Reprinted from
8 February 1974, Volume 183, pp. 482-488

SCIENCE

How Big Is a Chunk?

Herbert A. Simon

How Big Is a Chunk?

By combining data from several experiments, a basic human memory unit can be identified and measured.

Herbert A. Simon

During the past 15 years, substantial progress has been made toward understanding man's problem-solving and other complex cognitive processes—toward measuring the immense search spaces in problem-solving tasks, and identifying some of the heuristic principles that people use to reduce these spaces to manageable proportions. The understanding of problem-solving now being acquired suggests a new significance, and a new application, for the simpler cognitive tasks of the classical psychological laboratory.

A crucial role in problem-solving is played by man's short-term memory and the processes that transfer information from short-term to long-term memory (fixation processes). To continue the progress toward understanding complex cognitive behaviors, it is necessary to have good estimates of the basic parameters of short-term memory and of the memory fixation process. The classical laboratory tasks of experimental psychology provide efficient laboratory settings for estimating some of these parameters. Old experiments can be analyzed in new ways, and their findings can take on new significance, when the analyses are guided by knowledge of the complex processes in which these same parameters reappear.

In this article I examine some very simple experiments of a familiar kind, but examine them in a rather unfamiliar way. I seek to extract from earlier studies estimates of parameters that appear to be crucial to human performance in complex tasks and to illustrate how these parameter values predict behavior in a range of laboratory situations.

The author is Richard King Mellon Professor of Computer Science and Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

What Is an Experiment?

In psychology, the term "experiment" came to have a very specific meaning. An experiment required a dependent variable and one or more independent variables, the latter to be manipulated over a set of "experimental conditions." A null hypothesis was erected: that the mean values of the dependent variables were not significantly different for sets of subjects run under different experimental conditions. If the data led to rejection of the null hypothesis, one bit of information had been obtained—that the dependent variable was, apparently, affected by the independent variable.

Experiments conceived, executed, and published within this framework produced such facts as that the ease of learning nonsense syllables is related to their meaningfulness, or to their similarity. They paid little attention to the strength of the relation—whether running up the scale of meaningfulness from 0 to 100 reduced learning time by 5 percent, or 50 percent, or 500 percent.

Only in psychophysics (and, in a different way, in operant conditioning) was this orthodoxy ignored. When a subject is asked to compare the differences in pitch between two pairs of tones, the point of the experiment is to estimate the shape and parameters of a function in which physical pitch is the independent variable and "subjective" pitch the dependent variable. What is published is not the one-bit message that there is a relation between physical and subjective pitch, but the actual form of the function and the numerical values of its parameters.

What is done in the psychophysical laboratory does not fit the narrow definition of "experiment." No "control" condition is contrasted with the "experimental" condition. Instead of significance tests, there are reports of the standard or probable errors of measures—quite another matter. Probable errors are not intended to test whether a parameter may be different from zero, but to indicate the precision with which the measurements were carried out.

The Span of Immediate Recall

The methods of experimental psychology are now shifting from the narrow view of experiment bound up with hypothesis-testing to a view of experiment that puts its principal emphasis upon estimating parameters and the shapes of functions. It is now becoming possible to obtain replicable estimates of basic parameters that characterize human memory and to draw implications from these estimates for complex performance.

About 17 years ago, George A. Miller (*1*) introduced a "magic number"—the number of chunks that can be held in short-term memory for immediate recall. Of the studies he cites, at least 13 employ the parameter-estimation paradigm, and no more than two employ the standard hypothesis-testing paradigm.

In introducing the "chunk," Miller was artfully vague (*1*, p. 93):

> The contrast of the terms *bit* and *chunk* also serves to highlight the fact that we are not very definite about what constitutes a chunk of information. For example, the memory span of five words that Hayes obtained . . . might just as appropriately have been called a memory span of 15 phonemes, since each word had about three phonemes in it. Intuitively, it is clear that the subjects were recalling five words, not 15 phonemes, but the logical distinction is not immediately apparent. We are dealing here with a process of organizing or grouping the input into familiar units or chunks, and a great deal of learning has gone into the formation of these familiar units.

