judgement.
3. Call for correct answer.
4. Observe the "real random" sequence, and start again with 1.

In this way the task was paced by the subject. For each order of dependency 80 slides were judged in 5 sessions of 16 slides. Before and after the training period a sequence of 100 binary responses was generated by each subject. No control group was used as it may be assumed that sequential response bias will remain constant when the training sessions are omitted (see Experiment 6A, 6B).

Results

The results of the training period are shown in Figs. 5.6 and 5.7. The following aspects are noteworthy.

1. The subjects learned to select the correct sequences rather accurately.

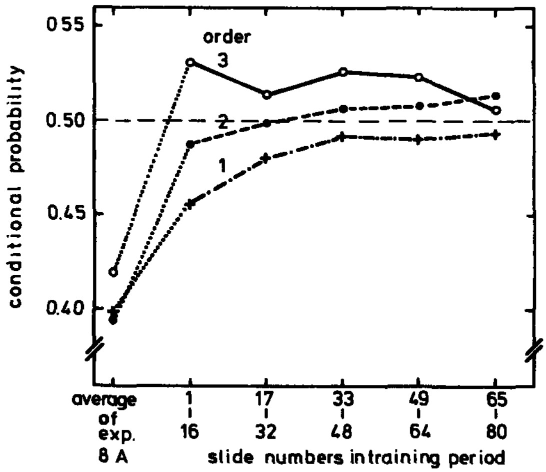


Fig. 5.6. Ideal points of randomness in the training sessions of Experiment 9.

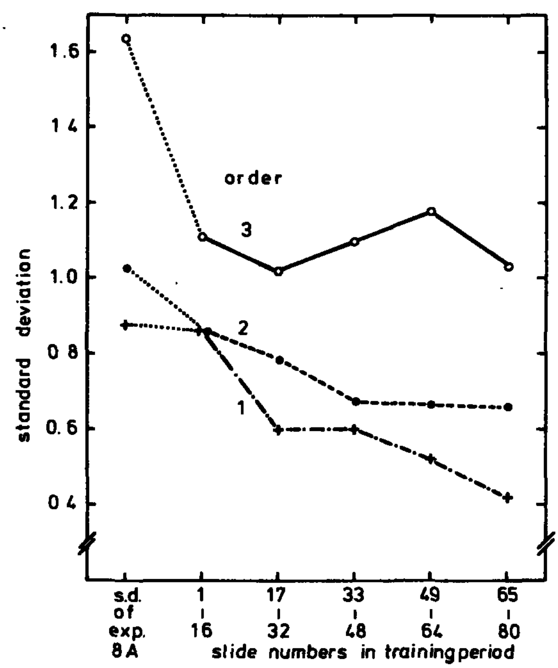


Fig. 5.7. Average standard deviation among successive judgements in the training sessions of Experiment 9. Note that the differences among the three orders of dependency remained.

2. The training on the first order sequences decreased the initial effects of second and third order non-randomness considerably, when compared with the average results of Experiment 8A. This points again to a common factor in all three orders of dependency.
3. The course of the discriminal dispersions during the training reveals that the differences in difficulty among the three orders of dependency remained.

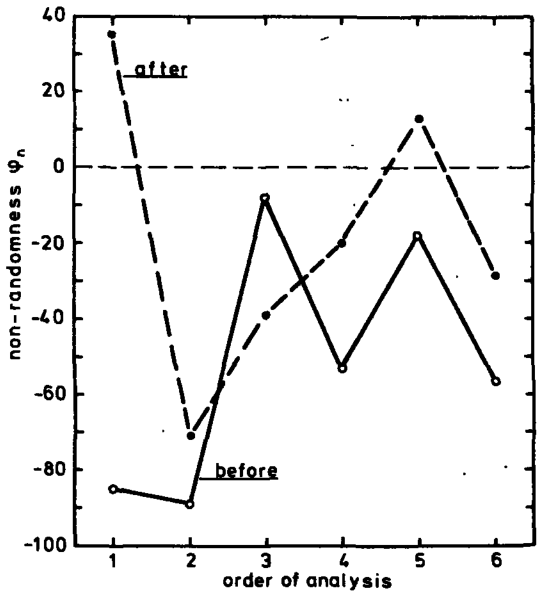


Fig. 5.8. Sequential response bias in the pre-test and post-test conditions of Experiment 9. The difference between first order effects is the only significant one.

The effects of sequential response bias in the randomization test before and after the training period are presented in Fig. 5.8. Separate t-tests for each order of analysis rendered a significant effect of the training *only for the first order bias* ($t = 3.64$, $df = 9$, $p < 0.01$).

Conclusion

Elimination of subjective randomness by training does only affect first order sequential response bias, which is a further confirmation of the previous finding that there is a close relation only between subjective randomness and first order sequential response bias.

§5. SYNOPSIS

The experiments described in the present chapter had only a limited scope: they were all concerned with binary sequences in which were introduced sequential dependencies of only one order at a time. The generalisations can therefore only be tentative. The main conclusions were

1. Subjective randomness of the first three orders of dependency is biased into the direction of negative recency.
2. The parameter judged by the subjects seems to be run structure rather than conditional probability.
3. Subjective randomness is weakly correlated to only first order sequential response bias.
4. Unlearning subjective randomness may change only first order sequential response bias.

It is difficult to assess whether a "subjective concept of randomness" is the only effect measured by the judgement paradigm. At any rate, the present findings are incompatible with Baddeley's statement (1966) that on the part of the subjects the "*concept* of randomness was not at fault". On the other hand the strict concept theory, supported more or less explicitly by Chapanis (1953), Mittenecker (1953, 1958), Rath (1966), Ross & Levy (1958), Skinner (1942), Teraoka (1963), and Zwaan (1964), is impaired since only a small part of sequential response bias could be related to the kind of performance measured in the experimental paradigm proposed by Mittenecker and Zwaan.

The general conclusion must be that the concept theory alone is no more capable of accounting for sequential response bias than the theories that attribute the effects to a limitation of a psychological function. In the next chapter an attempt will be made to reconcile some elements of these theories in order to make a first step towards a theory of sequential response bias.

CHAPTER 6

ELEMENTS OF A THEORY OF SEQUENTIAL RESPONSE BIAS

§1. INTRODUCTION

A theory of sequential response bias should consider two main questions: why do subjects show effects of sequential response bias at all, and why does sequential response bias take the form it actually does. These questions, though related, are not necessarily answered at the same time. For instance, each of the five theories on functional limitations, treated in Chapter 4, indicates why sequential response bias occurs, but they do not imply that the bias should be in the direction of negative recency. A complete theory should lead to a model which predicts the nature and extent of sequential response bias under various experimental conditions, and the problem is not solved until this theory has been outlined. However, the data gathered in the previous chapters are no sufficient bases for a complete theory and hence do not allow for the construction of a formal model. Therefore the present chapter will be limited to the indication of some elements which are to be incorporated in a theory and a model of sequential response bias.

