71-23,876

SIEGEL, William, 1943-
RETENTION-INTERVAL EFFECTS IN THE METHOD OF ABSOLUTE JUDGMENT.

The University of Michigan, Ph.D, 1971
Psychology, experimental

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

RETENTION-INTERVAL EFFECTS IN THE METHOD OF ABSOLUTE JUDGMENT

by

William Siegel

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
1971

Doctoral Committee:

Professor J.E. Keith Smith, Chairman
Professor Theodore G. Birdsall
Associate Professor Daniel J. Weintraub
Professor Robert B. Zajonc

PLEASE NOTE:

Some pages have indistinct print. Filmed as received.

UNIVERSITY MICROFILMS.

ABSTRACT

RETENTION-INTERVAL EFFECTS IN THE METHOD OF ABSOLUTE JUDGMENT

by

William Siegel

Chairman: J.E. Keith Smith

In the absolute-judgment paradigm, the retention interval varies from trial to trial. The retention interval is defined as concurrent changes in both the time and the number of intervening trials between successive occurrences of the same item. Moreover, the average retention interval is dependent upon the number of items being judged, and is thus confounded with stimulus information. A series of five experiments was carried out to investigate the role of memory in the identification of pitch. It was hypothesized that the absolute judgment of unidimensional stimuli is characterized by rapid forgetting, and that the well known channel-capacity phenomenon is a manifestation of this poor retention.

Experiment I was similar to the classic study of Pollack (1952), which demonstrated a channel-capacity effect in the absolute judgment of pitch, except that separate confusion matrices were generated for each retention interval. The frequency range from 700-1149 Hz was divided into nine equal-log steps, and the subjects judged sequences of the resulting nine tones with feedback. Subsets of five and seven alternatives were judged as well. As hypothesized, all six subjects showed extremely rapid forgetting in this study. At the briefest retention interval used here (2 sec./0 intervening trials), the performance of all subjects was virtually error-free. But after only one or two intervening trials, performance dropped to an asymptotic level which was similar to that thought to be typical of the

absolute-judgment task. It was concluded that, if there is a channel capacity on the identification of unidimensional stimuli, it is a limitation on memory.

Experiments II and III attempted to determine the limits of absolute-judgment accuracy under conditions where subjects were not required to store pitch information over time, and a modified task was developed which allowed the use of very brief retention intervals. With the new procedure, as in Experiment I, subjects showed rapid forgetting as the number of intervening items was increased. By the time just one intervening judgment trial occurred, accuracy dropped to asymptote. Moreover, with the briefest retention interval used in these studies (.00 sec./0 intervening trials), it was necessary to reduce the separation between adjacent tones to 1.3 Hz in order for the subjects to make a moderate number of errors, even with as many as nine alternative tone frequencies. Thus, when the retention interval is made comparable, absolute-judgment performance appears to be as precise as that found in relative-judgment tasks.

An additional purpose of Experiment III was to test several competing theories of the "end-anchoring" phenomenon, the fact that items at the ends of the stimulus continuum are judged more accurately than items in the middle. A review of the literature supported the hypothesis that end effects come about because subjects forget the middle items more rapidly. However, it was found here that end-anchoring effects were just as pronounced on trials where no mnemonic storage was required as at longer retention intervals. These results provide no support for a memory theory of the end-anchoring phenomenon, such as Eriksen and Hake's <u>subjective-standard</u> hypothesis.

In Experiment IV, an attempt was made to determine the degree to which retention effects in absolute judgment depend upon the mere passage of time or upon the interfering effects of intervening trials. When the stimulus repeated after no intervening trials, some forgetting occurred as the retention interval increased from 0-6 sec. In addition to this temporal decay, there was a large decrease in retention when the subject was required to listen to and judge other tones. Performance accuracy was reduced to an asymptotic level after the occurrence of just one intervening trial.

A final experiment examined retention-interval effects under conditions where feedback was omitted. It was hypothesized that feedback influences the decision rule upon which the subject bases his overt response, but not the information about pitch which is available to the subject. Four predictions derived from this hypothesis were tested in Experiment V, and each was confirmed.

ACKNOWLEDGEMENTS

It is with pleasure that I acknowledge a debt of gratitude to Profs. Keith Smith, Daniel Weintraub, and Robert Zajonc. The encouragement and good humor provided by these gentlemen did much to ease my passage through graduate school, and each has contributed materially to this dissertation.

I would also like to thank Dr. Sylvan Kornblum for providing laboratory facilities and financial support for this project, and Miss Barbara Friedman for her assistance with computer programming. The research reported here was carried out while the author held a Rackham Predoctoral Research Fellowship from the University of Michigan.

Above all, I must thank my wife, Jane Anna Siegel, for her aid and advice. Many of the arguments expressed in this paper were developed in discussions with Jane, and it is to her that I dedicate this paper.

TABLE OF CONTENTS

		Page
LIST OF TABLES		v
LIST OF FIGURES		vi
CHAPTER		
I.	INTRODUCTION	1
	The Sensory-Capacity Hypothesis	3
	The Memory-Capacity Hypothesis	4
	Partitioning of Sensory and Mnemonic Variance	5
	Sensory and Memory Effects in Comparison Tasks	5
	Sensory and Memory Effects in AJ	8
	The Effect of Number of Alternatives (N)	8
	The Effect of Ordinal Position	10
	Feedback	19
	Conclusion	19
II.	EXPERIMENTAL EVIDENCE	21
	Preliminary Experiment: Retention-Interval Effects in AJ	21
	Method	21
	Subjects	21
	Apparatus and Procedure	21
	Results	23
	Retention Interval	23
	Number of Alternatives	23
	Ordinal-Position Effects	25
	The Modified Method of Absolute Judgment	27
	General Procedure for Experiments II-V	29
	Experiment II: Retention-Interval Effects (Continued)	30
	Method	31
	Subjects	31
	Procedure	31
	Results	31
	Retention Interval	31
	Ordinal-Position Effects	31

Page

Experiment III: Minimizing Memory Effects in AJ 32
Method 34
Subjects 34
Practice Sessions 34
Experimental Conditions 35
Results 35
Retention Interval 36
Ordinal-Position Effects 36
Number of Alternatives 41
Experiment IV: Time and Intervening Trials as Determinants of Forgetting 41
Method 42
Subjects 42
Procedure 42
Results 42
ExperimentV: The Role of Feedback (FB) 44
Method 46
Results 46
Retention Interval 46
Feedback and the Decision Rule 46
Theoretical Analyses 49
Guessing Model 49
Forced-Choice Detection Model 50
Information Model 53
Information Theory and the Retention Interval 55

III. DISCUSSION 61

Retention-Interval Effects 61
The Effect of Number of Alternatives 64
The Effect of Ordinal Position 66
The Effect of Feedback 67
Conclusion 68

APPENDIX 70

Confusion Matrices for Repetition and Nonrepetition Trials, Experiment III.

BIBLIOGRAPHY 76

LIST OF TABLES

		Page
1.	Percent correct as a function of the retention interval: Experiment III.	37
2.	Proportion of trials in which subjects repeated the previous feedback, shown as a function of response category.	40

LIST OF FIGURES

		Page
1.	Increase in information transmission as a function of ordinal position in the stimulus set. An objective standard tone corresponded to either the first, fifth, or ninth ordinal position. (After Pollack, 1953)	13
2.	Accuracy of identification of tone frequencies for subjects given either "A-training" or normal AJ practice with feedback for all stimuli. (After Cuddy, 1968)	15
3.	Accuracy of identification of rectangles varying in height. The item in the first, eighth, or fifteenth ordinal position was a square. (After Weintraub, 1971)	17
4.	Identification accuracy as a function of the retention interval for each subject of Experiment I.	24
5.	Joint effects of the retention interval and the number of alternatives in Experiment I.	26
6.	Joint effects of the retention interval and ordinal position in Experiment I.	28
7.	Joint effects of the retention interval and ordinal position in Experiment II.	33
8.	Joint effects of the retention interval and ordinal position in Experiment III.	39
9.	The effects of time and number of intervening items on identification accuracy in Experiment IV.	43
10.	The effect of number of intervening items under feedback and no-feedback conditions.	47
11.	The effect of the number of alternatives being judged on (A) percentage of correct responses adjusted for "guessing", and on (B) the d' measure of detection theory.	51

Page

12. Joint effects of stimulus information and the retention interval on information transmitted in Experiment I. 57

13. Joint effects of stimulus information and the retention interval on information transmitted in Experiment III. 59

CHAPTER I: INTRODUCTION

The absolute-judgment (AJ) method has had a long history as a psychophysical task. In an early literature review, Wever and Zener (1928) traced the development of the AJ paradigm to the work of Fechner, Witmer, Major, and Cohn in the nineteenth century. The AJ method began as a variant of another psychophysical procedure, the well-known method of constant stimuli. The latter is a relative-judgment task in which two stimuli are presented on each trial and the subject is asked to compare them. With the AJ method the first stimulus, or standard was omitted, the items were presented one at a time in random order, and the subject was asked to base his response on an "absolute impression" of the stimulus. In practical terms, this meant that he rated the magnitude of his sensory experience with a set of discrete response categories, such as high - medium - low, or the numbers 1-5. Thus by the time of Wever and Zener's review, the AJ method was comparable to what we would now call a category-rating task.

The original focus of interest in the AJ paradigm was the psychophysical relation between the magnitude of sensation and changes in physical parameters of the external environment. With the advent of information theory, this task has been used to ask a somewhat different question: What is the largest number of stimuli that the human observer can categorize without error? The introduction of information theory led to several changes in the AJ method. First, the experimenter defined a "correct" mapping of responses onto stimuli. Second, the goal of the

subject was now to respond as accurately as possible, and informational feedback was often provided on every trial. These procedural modifications changed the emphasis in AJ from judgment to identification.

Miller (1956), Attneave (1959), and Garner (1962) have all reviewed AJ experiments in which the methodology of information theory was used. The most interesting finding is that when the stimuli vary along only one sensory dimension, identification performance is surprisingly poor. For a number of modalities, most subjects can identify only about 4-7 items without error. When the size of the stimulus set is further increased, the subjects begin to make confusions, but a measure of "information transmitted" (H_t) in their responses remains constant at a level equivalent to the perfect identification of 4-7 items. This asymptotic performance is referred to in the language of information theory as the channel capacity. The phenomenon of a fixed upper limit on H_t appears to have been demonstrated for a variety of sensory continua, including pitch (Pollack, 1952), loudness (Garner, 1953), hue (Chapanis and Halsey, 1956), size (Eriksen and Hake, 1955), and linear interpolation (Hake and Garner, 1951).

In relation to the thousands of words, faces, sounds, objects, and events which we can recognize both in and outside of the laboratory, the 4-7 stimuli which can be identified in the usual AJ experiment seems incredibly small. Unfortunately, the writers who have popularized the channel-capacity concept have been rather vague about its underlying mechanism. For example, in an influential paper Miller (1956) concluded that the limitation is "built into us either by learning or by the design of our nervous systems." Garner (1962) interprets the AJ data as demonstrating a "limitation of the perceptual process", and Attneave does

not speculate at all about the nature of the effect. In the remainder of this paper, I will discuss two plausible hypotheses concerning the channel capacity and will review evidence bearing on each one. Five experiments will be carried out in an attempt to demonstrate that the limitation on AJ accuracy results from poor perceptual memory for unidimensional stimuli.

The Sensory-Capacity Hypothesis. One possibility is that the sensory modalities act as information channels of limited capacity. When the input exceeds this capacity, the surplus information is lost. Thus, according to the sensory-capacity hypothesis, poor AJ performance results from a built-in upper limit on sensory resolving power. One is led to this hypothesis on the grounds that the AJ paradigm had its roots in psychophysics, and it is usually thought of as a sensory task. If this is the case, then it seems reasonable to assume that performance in the AJ task reflects for the most part the operating characteristics of the sensory systems.

A number of writers associate the pattern of errors in AJ with the discriminability of the stimuli. Wever and Zener (1928) and Pollack (1956) have argued that both AJ and discrimination tasks give roughly comparable measures of the subjects' sensitivity. Garner and Hake (1951) developed an equal-discriminability (ED) scale which derives measures of perceptual distance from confusions in AJ. According to Garner (1952), the ED procedure gives results which are directly related to the sensory scales based on discrimination data. As Torgerson (1958) points out, the ED technique is equivalent to the most restrictive form of Thurstone's (1927) theory of categorical judgment. Galanter and Messick (1961) used a less restrictive form of the Thurstone model to process AJ loudness data and attempted to relate the obtained scale values to a scale based on magnitude estimations of the same stimulus set. All of these examples

illustrate the common belief that the AJ method is primarily a sensory task.

On the other hand, Fitts and Posner (1967) suggest that something other than sensory variability must be responsible for poor AJ performance (this despite the fact that they discuss the AJ channel capacity in a section entitled "Sensory Capacities"). Their reasoning is based on the fact that, at least under some circumstances, subjects are able to differentiate much smaller stimulus differences in a comparison task than in AJ. In fact, the stimuli used in AJ are often spaced so far apart (e.g., Pollack, 1953) that they would be confused in a discrimination task only if the subject fell asleep. This last point makes it unlikely that the limitation on AJ accuracy results from a sensory channel capacity.

The Memory-Capacity Hypothesis. The main difference between AJ and discrimination paradigms is that the standard is omitted in the former task. In order to perform accurately in AJ, the subject may attempt to judge current input in relation to standards stored in memory. Several writers (Fitts and Posner, 1967; Holland and Lockhead, 1968; Holland, 1969) have suggested that the AJ channel capacity results in part from an inability to maintain precise mnemonic standards along a unidimensional continuum.

A further reason for suspecting that memory effects are important in AJ is that the retention interval is an uncontrolled variable in this task (Siegel and Siegel, 1968). In some cases the same stimulus is presented on two successive trials, and the subject must store sensory information in memory only during the interstimulus interval. More often however, repetitions of a given item are separated by a number of trials in which other stimuli are judged. In this case, the memory demands of the task are correspondingly greater. Thus the retention interval in AJ

involves concurrent changes in two variables: the time and the number of intervening judgment trials between successive presentations of the same stimulus.

If the series is random, then the probability of an item repeating after x intervening trials is equal to $p\ (1 - p)^x$, where p is the item's probability of occurrence. It is common practice to place restrictions on the randomness of the sequence, and this will influence the distribution of retention intervals somewhat. The fact remains, however, that the retention interval does vary in the AJ task even with such restrictions, and little is known about its effects.

Partitioning of Sensory and Mnemonic Variance

In deciding between competing theories of the channel capacity phenomenon, it would be helpful to have some information about the relative importance of sensory and memory effects in AJ. Much more progress has been made in partitioning sensory and mnemonic variance in comparison tasks. A brief review of the relevant literature on relative judgments is given in the hope that it will shed light on similar effects in AJ. A further question discussed below is whether variables known to be important in AJ, such as the number of alternatives, position in the stimulus set, and feedback influence the discrimination or retention of stimulus information.

Sensory and Memory Effects in Comparison Tasks

The discrimination or relative-judgment task becomes a recognition - memory paradigm if there is a time delay between the standard and comparison stimuli. Then the subject must base his decision on the

relationship between the perceived value of the comparison and the remembered value of the standard (Fechner, 1860; James, 1890). The retention interval has been varied in a number of studies investigating short-term memory for pitch (e.g., Harris, 1952; Bachem, 1954; Wickelgren, 1966; Wickelgren, 1969, Massaro, 1970a; Massaro, 1970b; Massaro, 1970c), loudness (Kinchla and Smyzer, 1967), the position of points of light (Kinchla and Smyzer, 1967), and kinesthetic position (Johnson, Chatelier, and Mocharnuk, 1969). In all of these experiments, the observers' sensitivity declined rapidly during the retention interval. For the pitch continuum, the only experiments which do not show forgetting are those in which a single standard is used (Harris, 1952; Wickelgren, 1969). Of course, when only one standard is employed the memory load on the subject is relatively light.

In the delayed-comparison task with small stimulus differences, the rate of forgetting is most rapid during the first few seconds after the offset of the standard. Moreover, this rapid forgetting takes the form of a spread of stimulus generalization along the unidimensional continuum. As the retention interval is increased, one must use stimuli which are less similar to one another in order to maintain a constant level of sensitivity (Shepard, 1958; Wickelgren, 1966; Kinchla and Smyzer 1967; Wickelgren, 1969).

When d'-type sensitivity measures are used in delayed-comparison tasks, performance appears to asymptote at a level above chance. Forgetting is too rapid at brief retention intervals and too slow at long retention intervals for good fit with a single exponential decay function. This last finding has led Wickelgren (1969) to argue that the memory trace has two components. The first, a short-term sensory trace, decays exponentially to zero with a time constant of 3-9 sec. In

addition to the short-term trace, Wickelgren feels that there is an intermediate-term trace which lasts for at least 40 sec. This trace may represent an imprecise verbal encoding of the stimulus. Several other writers (e.g., Kinchla and Smyzer, 1967; Durlach and Braida, 1969; Johnson et al., 1969) have also suggested that both sensory and verbal encoding are involved in psychophysical tasks.

