

The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information

by George A. Miller

originally published in *The Psychological Review*, 1956, vol. 63, pp. 81-97
(reproduced here, with the author's permission, by Stephen Malinowski)

Table of Contents

- **Information measurement**
 - **Absolute judgments of unidimensional stimuli**
 - **Absolute judgments of multidimensional stimuli**
 - **Subitizing**
 - **The span of immediate memory**
 - **Recoding**
 - **Summary**
 - **References**
-

My problem is that I have been persecuted by an integer. For seven years this number has followed me around, has intruded in my most private data, and has assaulted me from the pages of our most public journals. This number assumes a variety of disguises, being sometimes a little larger and sometimes a little smaller than usual, but never changing so much as to be unrecognizable. The persistence with which this number plagues me is far more than a random accident. There is, to quote a famous senator, a design behind it, some pattern governing its appearances. Either there really is something unusual about the number or else I am suffering from delusions of persecution.

I shall begin my case history by telling you about some experiments that tested how accurately people can assign numbers to the magnitudes of various aspects of a stimulus. In the traditional language of psychology these would be called experiments in absolute judgment. Historical accident, however, has decreed that they should have another name. We now call them experiments on the capacity of people to transmit information. Since these experiments would not have been done without the appearance of information theory on the psychological scene, and since the results are analyzed in terms of the concepts of information theory, I shall have to preface my discussion with a few remarks about this theory.

Information measurement

The "amount of information" is exactly the same concept that we have talked about for years under the name of "variance." The equations are different, but if we hold tight to the idea that anything that increases the variance also increases the amount of information we cannot go far astray.

The advantages of this new way of talking about variance are simple enough. Variance is always stated in terms of the unit of measurement -- inches, pounds, volts, etc. -- whereas the amount of information is a dimensionless quantity. Since the information in a discrete statistical distribution does not depend upon the unit of measurement, we can extend the concept to situations where we have no metric and we would not ordinarily think of using the variance. And it also enables us to compare results obtained in quite different experimental situations where it would be meaningless to compare variances based on different metrics. So there are some good reasons for adopting the newer concept.

The similarity of variance and amount of information might be explained this way: When we have a large variance, we are very ignorant about what is going to happen. If we are very ignorant, then when we make the observation it gives us a lot of information. On the other hand, if the variance is very small, we know in advance how our observation must come out, so we get little information from making the observation.

If you will now imagine a communication system, you will realize that there is a great deal of variability about what goes into the system and also a great deal of variability about what comes out. The input and the output can therefore be described in terms of their variance (or their information). If it is a good communication system, however, there must be some systematic relation between what goes in and what comes out. That is to say, the output will depend upon the input, or will be correlated with the input. If we measure this correlation, then we can say how much of the output variance is attributable to the input and how much is due to random fluctuations or "noise" introduced by the system during transmission. So we see that the measure of transmitted information is simply a measure of input-output correlation.

There are two simple rules to follow. Whenever I refer to "amount of information," you will understand "variance." And whenever I refer to "amount of transmitted information," you will understand "covariance" or "correlation."

The situation can be described graphically by two partially overlapping circles. Then the left circle can be taken to represent the variance of the input, the right circle the variance of the output, and the overlap the covariance of input and output. I shall speak of the left circle as the amount of input information, the right circle as the amount of output information, and the overlap as the amount of transmitted information.

In the experiments on absolute judgment, the observer is considered to be a communication channel. Then the left circle would represent the amount of information in the stimuli, the right circle the amount of information in his responses, and the overlap

the stimulus-response correlation as measured by the amount of transmitted information. The experimental problem is to increase the amount of input information and to measure the amount of transmitted information. If the observer's absolute judgments are quite accurate, then nearly all of the input information will be transmitted and will be recoverable from his responses. If he makes errors, the transmitted information may be considerably less than the input. We expect that, as we increase the amount of input information, the observer will begin to make more and more errors; we can test the limits of accuracy of his absolute judgments. If the human observer is a reasonable kind of communication system, then when we increase the amount of input information the transmitted information will increase at first and will eventually level off at some asymptotic value. This asymptotic value we take to be the *channel capacity* of the observer: it represents the greatest amount of information that he can give us about the stimulus on the basis of an absolute judgment. The channel capacity is the upper limit on the extent to which the observer can match his responses to the stimuli we give him.