Miller makes a fundamental distinction between a conventionally defined amount (numbers of words or phonemes) of material and a chunk of that material, which is a particular amount that has specific psychological significance. Thus, when measuring quantity of material conventionally, one may define either the word or the phoneme as the unit. The words in Miller's example are either one or three units in length, depending on whether the word or the phoneme, respectively, is the standard of measurement.

There is no such freedom with respect to chunks—else there would be no magic in the magic number. The significance of the magic number lies in the assertion that the capacity of short-term memory, measured in chunks, is independent of the material of which those chunks are manufactured—five chunks worth of words, five chunks of digits, five chunks of colors, five chunks of shapes, five chunks of poetry or prose (*2*). But unless there is a way of determining the chunk size of any given material independently of the measurement of memory span, the assertion that there is a fixed chunk span loses all empirical content.

Miller saves his proposition from tautology by using two methods for estimating chunk size, one depending on knowledge of the previous experiences of his subjects, the other depending on training procedures—on experience provided in the laboratory. With respect to the former, he again determines that words, not phonemes, are to be regarded as the chunks (*1*, p. 93):

> Intuitively, it is clear that the subjects were recalling five words, not 15 phonemes, but the logical distinction is not immediately apparent. We are dealing here with a process of organizing or grouping the input into familiar units or chunks, and a great deal of learning has gone into the formation of these familiar units.

With respect to the latter, altering the chunking of material by laboratory training, he says (*1*, p. 93):

> In order to speak more precisely, therefore, we must recognize the importance of grouping or organizing the input sequence into units or chunks. Since the memory span is a fixed number of chunks, we can increase the number of bits of information that it contains simply by building larger and larger chunks, each chunk containing more information than before.

He continues by describing (*1*, pp. 93–95) the now-famous experiment of Sidney Smith, who increased the number of binary (0 or 1) digits he could recall from about 12 to 40 by recoding each sequence of three binary digits into a single, octal (0 through 7) digit. Assuming that the digit could be equated with the chunk, the length of the sequence of digits that could be recalled should be independent of the size of the alphabet of digits, whether two or eight. And so it was.

Table 1. Span of immediate recall for words and phrases (with the author as subject).

Words and phrases	Span: Syllables	Span: Words	Span: Imputed chunks	Syllables (chunk)
1-syllable	7	7	7	1.0
2-syllable	14	7	7	2.0
3-syllable	18	6	6	3.0
2-word	22	9	4	5.5
8-word	26	22	3	8.7

The reality of the chunk can be pursued further by using words instead of digits and past experience instead of training in experiments analogous to Smith's. The span of immediate recall for words is roughly equal to the span for unrelated letters or for digits. This is the principal reason for concluding, as Miller did, that a word is a chunk. But the implications of the chunk hypothesis can be tested; if it is correct, the recall span for words should not depend on the number of syllables the words contain.

This prediction is easily checked. I made up lists of one-syllable, two-syllable, and three-syllable English nouns and tested my own span by later reading them aloud at about two items per second, then recalling them. My span was nearly seven words for the one-syllable and two-syllable nouns, and about six words for three-syllable nouns. Thus, if words are chunks, the span was not quite constant; there was a variation of some 15 percent between the extreme conditions. But what of the alternative—of treating syllables as chunks? The span for one-syllable nouns was 7 syllables; for three-syllable nouns, 18—a ratio of 2.5 to 1. One must conclude, therefore, that the syllable is not the invariant unit that measures short-term memory capacity, but that the word *may* be.

The chunking hypothesis does not assert that the word will always be the unit. Units much larger than words may be highly familiar, hence may serve as chunks. I tried to recall after one reading the following list of words: Lincoln, milky, criminal, differential, address, way, lawyer, calculus, Gettysburg. I had no success whatsoever. I should not have expected success, for the list exceeded my span of six or seven words. Then I rearranged the list a bit, as follows:

Lincoln's Gettysburg Address
Milky Way
Criminal lawyer
Differential calculus

I had no difficulty at all. Obvious? It is only obvious if one accepts the chunk hypothesis and if one knows that, in the culture in which I was raised, the four items in the list are, in fact, familiar chunks. If these premises are accepted, I have simply performed a variant of the Smith experiment.