Subjects are unable to randomize because they do not have a build-in randomizator.This statement seems to be trivial and yet it points to a basic property of the randomization task.

When a subject is requested to bend or to stretch his fingers, he can easily do this because he needs only to activate an already available, physical mechanism. Drawing a butterfly, on the other hand, though essentially effectuated by finger motions, is a much more difficult task, as there is no special purpose hardware device for drawing butterflies. In stead, the subject should first know what a butterfly looks like, and, secondly, he should be able to actualize this knowledge in a drawing. These aspects which belong to the domain of cognition and skill determine the final success.

Randomization is a task of this second type because there is no randomizing organ anywhere in the organism. A randomizing subject should have a more or less pronounced notion about the properties of a random sequence and he has to actualize these properties in the generated sequence. The difficulty of randomization is situated in these two subtasks. The first task was considered crucial by the supporters of a concept theory, while the second one comes close to theories that focus on functional limitations. It has never been attempted to combine both points of view, and that is exactly what will be proposed in the present chapter. On the basis of the experimental results presented in the previous chapter, it will be argued that sequential response bias stems really from two different sources at the same time. The evidence for this assertion will be presented in the next section.

§2. EVIDENCE FOR THE DUAL CHARACTER OF SEQUENTIAL RESPONSE BIAS

Sequential response bias at low orders of analysis showed a number of properties that were not found at the higher orders.

1. Large inter-individual differences at the first and second order of analysis were shown in Experiment 1 (Fig. 3.3). For the situation of Experiment 1 it can be calculated that the standard deviation of ϕ_n, under the hypothesis of no individual differences, should be about ± 29. For the two lowest orders of analysis this value was exceeded

Order of analysis	1	2	3	4	5	6
Standard deviation among subjects, averaged over five values of A (same data as in Fig. 3.3)	65.1	48.6	28.8	28.0	28.7	21.6

Table 6.1. From the results of Experiment 1. The standard deviation among subjects decreases to the theoretical value (29) with increasing order of analysis.

consistently (Table 6.1) but for the other orders the individual differences were not larger than expected by chance (it was already mentioned in Chapter 3, that individual differences were not significant above the second order of analysis).

The relative importance of individual differences at the first two orders was also revealed by the split-half reliabilities presented in the previous chapter where it was found that a subject may be characterized by his first and second order effects of sequential response bias. Moreover these two effects were significantly correlated (Table 5.4) in contrast with the other effects which were mutually independent.

2. Large intra-individual differences of sequential response bias were found again only at the first two orders of analysis. In the results of Experiment 6A it was tallied how often the standard deviation among the eight values of ϕ_n in each of the four conditions (massed-distributed and fast-slow) was larger than the standard deviation expected under the hypothesis that intra-individual variations were ruled by chance. Table 6.2 shows that significant intra-individual variations occurred only at the lower orders of analysis. The decrease of the frequency of large variations with increasing order of analysis was significant (Friedman-test: $\chi^2 = 11.6$, df = 5, $p < 0.05$). This trend was shown by all subjects to the same extent (Friedman-test: $\chi^2 = 9.11$, df = 5, $p > 0.10$). These results cannot be due to a differential ef-

Order of analysis	1	2	3	4	5	6
Number of standard deviations smaller than expected (out of 96)	24	21	39	38	45	47
p (sign test)	<0.01	<0.01	n.s.	n.s.	n.s.	n.s.

Table 6.2. From the results of Experiment 6A. The number of standard deviations smaller than expected under the hypothesis that intra-individual variations are governed by chance, increases with increasing order of analysis.

fect of task duration, since the decrease of response bias with increase of task duration was equal for all orders of analysis. The large intra-individual variations reveal that subjects are able to vary sequential response bias only at the first and second order of analysis.

3. Subjective randomness as measured in judgement experiments was cor-

related only with first and second order sequential response bias. (Experiment 8B). Unlearning subjective randomness does decrease only the first order bias (Experiment 9).

From the above mentioned evidence it may be concluded that there is a component in the lower order effects of sequential response bias, which is not present in the biases of higher orders. If there is any sense at all in speaking about a concept, then it is certainly in relation to this lower order component, which seems to be highly related to the specific intuition of the individual subject.

Apart from this individual component there is a more general effect operating equally at all orders of analysis, or only at the higher orders. This is illustrated by the results of Experiments 4, 6A, 6B and 7. The significant effects of conditions in these experiments (*i.e.*, artificial memory, monotony and secondary task) showed no interaction with order of analysis. In Experiments 1, 2, 3 and 5 only a higher order effect was reported, introduced by the number of alternatives, internal v.s. external representation of the set of alternatives, and vocalization. Obviously, this component which acts always upon higher-order effects and never exclusively upon lower-order effects is related to task variables rather than to individual variables. For the task component no individual differences are found. This was shown explicitly in Experiment 1 but it was also true for Experiments 2 to 7.

In the next sections the two components will be discussed separately. For as far as this is possible, an attempt will be made to show how both components originate, and why they behave as they do.

§3. THE INDIVIDUAL COMPONENT.

In the previous chapter strong indications were found that subjective randomness is related to a subjectively preferred runlength distribution. The same may hold for the first order sequential response bias, but this assertion is necessarily trivial, since the runlength distribution determines the number of runs, and therefore the first order bias. It is possible, however, that the runlength distribution determines higher order effects too, as is demonstrated in the second and third order stimuli of Experiment 8A. This would contradict the hypothesis that the independent individual component is nothing more than a subjectively preferred runlength distribution. A test for the presence of higher order effects in the runlength distributions was obtained by the analysis of artificially constructed sequences. These sequences had exactly the same runlength distributions as the sequences produced in Experiment 1, but the runs were mixed in a random order. The total number of responses generated in this way was, as in Experiment 1:

8 (subjects) x 5 (number of alternatives) x 4 (rate of production) x 106 (length of the sequence) = 16.960.

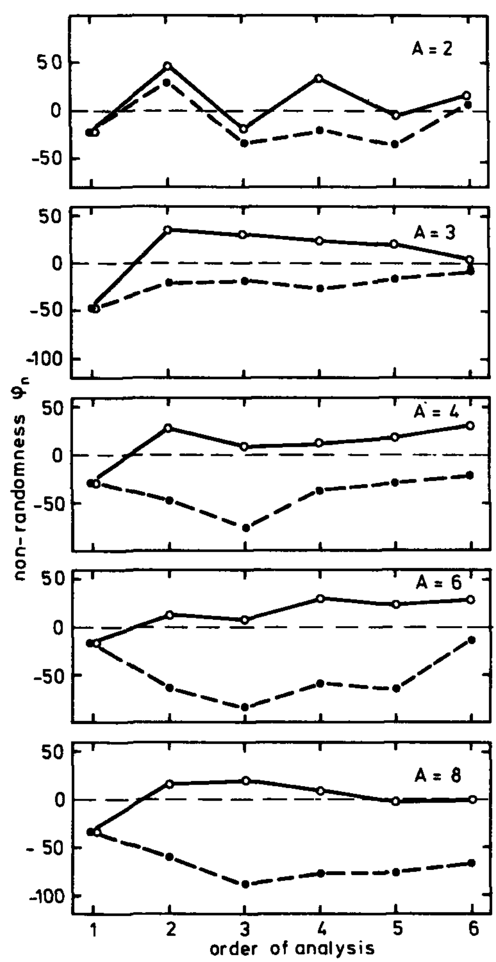


Fig. 6.1. Comparison of ϕ_n-scores in Experiment 1 (•---•) with ϕ_n-scores in artificial sequences with the same runstructure as the sequences in Experiment 1 (o—o).