The short-term trace appears to decay passively with time, since Kinchla and Smyzer, Johnson et al., and Massaro (1970a) all found rapid forgetting when the retention interval was blank. On the other hand, certain interference effects may occur as well. Massaro (1970a) found that short-term recognition memory for pitch was better with a blank retention interval than when the time between the standard and comparison was filled with noise or with pure tones. Moreover, when the interstimulus interval was filled with tonal stimuli, the number as well as the duration of the interpolated tones influenced retention (Massaro, 1970b). While short-term sensory memory appears to be sensitive to the nature of the retroactive stimuli, the effects described above do not seem to represent the associative interference which is thought to be important for verbal memory. Wickelgren (1966) found that performance was better if the interpolated tone was similar in pitch to the standard tone. The opposite would be expected if forgetting resulted from associative interference.

To summarize, relative-judgment performance reflects the discriminatory capacities of the sensory systems only when there is no temporal delay between the standard and comparison. If unidimensional stimuli are used in a delayed-comparison task, rapid forgetting occurs, and this forgetting is most rapid during the first few seconds after

the offset of the standard. Loss of information probably results from both passive temporal decay and from disruption of the trace when the subject attends to other stimulus events.

Sensory and Memory Effects in AJ

Pollack (1952) cites Stevens' conclusion that 350,000 tone frequencies can be differentiated with comparative-judgment procedures. Although this figure is inflated, Pollack was somewhat surprised to find that his subjects were capable of identifying only 4-5 tones without error in AJ. A number of other writers have also noted the great discrepancy between the sensitivity of relative judgments and the imprecision of AJ performance. The evidence reviewed here shows that relative-judgment accuracy is high only at brief retention intervals. If the subject is forced to store information in memory for even a few seconds, his sensitivity drops rapidly. This fact is notable, since on most AJ trials the retention interval is relatively long. If memory factors were made comparable it is possible that relative- and absolute-judgment performance would converge toward a common value. In any event, if one could devise a technique for minimizing memory variance in AJ, performance would more accurately reflect the discriminatory capacities of the sensory systems.

The Effect of Number of Alternatives (N)

When the size of the stimulus set is increased in AJ, the individual items are confused more frequently. Does this mean that, with increasing N, the individual items become less "distinctive" (Murdock, 1960), i.e., perceptually more similar, or does it mean that

the items are more difficult to remember? Most theorists would like to believe that the discrimination of stimulus differences is independent of the context in which they are judged. Thus a considerable effort has been expended to develop a theory whose sensitivity parameter is independent of N. Three models relevant to the AJ paradigm are discussed in some detail in Chapter III.

Siegel and Siegel (1968) have suggested that the effect of N results at least in part from memory factors since as N increases in the AJ paradigm, so does the average retention interval. It was shown earlier that with a random sequence, the probability of an item repeating itself after x intervening trials is equal to $p\ (1 - p)^{x}$, where p is the item's probability of occurrence. When the stimuli are equiprobable, as is usually the case in AJ, then the probability of a stimulus repeating after x intervening trials is specified by the following equation:

$$P\ (R_x) = 1/N\ \ (1 - 1/N)^{x} \qquad (1)$$

where N is the number of alternatives. The mean number of intervening items is equal to N - 1.

Since N and the mean retention interval are confounded, then at least some of the effect of N may result from changing the memory demands of the task. This would be the case if it could be demonstrated that AJ accuracy varies inversely with the retention interval, i.e., that forgetting occurs. If there is such a relation, then it would be desirable to vary N under conditions in which retention effects are held constant, or even better, where memory variance is minimal. This would be the first step towards determining whether N influences sensory discriminability, the rate of forgetting, both, or neither.

The Effect of Ordinal Position

In the AJ experiment, all members of the stimulus set are not judged with equal precision. Stimuli at the ends of the continuum are identified more accurately than those in the middle (e.g., Pollack, 1952; Garner, 1953; Eriksen and Hake, 1957; Alluisi and Sidorsky, 1958).

Data of Hartman (1954) demonstrate that the same tone frequency is confused much more often when it occurs in the middle of a stimulus set than when it is an end item. Thus the "end-anchoring" phenomenon is a genuine context effect. It depends upon ordinal position in the stimulus set rather than upon differences in discriminability along the physical continuum.

Theories of the Ordinal Position Effect. A number of writers have hypothesized about the nature of the ordinal-position effect. Of the current theories, perhaps the most simple is that subjects are biased toward the end responses, using them more frequently than the probability of occurrence of the end stimuli would warrant. Unfortunately, the relevant data are seldom published, and it is difficult to assess the degree to which ordinal-position effects obtained in previous studies are an artifact of response bias. However, an experiment of Holland (1969) shows striking end effects in AJ even though the marginal response distribution was virtually flat.

A second possibility is that ordinal position influences the perceived magnitude of the stimuli. This type of explanation assumes that the end stimuli are further apart on an underlying sensory continuum, and thus they are more salient, distinctive, or discriminable. While this theory seems plausible, there is no evidence for such context effects in the "sensory" scales based on most current scaling procedures.

Confusion scales derived from discrimination performance are generally logarithmic in form, while Stevens' direct-estimation procedures give rise to scales in which perceived magnitude is an exponential function of physical magnitude.

A third class of theories suggests that the ordinal-position effect results from an artifact of the response system in AJ. For example, Garner (1953) argued that the end stimuli are confused less often because errors can be made in only one direction at the ends and in two directions at the middle. This "unidirectional-error" hypothesis is also implicit in the Thurstone model of categorical judgment (Thurstone, 1927; Torgerson, 1958), at least in the simple form of the theory which is usually applied to AJ data.

Although the unidirectional-error hypothesis is quite simple and attractive, an experiment by Eriksen and Hake (1957) suggests that it is also incorrect. Eriksen and Hake used a series of hues in which the end-points were not clearly defined. Since extreme violet is more similar to extreme red than either of these hues is to, say, yellow, one can think of the hue continuum as forming a closed circle. This assumption seems reasonable in light of multidimensional scaling analyses of hue judgments (Shepard, 1962). The unidirectional-error hypothesis would not predict a bowed error distribution in this case, since there are no ends to the underlying continuum.

Eriksen and Hake had two groups of subjects judge a series of 20 Munsell hues, using the numbers 1-20 as responses. For one group, response "1" was assigned to a red stimulus, and for the second group the first response was assigned to a blue-green hue. For both groups, the remaining responses were assigned consecutively to the hues in the

set. The investigators found an ordinal-position effect in this experiment. But here, performance was most accurate for the stimuli at the ends of the response set, that is, for whichever stimuli were labelled 1 and 20. It is difficult to see how the unidirectional-error hypothesis would account for this result.

Eriksen and Hake put forward a theory of their own in which memory processes play an important part, and which they have named the subjective-standard hypothesis. According to this theory, some members of the stimulus set are stored in memory and used as standards for judging the remaining items. Typically, the subject selects the ends of the stimulus continuum for use as subjective standards. However, when the stimuli lack distinctive end-points, the subject may adopt a different mnemonic strategy. In Eriksen and Hake's experiment, for example, the subjects presumably stored the stimuli associated with the end responses for use as subjective standards.

Memory and the Ordinal Position Effect. Of the various theories outlined above, the subjective-standard hypothesis of Eriksen and Hake seems most plausible. According to this theory, the bowed ordinal-position curve comes about because the end items are stored more effectively in memory. If this analysis is correct, then operations which affect ease of retention will also change the shape of the ordinal-position curve. One such procedure is to provide the subject with known objective standards on every trial. Pollack (1953) used objective reference tones in a study involving absolute judgments of pitch. Nine tones varying in frequency comprised the judgmental series, and the reference tone was always the same frequency as one of the test tones. The results are shown in Figure 1. Improvements in AJ accuracy resulting

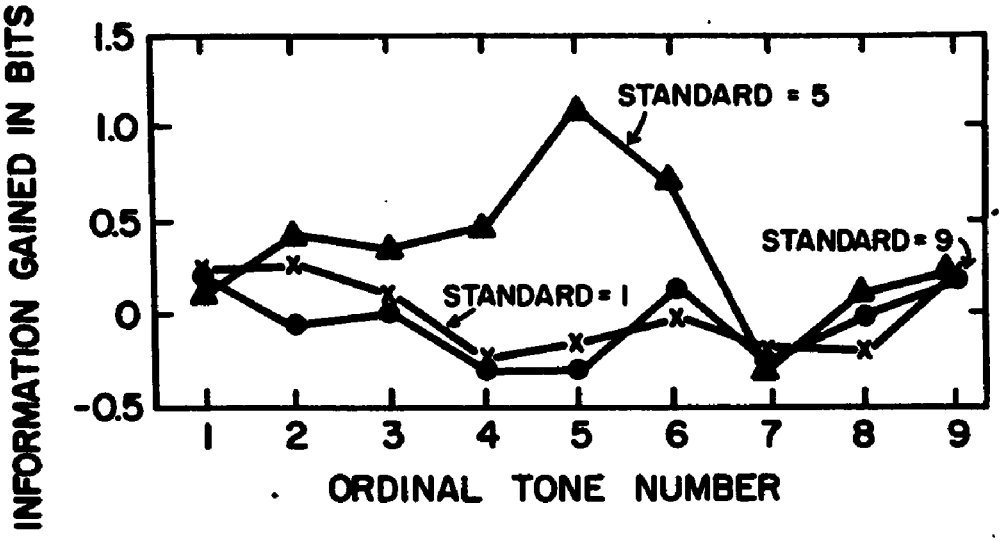


Fig. 1. Increase in information transmission as a function of ordinal position in the stimulus set. An objective standard tone corresponded to either the first, fifth, or ninth ordinal position. (After Pollack, 1953)

from the presence of the standard was found to depend upon the position of the standard in the stimulus set. The greatest gain occurred when the standard has a value near the middle of the series.

A second operation which should reduce the subject's memory load in the AJ task is extended training using members of the judgmental series. A study by Cuddy (1968) demonstrates that practice may produce dramatic alterations in the shape of the AJ ordinal-position curve. Two groups of subjects were given eight training sessions in which they judged 10 different tone frequencies. Note names were used as responses. One group was given typical AJ practice with feedback on all stimuli. The second group was given feedback on only one item, A_4, which was the middle frequency. Moreover, for the "A-training" group, the proportion of A_4's was progressively reduced from an initial level of .40 as the subjects became more proficient

The results of Cuddy's experiment are shown in Figure 2. On a pretest, both groups did equally well. But on a post-training test, the group given special A-training surpassed the other group in the identification of all tones, particularly A_4. A comparison of Cuddy's data with that of Pollack (Figure 1) suggests that A-training enabled subjects to encode A_4 into memory, so that it could be used as an acquired standard for judging other stimuli in the series.

The studies cited above demonstrate that the ordinal-position curve may be raised if the subject is given a memory crutch either by providing objective reference standards or by establishing subjective standards through training. A third procedure is to take advantage of the subject's prior learning history by using specific stimuli which have well-established associations in long-term memory. There are some

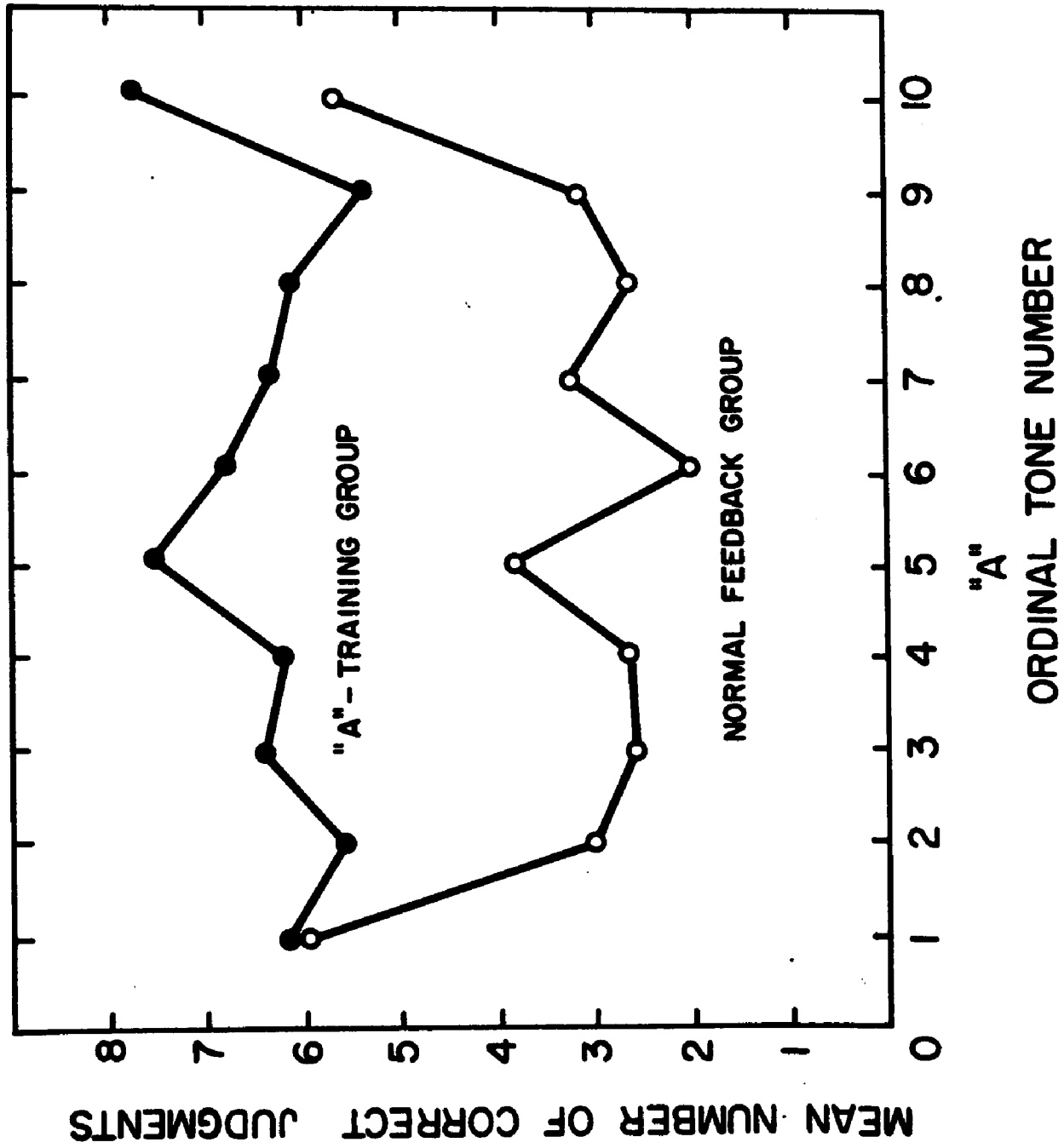


Fig. 2. Accuracy of identification of tone frequencies for subjects given either "A-training" or normal AJ practice with feedback for all stimuli. (After Cuddy, 1968)

continua for which subjects do not have overlearned associations. Most of us can name at least the vertical and horizontal when judging angle of inclination. There are distributions of color names associated with hue (Lenneberg, 1957; Beare, 1963), and these distributions depend on the culture one grows up in (Lenneberg, 1961). Individuals with "perfect pitch" use note names corresponding to the diatonic scale to identify tone frequencies.

It is interesting that the continua for which subjects have meaningful labels are characterized by irregular ordinal-position curves. Peaks or bumps in the performance curve have been observed in the judgment of hue (Eriksen and Hake, 1957; Volkman and Engen, 1961), angle of inclination (Muller, Sidorsky, Slivinske, Alluisi, and Fitts, 1955), and linear interpolation (Levett, 1952; Nash, 1964; Spitz, 1967), but are conspicuously absent for most other continua. Furthermore, unidimensional continua which are associated with names and labels also have the highest channel capacities.

An experiment by Weintraub (1971) provides an excellent example of meaningfulness effects in judgment. In a standard AJ situation, 15 rectangles varying only in height were used as stimuli with numbers as responses. One would expect to find the usual bow-shaped ordinal position curve under these conditions. However, in this experiment, one of the rectangles in each set was unique in that its height and width were equal. That rectangle could be uniquely labelled as a square. The experimenter did not bring the existence of the square to the subject's attention, and it was assigned a number response like any other rectangle.