The prediction that I should be able to recall lists of six familiar phrases of this sort is not substantiated. Four or five seems to be about the limit, indicating that something about the additional length of the material reduces the total number of imputed chunks that can be retained—just as in the comparison of three-syllable with one-syllable words. To substantiate further this gentle decline in capacity with imputed chunk length, I extrapolated from familiar two-word and three-word phrases to longer ones. Consider the list:

Four score and seven years ago
To be or not to be, that is the question
In the beginning was the word
All's fair in love and war

Lists of three such phrases were all I could recall with reliability, although I could sometimes retain four.

To summarize the results up to this point: as one moves from one-syllable to three-syllable words, then to familiar two-word and three-word phrases, then to familiar phrases of six to ten words, the memory span, measured in syllables, words, and imputed chunks, varies as shown in Table 1.

None of the measures remains constant, but the number of chunks retained declines only by a factor of two, while the number of words retained increases by a factor of three, and the number of syllables almost by a factor of four. I conclude that the "constant capacity in chunks" hypothesis is a rough first approximation of the true state of affairs but that it must be refined—perhaps by taking into account the additional time required to rehearse the longer passages—in order to achieve a fully satisfactory fit to the data.

Time to Learn

The experiments and data reported thus far still leave the chunk in an unsatisfactory status. Limiting the data to memory span experiments provides no evidence for the reality of the imputed chunks other than some agreement between the measured span and a priori notions of what the "familiar unit" actually is.

The difficulty arises because the number of chunks is not directly observable. Viewed in isolation, the hypothesis is not really an empirical statement at all, but a definition of "chunk": a chunk of any kind of stimulus material is the quantity that short-term memory will hold five of.

Simple examples from the physical sciences show, however, that difficulties of this sort can often be removed by compounding them. It is often observed that Newton's Second Law of Motion (force equals mass times acceleration) is not really a law, but a definition of force. By the same token, Hooke's Law, which states that the extensions of a spring are proportional to the forces applied to it, is also merely a definition of those forces. Taken together, however, the two laws can be tested empirically (for example, by whirling a weighted spring); one can determine whether the magnitudes of the forces determined by the one law (viewed as definition) agree with the magnitudes of the same forces determined by the other law.

A law of the form $y = am$, where m is an observable but y is not, can be used to estimate y, but observations can never refute the law. Suppose there is a second law, of the form $y = bp$, where p is also an observable. Taken together, the two laws imply $am = bp$, which is a testable proposition since the single free parameter, a/b, can be estimated from observations. Thus, one can use the first equation to estimate values of y and then see whether these satisfy the second equation.

In the case at hand, there is a quantity that is directly observable (number of syllables immediately recallable, say) and another, unobservable quantity that is postulated by the theory (number of syllables per chunk). Call these S and s, respectively. The hypothesis that short-term memory has a constant capacity of five chunks can be rendered as simply $S = 5s$. Given the measurement of S, the observable, this equation can be used to estimate chunk size: $s = S/5$. The theory is essentially untestable, however, because for any observed S there always exists an s that satisfies the equation.

Suppose, however, that there is another observable—the number of syllables that can be fixated per minute in a rote memory experiment. Consider the hypothesis that this number (F) is proportional to the number of syllables per chunk (s): $F = ks$. As before, the theory provides an equation for estimating an unobservable, but the theory is untestable.

Table 2. Spans of immediate recall [source: Brener (*13*)].

Test	Mean span
Digits	7.98
Nonsense syllables	2.49
Constants (visual)	7.30
Geometrical designs	5.31
Colors	7.06
Concrete words (visual)	5.76
Paired associates (pair)	2.50
Abstract words (visual)	5.24
Commands*	2.42
Sentences (six words)	1.75

* Each command involved a relation between two objects.

If, however, one puts the two hypotheses together, one can combine the two estimating equations for s, eliminating the unobserved s between them: $F = aS$, where a is a new constant parameter. This equation makes it possible to estimate the time required per syllable to learn any particular kind of stimulus material from the memory span for that same kind of material. The constant, a, can be estimated for any single kind of material, say common English words, thus reducing by 1 the degrees of freedom. Hence, the conjunction of the two hypotheses (the span of immediate recall is five chunks; the time required to memorize a chunk of material is k seconds) is testable, even though neither hypothesis taken separately is. Moreover, if the hypotheses satisfy the empirical test, either of the original equations can be used to measure chunk size (up to a constant of proportionality) for all kinds of stimulus material.