Now just a brief word about the *bit* and we can begin to look at some data. One bit of information is the amount of information that we need to make a decision between two equally likely alternatives. If we must decide whether a man is less than six feet tall or more than six feet tall and if we know that the chances are 50-50, then we need one bit of information. Notice that this unit of information does not refer in any way to the unit of length that we use -- feet, inches, centimeters, etc. However you measure the man's height, we still need just one bit of information.

Two bits of information enable us to decide among four equally likely alternatives. Three bits of information enable us to decide among eight equally likely alternatives. Four bits of information decide among 16 alternatives, five among 32, and so on. That is to say, if there are 32 equally likely alternatives, we must make five successive binary decisions, worth one bit each, before we know which alternative is correct. So the general rule is simple: every time the number of alternatives is increased by a factor of two, one bit of information is added.

There are two ways we might increase the amount of input information. We could increase the rate at which we give information to the observer, so that the amount of information per unit time would increase. Or we could ignore the time variable completely and increase the amount of input information by increasing the number of alternative stimuli. In the absolute judgment experiment we are interested in the second alternative. We give the observer as much time as he wants to make his response; we simply increase the number of alternative stimuli among which he must discriminate and look to see where confusions begin to occur. Confusions will appear near the point that we are calling his "channel capacity."

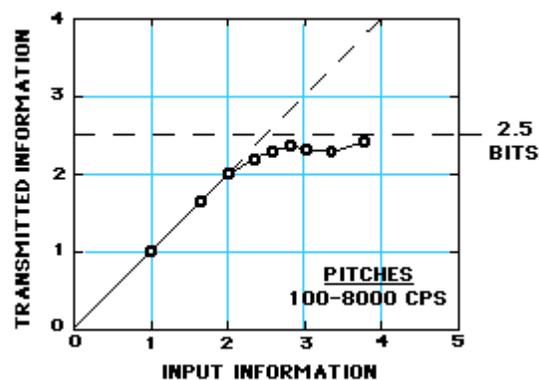
Absolute judgments of unidimensional stimuli

Now let us consider what happens when we make absolute judgments of tones. Pollack [17] asked listeners to identify tones by assigning numerals to them. The tones were different with respect to frequency, and covered the range from 100 to 8000 cps in equal

logarithmic steps. A tone was sounded and the listener responded by giving a numeral. After the listener had made his response, he was told the correct identification of the tone.

When only two or three tones were used, the listeners never confused them. With four different tones confusions were quite rare, but with five or more tones confusions were frequent. With fourteen different tones the listeners made many mistakes.

Figure 1. Data from Pollack [17, 18] on the amount of information that is transmitted by listeners who make absolute judgments of auditory pitch. As the amount of input information is increased by increasing from 2 to 14 the number of different pitches to be judged, the amount of transmitted information approaches as its upper limit a channel capacity of about 2.5 bits per judgment.



These data are plotted in Fig. 1. Along the bottom is the amount of input information in bits per stimulus. As the number of alternative tones was increased from 2 to 14, the input information increased from 1 to 3.8 bits. On the ordinate is plotted the amount of transmitted information. The amount of transmitted information behaves in much the way we would expect a communication channel to behave; the transmitted information increases linearly up to about 2 bits and then bends off toward an asymptote at about 2.5 bits. This value, 2.5 bits, therefore, is what we are calling the channel capacity of the listener for absolute judgments of pitch.

So now we have the number 2.5 bits. What does it mean? First, note that 2.5 bits corresponds to about six equally likely alternatives. The result means that we cannot pick more than six different pitches that the listener will never confuse. Or, stated slightly differently, no matter how many alternative tones we ask him to judge, the best we can expect him to do is to assign them to about six different classes without error. Or, again, if we know that there were N alternative stimuli, then his judgment enables us to narrow down the particular stimulus to one out of $N/6$.

Most people are surprised that the number is as small as six. Of course, there is evidence that a musically sophisticated person with absolute pitch can identify accurately any one

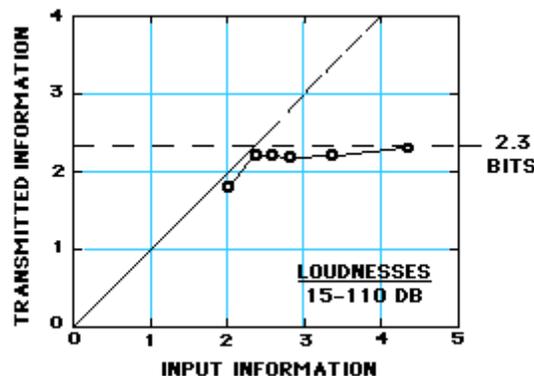
of 50 or 60 different pitches. Fortunately, I do not have time to discuss these remarkable exceptions. I say it is fortunate because I do not know how to explain their superior performance. So I shall stick to the more pedestrian fact that most of us can identify about one out of only five or six pitches before we begin to get confused.