Weintraub's data are shown in Figure 3. It is clear that the square tended to raise the ordinal position curve, and the greatest

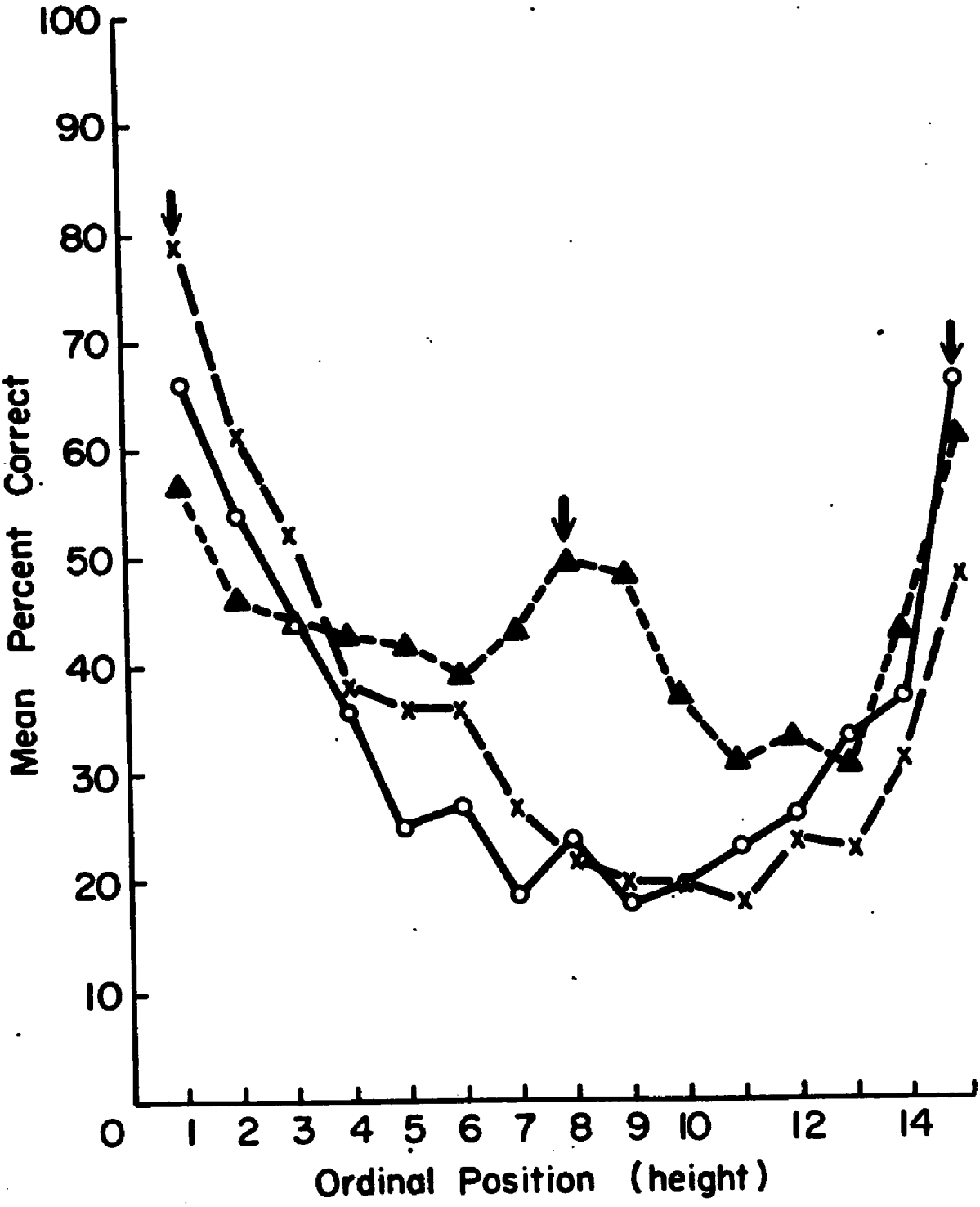


Fig. 3. Accuracy of identification of rectangles varying in height. The item in the first, eighth, or fifteenth ordinal position was a square. (After Weintraub, 1971)

improvement came when the square was the middle item. The results obtained by Weintraub bear a striking resemblance to those of Cuddy and Pollack. Taken together, the three studies suggest that subjective anchors may be provided in AJ either through training or by selecting stimuli which already have associations in long-term memory. These subjective anchors influence the ordinal-position effect in the same way as objective reference standards, by reducing the subject's memory load.

Finally, one can point to short-term memory effects in AJ which support the subjective-standard hypothesis. In the AJ paradigm, the subject judges a sequence of stimuli, and items which occur earlier in the series may be stored in memory to be used as standards for judging stimuli occurring later. Holland and Lockhead (1968) argue that the subject uses the immediately preceding stimulus as a subjective standard. In an experiment involving absolute judgments of loudness, they demonstrated that response variability on trial N depends upon the stimulus presented on trial N - 1. The more similar the stimuli on trials N and N - 1, the greater the reduction in response variance on trial N. Moreover, Holland and Lockhead found that the increased precision due to the preceding stimulus was greatest when the item on trial N - 1 was in the middle of the set. These findings fit in very nicely with the results of Pollack, Cuddy and Weintraub.

All in all, the evidence favoring a memory theory of the ordinal-position effect is quite solid. To test the theory critically, however, one would have to devise an AJ experiment in which the subject's reliance upon memory is minimal. If bowed error curves result from more rapid forgetting of the middle items, then when mnemonic variance is eliminated, the ordinal-position curve should be flat.

Feedback

As Gibson's (1953) review demonstrates, subjects perform less accurately when feedback is omitted in the AJ task. More recent papers confirm that without feedback the average error of responding is higher (Hartman, 1954; Ward and Lockhead, 1970), and information transmitted is lower (Hartman, 1954; Muller et al., 1955; Alluisi and Sidorsky, 1958). These findings could mean that feedback influences the ease with which subjects discriminate or retain information about sensory events. On the other hand, my hypothesis is that feedback operates on the decision rule used by the subject to map a set of overt response categories onto the internal sensory continuum. In particular, I suggest that with feedback, subjects adopt a response rule which will maximize accuracy according to the experimenter's criterion. When no knowledge of results is provided, the observers in effect provide their own feedback, and will attempt to be internally consistent. This possibility was suggested by Gibson (1953), who argued that judgments become more reliable in the absence of external reinforcement without becoming more veridical. In one of the following experiments I will attempt to demonstrate that, in the AJ task, subjects are just as consistent without feedback as they are accurate when correction is provided.

Conclusion

A number of questions remain about the role of the retention interval in AJ. One would like to have some information about how rapidly forgetting occurs and what proportion of the variance in the typical AJ experiment is memory variance. Does forgetting result from the interfering effect of intervening judgment trials or from the concurrent passage of time? Is the rate of forgetting dependent upon such factors

as the number of alternatives (stimulus information) and ordinal position? Do the results of absolute- and relative-judgment studies agree when memory effects are controlled? Evidence bearing on these questions is provided in the following series of five experiments, where an attempt was made to partition sensory and memory effects in AJ.

CHAPTER II: EXPERIMENTAL EVIDENCE

Preliminary Experiment: Retention-Interval Effects in AJ

The retention interval is an uncontrolled variable in AJ, and this experiment is a first attempt to assess the importance of its effects. Subjects judged sets of five, seven and nine tone frequencies presented in a random series. The frequency range and spacing, as well as the mode of feedback, were similar to those used in a well-known AJ pitch experiment showing the channel capacity effect (Pollack, 1952). An analysis program searched the data for trials corresponding to a given retention interval, and separate stimulus-response matrices were generated for each number of intervening items.

Method

Subjects. One male and five female students, aged 18 to 24, served in the experiment. Five subjects had volunteered for the paid subject pool and were given $1.75 per hour. JS served for science rather than for money. She had previously received 13 hours of AJ practice.

Apparatus and Procedure. A PDP-8 computer (Digital Equipment Corporation) was used to generate and program the stimuli, as well as to collect and analyze the responses. The computer's microsecond clock was used to time square waves which were amplified and delivered through Koss Pro-4 headphones at an amplitude of .12 v. RMS. The subjects sat in a quiet experimental room next door to the computer.

In any session, the subjects judged either five, seven or nine alternative tone frequencies. The set of nine alternatives was

constructed by dividing the range from 700-1149 Hz into nine steps which were spaced equally on a log scale. When fewer than nine tones were presented, the five and seven lowest frequencies of this set were used. The balancing of number of alternatives across sessions followed a Latin-square design.

The response system was a horizontal linear array of white plastic pushbuttons set into a black panel. The buttons were .6 in. square and were separated by a gap of .4 in. Responses were mapped onto stimuli consecutively so that the far-left button, labelled 1, corresponded to the lowest tone, and the far-right button, labelled 9, to the highest. When fewer than nine frequencies were used, the appropriate pushbuttons were masked.

Each block of trials began with a prerun in which the stimulus set was presented from low to high. This was followed by a test during which each frequency was presented 20 times. The order of presentation was random, except that each stimulus occurred equally often. This last restriction had a minute effect on the distribution of retention intervals.

The duration of each tone was 2 sec., but the stimulus was terminated by the subject's response if its latency was less than 2 sec. As soon as a choice was made, the correct pushbutton was illuminated to provide immediate feedback. Two sec. later, the next stimulus was presented for judgment. After a block of trials, the subjects were allowed to rest for five minutes. An experimental session lasted one hour. Each subject received nine blocks of trials for each of the three stimulus sets. The instructions were to be as accurate as possible, and to work at a brisk pace.

The sequence of stimuli and responses comprising each block of trials was recorded by the computer. A small amount of data in this and the following experiments was lost because of equipment failure.

Results

Retention Interval. In Figure 4, the percentage of correct responses for each subject, averaged across the three stimulus sets, is plotted as a function of the number of intervening items. On repetition trials (zero intervening items), the proportion of errors ranged from .00 for KM, JR, NC, and JS, to .07 for MM. This indicates that with the stimulus spacing used in this experiment, a spacing which is not atypical for the AJ paradigm, the subjects were capable of very accurate performance when the retention interval was brief. On repetition trials, there was virtually no variance to explain. However, as the retention interval increased, all six subjects showed substantial forgetting. There were large individual differences in the ability to identify tone frequency. An interesting trend is that the subjects with the best overall performance were able to store and utilize information about items further back in the series. For example, the three whose overall performance was lowest appeared to remember only the immediately preceding tone, while the other subjects showed less rapid forgetting. The most extensive intervening-items effect, as well as the best overall performance, was achieved by JS, who had practiced for 13 hours before participating in the experiment.

Number of Alternatives. It was demonstrated earlier that N is confounded with the mean retention interval. As N is increased, the probability of long retention intervals is increased as well. Since the accuracy of AJ performance is much poorer at long retention intervals

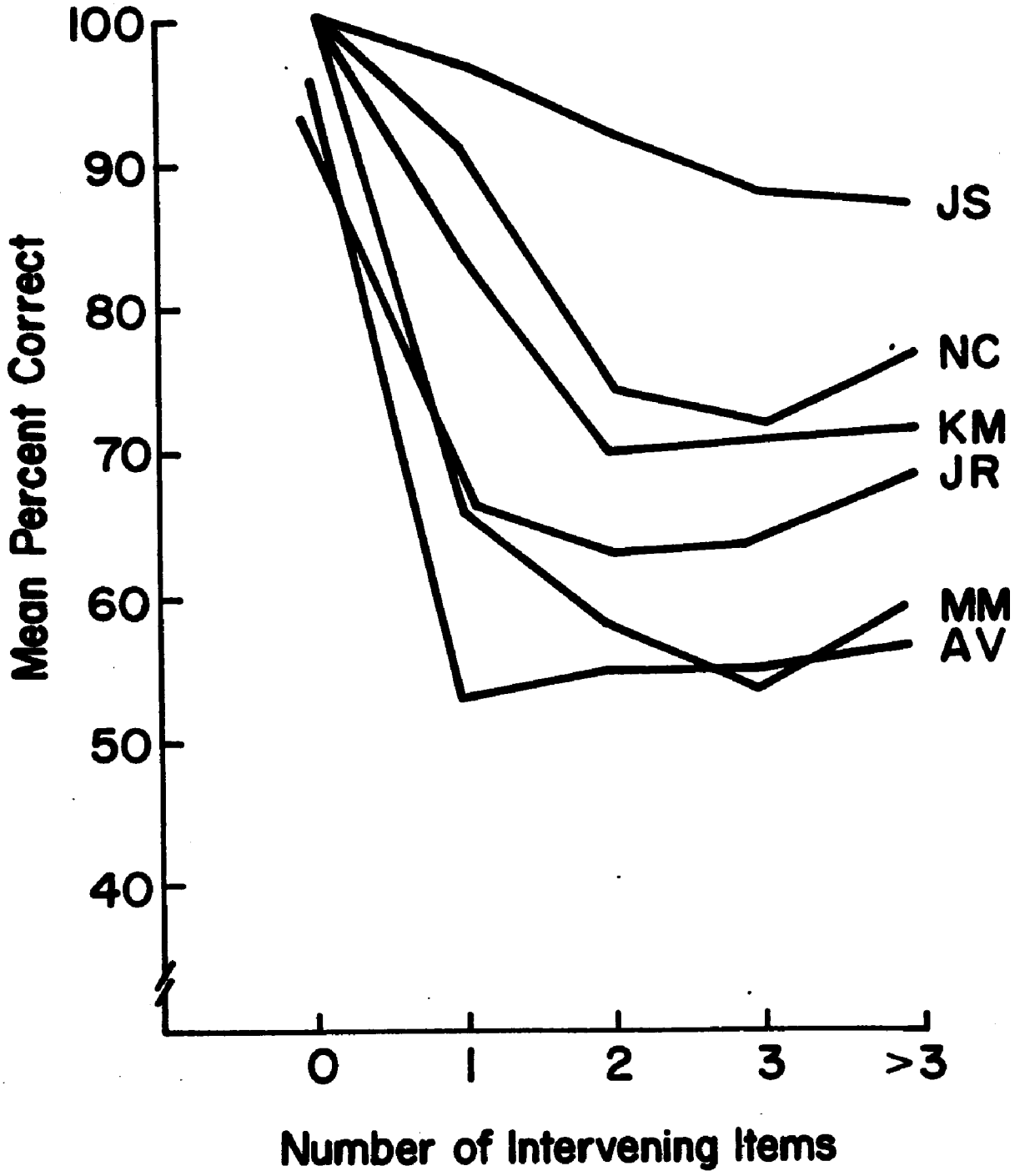


Fig. 4. Identification accuracy as a function of the retention interval for each subject of Experiment I.

(Figure 4), then increasing the size of the stimulus set will produce a decrement in AJ performance.

While N does exert an effect on AJ accuracy through its confounding with the average retention interval, the question remains as to whether this variable still has an effect when the retention interval is held constant. In an attempt to answer this question, the effect of N in this experiment was plotted in Figure 5 separately for each retention interval.

Figure 5 appears to demonstrate that N has an effect over and above that of the covarying mean retention interval. Except on repetition trials, five alternatives were judged more accurately than seven, and seven more accurately than nine. At zero intervening items, almost no errors were made, and the value of the Y intercept is independent of N. This suggests that the size of the stimulus set influences the rate of forgetting, and that N has no effect when the memory demands of the task are minimal. The force of this argument must, however, be tempered by two considerations which will be discussed more fully later. First, the lack of effect of N on repetition trials may be the artifact of a ceiling effect, since subjects approached 100% accuracy on repetitions. Second, the percentage of correct responses, while a simple and straightforward measure of response accuracy, may not be the most appropriate dependent variable for assessing the effect of N.

Ordinal-Position Effects. The most usual finding in AJ is that stimuli at the ends of the continuum being judged are identified more accurately than items in the middle. It was hypothesized that this bow-shaped ordinal-position effect results from more rapid forgetting of the middle stimuli. Thus on trials where memory effects are minimized,

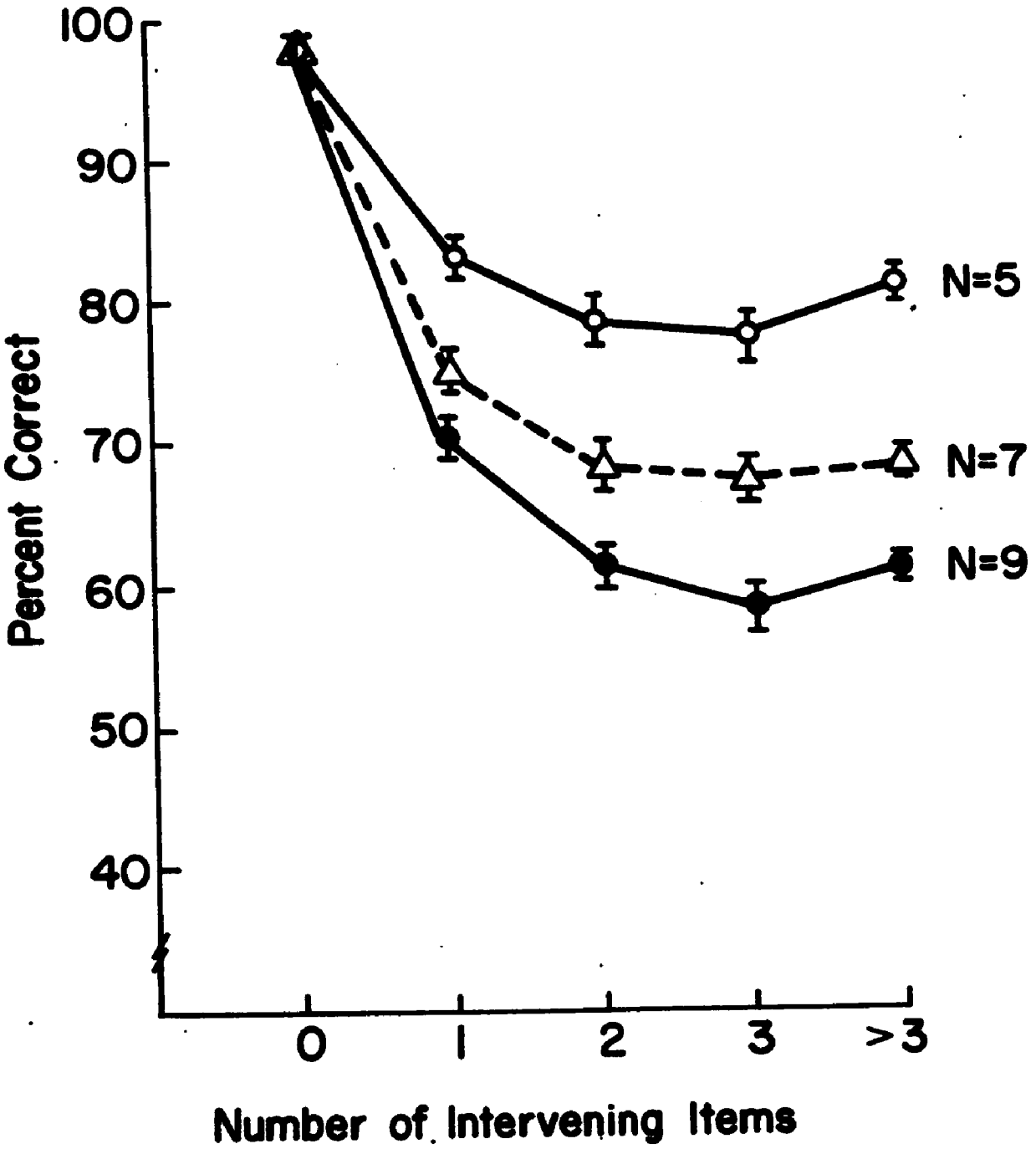


Fig. 5. Joint effects of the retention interval and the number of alternatives in Experiment I.

i.e., on repetition trials in this experiment, the ordinal-position curves should be flat.