The second of the two hypotheses above—that learning time is proportional to the number of chunks to be fixated—was introduced without any particular motivation. Before marshaling the data to test the two hypotheses, let me mention some of the evidence for the notion that quantity of material learned is proportional to time. The hypothesis rests on three legs—two empirical, the other theoretical.

Since most learning theories connected the learning of nonsense syllables with reinforcement, it was natural to measure ease or difficulty of learning by number of trials (that is, number of reinforcements) required to reach criterion. The following statement, published in 1942, is typical (*3*, pp. 105–106):

> When the presentation time of each syllable in a 12-syllable list is increased from 2 seconds to 4 seconds, the mean trials required to attain a criterion of 7 syllables correct out of 12 decreases from 6.05 to 3.28.

Of course, if one does not start with the theoretical presumption that the trial's the thing, there is a much more parsimonious way of reporting this experiment. One can say, simply:

> When the presentation time of each syllable in a 12-syllable list is increased from 2 seconds to 4 seconds, the mean length of time required to attain a criterion of 7 syllables correct out of 12 remains almost constant, increasing only from 12.1 seconds to 13.1 seconds.

A number of other experiments done before World War II support the hypothesis that the total learning time per unit of material of any particular kind is constant. This observation was an important clue that led Feigenbaum, in 1958, to use learning *time* rather than *trials* as the key variable in his EPAM (Elementary Perceiver and Memorizer) learning theory (*4*).

Apparently no experiments were run before 1960 with the deliberate aim of determining whether time, rather than trials, was the decisive variable in learning, although Wilcoxon, Wilson, and Wise (*5*) mentioned time constancy in 1961. The first experiment specifically designed to test the time-constancy hypothesis was conducted by Bugelski and published in 1962 (*6*). He found that the time required to learn a list was essentially independent of presentation speed over a wide range of speeds. Subsequent experiments have extended his result and have clarified the range of conditions under which the constancy may be expected to hold (*7*).

So much for the empirical evidence of time constancy in learning. The third leg of the stool is the EPAM theory of verbal learning, formulated as a simulation program for a digital computer (*4*). Since computers are serial devices, requiring time to carry out their processes, it was natural to hypothesize the same seriality in human beings, and hence to construct EPAM in such a way that amount of learning would be roughly proportional to time. As Feigenbaum and I stated the matter in an article on the serial position effect (*4*, p. 310):

> The fixation of an item on a serial list requires the execution of a sequence of information processes that requires, for a given set of experimental conditions, a

definite amount of processing time per syllable. The time per syllable varies with the difficulty of the syllables, the length of the list, the ability of the subject, and other factors.

While the experiments I have cited, as well as EPAM theory, support the idea that amount of material learned is proportional to learning time, they permit only comparisons of a single kind of learning material at different presentation speeds, not comparisons among different kinds of stimulus material. It still remains to define a unit quantity that permits the latter kind of comparison.

Testing the Chunking Hypothesis

I have now reviewed two basic hypotheses: that short-term memory holds a fixed number of chunks and that total learning time is proportional to the number of chunks to be assembled. The weakness of each hypothesis lies in its inability to provide an independent operational definition of the chunk. But by conjoining the two hypotheses, one removes the need for a priori assumptions about what constitutes a chunk.

In the EPAM theory, fixation is identified with assembling compound symbol structures from components—a familiar notion from association theory—and storing the compound structures in memory, appropriately "indexed." ("Indexing" simply means storing information that permits recovery of the compound structure upon recognition of its stimulus component.)

Thus, in paired-associate learning, a stimulus symbol and a response symbol can be assembled into a pair, indexed to the stimulus. But before this can happen, the stimulus and the response must each be assembled into a symbol compounded from their component letters (or phonemes, as the case may be). Under ordinary laboratory conditions of nonsense-syllable learning, the component letters can be assumed already to be unitary symbols at the outset of the experiments. For each pair in a set of paired-associate nonsense syllables of low association value, a total of about seven such compounding operations is required by EPAM for learning: three to familiarize the response, two to familiarize the stimulus (which need only be recognized, not recalled), and two to compound the pair (*8*). The corresponding number for a serial list is four compounding operations per syllable—three to familiarize the syllable, one to incorporate it in the list (*9*).

Table 3. Fixation times [source: Lyon (*14*)].