The ϕ_n-scores of these artificial data are compared with the results of Experiment 1 (see Fig. 6.1). In the artificial sequences the higher order effects of sequential response bias are too much in the direction of positive recency for all values of A. This means that the subjectively preferred runlength distributions account only for the first order bias. The biases of higher orders are due to the ordering of the runs and to the ordering of the run-contents (which is obtained by reducing all runs to a single element, indicating which alternative was in the run).

The effect of cycling, being a high order effect, was not present in the sequences generated on the basis of runlength distributions, as is shown in Fig. 6.2. Only for A = 2, a perfect correspondence between

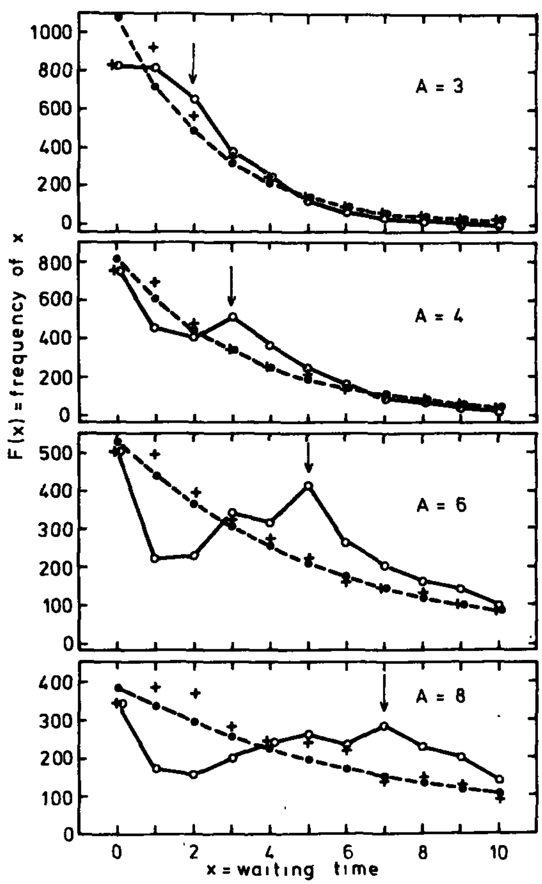


Fig. 6.2. Comparison of recurrence distributions in Experiment 1 (o—o) and in the artificial sequences (+ + +). •---• = theoretical distribution.

experimental and artificial recurrence distributions was obtained, which is caused by the fact that the recurrence and runlength distributions are identical for A = 2 and large N (see formulas 3.1 and 3.2). Hence in the binary case it is not justified to speak about cycling as a separate mechanism.

An analysis of variance on the first order ϕ_n-scores in Experiment 1 revealed that it is not possible to describe a subject with only one parameter value, since there was an interaction between subjects and number of alternatives ($F^*_{28,84} = 1.91$). The bias of some subjects is constant for all values of A or may go from alternation to repetition or *vice versa* (see Fig. 6.3). There does not seem to be any system in

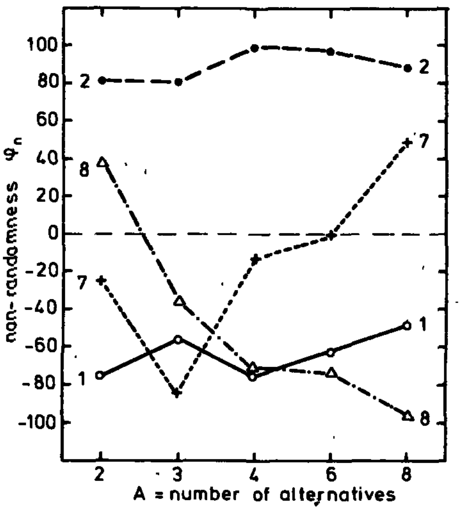


Fig. 6.3. First order ϕ_n-scores of Subject 1, 2, 7 and 8 in Experiment 1.

the individual differences at each level of A, which makes it difficult to include the individual component in a mathematical model for sequential response bias.

In conclusion, it can be said that any first order parameter, like number of alternations or number of runs, can be used to represent the individual component in sequential response bias. It is psychologically meaningful, however, to use the runlength distribution for this purpose. The present results show that this can be done without introducing any higher order effects.

§4. THE TASK COMPONENT

The task component contains the following two clearly distinguished effects:

1. An excessive matching of marginal frequencies within an interval of about six responses. This effect is independent of A.
2. A tendency to use each alternative once in a cycle of A responses, which of course heavily depends on A.

The contention suggest a similar procedure as used in the previous section, *i.e.* the generation of artificial sequences with the same frequency matching and cycling tendencies as were found in the experimental sequences, in order to test whether the two factors sufficiently account for the whole task component. In a sense such an artificial generator would be a model of a randomizing subject. Actually this test has not been attempted because a number of additional assumptions would be necessary. With respect to the matching tendency some function that weighs the importance of previous responses should be hypothesized. This function could be uniform (the six previous responses are weighted equally), linear but with a positive slope (recent responses have a larger contribution), U-shaped (as in short-term memory) or with any other shape. For the simulation of the cycling tendency, it should be realized that cycling cannot be complete, since in that case no run-length longer than 2 and no distance of repetition longer than 2A - 2 could occur. When cycling is incomplete it is again inevitable to make assumptions about the weighting function applied to the previous responses, with a maximum distance A - 1. Thus far no sound theoretical framework is available to arrive at reasonable assumptions concerning both types of weighting function, but some data provided in the previous chapter can be used as a starting point.