The data shown in Figure 6 are consistent with the hypothesis that the ordinal-position effect represents differential forgetting. In this figure, the percentage of correct responses is plotted as a function of position in the series for sets of five, seven and nine tones, and the parameter is number of intervening items. On repetition trials, accuracy was independent of ordinal position. The proportion of correct responses ranged from .96 to 1.00 for all stimuli. The end-anchoring phenomenon manifested itself only at longer retention intervals. In other words, Figure 6 appears to demonstrate that the end stimuli were stored more effectively in memory than items in the middle, and that the end-anchoring effects disappeared when the memory demands of the task were minimal. Again this conclusion must be qualified because of the ceiling effect on repetitions, and because percent correct may be an inappropriate response measure.

Interpretation of these end effects would be complicated if the subjects showed systematic biases or preferences for the different responses used in this task. However, in the present experiment, the marginal stimulus proportions did not differ from the corresponding response proportions by more than .01. It seems reasonable to conclude that, for all practical purposes, the data are free of response bias.

The Modified Method of Absolute Judgment

There were several features of the preliminary experiment which made it less than optimal for the study of retention effects. Since that

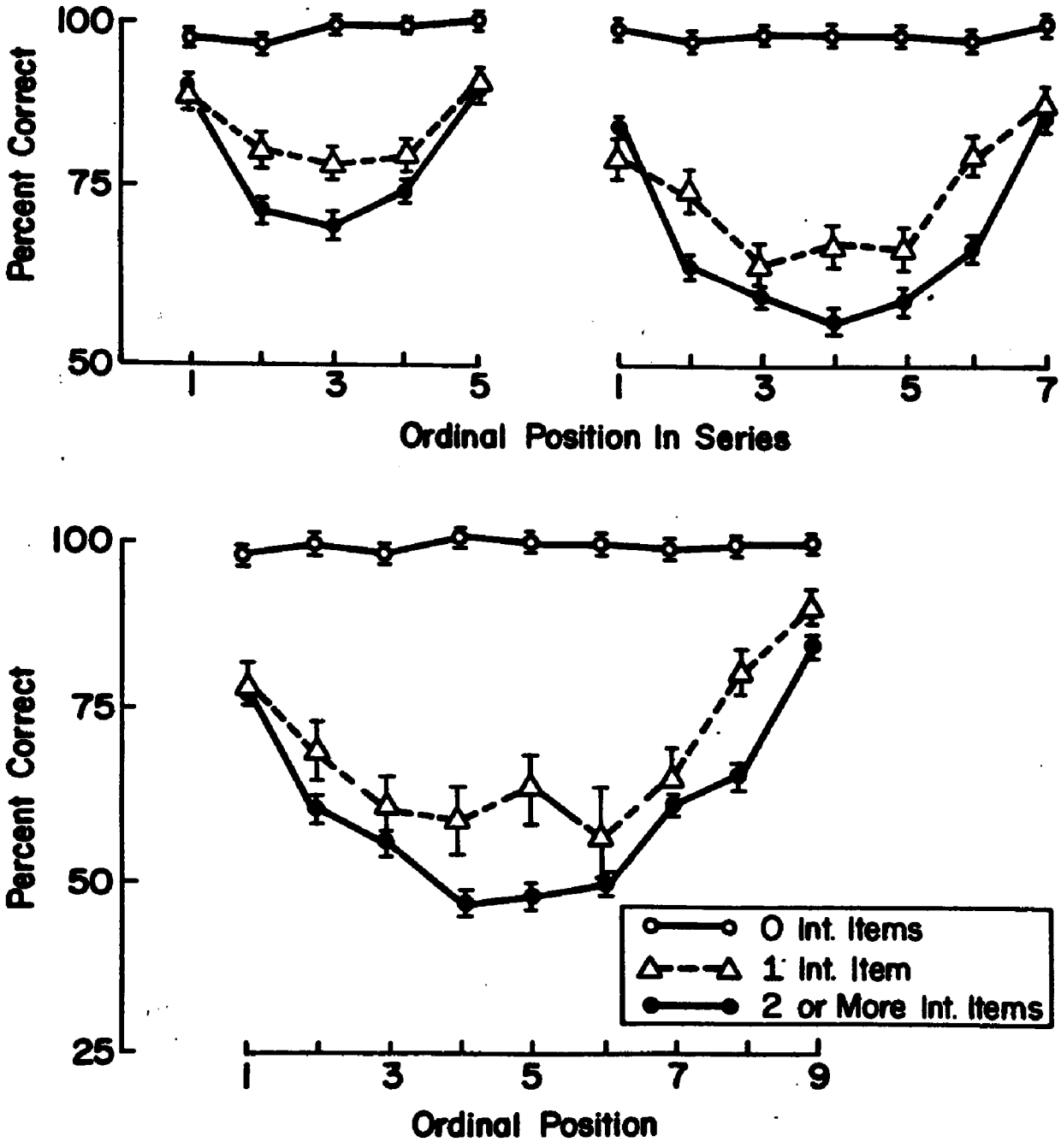


Fig. 6. Joint effects of the retention interval and ordinal position in Experiment I.

experiment was self-paced, and since the stimulus on any trial could be terminated by the subject's response, then neither the stimulus duration nor the interstimulus interval were completely under the experimenter's control. The data of Wickelgren (1966; 1969) obtained in a delayed-comparison task, suggest that both of these factors influence the retention of pitch information. In Experiment I, both of these variables depended upon the latency of the subject's response. Furthermore, the subjects often responded and always received feedback during the interstimulus interval. These distracting events may have impaired the subjects' memory for pitch.

General Procedure for Experiments II - V. Several procedural modifications were made to eliminate the above difficulties. First, the stimuli were presented at a fixed rate; both the tone time and the interstimulus interval were held constant. Second, the subjects were instructed to respond while the tone was on. The stimulus duration was made sufficiently long so that this was feasible. As a result, on any trial the subject could attend to the tone, make his response, and keep listening to the tone. In this way, the response and feedback were removed entirely from the retention interval. With this procedure, the proportion of no-responses was quite small (.002 in Experiment II; .003 in Experiment III), and the trials where the subject failed to respond were excluded from the analysis.

In each of the following experiments, the tone duration was 3 sec. and the feedback time was 400 msec. In other respects, except where noted, the procedure was identical to that used in the preliminary experiment.

A new method of generating the stimuli was used in Experiments II - V. In the preliminary experiment, the subjects listened to computer-generated square-wave tones. Now the computer was used to control the frequency of sine waves generated by a Wavetek model 110 oscillator. With this device, the frequency may be varied by applying a DC voltage across the appropriate terminals. The correct voltage could be obtained by programming the computer's 10-bit D:A converter. The stability of this system was found to be within .1 Hz after warmup, and the oscillator was left on throughout the experiments. Tones were presented binaurally at an amplitude of .03 v. RMS, and they were subjectively quite low in loudness.

Experiment II: Retention-Interval Effects (Continued)

In the preliminary experiment, it was found that on repetition trials, where the memory demands of the task were lowest, all subjects approached 100% accuracy. This suggests that with the stimulus spacing typically used in AJ, all of the variance is memory variance. However, because of the possibility of a ceiling effect, it was difficult to assess the effects of such variables as N and ordinal position on repetition trials. Thus a second experiment was carried out using stimuli which were much more similar to one another in an attempt to eliminate the ceiling effect. The modified method of AJ was used here, and as in the preliminary experiment, the minimum retention interval was 2 sec.

Method

Subjects. Nineteen undergraduate psychology majors participated in this experiment in order to fulfill a course requirement. None admitted to any hearing deficits.

Procedure. The stimulus set was made up of six tones ranging from 525-550 Hz with a step size of 5 Hz. The stimulus duration was 3 sec. and the interstimulus interval was 2 sec. Each subject received four blocks of trials with a rest period of several minutes between blocks. In a block of trials, each of the six tones was presented 20 times in random order. The first block of trials was considered to be practice and was not included in the analysis.

Results

Retention Interval. As in the preliminary experiment, the retention interval was found to be an important determinant of AJ accuracy. On trials where the memory load was lightest, at zero intervening items, the subjects averaged 74% correct. By one intervening item, this figure was reduced to 51%, and for two or more intervening items, 53% correct. In other words, forgetting was extremely rapid. On any trial, the subjects appeared to remember only the immediately preceding stimulus. After one intervening judgment trial, performance was reduced to its lowest level.

Ordinal-Position Effects. In the previous experiment, the ordinal-position curve was flat on repetition trials and bow-shaped on nonrepetition trials. These results suggested that U-shaped position curves are the result of more rapid forgetting of items in the middle of the stimulus set. However, in the preliminary experiment, the flat

ordinal-position curve on repetitions may have resulted from a ceiling effect.

In this experiment the ceiling effect was eliminated by using tones which were much closer together in pitch. Figure 7 demonstrates clearly that, under the conditions of this experiment, the ordinal-position curve on repetition trials is U-shaped. Subjects achieved 83% and 87% correct on the end tones, and only 62% correct on the fourth stimulus. Similar bow-shaped curves were obtained at the other retention intervals.

While this result is a little surprising, it does not deal a fatal blow to the memory hypothesis of the ordinal-position effect. For even though the memory demands were lowest on repetition trials, it is unlikely that memory variance was completely eliminated on repetitions. In this experiment, subjects were still required to store pitch information during the interstimulus interval, and thus the minimum retention interval was 2 sec. Several studies reviewed earlier demonstrate that recognition memory for unidimensional stimuli declines substantially over a 2-sec. blank interval. Thus it is possible that the ordinal-position effect on repetition trials obtained in this experiment resulted from more rapid forgetting of the middle items.

Experiment III: Minimizing Memory Effects in AJ

Experiments I and II demonstrated that number of intervening items has a large effect on AJ accuracy. In both of these studies the minimal retention interval was 2 sec. An attempt was made in the present experiment to eliminate retention factors as completely as

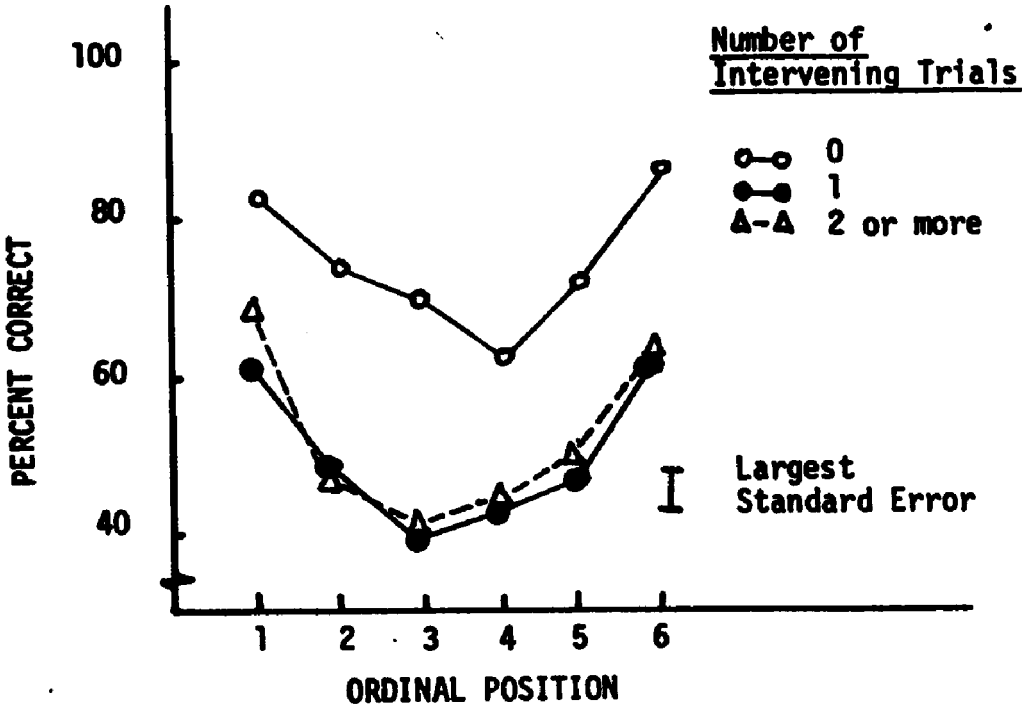


Fig. 7. Joint effects of the retention interval and ordinal position in Experiment II.

possible on repetition trials. To accomplish this, the time between successive tones was reduced to zero. On trials where the stimulus repeated, a continuous tone was presented. On nonrepetition trials, the tone modulated immediately to a higher or lower frequency without break. This was possible because, with the modified AJ method, the subject responds and is corrected while the stimulus is being presented. Thus there is no need to make the retention interval long enough to permit the subject to respond.

In this experiment the subjects received one week of AJ practice. During the practice sessions, the frequency difference between the tones was gradually reduced until the ceiling effect on repetitions was eliminated. This was followed by the main experiment, in which the same frequency difference was maintained, and the subjects judged sets of five, seven, and nine tone frequencies. The goals of this study were to determine the acuity of absolute judgments of pitch under conditions where memory variance was minimal and to examine the effects of N and ordinal position under these same conditions.

Method

Subjects. One male and five female students were selected from the departmental paid subject pool. None had participated in similar experiments previously, and none admitted to any hearing deficits. They received $1.75 per hour, and a bonus of $15.00 was offered to the student with the highest overall accuracy.

Practice Sessions. In order to minimize practice effects as completely as possible in the main experiment, seven hour-long training sessions were given to each subject. During that time, they were

familiarized with the paradigm, and a variety of tone stimuli were presented for judgment. The time between successive tones was .00 sec., and during the course of practice, the range of frequencies was gradually narrowed until the subjects were making a moderate number of errors on repetition trials. It was necessary to reduce the frequency step size to 1.3 Hz to eliminate the ceiling effect on repetitions.

Experimental Conditions. This experiment was run with zero delay between tones, and a frequency step of 1.3 Hz. Each subject was tested with sets of five, seven and nine tone frequencies. Response 1 was assigned to the lowest tone, which was 525 Hz in all cases, and the five lowest tones were identical in all three sets. The order in which the subjects received each number of alternatives was balanced by a Latin-square design. An attempt was made to have the total number of repetition trials constant across the experimental conditions. Thus each subject received 10 blocks of 5 x 25 trials in the five-alternative condition, 17 blocks of 7 x 20 trials in the seven-alternative condition, and 27 blocks of 9 x 15 trials in the nine-alternative condition. The subjects participated at their own convenience for no more than two hours per day.

Results

As in the previous experiments, separate stimulus-response matrices were generated for each retention interval. The matrices for repetition and nonrepetition trials, based on five, seven and nine alternatives, are shown in the Appendix. In these matrices, the data of all six subjects were pooled.

Retention Interval. Table 1 shows the effect of number of intervening items on the percentage of correct responses for each number of alternatives. With all three sets of stimuli, the subjects were most accurate on repetition trials. As in the previous experiments, forgetting was extremely rapid, and there appeared to be little difference among the nonrepetitions. If anything, there was a small tendency for the percentage of correct responses to increase at longer retention intervals. This "reminiscence" effect was also observed to a slight degree in Experiment II and with five out of six subjects in Experiment I. The phenomenon, if it is replicable, is quite small relative to the repetition effect, and it is difficult to explain.

Before concluding that the repetition effect results from a lesser amount of memory variance at brief retention intervals, one must rule out the possibility that it is an artifact of response strategies. In this experiment the subjects were informed of the correct response on each trial. In order to get 100% correct on repetitions, they could merely adopt the strategy of choosing on trial N the response that was correct on trial N - 1. Holland and Lockhead (1968) found that, in an AJ loudness study, subjects repeated the previous feedback significantly more often than chance. An analysis of the present data, however, provides no evidence for such a strategy. The proportion of trials on which subjects repeated the previous feedback was .20 with five alternatives, .14 with seven alternatives, and .11 with nine alternatives. Of course, chance level was .20, .14, and .11 in these three conditions.