Material	Unit	Time (second/ unit)
Nonsense syllables	Syllable	27.9
Digits	Digit	25.5
Prose	Word	7.2
Poetry	Word	3.0

Analogous to Miller's encoding assumptions, which allowed him to predict the span of recall for recoded digits, are encoding assumptions for rote learning that enable us to predict the relative learning times for different materials. Thus, by making the a priori (but plausible) assumption that unfamiliar nonsense syllables are initially encoded as three chunks, while familiar syllables and one-syllable words are encoded as single chunks, one can predict the relative learning times for these materials.

The predictions that have been made on this basis, and tests of these predictions, are mainly reported in two articles (*10, 11*). The EPAM theory predicts, correctly, that lists of syllables of low familiarity will take nearly three times as long to learn as lists of highly familiar syllables. In fact, syllables of low familiarity take about 2.5 times as long to learn as syllables of high familiarity (*10*). The EPAM theory also predicts accurately the circumstances under which learning will have a one-trial character, and those under which it will be incremental (*11*).

If data from experiments on immediate recall could be directly compared with data from experiments on rote learning, a priori assumptions about chunk size would be unnecessary. There is considerable consistency in measurements of the relative memory spans for different kinds of materials—for example, the ratio of the memory span for digits to the memory span for words. No a priori assumptions about chunk size enter into this ratio. Similarly, there is fairly good consistency in the relative learning times reported for different kinds of materials—for example, the ratio of the learning time for nonsense syllables to the learning time for simple words. Again, this ratio is independent of assumptions about chunk size. If the theory proposed here is correct, the ratios obtained by these different and independent experimental operations should be the same for the same pairs of experimental materials (*12*).

With respect to memory span, there is a representative set of data in an experiment conducted by Brener in 1940 (*13*). Table 2, taken from Brener's study, shows memory spans for ten different kinds of stimulus material, ranging from digits and colors to six-word sentences. The ten mean values fall into four groups: spans of 10 (words in six-word sentences), around 7.5 (digits, consonants, colors), 5.0 (geometric designs, concrete nouns, abstract nouns), and 2.5 (nonsense syllables, paired associates, and simple commands). The task now is to compare the ratios of these spans with ratios of learning times for the same materials.

Unfortunately, data on learning times in serial or paired-associate paradigms are available for only a few of the materials for which digit spans have been measured. Those ratios that have been measured are reasonably consistent from one experiment to another. I have already mentioned the commonly observed 2.5 to 1 advantage in learning simple words over nonsense syllables. Averaging Brener's data for abstract and concrete words (the difference is only about 10 percent), one finds a span of 5.5 for words, as compared with 2.49 for nonsense syllables—a ratio of 2.2. Thus the two operations give us estimates—2.5 and 2.2, respectively—that differ by only about 15 percent.

Lyon's 1914 experiments, with himself as subject, in memorizing lists of hundreds of nonsense syllables, digits, and passages of prose and poetry provide a second source of data (*14*). Table 3 shows the time, in seconds per unit of material, it took him to memorize 200 units of material by reciting them once each day.

From the Brener data (*13*), the ratio of the span for sentences (measured in words) to the span for nonsense syllables is 10.5 to 2.49, or 4.2. From the Lyon data (*14*), the ratio of learning times for nonsense syllables and prose (per word) is 27.9 to 7.2, or 3.9. Again, the two ratios agree within about 10 percent.

There is no such happy agreement when memory spans and learning times for nonsense syllables and digits are compared. In the Brener data, the ratio of spans for the two kinds of stimuli is 7.98 to 2.49, or 3.2. In the Lyon data, the ratio of learning times is 27.9 to 25.5, or 1.1. No significance test is needed to show that something is wrong. The theory is certainly not entirely accurate.

Lyon himself argues that the excessive difficulty in learning the digit list arose from high intralist similarity. In a list of 200 digits, each digit will appear about 20 times, and each pair, on the average, about twice. However, in experiments involving the learning of nonsense syllables, where similarity is manipulated as the independent variable, the difference in learning times between high and low similarity conditions is about 30 percent—very far from the ratio of 3 to 1 that appears in the Simon and Feigenbaum data (*10*).