It is interesting to consider that psychologists have been using seven-point rating scales for a long time, on the intuitive basis that trying to rate into finer categories does not really add much to the usefulness of the ratings. Pollack's results indicate that, at least for pitches, this intuition is fairly sound.

Next you can ask how reproducible this result is. Does it depend on the spacing of the tones or the various conditions of judgment? Pollack varied these conditions in a number of ways. The range of frequencies can be changed by a factor of about 20 without changing the amount of information transmitted more than a small percentage. Different groupings of the pitches decreased the transmission, but the loss was small. For example, if you can discriminate five high-pitched tones in one series and five low-pitched tones in another series, it is reasonable to expect that you could combine all ten into a single series and still tell them all apart without error. When you try it, however, it does not work. The channel capacity for pitch seems to be about six and that is the best you can do.

While we are on tones, let us look next at Garner's [7] work on loudness. Garner's data for loudness are summarized in Fig. 2. Garner went to some trouble to get the best possible spacing of his tones over the intensity range from 15 to 110 dB. He used 4, 5, 6, 7, 10, and 20 different stimulus intensities. The results shown in Fig. 2 take into account the differences among subjects and the sequential influence of the immediately preceding judgment. Again we find that there seems to be a limit. The channel capacity for absolute judgments of loudness is 2.3 bits, or about five perfectly discriminable alternatives.

Figure 2. Data from Garner [7] on the channel capacity for absolute judgments of auditory loudness.



Since these two studies were done in different laboratories with slightly different techniques and methods of analysis, we are not in a good position to argue whether five

loudnesses is significantly different from six pitches. Probably the difference is in the right direction, and absolute judgments of pitch are slightly more accurate than absolute judgments of loudness. The important point, however, is that the two answers are of the same order of magnitude.

The experiment has also been done for taste intensities. In Fig. 3 are the results obtained by Beebe-Center, Rogers, and O'Connell [1] for absolute judgments of the concentration of salt solutions. The concentrations ranged from 0.3 to 34.7 gm. NaCl per 100 cc. tap water in equal subjective steps. They used 3, 5, 9, and 17 different concentrations. The channel capacity is 1.9 bits, which is about four distinct concentrations. Thus taste intensities seem a little less distinctive than auditory stimuli, but again the order of magnitude is not far off.

Figure 3. Data from Beebe-Center, Rogers, and O'Connell [1] on the channel capacity for absolute judgments of saltiness.

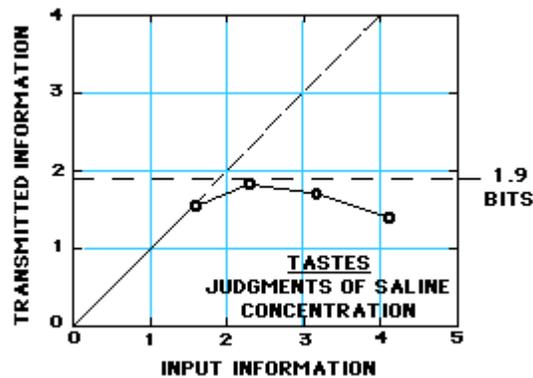
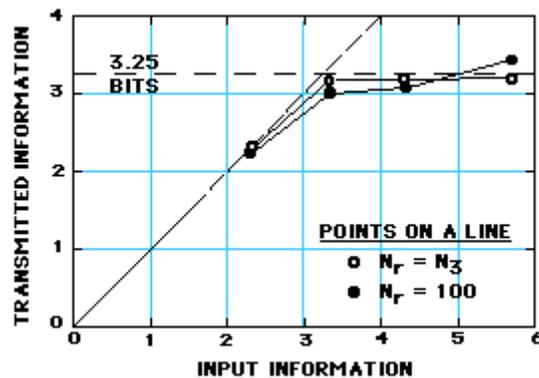


Figure 4. Data from Hake and Garner [8] on the channel capacity for absolute judgments of the position of a pointer in a linear interval.