Ordinal-Position Effects. It was assumed that, with the interstimulus interval set at zero sec., memory variance will be virtually eliminated on repetition trials. If the bowed error curves

Table 1

PERCENT CORRECT AS A FUNCTION OF THE RETENTION INTERVAL: EXPERIMENT III

	Number of Intervening Items				
	0	1	2	3	>3
Five Alternatives					
%	74	65	62	67	67
n	1528	1123	927	735	2865
Seven Alternatives					
%	68	53	54	54	55
n	2015	1632	1485	1169	7225
Nine Alternatives					
%	65	47	49	50	51
n	2318	1972	1812	1592	12,677

usually found in AJ are the result of more rapid forgetting of the middle items, then in this experiment the ordinal-position curves on repetition trials should be flat.

In Figure 8, the percentage of correct responses on repetition and nonrepetition trials is plotted as a function of ordinal position. With all three sets of stimuli, the end items were judged more accurately than those in the middle. This effect was shown on both repetitions and nonrepetitions by all six subjects. The phenomenon was not an artifact of the particular frequencies used in this study, since the five lowest tones in all three sets were identical. The fifth tone was judged quite accurately in the context of the five-alternative set, where it was an end-item, and relatively poorly when it occurred in the middle of the nine-alternative set.

The subjects could have raised the percentage of correct responses for the end items by adopting a response strategy which favored the end categories. The relevant data are shown in Appendix I. There is no evidence that the end categories were used systematically more often than chance level. For example, on repetition trials, the proportion of use of the two end responses was .20 and .21, .13 and .15, .11 and .11 for five, seven, and nine alternatives, respectively.

The subjects could have adopted a different strategy which would result in an artifactual end-effect on repetitions. It was demonstrated above that a bias towards repeating the previous feedback would enhance the percent correct on repetitions. While the overall data showed no evidence for this type of response bias, it is still possible that subjects repeat the previous feedback too often when an end response is correct on trial N - 1. In Table 2, the proportion of trials

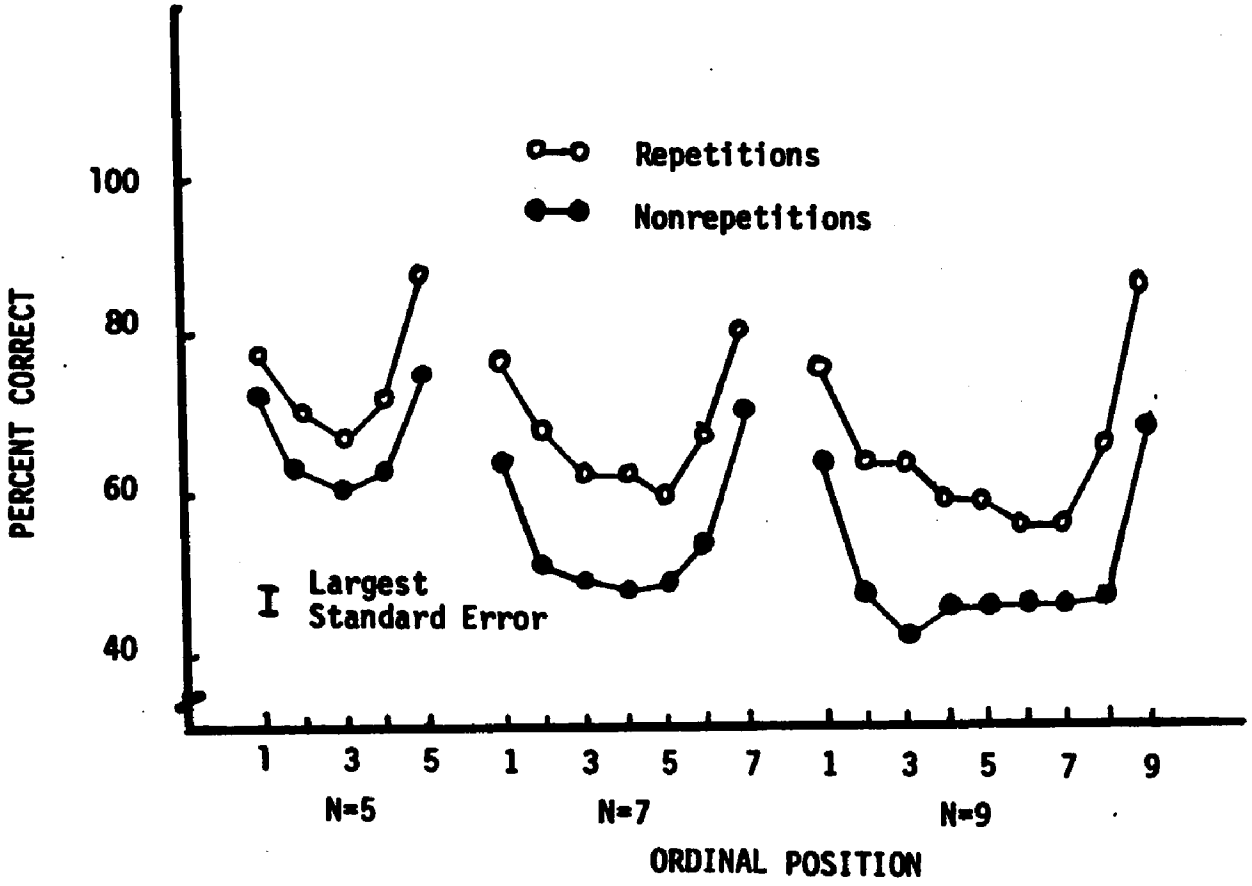


Fig. 8. Joint effects of the retention interval and ordinal position in Experiment III.

Table 2

PROPORTION OF TRIALS IN WHICH SUBJECTS REPEATED THE PREVIOUS FEEDBACK, SHOWN AS A FUNCTION OF RESPONSE CATEGORY

	Response Category								
	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-
Five Alternatives (n = 1478)	.20	.22	.19	.18	.22				
Seven Alternatives (n = 2024)	.14	.16	.14	.13	.12	.16	.15		
Nine Alternatives (n = 2400)	.11	.12	.12	.10	.09	.10	.10	.14	.11

on which subjects repeated the previous feedback is shown separately for each category. There is no evidence favoring the existence of this type of response strategy.

It is surprising to find end-effects on repetition trials that are approximately as large as those on nonrepetitions, even under conditions which require virtually no temporal retention on the part of the subjects. This finding directly contradicts the subjective-standard hypothesis of Eriksen and Hake, and makes a memory theory of the ordinal position effect difficult to justify.

Number of Alternatives. As shown in Table 1, when the size of the stimulus set is increased, the percentage of correct responses is lowered at each retention interval. Even on repetition trials, where memory variance is minimal, the percent correct dropped from 74% with five alternatives to 65% with nine alternatives. There is no a priori reason why the discriminability of the stimuli should depend upon the context in which they are judged. One would thus like to find a theory of AJ whose sensitivity parameter is independent of N. Several theories of categorical judgment are described later in this paper, and an attempt is made to see how they interpret the present data.

Experiment IV: Time and Intervening Trials as Determinants of Forgetting

In the preceding experiments, the retention interval was defined as the number of trials intervening between the present stimulus and its last occurrence. When the number of intervening items is varied, the time between successive occurrences changes as well. Thus it is

not clear whether the forgetting that occurs in AJ results from passive temporal decay or from the disrupting effect of intervening judgment trials. In this experiment, the subjects judged seven tone frequencies, and the interstimulus interval varied from 0-6 sec. By varying the time between tones in this way it was possible to manipulate the temporal retention interval while holding the number of intervening items constant.

Method

Subjects. The five subjects in this study had participated in Experiment III as well. The subject with the lowest accuracy score in that experiment was dropped here in order to complete the design.

Procedure. The stimuli were seven tones spaced 10 Hz apart, ranging from 500-560 Hz. The stimulus duration was 3 sec., and each subject judged the stimulus set at five different rates of presentation. In order to vary rate of presentation, the interstimulus interval was made 0, .5, 1, 3, or 6 sec. In a block of trials, each tone frequency occurred 15 times, and the observers received four blocks at a given delay before the next condition began. Order effects were balanced by a Latin-square design. Other procedural details were the same as in Experiment III.

Results

In the following analyses, the data of all five subjects were pooled. Figure 9 shows the percentage of correct responses as a function of time, and the parameter is number of intervening items. The effect of the temporal retention interval, with number of inter-

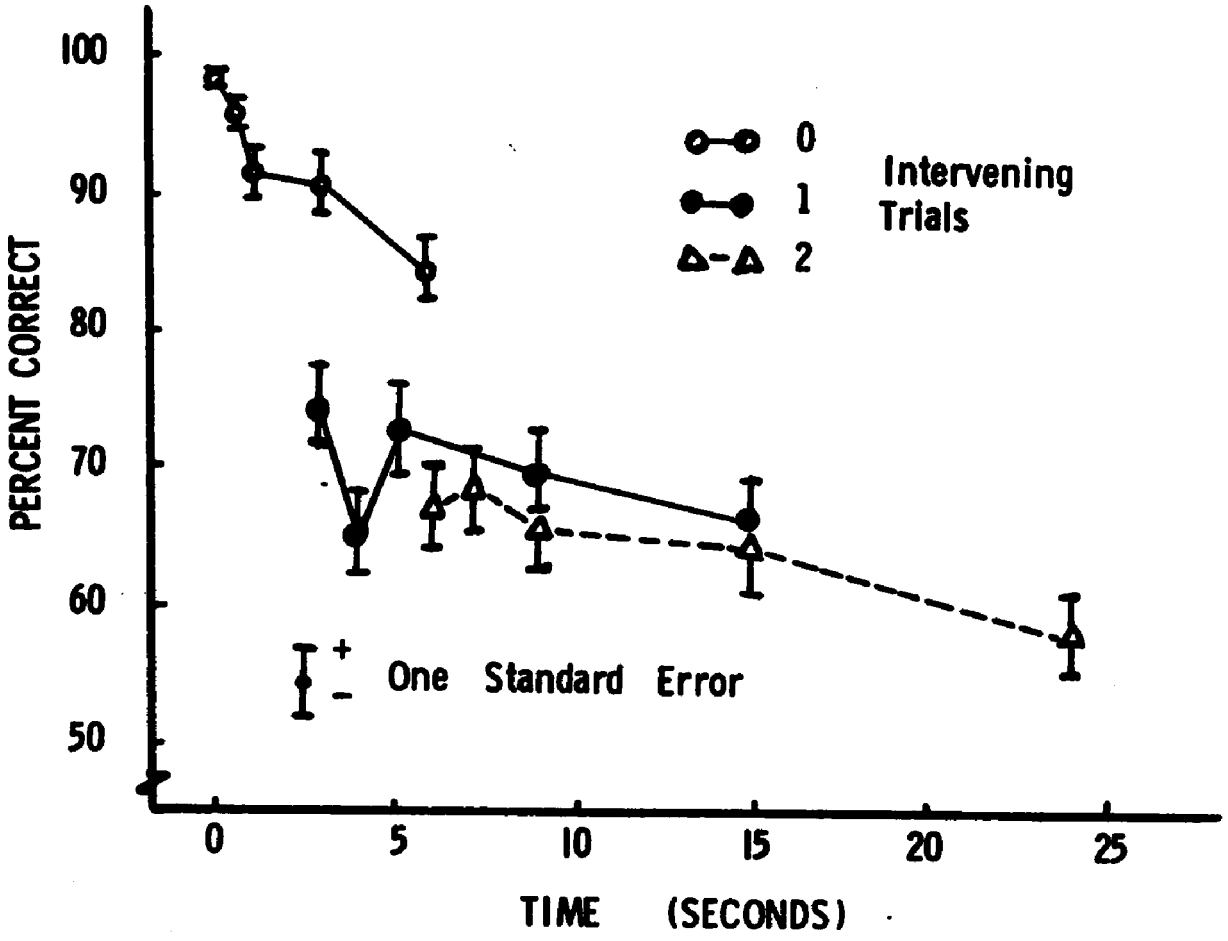


Fig. 9. The effects of time and number of intervening items on identification accuracy in Experiment IV.

vening items held constant, is clear-cut. On repetition trials, represented by the open circles, performance decreased monotonically from 98% correct at zero delay to 84% correct at a 6-sec. delay. Thus substantial forgetting occurred with the mere passage of time. At one and two intervening items, there was a slight time effect, but this effect was small relative to the confidence intervals.

Number of intervening items also appears to be an important determinant of AJ accuracy. On repetition trials performance was considerably above that on nonrepetitions. For example at a 3 sec. retention interval the subjects achieved 90% correct on repetitions and only 74% correct with one intervening item. At a 6-sec. delay, they achieved 84% correct on repetitions and 67% correct with two intervening items. As in the previous experiments, most of the effect of intervening items was due to the difference between repetition and nonrepetition trials.

To summarize these results, the mere passage of time produced substantial forgetting during the retention interval of 0-6 sec. studied here. In addition, there was a large decrease in retention when the subject was required to listen to and judge other tones. The occurrence of one intervening trial was highly interfering, reducing performance to a level approaching asymptote.

Experiment V: The Role of Feedback (FB)

In all of the preceding experiments, the subjects were informed of the correct response immediately after their choice on each trial. While this type of informative FB is often used in AJ, it is also common to run

the experiment under no-FB conditions. The purpose of the present study was to examine AJ retention-interval effects when FB is omitted.

It is hypothesized that FB influences the decision rule employed by the subject, but not his sensitivity to pitch differences. In AJ, the goal of the subject is to be correct as often as possible with respect to the experimenter's criterion. Without FB, the subject may instead adopt the strategy of being as consistent as possible. When FB is provided and the subject perceives the current stimulus to be the same as on the previous trial, he should attempt to make the correct response; that is, he should use the response which, according to the FB, was correct on trial N - 1. In the no-FB condition, on the other hand, even if the subject correctly judges that the stimuli on trials N and N - 1 are identical, he may still be unable to remember which response is correct. Under these circumstances, my guess is that he will be consistent and repeat his previous response.

This hypothesis leads to several predictions:

(1) The percentage of correct responses should be lower when FB is omitted;

(2) On repetition trials, subjects should repeat their previous response in the no-FB condition as often as they make a correct response with FB;

(3) With FB, the subject will tend to repeat his previous response on repetition trials when he was correct on trial N - 1, and he will tend to change his response if he was incorrect on the previous trial;

(4) Without FB, on repetition trials the subject will repeat his previous response equally often whether or not he was correct on trial N - 1.

Method

The five subjects of Experiment IV participated in this study as well. Here the five-alternative condition of Experiment III was replicated in all details, except that the FB lights were disabled. Data for the same five subjects in a FB condition were taken from Experiment III.

Results

Retention Interval. The effect of number of intervening items is shown for both FB and no-FB conditions in Figure 10. When knowledge of results was omitted, all five subjects were less accurate at each retention interval. Moreover, without FB the effect of the retention interval was opposite to that found in Experiment III. The percent correct was actually lowest on repetition trials, rising slightly at longer retention intervals. All five subjects were correct most often on repetitions when FB was given, but in the no-FB condition, none of the same five subjects was most accurate on repetition trials. One interpretation of these data is that without FB, the subject's sensitivity to pitch differences is impaired and his sensitivity is actually higher at longer retention intervals.

Feedback and the Decision Rule. A second possibility is that FB has no effect on information about pitch, and that when knowledge of results is omitted, the subject adopts a different response rule. If this is the case, then in the no-FB condition, percent correct is an inappropriate measure of sensitivity.