Another possible explanation is provided by McLean and Gregg (*15*), who showed that sequences of letters were learned about twice as rapidly when the letters were presented in groups of three or more as when they were presented one at a time. Their interpretation (an extension of the EPAM theory) was that, in the absence of cues from the experimenter, the subject was unable to group the letters consistently from one trial to the next, hence was forced to learn unnecessary additional groups. This hypothesis would account for two-thirds of the discrepancy in the Lyon data on digits.

That the length of the series learned by Lyon has something to do with the problem is shown by the fact that shorter strings of digits were learned much more rapidly than shorter strings of syllables. For example, the ratio of learning times for 16 syllables and 16 digits was almost exactly 2 to 1—still only two-thirds of the ratio predicted by the simple version of the theory.

These explanations are hardly satisfactory. The hypothesis simply does not work well with material in which there is frequent repetition of the same chunks. The difficulty in carrying out rote learning experiments with materials like digits, simple geometric designs, or colors is that, if one uses long series, one must repeat symbols; if one uses short series, one is in danger of confounding short-term with long-term memory, and hence not obtaining an independent measure of the parameters associated with the latter.

To summarize, the estimates of relative chunk size for nonsense syllables, words, and prose obtained from immediate recall experiments agree very well with the estimates obtained from rote learning experiments. There is serious disagreement, however, between the two estimates of digit chunk size; data for estimating chunk size for colors and geometric figures are apparently not available from the rote learning paradigm.

Table 4. Stanford-Binet norms for digit span [source: Woodworth and Schlosberg (*16*)].

Age (years)	Digits
2.5	2
3.0	3
4.5	4
7.0	5
10.0	6
College	8

The theory had some successes, but also a clear-cut (although perhaps temporary) failure. The failure is as instructive as the successes. It did not arise from either experiment taken in isolation from the other. Each was perfectly consistent within itself—it provided the one bit of information that it was capable of giving within the classical paradigm for each experimental condition. It does not make much sense to ask, within the context of Lyon's experiment alone, whether digits "should" have been learned faster than nonsense syllables. After an independent experiment has predicted a 3-to-1 advantage of digits over syllables, this "should" becomes something that must be taken seriously.

The main importance of invariants lies in their power to strip away the complexity and diversity of a whole range of phenomena and to reveal the simplicity and order underneath. Invariants, however, not only provide explanations for simple cognitive phenomena, they are also needed in the explanation of more complicated phenomena of thinking and problem-solving. Having discovered what a chunk is—if we have—it remains to be seen how it can be used in predicting human cognitive behavior in complex settings.

Significance of the Chunk in Cognition

The examples I use refer to three very different situations. The first is a modest extrapolation from the immediate recall experiments I have already examined: Does the theory of chunking have implications for the change in memory span with age? The second is an extrapolation to a relatively structured task that one might not even want to call "problem-solving"—mental multiplication of relatively large numbers. The third is an extrapolation to the initial stages of problem-solving, the period during which the subject characterizes for himself the problem that has been placed before him. In all three cases, the extrapolation depends not merely on having a general hypothesis that some independent variable affects some dependent variable, but on having quantitative estimates, derived from the simpler situations, of the values of parameters.

Because digit span increases with age, tests of digit span are included in most standard instruments for measuring mental age. Span is measured, of course, in common units (that is, digits) that might or might not represent the same number of chunks at different ages. In fact, the chunking hypothesis forces one to conclude that, with cumulative experience with numbers, children should learn to encode digits in larger and larger chunks, so that an increasing number of digit pairs and even triplets might become recognizable as a single chunk.

If the capacity of short-term memory is five chunks, and if the growth in digit span is due to learning, I should be able to make at least one quantitative prediction about absolute digit span as a function of age. Specifically, a digit should be equivalent to almost exactly one chunk at an age where the child knows the individual digits well, but has not had much arithmetic practice in combining or manipulating them—that is, at about the age the child enters school.

Table 4 gives the revised Stanford-Binet norms for digit span (*16*, p. 704). It shows that, in fact, the norm for digit span is five at age seven—the age of first or second grade children—while it is only four at age four and a half—the age of prekindergarten children. The digit span for college students is slightly below the value it would have if they handled pairs of digits as chunks. Both of these facts are consistent with the hypothesis that the change in digit span with age is due to the shortening of the encoded strings by the use of learned chunks. Also consistent with this hypothesis are experiments which show that digit span can be increased substantially (for example, from 4.4 to 6.4 among kindergartners, from 10 to 14 among college students) with persistent practice.