On the other hand, the channel capacity for judgments of visual position seems to be significantly larger. Hake and Garner [8] asked observers to interpolate visually between two scale markers. Their results are shown in Fig. 4. They did the experiment in two ways. In one version they let the observer use any number between zero and 100 to describe the position, although they presented stimuli at only 5, 10, 20, or 50 different positions. The results with this unlimited response technique are shown by the filled circles on the graph. In the other version the observers were limited in their responses to reporting just those stimulus values that were possible. That is to say, in the second version the number of different responses that the observer could make was exactly the same as the number of different stimuli that the experimenter might present. The results with this limited response technique are shown by the open circles on the graph. The two functions are so similar that it seems fair to conclude that the number of responses available to the observer had nothing to do with the channel capacity of 3.25 bits.

The Hake-Garner experiment has been repeated by Coonan and Klemmer. Although they have not yet published their results, they have given me permission to say that they obtained channel capacities ranging from 3.2 bits for very short exposures of the pointer position to 3.9 bits for longer exposures. These values are slightly higher than Hake and Garner's, so we must conclude that there are between 10 and 15 distinct positions along a linear interval. This is the largest channel capacity that has been measured for any unidimensional variable.

At the present time these four experiments on absolute judgments of simple, unidimensional stimuli are all that have appeared in the psychological journals. However, a great deal of work on other stimulus variables has not yet appeared in the journals. For example, Eriksen and Hake [6] have found that the channel capacity for judging the sizes of squares is 2.2 bits, or about five categories, under a wide range of experimental conditions. In a separate experiment Eriksen [5] found 2.8 bits for size, 3.1 bits for hue, and 2.3 bits for brightness. Geldard has measured the channel capacity for the skin by placing vibrators on the chest region. A good observer can identify about four intensities, about five durations, and about seven locations.

One of the most active groups in this area has been the Air Force Operational Applications Laboratory. Pollack has been kind enough to furnish me with the results of their measurements for several aspects of visual displays. They made measurements for area and for the curvature, length, and direction of lines. In one set of experiments they used a very short exposure of the stimulus -- 1/40 second -- and then they repeated the measurements with a 5-second exposure. For area they got 2.6 bits with the short exposure and 2.7 bits with the long exposure. For the length of a line they got about 2.6 bits with the short exposure and about 3.0 bits with the long exposure. Direction, or angle of inclination, gave 2.8 bits for the short exposure and 3.3 bits for the long exposure. Curvature was apparently harder to judge. When the length of the arc was constant, the result at the short exposure duration was 2.2 bits, but when the length of the chord was constant, the result was only 1.6 bits. This last value is the lowest that anyone has measured to date. I should add, however, that these values are apt to be slightly too low

because the data from all subjects were pooled before the transmitted information was computed.

Now let us see where we are. First, the channel capacity does seem to be a valid notion for describing human observers. Second, the channel capacities measured for these unidimensional variables range from 1.6 bits for curvature to 3.9 bits for positions in an interval. Although there is no question that the differences among the variables are real and meaningful, the more impressive fact to me is their considerable similarity. If I take the best estimates I can get of the channel capacities for all the stimulus variables I have mentioned, the mean is 2.6 bits and the standard deviation is only 0.6 bit. In terms of distinguishable alternatives, this mean corresponds to about 6.5 categories, one standard deviation includes from 4 to 10 categories, and the total range is from 3 to 15 categories. Considering the wide variety of different variables that have been studied, I find this to be a remarkably narrow range.

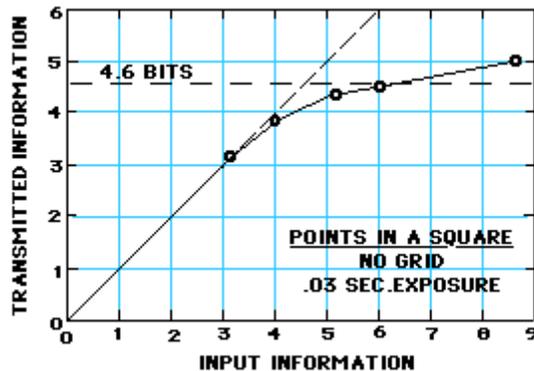
There seems to be some limitation built into us either by learning or by the design of our nervous systems, a limit that keeps our channel capacities in this general range. On the basis of the present evidence it seems safe to say that we possess a finite and rather small capacity for making such unidimensional judgments and that this capacity does not vary a great deal from one simple sensory attribute to another.