It was hypothesized that when FB is given, the subject attempts to maximize percent correct, while without FB the goal is to maximize

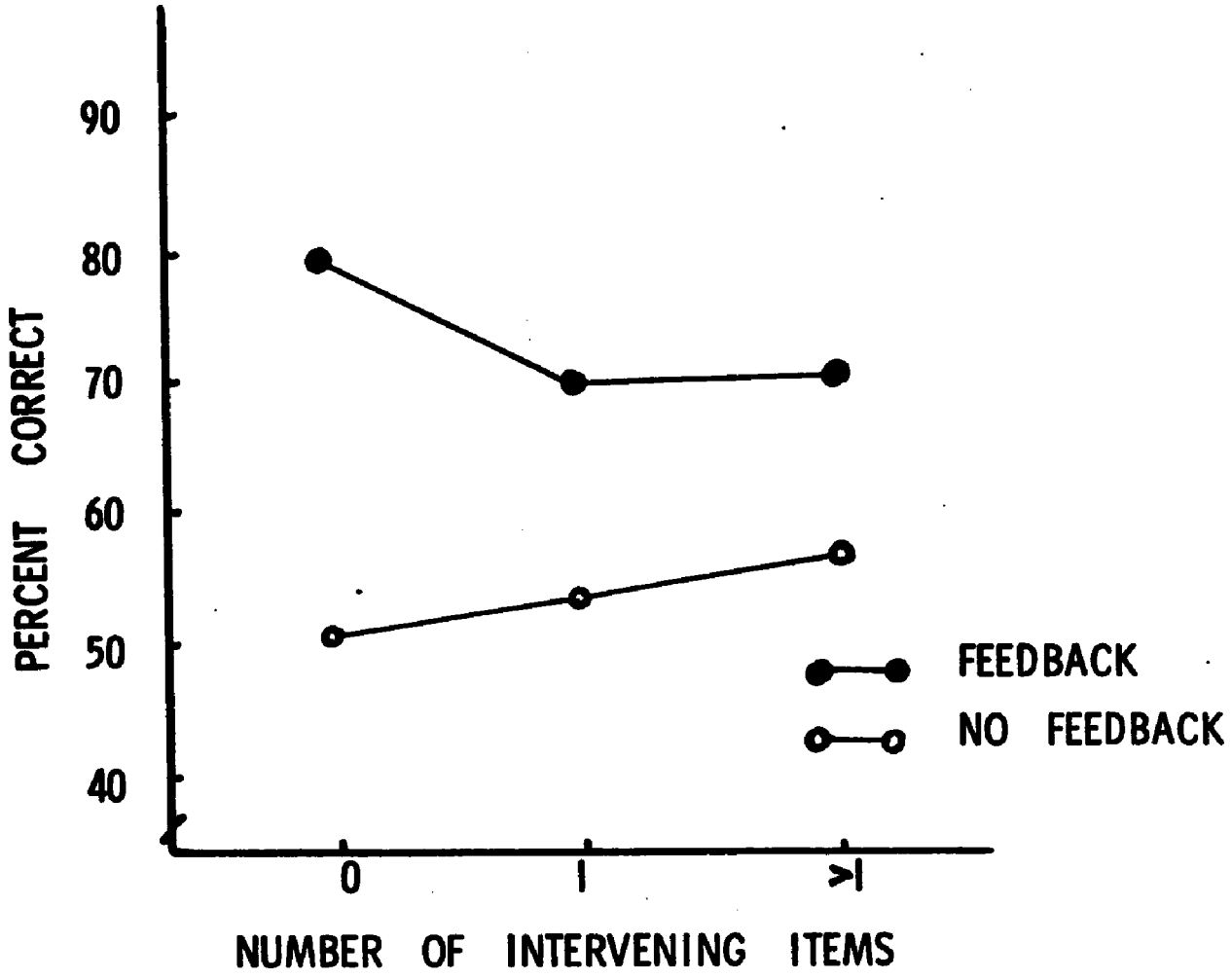


Fig. 10. The effect of number of intervening items under feedback and no-feedback conditions.

internal consistency, and this hypothesis led to a number of predictions:

(1) Percent correct should be lower in the no-FB condition. As noted above, all five subjects had a lower percentage of correct responses when FB was omitted.

(2) On repetition trials, the probability of a correct response with FB should equal the probability of a response repetition without FB. When the stimulus repeated, the mean proportion of correct responses with FB was found to be .79, while in the no-FB condition, the average proportion of response repetitions was .74. As predicted, the difference between these means was not statistically significant ($t_4 = 1.80$, $p > .10$)

(3) With FB, the subject should tend to repeat his previous response on repetition trials when he was correct on trial N - 1, and he should tend to change his response if he was incorrect on the previous trial. It was found that when the subject was correct and then the stimulus repeated, the response made on the previous trial was repeated 82% of the time. However, when the response on trial N - 1 was incorrect, response repetitions occurred on only 11% of trials. This difference was highly significant ($t_4 = 14.20$, $p < .001$), and thus with FB, the response on trial N depended markedly on whether the subject was correct on trial N - 1.

(4) On repetition trials, in the no-FB condition, the subject should repeat his previous response equally often whether or not he was correct on trial N - 1. Under these conditions the proportion of response repetitions was found to be .77 when the previous response was correct, and .69 when the previous response was incorrect. This difference was not statistically significant ($t_4 = 2.27$, $p > .05$).

All of the data presented here support the contention that FB influences the decision rule but not the subjects' sensitivity to pitch differences. These results suggest that subjects are just as accurate according to their own criterion without FB as they are in reference to the experimenter's criterion when FB is provided.

Theoretical Analyses

In Experiment III, AJ data were collected under conditions where memory variance was minimal (repetition trials), and also under conditions where considerable forgetting occurred (nonrepetitions). Both N and ordinal position were variables in this experiment, and it would be desirable to know whether these factors influence the discrimination or retention of pitch information. In order to provide a definitive answer, one would first have to develop a correct theory of AJ performance. Lacking such a theory, I will attempt to discuss how several current models interpret the present data.

Guessing Model

Error data in identification tasks is sometimes transformed by a "correction for chance", a procedure which derives from a high-threshold theory of judgment performance (Green and Swets, 1966). The model states that the subject recognizes the stimulus event on a proportion p of the trials, and on the remaining trials he has no information at all about the stimulus. In the latter state he guesses. If his responses are unbiased, he will be correct by chance with a probability of 1/N, where N is the number of alternatives judged. Thus the obtained percent correct is made up of two components, correct recognitions and correct guesses.

The guessing model was applied to the data of Experiment III, and the corrected accuracy scores are plotted as a function of N in Figure 11A. On repetition trials, where memory effects were minimal, PC' decreased monotonically with N. Thus according to the guessing model, discriminability varies inversely with N. On nonrepetition trials, where the subjects were required to store pitch information in memory, PC' also declines with increasing N, but the slope of the curve is steeper than that for repetitions. This last finding indicates that, if the model is correct, more forgetting occurs with a larger stimulus set.

According to the guessing model, the subject has no information on trials where his response is incorrect, and thus his choice of error responses should be random. The data shown in the Appendix clearly demonstrate that the distribution of errors is not rectangular. Rather, as is typically the case in AJ, the incorrect responses are usually near misses, and extreme errors occur on only a small percentage of trials. Thus the guessing model does not provide an adequate account of the data of Experiment III.

Forced-Choice Detection Model

A second theory which may be applicable to the present data was originally developed for the forced-choice detection task (see Green and Swets, 1966). In this paradigm, the subject judges which of N temporal intervals contains a signal embedded in background noise. Noise alone is presented in all but one of the N intervals. In the detection model, the sensory effects of signal plus noise are represented by a normal variable with mean d' and unit variance; each interval containing noise alone is represented by an independent Gaussian variable with a mean of zero and unit variance. For each temporal interval on

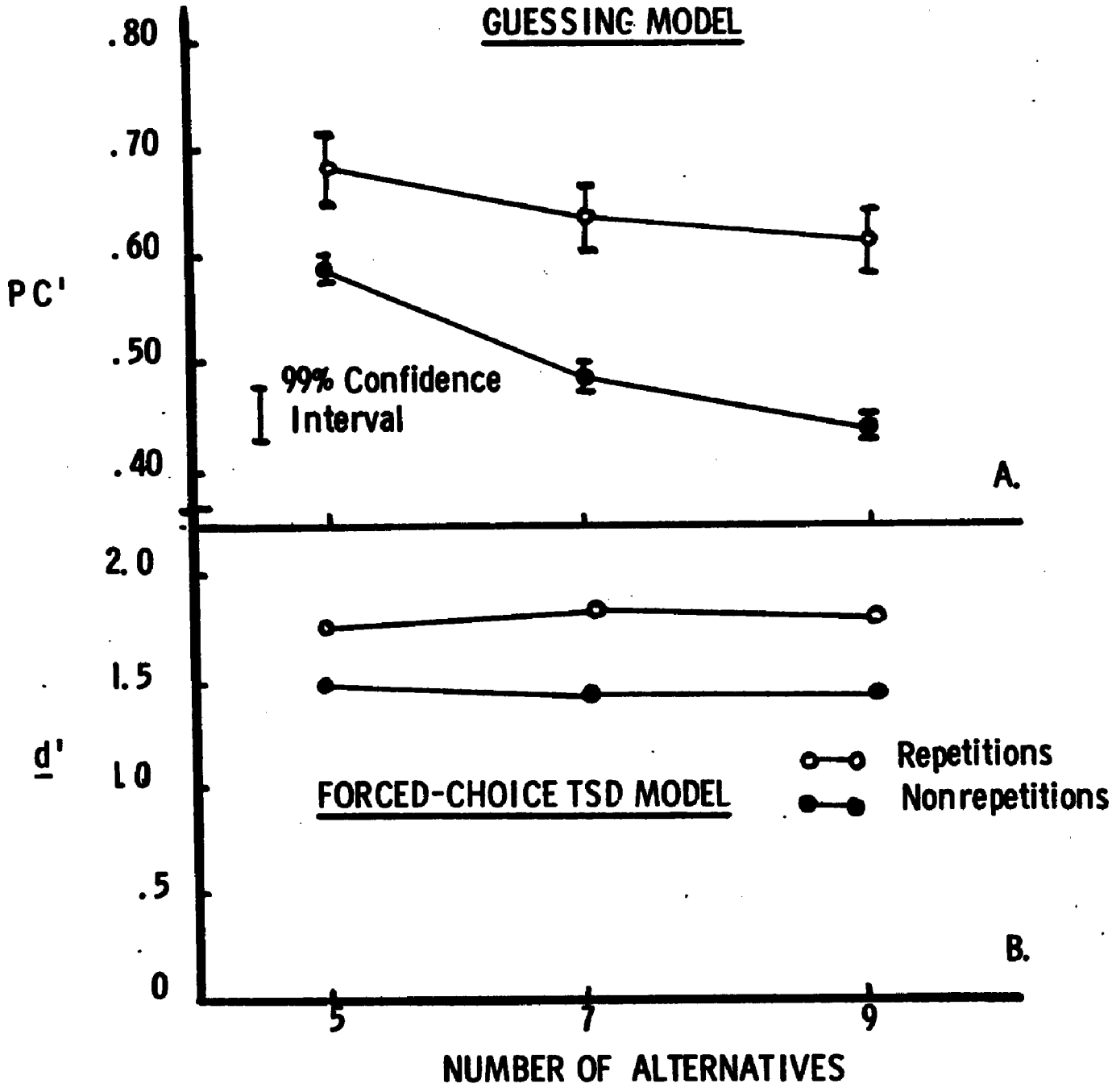


Fig. 11. The effect of the number of alternatives being judged on (A) percentage of correct responses adjusted for "guessing", and on (B) the d' measure of detection theory.

every trial, the subject is assumed to estimate the likelihood that a signal has occurred. He is required to judge which interval contains the signal, and the optimum choice is the interval which gives rise to the largest likelihood ratio.

While this model was developed for the forced-choice detection task, it has been applied to the AJ paradigm as well (Green and Birdsall, 1964). In this case it is assumed that each member of the stimulus set is represented in memory, and on any trial the subject attempts to match or correlate the input with every stored representation. The obtained correlations are represented as independent normal variables with unit variance. One correlation has a mean of d', and all others have a mean of zero. The optimal choice is to select the item stored in memory which correlates most highly with the input.

Green and Birdsall (1964) used the forced-choice model to reanalyze the results of a study by Miller, Heise, and Lichten in which observers attempted to identify words presented in a noise background. It was found that the d'-parameter of detection theory, which is assumed to reflect the subjects' sensitivity to stimulus differences, remained invariant as the vocabulary size was manipulated. This is interesting since parameter invariances are rare in psychology, especially those that make a certain amount of theoretical sense.

In light of Green and Birdsall's success with the forced-choice model, this theory was applied to the present data to see if the d'-parameter remains constant over changes in the size of the stimulus set. The percentage of correct responses for relevant conditions in Experiment III was transformed to d' values using Elliott's tables (Swets, 1964). Advantage was taken of the fact that d' is an approximately linear function

of percent correct when plotted on normal probability paper, and the scores shown here, in Figure 11B, were estimated by graphic interpolation.

The forced-choice detection model interprets the data of Experiment III in a clear and simple manner. According to this model, AJ accuracy is independent of N on both repetition and nonrepetition trials. For example, the d' on nonrepetitions was 1.48, 1.43, and 1.43 for five, seven, and nine alternatives respectively. These striking results indicate that, if the model is correct, both sensory discriminability and memory performance in AJ are invariant over changes in N.

Unfortunately, it appears that the model is not correct, at least with respect to the data of Experiment III. The forced-choice detection theory, like the guessing model, predicts that the distribution of error responses is rectangular. This is because, in the detection model, the correlations between the input and the stored representations corresponding to the incorrect choices all have the same mean and variance. Therefore this simple version of the detection model must be rejected as inappropriate to the present task.

Information Model

According to information theory, the human perceptual system acts as an analyzer of the information content of the stimulus. In a random sequence of equiprobable events, stimulus information (H_s) is defined as $\log_2 N$, where N is the number of alternative items. A number of AJ experiments appear to show that when H_s is increased, a measure of information transmitted in the subjects' responses (H_t) first increases at a maximal rate, and then remains invariant with further changes in H_s (Miller, 1956; Attneave, 1959; Garner, 1962). As was noted earlier, this asymptotic performance has been interpreted as evidence for a channel capacity, or a fixed upper limit on AJ accuracy. The proponents

of information theory have stressed the stability of the H_t measure over changes in a number of experimental variables, including the range and spacing of stimuli, practice, and feedback, as well as the number of alternatives.

A series of experiments conducted by Pollack (1952, 1953) provided the evidence for a fixed channel capacity on absolute judgments of pitch, the sensory continuum investigated in this paper. From 2-14 different tones were presented in an AJ task with feedback, and the tones were equidistant from one another on a log scale. The most similar set of tone frequencies ranged from 100-500 Hz, while the largest range used was 60-14,000 Hz. Pollack plotted information transmission as a function of H_s and fitted curves which asymptote at 2.0 to 2.3 bits. Thus his subjects were capable of the accurate identification of only 4-5 tones. This value did not change appreciably as a function of the range of frequencies used. In one experiment Pollack used eight tones of which the lowest was always 100Hz, while the highest varied from 500-8000 Hz. Under these conditions, even though the stimulus range changed by a factor of 20, H_t was quite stable, varying from only 1.8 to 2.0 bits.

The H_t measure may not be as insensitive to variations in the range of stimuli as the data of Pollack suggest. Hartman (1954) had four groups of subjects judge nine tone frequencies separated by a pitch distance of 50, 100, 200, or 300 mels. The smallest frequency range, for the 50-mel group, was from 1330-2104 Hz. During the first week of testing, Hartman's results were similar to those of Pollack, as the frequency range had little effect. H_t varied from 1.0 to 1.3 bits for the four groups. By the seventh week, however, the 300-mel group was transmitting 2.2 bits, while the 50-mel group managed only 1.4 bits of information.

MacRae (1970) has made the more serious criticism that the H_t measure is not independent of the information in the stimulus. His rationale is that published values of H_t are often too large because of bias in estimating this parameter. Bias is especially problematical when the number of stimulus and response categories is large, and when the sample size is small. MacRae reviewed a number of AJ studies, and he argued that when bias is taken into account, H_t first increases and then declines as H_s is made larger. Pollack's (1952) experiment provides a good illustration. In one condition, H_t was estimated from data in which up to 14 different tones were presented 20 times each. Here H_t was relatively invariant, but bias was quite considerable, especially at the large values of H_s. In a second study, a maximum of 10 tones were presented 30 times each, and here bias was much smaller. Under these conditions, Pollack's data points rose to a maximum in the vicinity of seven items and declined beyond this point. Thus, as MacRae points out, the concept of a fixed channel capacity may not be appropriate when sampling bias is taken into consideration.

Information Theory and the Retention Interval. It was demonstrated by Siegel and Siegel (1968) that the H_s measure of information theory is confounded with the average retention interval in the AJ paradigm. In Chapter I, I showed that with a random sequence of equiprobable stimuli, the probability with which x trials intervene between successive occurrences of the same stimulus is equal to $1/N\ (1 - 1/N)^x$, where N is the number of stimulus alternatives. The mean retention interval is equal to $N - 1$. The confounding of stimulus information with the mean retention interval results from the fact that, under these conditions, H_s is equal to the logarithm of the number of alternatives. Kornblum (1968; 1969) has recently pointed out that, in the choice reaction-time task, H_s is

confounded with the probability of a repetition trial, or zero intervening items in my terminology. It has been demonstrated here that each number of intervening items depends upon N and therefore on H_s. Since H_s and the mean retention interval covary, and since AJ accuracy is lower at long retention intervals, then at least some of the information effect results from changing the average memory load placed on the subject.

This problem could be eliminated by examining the effects of H_s under conditions where the retention interval is held constant. The data of Experiments I and III are relevant, since in these studies the number of alternatives was varied, and separate confusion matrices were generated for each number of intervening items. From these confusion matrices, H_t was estimated using standard procedures, and the data from all subjects were pooled to minimize bias.

In Figure 12, H_t scores for Experiment I are plotted as a function of H_s and the parameter is number of intervening items. This experiment was run using a stimulus spacing that is not atypical for the AJ task, and in fact the range was considerably narrower than the extreme values used by Pollack and Hartman. On the basis of Pollack's (1952) study, one would expect to find that H_t remains at least roughly invariant as H_s is increased. For the overall data, shown by the heavy line, this was indeed the case, and performance appeared to asymptote at about 1.6 bits. However, when the retention interval was introduced into the analysis, the invariance of H_t disappeared. On repetition trials, transmission was virtually perfect at all values of H_s. With one intervening item, H_t increased from 1.54 to 1.85 bits as H_s varied from 2.32 to 3.17 bits. Only at longer retention intervals was H_s roughly constant.