I next turn to a task where immediate recall and rote learning are only components of a process. In an endeavor to explain the relative lengths of time required for subjects to do mental multiplications of pairs of numbers, Dansereau (*17*) constructed a simulation model of the process, assigning specific time parameters to each of the subprocesses and capacity parameters to short-term memory.

Since Dansereau's model was more detailed and complete than the one I have used informally throughout this article, he needed more parameters than the two I have discussed. He needed to specify the capacity of the short-term visual memory and the short-term auditory memory; he also needed to specify the times required to transfer symbols from the external stimulus to internal memories, and from each internal memory (visual, auditory, long-term) to each of the others. An important constraint that Dansereau imposed on his model was that these parameters not be selected simply to fit his data on mental multiplication speeds, but that they be consistent with estimates of the same parameters derived from simpler component tasks. By drawing on data from others' experiments as well as experiments on component tasks that he himself carried out, Dansereau greatly reduced the degrees of freedom available for fitting his mental multiplication data to his processing model.

For example, he specified 2 seconds per digit as the time required to transfer symbols from short-term to long-term memory. He based this specification on times of 5 seconds per chunk, derived from the experiments of Bugelski and others (*5–7*), together with the assumption that chunks averaged three digits each. These specifications are perhaps biased on the low side, and I might want to quarrel with the details, but the important points are (i) that it is a definite enough theoretical structure to be quarreled about meaningfully and (ii) that the outcome of the quarrel could hardly change the estimate by as much as a factor of 2.

I cannot summarize Dansereau's results here. Rather, I cite his study as another example of the strategy of using parameters estimated from experiments on simple tasks to predict performance on complex tasks. Dansereau undertook to explain performance in mental multiplication on the basis of the same component processes and the same system parameters as those already encountered in laboratory experiments with simpler component tasks.

As a final example, I should like to mention some perceptual phenomena that have been studied a great deal in the past few years. Experiments by de Groot (*18*), Jongman (*19*), and others on the ability of subjects to reproduce the pattern of pieces on a chessboard after an exposure of 5 to 10 seconds, have yielded the following results:

If the pieces represent a position from an actual game (unknown to the subjects), then grandmasters and masters will generally reproduce the position (about 20 to 25 pieces) almost without error, while ordinary players will generally be able to place only a half-dozen pieces correctly. If the same number of pieces is placed on the board in a random pattern, grandmasters and ordinary players alike will be able to place only a half-dozen pieces correctly.

The grandmasters' performance in the first situation could be explained by attributing to them some extraordinary perceptual capability. In the second situation, however, this capability disappears. A more parsimonious explanation would be that the same number of chunks was being retained in memory by both sets of subjects in both situations. To complete this explanation, one would then have to show how a chess position composed of 24 pieces could be recoded into a half-dozen chunks by a master.

This hypothesis has been explored in a series of studies by Barenfeld, Charness, Chase, Gilmartin, and myself, with generally positive results (*20*). Direct evidence for the chunking hypothesis was obtained, for example, by timing how rapidly pieces were replaced on a chessboard from memory. The longer pauses occurred when two unrelated pieces were placed in sequence, while the shorter pauses occurred when closely related pieces were placed in sequence. Interpreting the longer pauses as chunk boundaries, it was found that more than half of the variance between the numbers of pieces remembered by strong and weak players, respectively, could be attributed to the larger average chunk size of the former. The explanation for the remaining variance is still being sought.

Using simple probability models, as well as a computer simultation of the chess perception processes, quantitative estimates were made of the "vocabulary" of familiar chunks in a master's memory. The estimates obtained by several different procedures all fall in the range of 25,000 to 100,000 chunks—that is, a vocabulary of roughly the same size as the vocabulary of an educated adult in his native language. Here, again, the combination of approximate measurements of a few basic parameters and a detailed process theory permits one to make far-reaching predictions and extrapolations.

Conclusion

I have explored some of the interactions between research on higher mental processes over the past decade or two and laboratory experiments on simpler cognitive processes. I have shown that, by viewing experimentation in a parameter-estimating paradigm instead of a hypothesis-testing paradigm, one can obtain much more information from experiments—information that, combined with contemporary theoretical models of the cognitive processes, has implications for human performance on tasks quite different from those of the original experiments.