The basic speculation, then, is that all high order effects of sequential response bias are to be explained by a tendency to match marginal frequencies of the alternatives throughout the sequence. The two ways in which this can be done without a counting or tallying strategy are matching marginal frequencies in small intervals and cycling. When the matching interval contains only few responses, matching will become increasingly difficult with increase of A. The more A approaches the number of responses in the matching interval, the more a subject will be thrown upon cycling, and the heavier the biases that occur. Some support for these speculations is obtained by looking at the degree of matching and cycling in Experiments 1 and 2. Fig. 6.4 shows that the ratio of observed and theoretical values at the top of the sampling distributions (Fig. 3.5) is fairly constant for different values of A,

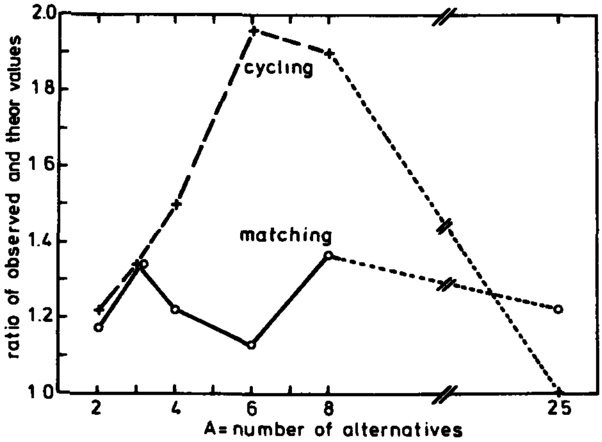


Fig. 6.4. The effect of increase of the number of alternatives on recurrence and sampling distributions. For the recurrence distributions the values are ratios between observed and theoretical values of f(x) at x = A - 1 (see Fig. 3.4). For the sampling distributions the numbers are ratios between observed and theoretical values of f(x) at the top of the theoretical distributions (see Fig. 3.5).

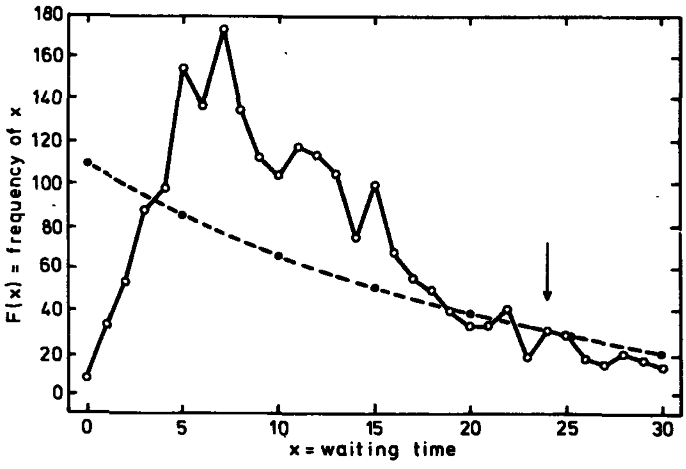


Fig. 6.5. Recurrence distributions for the slow condition with external representation of 25 alternatives in Experiment 2.

which suggests that the subjects tried to maintain a constant matching criterion. The cycling effect, however, as measured by the ratio of observed and theoretical values at the point $x = A - 1$ in the recurrence distributions (see Fig. 3.4), increase when A goes from two to six. When A exceeds a value of six the cycling effect decreases again. Fig. 6.5 shows that cycling is even completely absent when A = 25. In this case it is rather found that subjects tend to maintain a cycle period of six to eight responses, which is further support for the idea that only a limited set of alternatives can be taken into account.

§5. A FUNCTIONAL LIMITATION THEORY, AFTER ALL?

The above outlined speculative theory contains some elements that come very close to a memory theory, because of the assumption that matching of marginal probabilities is limited by the inability to recall more than six to eight previous responses. An illustration of this limitation is provided by a small control experiment.

Experiment 10

Procedure

Five subjects were requested to generate random sequences consisting of 1500 responses by pressing two, four or six pushbuttons, labelled with digits from one to six. (It was demonstrated in Experiment 3 that labelling the buttons does not change randomization behaviour). Rate of production was one response in two sec. During the sessions generation was stopped 28 times. The subjects were then requested to write down as many of the previous responses as they could. The intervals between two stops varied from 1.0 to 2.5 min. It was permitted to leave one or more places open for responses not recalled. Subjects were the same as in the control condition of Experiment 2. Therefore, it was possible in the conditions with four alternatives to check whether the additional task of recalling previous responses changed the subject's strategies, notwithstanding the explicit instruction that the emphasis was on randomization and not on recall. The total number of responses was

5 (subjects) x 3 (number of alternatives) x 1.500 (responses) = 22.500.

Results and discussion

The ϕ_n-scores, calculated for the four-alternative case (Fig. 6.6) show that the strategy did not noticeably change as a result of the re-

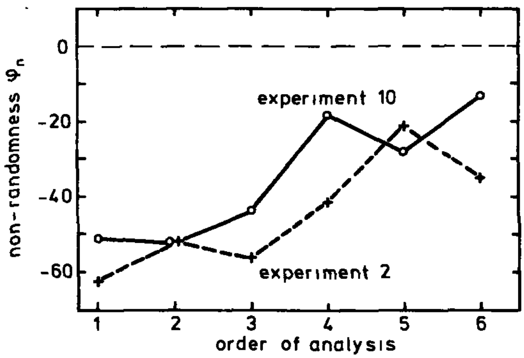


Fig. 6.6. Comparison of ϕ_n-scores in the four-alternative sequences of Experiment 10 and the corresponding sequences in the control condition of Experiment 2.

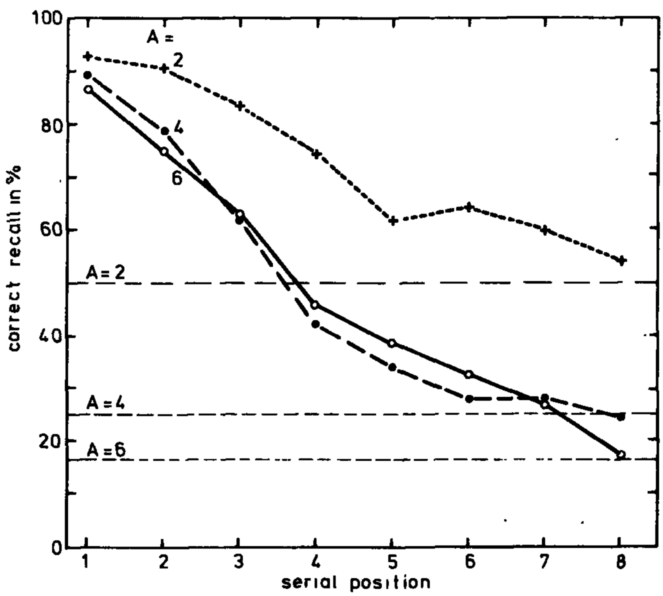


Fig. 6.7. Recall scores in Experiment 10, as a function of the serial position (1 = most recent response).

call task ($F_{1,5} = 2.13$). This was also true for the interactions with order of analysis ($F_{5,25} < 1$) and with subjects ($F_{5,25} = 1.51$).

The recall scores, presented in Fig. 6.7, reveal that subjects were not able to recall more than seven responses. The approach to chance performance occurs at the response eight places back for each value of A. A sign test showed that recall is significantly better than chance for the five last responses.