Absolute judgments of multidimensional stimuli

You may have noticed that I have been careful to say that this magical number seven applies to one-dimensional judgments. Everyday experience teaches us that we can identify accurately any one of several hundred faces, any one of several thousand words, any one of several thousand objects, etc. The story certainly would not be complete if we stopped at this point. We must have some understanding of why the one-dimensional variables we judge in the laboratory give results so far out of line with what we do constantly in our behavior outside the laboratory. A possible explanation lies in the number of independently variable attributes of the stimuli that are being judged. Objects, faces, words, and the like differ from one another in many ways, whereas the simple stimuli we have considered thus far differ from one another in only one respect.

Fortunately, there are a few data on what happens when we make absolute judgments of stimuli that differ from one another in several ways. Let us look first at the results Klemmer and Frick [13] have reported for the absolute judgment of the position of a dot in a square. In Fig. 5 we see their results. Now the channel capacity seems to have increased to 4.6 bits, which means that people can identify accurately any one of 24 positions in the square.

Figure 5. Data from Klemmer and Frick [13] on the channel capacity for absolute judgments of the position of a dot in a square.



The position of a dot in a square is clearly a two-dimensional proposition. Both its horizontal and its vertical position must be identified. Thus it seems natural to compare the 4.6-bit capacity for a square with the 3.25-bit capacity for the position of a point in an interval. The point in the square requires two judgments of the interval type. If we have a capacity of 3.25 bits for estimating intervals and we do this twice, we should get 6.5 bits as our capacity for locating points in a square. Adding the second independent dimension gives us an increase from 3.25 to 4.6, but it falls short of the perfect addition that would give 6.5 bits.

Another example is provided by Beebe-Center, Rogers, and O'Connell [1]. When they asked people to identify both the saltiness and the sweetness of solutions containing various concentrations of salt and sucrose, they found that the channel capacity was 2.3 bits. Since the capacity for salt alone was 1.9, we might expect about 3.8 bits if the two aspects of the compound stimuli were judged independently. As with spatial locations, the second dimension adds a little to the capacity but not as much as it conceivably might.

A third example is provided by Pollack [18], who asked listeners to judge both the loudness and the pitch of pure tones. Since pitch gives 2.5 bits and loudness gives 2.3 bits, we might hope to get as much as 4.8 bits for pitch and loudness together. Pollack obtained 3.1 bits, which again indicates that the second dimension augments the channel capacity but not so much as it might.

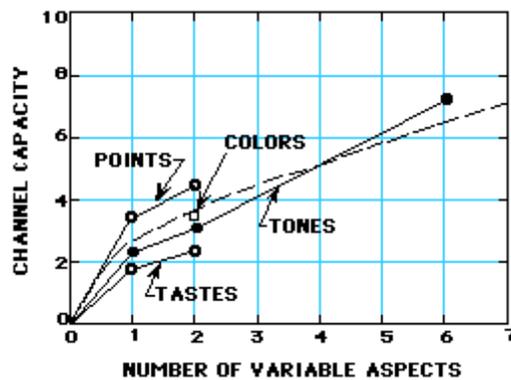
A fourth example can be drawn from the work of Halsey and Chapanis [9] on confusions among colors of equal luminance. Although they did not analyze their results in informational terms, they estimate that there are about 11 to 15 identifiable colors, or, in our terms, about 3.6 bits. Since these colors varied in both hue and saturation, it is probably correct to regard this as a two-dimensional judgment. If we compare this with Eriksen's 3.1 bits for hue (which is a questionable comparison to draw), we again have something less than perfect addition when a second dimension is added.

It is still a long way, however, from these two-dimensional examples to the multidimensional stimuli provided by faces, words, etc. To fill this gap we have only one

experiment, an auditory study done by Pollack and Ficks [19]. They managed to get six different acoustic variables that they could change: frequency, intensity, rate of interruption, on-time fraction, total duration, and spatial location. Each one of these six variables could assume any one of five different values, so altogether there were 5^6 , or 15,625 different tones that they could present. The listeners made a separate rating for each one of these six dimensions. Under these conditions the transmitted information was 7.2 bits, which corresponds to about 150 different categories that could be absolutely identified without error. Now we are beginning to get up into the range that ordinary experience would lead us to expect.

Suppose that we plot these data, fragmentary as they are, and make a guess about how the channel capacity changes with the dimensionality of the stimuli. The result is given in Fig. 6. In a moment of considerable daring I sketched the dotted line to indicate roughly the trend that the data seemed to be taking.

Figure 6. The general form of the relation between channel capacity and the number of independently variable attributes of the stimuli.