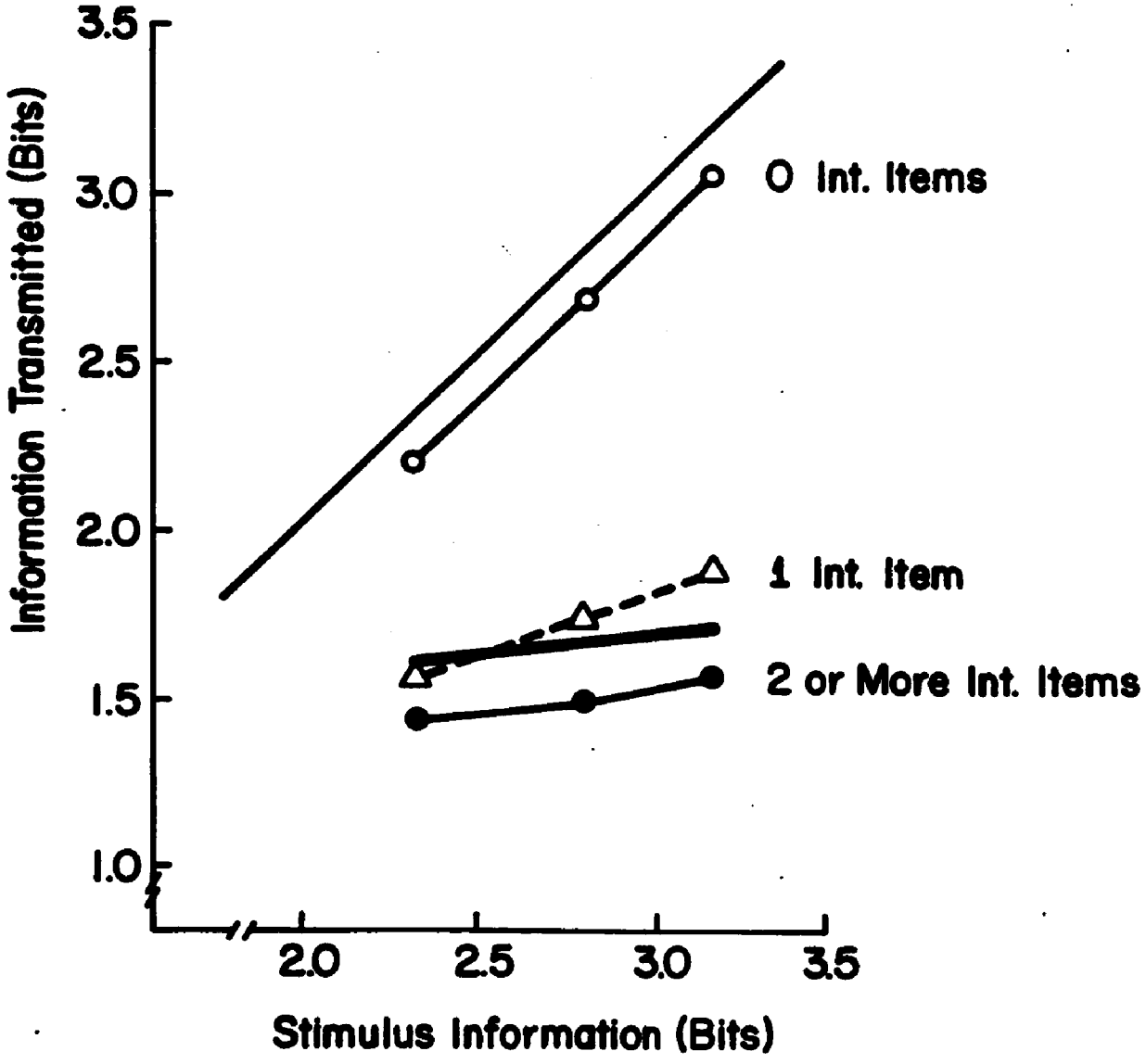


Fig. 12. Joint effects of stimulus information and the retention interval on information transmitted in Experiment I.

Bias in estimating H_t should not play a major role in the present study since the maximum number of stimuli was nine, and since the sample size was relatively large. The number of trials per stimulus was smallest with nine alternatives, and even here each item occurred an average of 123 times on repetition trials, 84 times at one intervening item, and 833 times at two or more intervening items. MacRae (1970) found for a number of AJ studies that, when bias is controlled, H_t first increases and then declines as H_s is made larger. In all conditions of Experiment I, H_t appears to be an increasing function of H_s, although the increase is slight at longer retention intervals. The fact that no terminal decrease was found here is not surprising, since no large values of H_s were employed in this study.

In Experiment III, five, seven, and nine alternative tones were used and the interstimulus interval was zero. The stimuli were spaced only 1.3 Hz apart, and thus the frequency range was much narrower than in the studies of Pollack and Hartman. Transmission scores for this experiment are plotted in Figure 13. On repetition trials, where performance was limited by the sensory discriminability of the stimuli, H_t rose from 1.22 bits with five alternative tones to 1.83 bits with nine alternatives. On nonrepetition trials, where the subjects were required to store pitch information in memory, the H_t values were considerably lower. As in Experiment I, transmission scores were an increasing function of H_s, and the rate of increase was again greatest on repetition trials.

In any AJ experiment the retention interval varies from trial to trial. The results of Experiments I and III demonstrate that the H_t measure of information theory does not remain constant as the retention interval is changed, but rather it declines rapidly as the number of

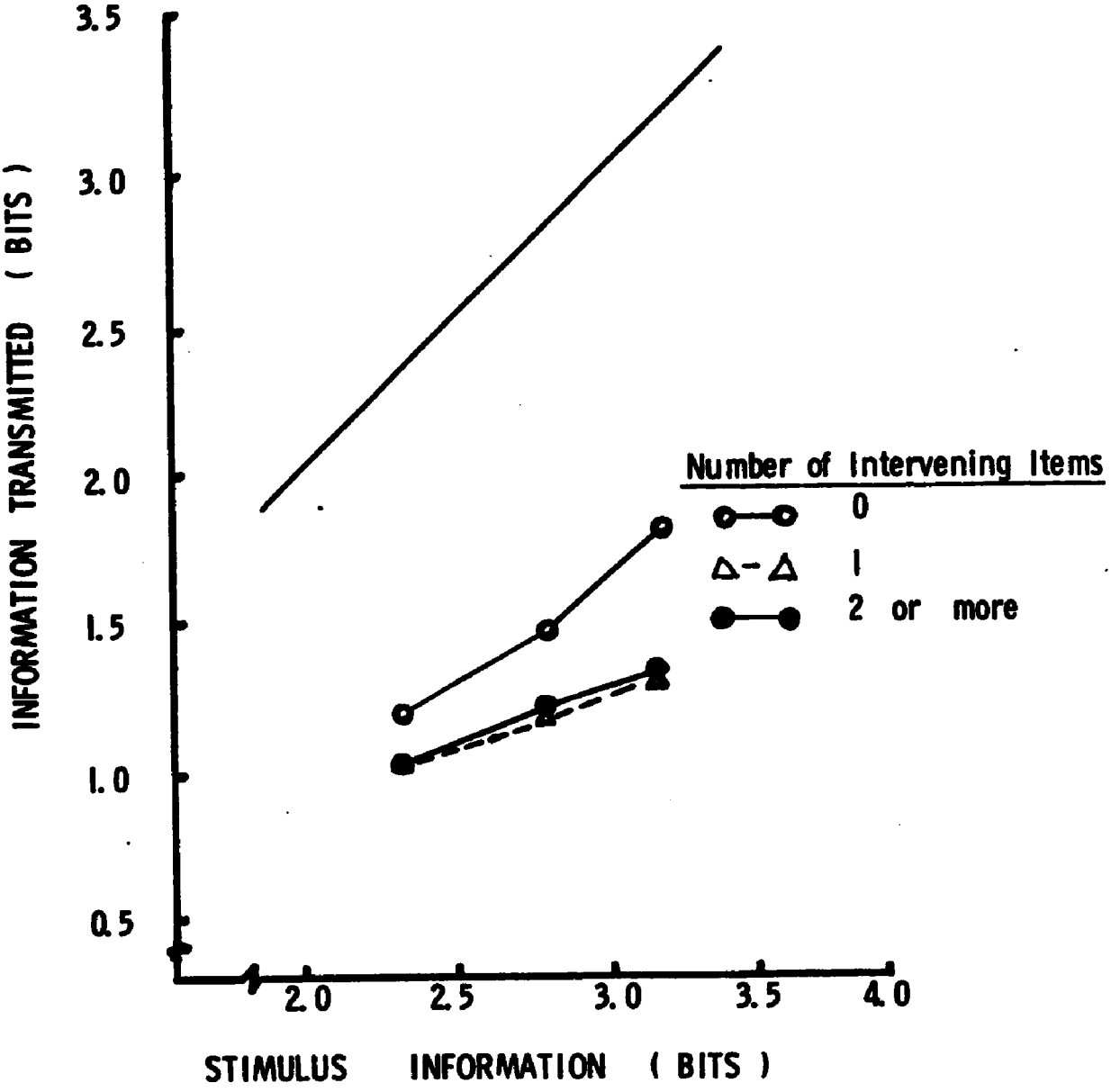


Fig. 13. Joint effects of stimulus information and the retention interval on information transmitted in Experiment III.

intervening items is increased. With the stimulus spacing usually used in AJ, transmission is virtually perfect on trials where the memory demands are light. The type of performance which has given rise to the concept of a channel capacity is found only at longer retention intervals. These findings demonstrate clearly that AJ performance is limited by the observer's inability to store information about unidimensional stimuli in memory.

CHAPTER III: DISCUSSION

Retention-Interval Effects

The retention interval changes from trial to trial in the AJ task, and in the preceding series of experiments, I attempted to determine the extent of this variable's effects. From the results of these studies several conclusions may be drawn. First, all subjects in all experiments showed forgetting as the retention interval was increased. Moreover, this forgetting was extremely rapid, since most subjects could remember only the immediately preceding stimulus in the series. With the wide stimulus spacing commonly used in AJ, performance was almost error-free at brief retention intervals. Thus in the typical AJ experiment, virtually all of the variance is attributable to forgetting. When retention effects are minimal, subjects' acuity in AJ appears to be as fine as it is in discrimination tasks. Just as in delayed-comparison studies, forgetting of pitch information is influenced both by the mere passage of time and by the nature of the retroactive events.

In Experiment I the stimulus range was fairly typical for the AJ task, and it was somewhat narrower than the extreme values used by Pollack (1952; 1953) and by Hartman (1954). On repetition trials, both percent correct and information transmitted were close to the maximal value, but these two measures dropped rapidly as the retention interval was increased. This finding suggests that AJ accuracy is limited by poor perceptual memory for unidimensional stimuli. For a large proportion of subjects, forgetting was complete after the occurrence of just one intervening judgment trial. On the other hand, when multidimensional verbal stimuli are used in an identification task (e.g., Brelsford,

Keller, Shiffrin, and Atkinson, 1966), observers are able to remember much further back into the series. Thus the rapid forgetting observed in the present series of experiments appears to be peculiar to sensory stimuli and if there is a channel capacity on absolute judgments, it is indeed a limitation on memory.

The poor memory for pitch observed here agrees very nicely with similar findings obtained in comparison tasks (e.g., Wickelgren, 1969; Massaro, 1970a). This suggests a resolution of the controversy over why AJ performance is usually so much poorer than that obtained in relative-judgment studies. The reason is that in AJ, the average retention interval is typically much longer than that employed in comparison tasks, and in addition it is filled with highly disrupting events. The "threshold" for pitch is generally accepted to be in the neighborhood of 1-3 Hz for the range of frequencies used in the present studies (Woodworth and Schlosberg, 1954). In Experiment III, when memory variance was minimal, it was necessary to reduce the frequency step between adjacent stimuli to 1.3 Hz in order to get a moderate number of confusions. There is some unpublished data collected by Jane Siegel on apparatus similar to mine which indicates that in a forced-choice pitch-discrimination task, where the retention interval was zero, it was also necessary to reduce the frequency step to 1.3 Hz to get a reasonable number of errors. Thus, when memory factors are made comparable in absolute- and relative-judgment tasks, performance appears to converge toward a common value.

When subjects are required to make relative judgments of pitch, performance deteriorates when both time and the number of interpolated stimuli increases (Massaro, 1970b). In AJ these two variables are

confounded, since the retention interval involves concurrent changes in the time and the number of intervening trials since the last occurrence of the present item. In Experiment IV each of these variables was manipulated independently, and both were found to influence AJ accuracy. On repetition trials, where the retention interval was blank and the subjects were free to rehearse, performance accuracy decreased nonetheless as the retention interval varied from 0-6 sec. In addition to this temporal decay, the occurrence of just one intervening judgment trial was sufficient to reduce performance to approximately its lowest level. Thus, as in relative-judgment tasks, both decay and interference effects influence short-term memory for pitch in AJ.

In discussing retroactive interference in AJ, I have until now considered only the number of intervening trials, but obviously performance may depend upon which particular items occurred during the retention interval. Holland and Lockhead (1968) have examined the effect of the immediately preceding stimulus on the current response in AJ, and their results indicate that response variability is lower when the items on trials N and N - 1 are similar in value to one another. Although I did not analyze this effect formally, inspection of my data reveals that the findings of Holland and Lockhead are clearly replicated in the present experiments. This effect of similarity is the same as that observed by Wickelgren (1969) and Massaro (1970a, Experiment I) in studies involving delayed comparison of pitch, and thus there appears to be another phenomenon which is characteristic of both relative- and absolute-judgment performance.

The Effect of Number of Alternatives

The size of the stimulus set has been a theoretically important variable partly because it is monotonically related to the H_s measure of information theory. As was noted previously, N may influence AJ accuracy in two ways. First, there is an indirect effect resulting from the confounding of N with the mean retention interval. As the number of stimuli is increased, the subject must on the average store information in memory over a greater period of time and a larger number of intervening trials. Second, N might affect performance even when the retention interval is held constant.

When the dependent variable is the percentage of correct responses, performance in Experiment III was less accurate as N was increased from five to nine alternatives. This effect was observed on repetition trials, where memory variance was minimal, but the influence of N was even greater on nonrepetition trials. Thus, if these results are taken literally, both the discrimination and retention of pitch vary inversely with N in the AJ task. However, most investigators would agree that percent correct is not a particularly apt measure when dealing with the effects of N, and it is impossible to assess the role of this variable without adopting a particular theory of the judgment process. In order to illustrate this point, the sensitivity parameters of several theoretical models were estimated from the present data.

First, the results of Experiment III were adjusted by the traditional "correction for chance" which derives from high-threshold theory. It seems that this guessing model interprets the data in much the same way as the simple percent-correct measure. According to the guessing model, the discriminability of the stimuli is lower and forgetting is

more extensive as N is made larger. A second theory, the forced-choice detection model provides a more simple and satisfactory explanation of the results of Experiment III. When this model is employed, the d' sensitivity parameter is invariant over changes in N on both repetition and nonrepetition trials. Thus, according to the detection model, both the discrimination and retention of pitch information are independent of N. Unfortunately, both this theory and the guessing model must be rejected because they fail to predict other aspects of the data.

It is usually found that when the information-theoretic model is applied to AJ data, the H_t measure remains roughly constant as N is made large. MacRae (1970) has suggested that this finding is an artifact of bias in published estimates of H_t. When bias is minimal, H_t appears to first increase and then drop as the size of the stimulus set is made greater. In the present experiments, the sample size was large and the maximum value of H_s was relatively small, and so bias should not play an important role. Under all conditions, it was found that H_t was an increasing function of N. Moreover the rate of increase was greatest on repetition trials, where memory effects were minimal. Presumably, no terminal decrease in H_t was found because only the rising part of the function was sampled.

The information model interprets the present data as demonstrating that both the discrimination and retention of pitch information are better as N is made larger. This interpretation is exactly the opposite of the one provided by the guessing model. And so, while one would like to provide an unequivocal answer about the effects of N in the AJ task, it is clear that the conclusions which one draws depend

upon the particular theory of judgment which is adopted. At the moment it is unlikely that any existing theory of AJ performance is the correct one. In any event, the experiments conducted here provide data which must be accounted for by any adequate theoretical model.

The Effect of Ordinal Position

It is usually found in AJ that the function relating performance accuracy to ordinal position in the stimulus set is U-shaped. Earlier, I reviewed evidence favoring a memory theory of the ordinal-position effect, such as Eriksen and Hake's (1957) subjective-standard hypothesis. The latter theory states that the observer stores the end items in memory, and uses them as references for judging other members of the stimulus set. This theory was supported by a number of experiments demonstrating that factors which influence the memory system are able to alter the shape of the ordinal-position curve. However, there is one critical test of the hypothesis that subjects are more accurate in judging the end items because these stimuli are stored more effectively in memory: When AJ data are collected under conditions where the subject is not required to store information in memory, the ordinal-position curve should be flat.

On repetition trials in Experiment III, retention effects were at a minimum since there was no temporal delay between successive tones. It was expected that ordinal position would have no effect under these circumstances, but this prediction was completely contradicted by the data. When the subjects judged five, seven, or nine alternative tone frequencies, a U-shaped ordinal-position curve was found on both repetition and nonrepetition trials. Moreover, the size

of the effect was approximately as great on repetition trials as on nonrepetitions, even though the retention interval was zero when the stimulus repeated. Several response-bias interpretations were considered and rejected because they were not supported by the data. Thus the present findings are not in agreement with a memory theory of the ordinal-position effect, such as the subjective-standard hypothesis of Eriksen and Hake.