Apart from the fact that subjects might have introduced some undetected systematic trends in order to facilitate recall, there is the possibility that the instruction to recall induced subjects to memorize more responses than in a normal randomization experiment. Therefore the span of seven recalled responses should be considered as an upper limit. A possible enlargement of this span may be obtained when subjects try to recall which items did not occur in the past sequence (missing item span: Buschke, 1963). The cycling data for A=25 do not support this idea.

Conclusion

The recall of previous responses is limited at most to the last seven responses.

Notwithstanding the fact that the limitation of memory is supposed to be an important factor in the origin of sequential response bias, it is not justified to call the theory outlined in the previous section a strict memory theory. This may be best demonstrated from the effect of the number of alternatives. A strict memory theory predicts that there should exist no relation between A and the degree of sequential response bias, if the memory span is independent of the number of alternatives. The matching hypothesis, in turn, states that sequential response bias will increase with A, especially when recall is independent of A, because of the fact that a constant tendency for matching within a constant interval will increasingly limit the frequency of the separate alternatives in that interval, when A increases. In the same way an explanation for the effect of providing an artificial memory can be formulated. The artificial span did not increase the matching interval beyond the limit of seven responses, but only facilitated matching by inducing a 100% recall. As a consequence, matching can be more strict, especially in the difficult conditions, *i.e.* with many alternatives.

The proposed tentative theory cannot be classified as a strict theory on functional limitations, because matching is not a separate mental function. Rather, a number of mental functions is involved among which are memory and attention. Therefore the predictions of both negative memory and negative attention theory may be also valid for the matching theory. Although the results obtained in Experiments 1 to 7 are not in contradiction with this theory it would be precipitate to consider these data as a plain confirmation, since many other theories may predict similar results.

The present speculations are committed to paper with a great deal of hesitation, because much effort and experimental ingenuity will be required in order to find experimental paradigms which allow for a conclusive test. The single elements of the theory, however, constitute a sound base which should be an indispensable part of any theory of sequential response bias.

SUMMARY

This study is mainly devoted to the phenomenon of sequential response bias in randomization experiments. When a human subject is asked to produce a random sequence of responses, he continuously shows biases in favour of certain alternatives, depending on his previous responses. Hence man appears to be a bad randomizer, and it is the aim of this study to trace some backgrounds of this phenomenon. The experiments will be limited to situations in which normal human adults are explicitly instructed to produce a random sequence of responses, without any stimulus or feedback, whereas the sequences contain at least 100 responses.

In Chapter 1 a survey of literature is presented. An analysis of 15 experimental studies shows a striking divergence of experimental procedures, disagreement among experimental results and consequently a number of often contradictory theories. The *experimental procedures* differ with respect to the number of alternatives to be randomized, the nature of the alternatives, the sequence length, the presentation mode of the choice-set, the availability of previous responses and the rate of production. Apart from these factors little standardization is found concerning the criterion for calling a sequence random or non-random. Here, the problems center around two questions: how is a sequential dependency to be measured and how many previous responses should be taken into account. The *experimental results* only agree with respect to the existence of sequential response bias but the nature and size of the effect appear to be still under discussion. The existing *theories* can be divided in two main groups: theories that assume an incorrect "concept of randomness", and

theories that attribute sequential response bias to the limiting effect of some psychological function.

In Chapter 2 a method is proposed for the measurement of non-randomness. This measure of non-randomness does not require the production of very long sequences, it provides independent values for separate orders of analysis, it distinguishes between negative recency (too many alternations) and positive recency (too many repetitions) and it is able to deal with sequences of different lengths and different numbers of alternatives. As all other measures of non-randomness it carries some danger of overlooking certain effects of sequential dependency.

In Chapter 3 an introductory experiment is described. The results show a very consistent effect of negative recency for A (= number of alternatives) ranging from 2 to 8. Two separate trends are revealed by recurrence and sampling distributions, *viz.* a tendency to follow a cycle with length A, and a tendency to match the frequencies of the various alternatives within an interval of about seven responses. In addition, it appears that sequential response bias neither increases nor decreases as a function of rate of production, which is in contradiction with many theories. On the other hand sequential response bias increases with the number of alternatives. Finally, individual differences are limited to first and second order effects.

In Chapter 4 sequential response bias is studied as a function of various experimental conditions. Predictions are tested of five different theories that attribute sequential response bias to the operation of some function, *e.g.* memory, attention. The experimental conditions studied are for the greater part the same as listed in Chapter 1. The major outcome is that sequential response bias changes with a change in external experimental conditions, which explains why experimental evidence in the literature has been so inconclusive. None of the five theories can predict these changes in a consistent way.

Chapter 5 is devoted to a related topic: subjective randomness in judgement experiments. Theories that assume an incorrect concept of randomness predict that in a judgement situation subjects will mistake certain biased sequences for random ones. Since the literature on this issue is unsystematic and confusing, it was felt necessary to include some experiments on this topic. The experiments reveal that subjective randomness of at least the first three orders is biased in the direction of negative recency. Yet the concept theory is only partially supported since results of 110 subjects yield a rather weak relation between subjective randomness in judgement and sequential response bias in a randomization experiment. The relation is limited to the first order effect of sequential response bias, which is further illustrated by the fact that unlearning subjective randomness changes only first order effects of sequential response bias. The general conclusion of this chapter is that the concept theory is no more capable of accounting for sequential response bias than the theories that attributed the effects to a limitation of a psychological function.

In Chapter 6, finally, it is attempted to reconcile some elements of the two main theoretical directions. It would be premature to present a complete theory of sequential response bias, but it is certainly possible to list a number of elements that such a theory should contain. Firstly the dual character of sequential response bias is stressed: some tendencies are limited to low order effects only, as there are: significant inter-individual and intra-individual differences and the relation between sequential response bias and subjective randomness. Consequently, two separate components are distinguished: an individual component and a task component. The low order individual component can be characterized by the runlength distribution without introducing any higher order effects. The task component, at the other hand, contains two separate unrelated effects, *viz.* matching of marginal frequencies within an interval of about seven responses and a tendency to use each alternative once in a cycle of A responses.

Finally a brief outline of a tentative theory is presented.

SAMENVATTING

Deze studie is hoofdzakelijk gewijd aan het verschijnsel van sequentiële antwoordvoorkeur bij het produceren van toevallige reeksen. Wanneer een proefpersoon wordt gevraagd om een toevallige antwoordreeks te produceren treedt er wisselende voorkeur voor bepaalde antwoordmogelijkheden op, die afhangt van de vorige antwoorden. Het blijkt dus dat mensen niet in staat zijn om toevallige antwoordreeksen te produceren. Het doel van deze monografie is om de achtergronden van dit verschijnsel wat duidelijker te maken. De experimenten zijn beperkt tot situaties waarin normale volwassen proefpersonen de uitdrukkelijke opdracht krijgen om een toevallige reeks van antwoorden te produceren. Er is geen enkele stimulus of terugmelding, terwijl de lengte van de reeks tenminste 100 antwoorden bedraagt.