Clearly, the addition of independently variable attributes to the stimulus increases the channel capacity, but at a decreasing rate. It is interesting to note that the channel capacity is increased even when the several variables are not independent. Eriksen [5] reports that, when size, brightness, and hue all vary together in perfect correlation, the transmitted information is 4.1 bits as compared with an average of about 2.7 bits when these attributes are varied one at a time. By confounding three attributes, Eriksen increased the dimensionality of the input without increasing the amount of input information; the result was an increase in channel capacity of about the amount that the dotted function in Fig. 6 would lead us to expect.

The point seems to be that, as we add more variables to the display, we increase the total capacity, but we decrease the accuracy for any particular variable. In other words, we can make relatively crude judgments of several things simultaneously.

We might argue that in the course of evolution those organisms were most successful that were responsive to the widest range of stimulus energies in their environment. In order to survive in a constantly fluctuating world, it was better to have a little information about a lot of things than to have a lot of information about a small segment of the environment. If a compromise was necessary, the one we seem to have made is clearly the more adaptive.

Pollack and Ficks's results are very strongly suggestive of an argument that linguists and phoneticians have been making for some time [19]. According to the linguistic analysis of the sounds of human speech, there are about eight or ten dimensions -- the linguists call them *distinctive features* -- that distinguish one phoneme from another. These distinctive features are usually binary, or at most ternary, in nature. For example, a binary distinction is made between vowels and consonants, a binary decision is made between oral and nasal consonants, a ternary decision is made among front, middle, and back phonemes, etc. This approach gives us quite a different picture of speech perception than we might otherwise obtain from our studies of the speech spectrum and of the ear's ability to discriminate relative differences among pure tones. I am personally much interested in this new approach [15], and I regret that there is not time to discuss it here.

It was probably with this linguistic theory in mind that Pollack and Ficks conducted a test on a set of tonal stimuli that varied in eight dimensions, but required only a binary decision on each dimension. With these tones they measured the transmitted information at 6.9 bits, or about 120 recognizable kinds of sounds. It is an intriguing question, as yet unexplored, whether one can go on adding dimensions indefinitely in this way.

In human speech there is clearly a limit to the number of dimensions that we use. In this instance, however, it is not known whether the limit is imposed by the nature of the perceptual machinery that must recognize the sounds or by the nature of the speech machinery that must produce them. Somebody will have to do the experiment to find out. There is a limit, however, at about eight or nine distinctive features in every language that has been studied, and so when we talk we must resort to still another trick for increasing our channel capacity. Language uses sequences of phonemes, so we make several judgments successively when we listen to words and sentences. That is to say, we use both simultaneous and successive discriminations in order to expand the rather rigid limits imposed by the inaccuracy of our absolute judgments of simple magnitudes.

These multidimensional judgments are strongly reminiscent of the abstraction experiment of Külpe [14]. As you may remember, Külpe showed that observers report more accurately on an attribute for which they are set than on attributes for which they are not set. For example, Chapman [4] used three different attributes and compared the results obtained when the observers were instructed before the tachistoscopic presentation with the results obtained when they were not told until after the presentation which one of the three attributes was to be reported. When the instruction was given in advance, the judgments were more accurate. When the instruction was given afterwards, the subjects presumably had to judge all three attributes in order to report on any one of them and the accuracy was correspondingly lower. This is in complete accord with the results we have

just been considering, where the accuracy of judgment on each attribute decreased as more dimensions were added. The point is probably obvious, but I shall make it anyhow, that the abstraction experiments did not demonstrate that people can judge only one attribute at a time. They merely showed what seems quite reasonable, that people are less accurate if they must judge more than one attribute simultaneously.

Subitizing

I cannot leave this general area without mentioning, however briefly, the experiments conducted at Mount Holyoke College on discrimination of number [12]. In experiments by Kaufman, Lord, Reese, and Volkmann random patterns of dots were flashed on a screen for 1/5 of a second. Anywhere from 1 to more than 200 dots could appear in the pattern. The subject's task was to report how many dots there were.

The first point to note is that on patterns containing up to five or six dots the subjects simply did not make errors. The performance on these small numbers of dots was so different from the performance with more dots that it was given a special name. Below seven the subjects were said to *subitize*; above seven they were said to *estimate*. This is, as you will recognize, what we once optimistically called "the span of attention."