If the U-shaped ordinal position curve does not reflect more effective retention of the end items, then it may represent a genuine perceptual effect. At any rate, this is the one phenomenon which does not appear to have any counterpart in a discrimination task. In the latter situation, one does not normally find striking end-effects, and so scales based on confusions in the relative- and absolute-judgment paradigms will not agree. The data obtained here are especially surprising since, in all other respects my results correspond very nicely to those found in comparison tasks. Clearly further research is needed to clarify the nature of the ordinal-position effect in AJ.

The Effect of Feedback

It has been shown by a number of investigators that AJ performance is less accurate when knowledge of results is omitted. I argued earlier that this does not result from an effect of FB on subjects' ability to discriminate stimuli from one another or to store sensory information over time. Rather, my hypothesis was that FB influences the decision rule mapping the set of overt responses which the subject is required to make in the AJ task onto internal sensory events. In particular, I suggested that when FB is provided, subjects will attempt to maximize

their accuracy of responding, while without FB they will maximize internal consistency. Much the same point was made by Gibson (1953) and more recently by investigators working with the yes-no detection task (e.g., Tanner, Haller, and Atkinson, 1967; Friedman, Carterette, Nakatani, and Ahumada, 1968). The latter task is a two-choice AJ paradigm.

Four predictions were derived from my hypothesis and all four were supported by the data of Experiment V. The most important was that subjects should be as consistent without FB as they are accurate with FB. Indeed it was found that, even though the percentage of correct responses was lower in the no-FB condition, subjects did use their responses quite consistently in that condition. When the stimulus repeated, the observers repeated their previous response in the no-FB condition to approximately the same degree that they made a correct response when FB was given. When knowledge of results was provided, subjects tended to repeat their previous response on repetition trials only when that response was correct. Thus Experiment V demonstrates that FB exerts its effect on the subjects' decision rule, and the reliability of responding without FB is roughly equal to the veridicality of responding with FB.

Conclusion

The AJ paradigm originated as a psychophysical task, and it has been used by a number of investigators for developing "sensory" scales. Despite this fact, the experiments conducted here have shown that in the typical AJ study, almost all of the variance is attributable to the effects of memory rather than to the operation of the sensory systems. AJ performance is characterized by rapid forgetting, and this

is especially interesting since in this task the retention interval varies from trial to trial. When memory effects are minimized, subjects make very few errors and AJ accuracy is on a level with that found in discrimination tasks. All in all, the results of the five experiments conducted here demonstrate that the limitation of AJ performance is due to poor perceptual memory for unidimensional stimuli.

APPENDIX

CONFUSION MATRICES FOR REPETITION AND NONREPETITION TRIALS OF EXPERIMENT III

FIVE ALTERNATIVES, REPETITIONS

	Response						
Stimulus	1	2	3	4	5	n	p
-1-	248	63	6	-	2	319	.212
	.777	.197	.019	-	.006		
-2-	40	227	50	5	3	325	.216
	.123	.698	.154	.015	.009		
-3-	8	34	196	47	6	291	.193
	.027	.117	.674	.162	.021		
-4-	-	3	36	195	36	270	.179
	-	.011	.133	.722	.133		
-5-	-	2	3	30	268	303	.201
	-	.007	.010	.099	.884		
n	296	329	291	277	315	1508	
p	.196	.218	.193	.184	.209		

CONFUSION MATRICES FOR REPETITION AND NONREPETITION TRIALS OF EXPERIMENT III

FIVE ALTERNATIVES, NONREPETITIONS

	Response						
Stimulus	1	2	3	4	5	n	p
-1-	804	262	32	7	-	1105	.198
	.728	.237	.030	.006	-		
-2-	243	664	111	35	4	1057	.190
	.230	.628	.105	.033	.004		
-3-	25	180	673	232	20	1130	.203
	.022	.159	.596	.205	.018		
-4-	4	32	191	727	202	1156	.207
	.003	.028	.165	.629	.175		
-5-	2	6	22	252	844	1126	.202
	.002	.005	.020	.224	.750		
n	1078	1144	1029	1253	1070	5574	
p	.193	.205	.185	.225	.192		

CONFUSION MATRICES FOR REPETITION AND NONREPETITION TRIALS OF EXPERIMENT III

SEVEN ALTERNATIVES, REPETITIONS

	Response								
Stimulus	1	2	3	4	5	6	7	n	p
-1-	215	59	6	1	-	-	-	281	.140
	.765	.210	.021	.004	-	-	-		
-2-	47	191	37	4	1	-	-	280	.140
	.168	.682	.132	.014	.004	-	-		
-3-	5	42	171	48	8	2	-	276	.138
	.018	.152	.620	.174	.029	.007	-		
-4-	-	3	43	182	58	8	-	294	.146
	-	.010	.146	.619	.197	.027	-		
-5-	-	1	9	39	174	69	3	295	.147
	-	.003	.031	.132	.590	.234	.010		
-6-	-	-	2	9	23	194	60	288	.143
	-	-	.007	.031	.080	.674	.208		
-7-	-	1	-	2	5	47	238	293	.146
	-	.003	-	.007	.017	.160	.812		
n	267	297	268	285	269	320	301	2007	
p	.133	.148	.134	.142	.134	.159	.150		

CONFUSION MATRICES FOR REPETITION AND NONREPETITION TRIALS OF EXPERIMENT III

SEVEN ALTERNATIVES, NONREPETITIONS

	Response								
Stimulus	1	2	3	4	5	6	7	n	p
-1-	1054	462	90	25	3	2	1	1637	.143
	.644	.282	.055	.015	.002	.001	.001		
-2-	411	820	284	60	17	5	1	1598	.140
	.257	.513	.178	.038	.011	.003	.001		
-3-	61	361	803	337	59	19	3	1643	.144
	.037	.220	.489	.205	.036	.012	.002		
-4-	22	97	334	771	300	85	14	1623	.143
	.014	.060	.206	.475	.185	.052	.009		
-5-	4	5	45	234	798	468	85	1639	.144
	.002	.003	.027	.143	.487	.286	.052		
-6-	1	3	11	50	272	881	425	1643	.144
	.001	.002	.007	.030	.166	.536	.259		
-7-	-	2	3	16	63	386	1166	1636	.143
	-	.001	.002	.010	.039	.236	.713		
n	1553	1750	1570	1493	1512	1846	1695	11,418	
p	.136	.153	.138	.131	.132	.162	.148		

CONFUSION MATRICES FOR REPETITION AND NONREPETITION TRIALS OF EXPERIMENT III

NINE ALTERNATIVES, REPETITIONS

	Response										
Stimulus	1	2	3	4	5	6	7	8	9	n	p
-1-	188	52	7	-	-	-	-	-	-	247	.107
	.761	.211	.028	-	-	-	-	-	-		
-2-	55	147	22	5	-	1	-	-	-	230	.100
	.239	.639	.096	.022	-	.004	-	-	-		
-3-	3	53	179	39	6	-	-	-	-	280	.123
	.011	.189	.639	.139	.021	-	-	-	-		
-4-	-	2	47	139	42	5	2	-	-	237	.103
		.008	.198	.586	.177	.021	.008	-	-		
-5-	-	-	6	41	149	46	9	-	1	252	.110
	-	-	.024	.163	.591	.183	.036	-	.004		
-6-	-	-	-	8	48	167	62	10	1	296	.130
	-	-	-	.027	.162	.564	.209	.034	.003		
-7-	-	-	-	-	3	22	140	80	4	249	.108
	-	-	-	-	.012	.088	.562	.321	.016		
-8-	-	-	-	-	-	4	33	183	57	277	.121
	-	-	-	-	-	.014	.119	.661	.206		
-9-	-	1	-	-	-	-	1	30	199	230	.100
	-	.004	-	-	-	-	.004	.130	.865		
n	246	254	261	232	248	245	247	303	262	2298	
p	.107	.111	.114	.101	.108	.107	.107	.132	.114		

CONFUSION MATRICES FOR REPETITION AND NONREPETITION TRIALS OF EXPERIMENT III

NINE ALTERNATIVES, NONREPETITIONS

	Response										
Stimulus	1	2	3	4	5	6	7	8	9	n	p
-1-	1282	554	121	36	10	5	2	1	1	2012	.112
	.637	.275	.060	.018	.005	.002	.001	-	-		
-2-	561	971	384	73	28	5	1	-	-	2023	.113
	.277	.480	.190	.036	.014	.002	-	-	-		
-3-	148	462	838	412	92	22	5	2	2	1983	.110
	.075	.233	.423	.208	.046	.011	.003	.001	.001		
-4-	31	137	506	910	291	90	19	4	3	1991	.111
	.016	.069	.254	.457	.146	.045	.010	.002	.002		
-5-	7	31	78	365	918	446	104	22	11	1982	.110
	.004	.016	.039	.184	.463	.225	.052	.011	.006		
-6-	3	2	27	77	334	893	497	109	22	1964	.109
	.002	.001	.014	.039	.170	.455	.253	.055	.011		
-7-	-	-	5	19	74	284	925	573	128	2008	.112
	-	-	.002	.009	.037	.141	.461	.285	.064		
-8-	-	2	2	6	22	85	420	931	517	1985	.110
	-	.001	.001	.003	.011	.043	.212	.469	.260		
-9-	1	2	-	2	13	36	109	471	1399	2033	.113
	-	.001	-	.001	.006	.018	.054	.232	.688		
n	2033	2161	1961	1900	1782	1866	2082	2113	2083	17981	
p	.113	.120	.109	.106	.099	.104	.116	.118	.116		

REFERENCES

Alluisi, E. A., & Sidorsky, R. C. The empirical validity of equal discriminability scaling. Journal of Experimental Psychology, 1958, 55, 86-95.

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

Beare, A. C. Color name as a function of wave-length. American Journal of Psychology, 1963, 76, 248-256.

Brelsford, J. W., Keller, L., Shiffrin, R. M., & Atkinson, R. C. Short-term recall of paired-associates as a function of the number of interpolated pairs. Psychonomic Science, 1966, 4, 73-74.

Chapanis, A., & Halsey, R. Absolute judgments of spectrum colors. Journal of Psychology, 1956, 42, 99-103.

Cuddy, L. L. Practice effects in absolute judgments of pitch. Journal of the Acoustical Society of America, 1968, 43, 1069-1076.

Durlach, N. I., & Braida, L. D. Intensity of perception. I. Preliminary theory of intensity resolution. Journal of the Acoustical Society of America, 1969, 46, 372-383.

Eriksen, C. W., & Hake, H. W. Multidimensional stimulus differences and accuracy of discrimination. Journal of Experimental Psychology, 1955, 50, 153-160.

Eriksen, C. W., & Hake, H. W. Anchor effects in absolute judgments. Journal of Experimental Psychology, 1957, 53, 132-138.

Fechner, G. T. Elemente der psychophysik. Leipzig: Breitkoph & Härtel, 1860.

Fitts, P. M., & Posner, M. I. Human performance. Belmont, Calif.: Brooks-Cole, 1967.

Friedman, M. P., Carterette, E. C., Nakatani, L., & Ahumada, A. Comparisons of some learning models for response bias in signal detection. Perception and Psychophysics, 1968, 3, 5-10.

Galanter, E., & Messick, S. The relation between category and magnitude scales of loudness, Psychological Review, 1961, 68, 363-372.

Garner, W. R. An equal discriminability scale for loudness judgments. Journal of Experimental Psychology, 1952, 43, 232-238.

Garner, W. R. An informational analysis of absolute judgments of loudness. Journal of Experimental Psychology, 1953, 46 373-380.

Garner, W. R. Uncertainty and structure as psychological concepts. New York: John Wiley & Sons, 1962.

Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. Psychological Review, 1951, 58, 446-459.

Gibson, E. J. Improvement in perceptual judgments as a function of controlled practice or training. Psychological Bulletin, 1953, 50, 401-431.

Green, D. M., & Birdsall, T. G. The effect of vocabulary size on articulation score. In J. A. Swets (Ed.), Signal detection and recognition by human observers. New York: Wiley, 1964.

Green, D. M., & Swets, J. A. Signal detection theory and psychophysics. New York: John Wiley & Sons, 1966.

Hake, H. W., & Garner, W. R. The effect of presenting various numbers of discrete steps on scale reading accuracy. Journal of Experimental Psychology, 1951, 42, 358-366.

Hartman, E. B. The influence of practice and pitch-distance between tones on the absolute identification of pitch. American Journal of Psychology, 1954, 67, 1-14.

Holland, M. Channel capacity and sequential effects: The influence of the immediate stimulus history on recognition performance. (Doctoral dissertation, Duke University) Ann Arbor, Mich.: University Microfilms, 1969, No. 69-16, 758.

Holland, M. K., & Lockhead, G. R. Sequential effects in absolute judgments of loudness. Perception and Psychophysics, 1968, 3, 409-414.

James, W. Principles of psychology. New York: Holt, 1890.

Johnson, P. J., Chatelier, P., & Mocharnuk, J. Factors influencing short-term kinesthetic recall. Paper presented at the meetings of the Psychonomic Society, St. Louis, November, 1969.

Kinchla, R. A., & Smyzer, F. A diffusion model of perceptual memory. Perception and Psychophysics, 1967, 2, 219-229.

Kornblum, S. Serial-choice reaction time: Inadequacies of the information hypothesis. Science, 1968, 159, 432-434.

Kornblum, S. Sequential determinants of information processing in serial and discrete choice reaction time. Psychological Review, 1969, 76, 113-131.

Lenneberg, E. H. A probabilistic approach to language learning. Behavioral Science, 1957, 2, 1-13.

Lenneberg, E. H. Color naming, color recognition, color discrimination. A re-appraisal. Perceptual and Motor Skills, 1961, 2, 375-382.

Levett, C. M., Jr. Errors of interpolation in instrument reading and setting. Journal of Applied Psychology, 1952, 36, 49-52.

MacRae, A. W. Channel capacity in absolute judgment tasks: An artifact of information bias? Psychological Bulletin, 1970, 73, 112-121.

Massaro, D. W. Retroactive interference in short-term recognition memory for pitch. Journal of Experimental Psychology, 1970, 83, 32-39. (a)

Massaro, D.W. Forgetting: Interference or decay? Journal of Experimental Psychology, 1970, 83, 238-243. (b)

Massaro, D.W. Consolidation and interference in the perceptual memory system. Perception and Psychophysics, 1970, 7, 153-156. (c)

Miller, G. A. The magical number seven, plus or minus two. Psychological Review, 1956, 63, 81-97.

Muller, P. F., Jr., Sidorsky, R. C., Slivinske, A. J., Alluisi, E.A., & Fitts, P. M. The symbolic coding of information on cathode ray tubes and similar displays. USAF WADC Technical Report, 1955, No. 55-375.

Murdock, B. B. The distinctiveness of stimuli. Psychological Review, 1960, 67, 16-31.

Nash, H. The judgment of linear proportions. American Journal of Psychology, 1964, 77, 480-484.

Pollack, I. The information of elementary auditory displays. Journal of the Acoustical Society of America, 1952, 24, 745-749.

Pollack, I. The information of elementary auditory displays, II. Journal of the Acoustical Society of America, 1953, 25, 765-769.

Pollack, I. Identification and discrimination of components of elementary auditory displays. Journal of the Acoustical Society of America, 1956, 28, 906-909.

Shepard, R. N. Stimulus and response generalization: Deduction of the generalization gradient from a trace model. Psychological Review, 1958, 65, 242-256.

Shepard, R. N. The analysis of proximities. Multidimensional scaling with an unknown distance function. II. Psychometrika, 1962, 27, 219-246.

Siegel, W., & Siegel, J. A. Memory effects in absolute judgment: Implications for information measurement and psychophysical scaling. University of Michigan, Mental Health Research Institute Communication #234, 1968.

Spitz, H. H. Information transmission in an absolute judgment task with feedback using normal and retarded subjects. Journal of Comparative and Physiological Psychology, 1967, 64, 85-92.

Swets, J. A. (Ed.) Signal detection and recognition by human observers. New York: Wiley, 1964.

Tanner, T. A., Jr., Haller, R. W., & Atkinson, R. C. Signal recognition as influenced by presentation schedules. Perception and Psychophysics, 1967, 2, 349-358.

Thurstone, L. L. Psychophysical analysis. American Journal of Psychology, 1927, 38, 368-389.

Torgerson, W. S. Theory and methods of scaling. New York: Wiley, 1958.

Volkman, F. C., & Engen, T. Three types of anchoring effects in the absolute judgment of hue. Journal of Experimental Psychology, 1961, 61, 1-17.

Ward, L. M., & Lockhead, G. R. Sequential effects and memory in category judgments. Journal of Experimental Psychology, 1970, 84, 27-34.

Weintraub, D. J. Rectangle discriminability: Perceptual relativity and the Law of Pragnanz. Journal of Experimental Psychology, 1971, in press.

Wever, E. G., & Zener, K.E. The method of absolute judgment in psychophysics. Psychological Review, 1928, 35, 466-493.

Wickelgren, W. A. Consolidation and retroactive interference in short-term recognition memory for pitch. Journal of Experimental Psychology, 1966, 72, 250-259.

Wickelgren, W. A. Associative strength theory of recognition memory for pitch. Journal of Mathematical Psychology, 1969, 6, 13-61.

Woodworth, R. S., & Schlosberg, H. Experimental psychology. (Rev. ed.) New York: Henry Holt, 1954.