In Hoofdstuk 1 wordt een overzicht gegeven van de bestaande literatuur. Een analyse van de 15 belangrijkste experimenten leert dat een veelheid van experimentele procedures is toegepast, dat de experimentele resultaten weinig onderlinge overeenstemming vertonen en, als gevolg daarvan, dat een aantal theorieën is ontwikkeld die elkaar veelal tegenspreken. De *experimentele procedures* verschillen op het punt van het aantal alternatieven waaruit de toevallige reeks moet worden opgebouwd, de aard van de alternatieven, de reekslengte, de wijze waarop de set alternatieven wordt aangeboden, de beschikbaarheid van vorige antwoorden en de productiesnelheid. Behalve deze factoren is ook weinig standaardisering aanwezig met betrekking tot het criterium om een reeks toevallig of niet-toevallig te noemen. De problemen concentreren zich hierbij op twee vragen: hoe moet een sequentiële afhankelijkheid worden gemeten, en hoeveel voorgaande antwoorden moeten in rekening worden gebracht. Hoewel alle *experimentele resultaten* laten zien dat sequentiële antwoordvoorkeur bestaat, is de aard en de omvang van het effect een punt van discussie. De bestaande *theorieën* zijn in twee hoofdgroepen te verdelen: theorieën die veronderstellen dat de proefpersonen een foutief "concept van toeval" hanteren, en theorieën die het verschijnsel wijten aan de beperkende werking van een psychologische functie.

In Hoofdstuk 2 wordt een methode voor het meten van afwijkingen van toeval voorgesteld. Deze methode vereist geen uitzonderlijk lange antwoordreeksen; er worden onafhankelijke waarden voor de afzonderlijke ordes van analyse verkregen; er wordt onderscheid gemaakt tussen negatieve effecten (te veel afwisselingen) en positieve effecten (te veel herhalingen) terwijl reeksen van verschillende lengte of opgebouwd uit verschillende aantallen alternatieven vergelijkbare grootheden opleveren. Zoals bij alle andere maten voor afwijkingen van toeval, bestaat er een kans dat bepaalde effecten van sequentiële afhankelijkheid over het hoofd worden gezien.

In Hoofdstuk 3 wordt een inleidend experiment beschreven. De belangrijkste uitkomsten betreffen een zeer consistent negatief effect van antwoordvoorkeur voor waarden van A (aantal alternatieven) die variëren van 2 tot 8. Een nadere analyse toont twee afzonderlijke componenten, n.l. een neiging om een cyclus van A antwoorden te volgen en een neiging om de verschillende alternatieven even vaak te laten voorkomen binnen een interval van ongeveer zeven antwoorden. Bovendien blijkt dat sequentiele antwoordvoorkeur niet toe- of afneemt bij verandering van de productiesnelheid, hetgeen in tegenspraak is met de meeste theorieën. Anderzijds is er wel een duidelijk verband met het aantal alternatieven. Tenslotte blijkt ook dat individuele verschillen tussen proefpersonen beperkt zijn tot eerste- en tweede-orde effecten.

In Hoofdstuk 4 wordt een onderzoek gedaan naar de invloed die verscheidene experimentele condities op sequentiële antwoordvoorkeur hebben. Voorspellingen worden geverifieerd van vijf theorieën die sequentiële antwoordvoorkeur wijten aan de werking van een functie als geheugen of aandacht. De experimentele condities die worden onderzocht zijn goeddeels degene die in het eerste hoofdstuk reeds zijn genoemd. Het belangrijkste resultaat is wel dat sequentiële antwoordvoorkeuren veranderen wanneer de externe experimentele condities zich wijzigen. Dit verklaart waarom er tot nu toe zo weinig overeenstemming bestond tussen de resultaten van verschillende, uit de literatuur bekende, onderzoeken. Geen van de vijf theorieën is volledig in staat om de veranderingen te voorspellen.

Hoofdstuk 5 is gewijd aan een onderwerp dat verband houdt met sequentiële antwoordvoorkeuren: subjectieve toevalligheid in beoordelingsexperimenten. Een theorie die aanneemt dat proefpersonen een foutief concept van toeval hebben, voorspelt dat proefpersonen niet-toevallige reeksen met toevallige zullen verwarren, wanneer zij toevalligheid in een reeks moeten beoordelen. Omdat in de literatuur op dit punt weinig overeenstemming bestaat, leek het noodzakelijk om ook enige experimenten over dit onderwerp uit te voeren. Deze experimenten laten zien dat subjectieve toevalligheid zich ook uit door voorkeur voor afwisselingen. De concept-theorie wordt echter slechts gedeeltelijk bevestigd, aangezien in de resultaten van 110 proefpersonen een zwak verband tussen sequentiele antwoordvoorkeur en subjectieve toevalligheid aanwezig is. Het verband is bovendien nog beperkt tot effecten van de eerste orde. Dit wordt verder geïllustreerd door het gegeven dat het afleren van subjectieve toevalligheid alleen invloed heeft op eerste-orde effecten van antwoordvoorkeur.

De algemene conclusie van het hoofdstuk is dat de concept-theorie niet beter in staat is om sequentiële antwoordvoorkeur te verklaren dan theorieën die het effect wijten aan een functionele beperking.

In Hoofdstuk 6 is tenslotte een poging ondernomen om elementen van de twee theoretische hoofdgroepen te herenigen. De presentatie van een volledige theorie over sequentiële antwoordvoorkeuren zou voorbarig zijn, maar het is zeker mogelijk een aantal elementen te noemen die in een dergelijke theorie aanwezig zouden moeten zijn.

Ten eerste is nadruk gelegd op het tweeledige karakter van sequentiele antwoordvoorkeur: sommige aspecten zijn beperkt tot effecten van lagere orde, zoals inter-individuele en intra-individuele verschillen en het verband tussen antwoordvoorkeur en subjectieve toevalligheid. Daarom kan men twee componenten onderscheiden: een individuele component en een taak-component. De lagere-orde individuele component kan gekarakteriseerd worden door de verdeling van de lengte van ononderbroken series van identieke elementen, zonder dat hierdoor hogere-orde effecten worden geintroduceerd. De taak-component bestaat weer uit twee afzonderlijke onderling onafhankelijke effecten, te weten: het gelijkhouden van de frequentie van de verschillende alternatieven binnen een interval van ongeveer zeven antwoorden, en de neiging om ieder alternatief eenmaal in een cyclus van A antwoorden te gebruiken. Tenslotte wordt in het kort een voorlopige theorie geschetst.