This discontinuity at seven is, of course, suggestive. Is this the same basic process that limits our unidimensional judgments to about seven categories? The generalization is tempting, but not sound in my opinion. The data on number estimates have not been analyzed in informational terms; but on the basis of the published data I would guess that the subjects transmitted something more than four bits of information about the number of dots. Using the same arguments as before, we would conclude that there are about 20 or 30 distinguishable categories of numerosness. This is considerably more information than we would expect to get from a unidimensional display. It is, as a matter of fact, very much like a two-dimensional display. Although the dimensionality of the random dot patterns is not entirely clear, these results are in the same range as Klemmer and Frick's for their two-dimensional display of dots in a square. Perhaps the two dimensions of numerosness are area and density. When the subject can subitize, area and density may not be the significant variables, but when the subject must estimate perhaps they are significant. In any event, the comparison is not so simple as it might seem at first thought.

This is one of the ways in which the magical number seven has persecuted me. Here we have two closely related kinds of experiments, both of which point to the significance of the number seven as a limit on our capacities. And yet when we examine the matter more closely, there seems to be a reasonable suspicion that it is nothing more than a coincidence.

The span of immediate memory

Let me summarize the situation in this way. There is a clear and definite limit to the accuracy with which we can identify absolutely the magnitude of a unidimensional stimulus variable. I would propose to call this limit the *span of absolute judgment*, and I

maintain that for unidimensional judgments this span is usually somewhere in the neighborhood of seven. We are not completely at the mercy of this limited span, however, because we have a variety of techniques for getting around it and increasing the accuracy of our judgments. The three most important of these devices are (a) to make relative rather than absolute judgments; or, if that is not possible, (b) to increase the number of dimensions along which the stimuli can differ; or (c) to arrange the task in such a way that we make a sequence of several absolute judgments in a row.

The study of relative judgments is one of the oldest topics in experimental psychology, and I will not pause to review it now. The second device, increasing the dimensionality, we have just considered. It seems that by adding more dimensions and requiring crude, binary, yes-no judgments on each attribute we can extend the span of absolute judgment from seven to at least 150. Judging from our everyday behavior, the limit is probably in the thousands, if indeed there is a limit. In my opinion, we cannot go on compounding dimensions indefinitely. I suspect that there is also a *span of perceptual dimensionality* and that this span is somewhere in the neighborhood of ten, but I must add at once that there is no objective evidence to support this suspicion. This is a question sadly needing experimental exploration.

Concerning the third device, the use of successive judgments, I have quite a bit to say because this device introduces memory as the handmaiden of discrimination. And, since mnemonic processes are at least as complex as are perceptual processes, we can anticipate that their interactions will not be easily disentangled.

Suppose that we start by simply extending slightly the experimental procedure that we have been using. Up to this point we have presented a single stimulus and asked the observer to name it immediately thereafter. We can extend this procedure by requiring the observer to withhold his response until we have given him several stimuli in succession. At the end of the sequence of stimuli he then makes his response. We still have the same sort of input-output situation that is required for the measurement of transmitted information. But now we have passed from an experiment on absolute judgment to what is traditionally called an experiment on immediate memory.

Before we look at any data on this topic, I feel I must give you a word of warning to help you avoid some obvious associations that can be confusing. Everybody knows that there is a finite span of immediate memory and that for a lot of different kinds of test materials this span is about seven items in length. I have just shown you that there is a span of absolute judgment that can distinguish about seven categories and that there is a span of attention that will encompass about six objects at a glance. What is more natural than to think that all three of these spans are different aspects of a single underlying process? And that is a fundamental mistake, as I shall be at some pains to demonstrate. This mistake is one of the malicious persecutions that the magical number seven has subjected me to.

My mistake went something like this. We have seen that the invariant feature in the span of absolute judgment is the amount of information that the observer can transmit. There is

a real operational similarity between the absolute judgment experiment and the immediate memory experiment. If immediate memory is like absolute judgment, then it should follow that the invariant feature in the span of immediate memory is also the amount of information that an observer can retain. If the amount of information in the span of immediate memory is a constant, then the span should be short when the individual items contain a lot of information and the span should be long when the items contain little information. For example, decimal digits are worth 3.3 bits apiece. We can recall about seven of them, for a total of 23 bits of information. Isolated English words are worth about 10 bits apiece. If the total amount of information is to remain constant at 23 bits, then we should be able to remember only two or three words chosen at random. In this way I generated a theory about how the span of immediate memory should vary as a function of the amount of information per item in the test materials.