The work of identifying and measuring the basic parameters of the human information processing system has just begun, but already important information has been gained. The psychological reality of the chunk has been fairly well demonstrated, and the chunk capacity of short-term memory has been shown to be in the range of five to seven. Fixation of information in long-term memory has been shown to take about 5 or 10 seconds per chunk.

Some other "magical numbers" have been estimated—for example, visual scanning speeds and times required for simple grammatical transformations—and no doubt others remain to be discovered. But even the two basic constants discussed in this article—short-term memory capacity and rate of fixation in long-term memory—organize, systematize, and explain a wide range of findings, about both simple tasks and more complex cognitive performances that have been reported in the psychological literature over the past 50 years or more.

References and Notes

1. G. A. Miller, *Psychol. Rev.* **63**, 81 (1956).
2. The magic number proposed here is five, rather than Miller's seven, because the estimates that I derive from data in the literature are closer to the former number than to the latter. But most of my analysis will depend only on the hypothesis that the capacity of short-term memory is a *constant* number of chunks, the same for all kinds of stimuli; the exact numerical value of the constant will not be crucial.
3. J. A. McGeogh, *The Psychology of Human Learning* (Longmans Green, New York, 1942).

4. E. A. Feigenbaum, thesis, Carnegie Institute of Technology (1959). The hypothesis is used in our joint paper, "A theory of the serial position effect" [*Brit. J. Psychol.* **53**, 307 (1962)], which was first circulated as a working paper in 1958.
5. H. C. Wilcoxon, W. R. Wilson, D. A. Wise, *J. Exp. Psychol.* **61**, 283 (1961).
6. B. R. Bugelski, *ibid.* **63**, 409 (1962).
7. E. H. Cooper and A. J. Pantle, *Psychol. Bull.* **68**, 221 (1967).
8. Thus, to learn the pair CEF-DAX, the letters D, A, and X must be compounded into the response; then C and F, say, must be compounded into the recognition cue (C-F) for the stimulus; and finally, stimulus cue and response must each be incorporated in the pair structure: (C-F)–(DAX).
9. Similarly, if DAX (*8*) is to be learned as an item in a list, the D, A, and X must be compounded into a familar syllable (DAX) and this syllable incorporated in the list.
10. H. A. Simon and E. A. Feigenbaum, *J. Verb. Learn. Verb. Behav.* **3**, 385 (1964).
11. L. W. Gregg and H. A. Simon, *ibid.* **6**, 780 (1967).
12. Using the equation introduced previously and denoting by the subscripts i and j two different kinds of stimulus material, we will have, from $S_i = 5s_i$ and $S_j = 5s_j$, $S_i/S_j = s_i/s_j$. Similarly, from $F_i = ks_i$ and $F_j = ks_j$, we will have $F_i/F_j = s_i/s_j$; whence $S_i/S_j = F_i/F_j$.
13. R. Brener, *J. Exp. Psychol.* **26**, 467 (1940).
14. D. O. Lyon, *J. Educ. Psychol.* **5**, 1, 85, 155 (1914).
15. R. S. McLean and L. W. Gregg, *J. Exp. Psychol.* **74**, 455 (1967).
16. R. S. Woodworth and H. Schlosberg, *Experimental Psychology* (Holt, New York, ed. 2, 1954).
17. D. F. Dansereau, thesis, Carnegie-Mellon University (1969).
18. A. D. de Groot, *Thought and Choice in Chess* (Mouton, The Hague, 1965).
19. R. W. Jongman, *Het Oog van de Meester* (van Gorcum, Amsterdam, 1968).
20. For general summaries, see W. G. Chase and H. A. Simon, in *Visual Information Processing* (Academic Press, New York, 1973), pp. 215–281; H. A. Simon and W. G. Chase, *Am. Sci.* **61**, 394 (1973).
21. This work was supported by PHS research grant MH-07722 from the National Institutes of Health. The thesis set forth here has benefited from many discussions with my colleagues William G. Chase, Lee W. Gregg, and Allen Newell. An earlier version of this article was given as an invited address at the meeting of the Eastern Psychological Association, 4 April 1970.