FOOTNOTES

1. Part of this section is accepted for publication in the *Psychological Bulletin* (Wagenaar, 1972).

2. With permission of UPI © 1968, United Feature Syndicate Inc.,

3. In information theory the zero-order of dependency is usually called the first-order of redundancy (Attneave, 1959).

4. Production as fast as possible.

5. Negative when subjects are briefed about expected frequency of runs.

6. The essential parts of the following sections were reported before by Wagenaar & Truijens (1970).

7. Proof for the case A = 2. One of the two elements has a frequency N/2 + x, where x is an integer between -N/2 and +N/2.
   The expected frequency of repetitions is:

$$f_{exp}(rep) = \frac{(N/2 + x)^2}{N} + \frac{(N/2 - x)^2}{N} = N/2 + \frac{2x^2}{N}$$

   The last term is always zero or positive.

8. The random generator used in this program was of the additive type with a running memory of 16 numbers and a wordlength of 18 bits

(For a detailed description see Green, 1963).

9. Part of this section was published before in *Acta Psychologica* (Wagenaar, 1970a).

10. In a pilot study not reported here, it was found that the use of two hands yielded sequences that were dramatically different from those produced with one hand.

11. Part of this section was published before in *Acta Psychologica* (Wagenaar, 1971a).

12. Considering Table 5.1 it must be concluded that Cook meant the proportions to be 45-55, 25-75 and 10-90 for list 2, 3 and 4.

13. Some parts of this section were published before in *Acta Psychologica* (Wagenaar, 1970b).

14. There is a possible conflict in this instruction, connected with the paradox mentioned in §1, Chapter 2. All of the 2^{100} possible binary sequences have the same probability of occurrence and are therefore "random" to the same extent. However, the hypothesis that the generating source was an ideal randomizer is most easily accepted on the basis of sequences with conditional probabilities close to 0.5. Actually no subject ever had any problem with the given instruction.

15. The author is largely indebted to Drs. J. de Klerk of the Dutch Army Selection Centre for his generous cooperation.

LITERATURE

ADAMS, J.A., *Human Memory*, New York: McGraw-Hill, 1967.

ATTENEAVE, F., *Applications of information theory to psychology*, New York: Holt, 1959.

BADDELEY, A.D., Some factors influencing the generation of random letter sequences, *Med. Res. Council Rep., Appl. Psychol. Unit*, 1962, No. 422/62.

BADDELEY, A.D., The capacity for generating information by randomization, *Quart. J. exp. Psychol.*, 1966, *18*, 119-129.

BAKAN, P., Response tendencies in attempts to generate random binary series, *Amer. J. Psychol.*, 1960, *73*, 127-131.

BENDIG, A.W., The effect of reinforcement on the alternation of guesses, *J. exp. Psychol.*, 1951, *41*, 105-107.

BENNET, G.K., *Test of mechanical comprehension, Form AA, Manual*, New York: Psychological Corporation, 1947.

BRUNER, J.S., WALLACH, M.A., & GALANTER, E.H., The identification of recurrent regularity, *Amer. J. Psychol.*, 1959, *72*, 200-209.

BRUNSWIK, E., In defence of probabilistic functionalism: a reply. *Psychol. Rev.*, 1955, *62*, 236-242.

CHAPANIS, A., Random-number guessing behavior, *Amer. Psychologist*, 1953, *8*, 332.

COHEN, J., & HANSEL, C.E.M., The idea of a distribution, *Brit, J. Psychol.*, 1955, *66*, 111-121 (a).

COHEN, J., & HANSEL, C.E.M., The idea of independence, *Brit. J. Psychol.*, 1955, *66*, 178-190 (b).

CONRAD, R., & HULL, A.J., Information, acoustic confusion and memory span, *Brit. J. Psychol.*, 1964, *55*, 429-432.

CONRAD, R., & HULL, A.J., Input modality and the serial position curve in short-term memory, *Psychonomic Science*, 1968, *10*, 135-136.

COOK, A., Recognition of bias in strings of binary digits, *Perc. mot. Skills*, 1967, *24*, 1003-1006.

COURNOT, A.A., *Mathematical Principals of the Theory of Wealth*, (Bacon Translation), 1929, Ch. 18.

CROLL, W.L., Children's response alternation as a function of stimulus duration, inter-trial interval and trials, *Psychonomic Science*, 1966, *6*, 247-248.

CRONBACH, L.J., *Essentials of psychological testing*, 2nd ed., New York: Harper, 1960.

CROSSMAN, E.R.F.W., Entropy and choice times: the effect of frequency unbalance on choice responses, *Quart. J. exp. Psychol.*, 1953, *5*, 41-51.

CROWDER, R.G., & MORTON, J., Precategorical acoustic storage (PAS), *Perception & Psychophysics*, 1969, *5*, 363-373.

DAY, W.F., Serial non-randomness in auditory differential thresholds as a function of interstimulus interval, *Amer. J. Psychol.*, 1957, *70*, 238-242.

DERKS, P.L., & HOUSE, J.L., Effect of event run structure on prediction of recursive binary sequences, *Psychol. Rep.*, 1965, *17*, 447-456.

DERKS, P.J., & HOUSE, J.L., The prediction and acquisition of sequential patterns of binary events, *Brit. J. Psychol.*, 1970, *61*, 23-31.

DOORNE, H. van, & SANDERS, A.F., PSARP, a programmable signal and response processor, *Behav. Res. Meth. & Instr.*, 1968, *1*, 29-32.

FERNBERGER, S.W., On the relation of the method of just perceptible differences and constant stimuli, *Psychol. Monogr.*, 1913, *14* (4, whole No. 61).

GALANTER, E.H., & SMITH, W.A.S., Some experiments on a simple thought problem, *Amer. J. Psychol.*, 1958, *71*, 359-366.

GARNER, W.R., *Uncertainty and structure as psychological concepts*, New York: Wiley, 1962.

GEMELLI, A., & ALBERONI, F., Experimental studies of the concept of chance, *J. Gen. Psychol.*, 1961, *65*, 3-24.

GOODFELLOW, L.D., A psychological interpretation of the results of the Zenith radio experiments in telepathy, *J. exp. Psychol.*, 1938, *23*, 601-632.

GOODFELLOW, L.D., The human element in probability, *J. Gen. Psychol.*, 1940, *23*, 201-205.

GOODNOW, J.J., & PETTIGREW, T.F., Effect of prior patterns of experience upon strategies and learning sets, *J. exp. Psychol.*, 1955, *49*, 381-398.

GREEN, B.F., *Digital computers in research*, New York: McGraw-Hill, 1963.

GREEN, R.T., Factors affecting inductive prediction, *Brit. J. Psychol.*, 1958, *49*, 97-105.

HAKE, H.W., & HYMAN, R., Perception of the statistical structure of a random series of binary symbols, *J. exp. Psychol.*, 1953, *45*, 64-75.

JEFFREY, W.E., & COHEN, L.B., Response tendencies of children in a two-choice situation, *J. exp. Child Psychol.*, 1965, *2*, 248-254.

KELLER, L., Run structure and the learning of periodic sequences, *Unpublished doctoral dissertation*, Indiana Univer., 1963. Cited by Restle (1967).

LESTER, D., Response alternation by rates in mazes, *J. of Psychol.*, 1968, *70*, 51-53.

LIEBERMANN, B., How people predict three balls fall into three cells, *Psychol. Rep.*, 1963, *13*, 711-716.

LINCOLN, R.S., & ALEXANDER, L.T., Preferred patterns of motor and verbal responses, *J. exp. Psychol.*, 1955, *50*, 106-112.

LUDVIGSON, H.W., Response units in the prediction of simple event patterns, *J. exp. Psychol.*, 1966, *72*, 355-360.

MACKWORTH, J.F., Paced memorizing in a continuous task, *J. exp. Psychol.*, 1959, *58*, 193-198.

MANLEY, S., & MILLER, F.D., Factors affecting children's alternation and choice behaviors, *Psychonomic Science*, 1968, *13*, 65-66.

MILLER, F.D., DEE-WANG, TU, J., MOFFAT, G.H., & MANLEY, S., Children's response alternation as a function of stimulus duration, age and trials, *Psychonomic Science*, 1969, *15*, 199-200.

MILLER, G.A., Note on the bias of information estimates. In Quastler, H., (ed.), *Information Theory in Psychology*, Glencoe, Illinois: Free Press, 1955.

MITTENECKER, E., Perseveration und Persönlichkeit, I u. II, *Z. exp. ang. Psychol.*, 1953, *1*, 5-31, 265-284.

MITTENECKER, E., Die Analyse "zufälliger" Reaktionsfolgen, *Z. exp. ang. Psychol.*, 1958, *5*, 45-60.

MOOD, A.M., The distribution theory of runs, *Annals of Math. Stat.*, 1940, *11*, 367-392.

MOSS, S.M., Probe memory for vocalized lists, 1971, *in preparation.*

MURDOCK, B.B., Effects of a subsidiary task on short-term memory, *Brit. J. Psychol.*, 1965, *56*, 413-419.

MURRAY, D.J., Vocalization-at-presentation and immediate recall, with varying recall methods, *Quart. J. exp. Psychol.*, 1966, *18*, 9-18.

NICKS, D.C., Prediction of sequential two-choice decisions from event runs, *J. exp. Psychol.*, 1959, *57*, 105-114.

OTIS, A.S., *Otis Quick-Scoring mental ability tests*, New York: World Book Company, 1954.

PETERSON, C.R., & BEACH, L.R., Man as an intuitive statistician, *Psychol. Bull.*, 1967, *68*, 29-46.

PETERSON, C.R., DUCHARME, W.M., & EDWARDS, W., Sampling distributions and probability revisions, *J. exp. Psychol.*, 1968, *76*, 236-243.

PÖPPEL, E., Signifikanz-Artefakte in der experimentellen Parapsychologie, *Z. Parapsycholog. Grenzgeb. Psychol.*, 1967, *10*, 63-72.

POLLACK, I., & JOHNSON, L., Continuing memory span for digits, *Percept. mot. Skills*, 1963, *17*, 731-734.

POSNER, M.I., Immediate memory in sequential tasks, *Psychol. Bull.*, 1963, *60*, 333-349.

POSNER, M.I., Rate of presentation and order of recall in immediate memory, *Brit. J. Psychol.*, 1964, *55*, 303-306.

RABINOWITZ, F.M., Sequential dependencies and children's stimulus alternation, *Psychonomic Science*, 1969, *15*, 189-191.

RATH, G.J., Randomization by humans, *Amer. J. Psychol.*, 1966, *79*, 97-103.

REICHENBACH, H., *The theory of probability: An enquiry into the logical and mathematical foundations of the calculus of probability*, Berkeley: Univ. California Press, 1949.

RESTLE, F., *Psychology of judgment and choice*, New York: Wiley, 1961.

RESTLE, F., Grammatical analysis of the prediction of binary events, *J. verb. Learn. verb. Behav.*, 1967, *6*, 17-25.

RIEBER, M., Response alternation in children under different reinforcement schedules, *Psychonomic Science*, 1966, *4*, 149-150.

ROSS, B.M., Randomization of a binary series, *Amer. J. Psychol.*, 1955, *68*, 136-138.

ROSS, B.M., & LEVY, N., Patterned predictions of chance events by children and adults, *Psychol. Rep.*, 1958, *4*, 87-124.

ROSS, B.M., & WEINER, S.G., On making a random pattern, *Percept. mot. Skills*, 1963, *17*, 587-600.

ROUTH, D.A., "Trace strength", modality and the serial position curve in immediate memory, *Psychonomic Science*, 1970, *18*, 355-357.

SANDERS, A.F., *De psychologie van de informatieverwerking*, Arnhem: Van Longhum Slaterus, 1967.

SANDERS, A.F., Short-term memory for spatial positions, *Ned. Tijdschr. Psychol.*, 1968, *23*, 1-15.

SANDERS, A.F., & BORSELEN, J.W. van, Continuing memory and information processing, *J. exp. Psychol.*, 1966, *71*, 844-848.

SENDERS, V.L., Further analysis of response sequences in the setting of a psychophysical experiment, *Amer. J. Psychol.*, 1953, *66*, 358-374.

SENDERS, V.L., & SOWARDS, A., Analysis of response sequences in the setting of a psychophysical experiment, *Amer. J. Psychol.*, 1952, *65*, 358-374.

SKINNER, B.F., The processes involved in the repeated guessing of alternatives, *J. exp. Psychol.*, 1942, *30*, 495-503.

SPENCE, D.P., WOLITZKY, D.L., & PEZENIK, J., Random generation of digits as a measure of attention: relation to text redundancy and recall, *J. verb. Learn. verb. Behav.*, 1969, *8*, 9-15.

TERAOKA, T., Some serial properties of "subjective randomness", *Japan. Psychol. Res.*, 1963, *5*, 120-128.

TUNE, G.S., Response preferences: a review of some relevant literature, *Psychol. Bull.*, 1964, *61*, 286-302 (a).

TUNE, G.S., A brief survey of variables that influence random generation, *Percept. mot. Skills*, 1964, *18*, 705-710 (b).

VITZ, P.C., & TODD, T.C., A model of learning for simple repeating binary patterns, *J. exp. Psychol.*, 1967, *75*, 108-117.

VITZ, P.C., & TODD, T.C., A coded element model of the perceptual processing of sequential stimuli, *Psychol. Rev.*, 1969, *76*, 433-449.

WAGENAAR, W.A., Sequential response bias in psychophysical experiments, *Perception & Psychophysics*, 1968, *3*, 364-366.

WAGENAAR, W.A., Subjective randomness and the capacity to generate information. In: Sanders, A.F., (ed.), *Attention and performance III, Acta Psychol.*, 1970, *33*, 233-242 (a).

WAGENAAR, W.A., Appreciation of conditional probabilities in binary sequences, *Acta Psychol.*, 1970, *34*, 348-356 (b).

WAGENAAR, W.A., Serial non-randomness as a function of duration and monotony of a randomization task, *Acta Psychol.*, 1971, *35*, 78-87 (a).