The measurements of memory span in the literature are suggestive on this question, but not definitive. And so it was necessary to do the experiment to see. Hayes [10] tried it out with five different kinds of test materials: binary digits, decimal digits, letters of the alphabet, letters plus decimal digits, and with 1,000 monosyllabic words. The lists were read aloud at the rate of one item per second and the subjects had as much time as they needed to give their responses. A procedure described by Woodworth [20] was used to score the responses.

The results are shown by the filled circles in Fig. 7. Here the dotted line indicates what the span should have been if the amount of information in the span were constant. The solid curves represent the data. Hayes repeated the experiment using test vocabularies of different sizes but all containing only English monosyllables (open circles in Fig. 7). This more homogeneous test material did not change the picture significantly. With binary items the span is about nine and, although it drops to about five with monosyllabic English words, the difference is far less than the hypothesis of constant information would require.

Figure 7. Data from Hayes [10] on the span of immediate memory plotted as a function of the amount of information per item in the test materials.

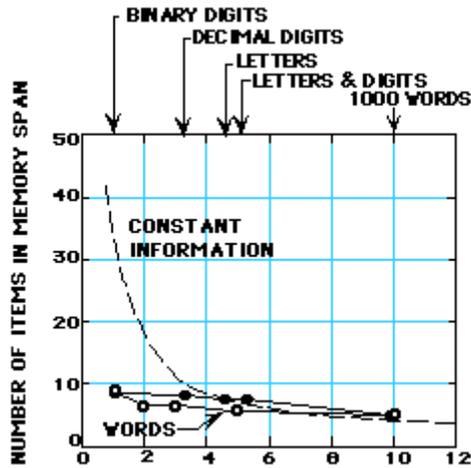
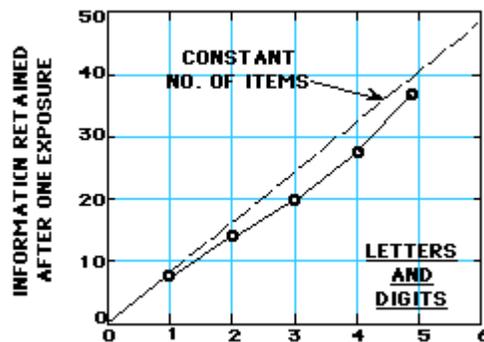


Figure 8. Data from Pollack [16] on the amount of information retained after one presentation plotted as a function of the amount of information per item in the test materials.



There is nothing wrong with Hayes's experiment, because Pollack [16] repeated it much more elaborately and got essentially the same result. Pollack took pains to measure the amount of information transmitted and did not rely on the traditional procedure for scoring the responses. His results are plotted in Fig. 8. Here it is clear that the amount of information transmitted is not a constant, but increases almost linearly as the amount of information per item in the input is increased.

And so the outcome is perfectly clear. In spite of the coincidence that the magical number seven appears in both places, the span of absolute judgment and the span of immediate memory are quite different kinds of limitations that are imposed on our ability to process information. Absolute judgment is limited by the amount of information. Immediate memory is limited by the number of items. In order to capture this distinction in somewhat picturesque terms, I have fallen into the custom of distinguishing between *bits*

of information and *chunks* of information. Then I can say that the number of bits of information is constant for absolute judgment and the number of chunks of information is constant for immediate memory. The span of immediate memory seems to be almost independent of the number of bits per chunk, at least over the range that has been examined to date.

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 when each word was drawn at random from a set of 1,000 English monosyllables 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.

Recoding

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.

A man just beginning to learn radio-telegraphic code hears each *dit* and *dah* as a separate chunk. Soon he is able to organize these sounds into letters and then he can deal with the letters as chunks. Then the letters organize themselves as words, which are still larger chunks, and he begins to hear whole phrases. I do not mean that each step is a discrete process, or that plateaus must appear in his learning curve, for surely the levels of organization are achieved at different rates and overlap each other during the learning process. I am simply pointing to the obvious fact that the dits and dahs are organized by learning into patterns and that as these larger chunks emerge the amount of message that the operator can remember increases correspondingly. In the terms I am proposing to use, the operator learns to increase the bits per chunk.

In the jargon of communication theory, this process would be called *recoding*. The input is given in a code that contains many chunks with few bits per chunk. The operator recodes the input into another code that contains fewer chunks with more bits per chunk. There are many ways to do this recoding, but probably the simplest is to group the input events, apply a new name to the group, and then remember the new name rather than the original input events.

Since I am convinced that this process is a very general and important one for psychology, I want to tell you about a demonstration experiment that should make perfectly explicit what I am talking about. This experiment was conducted by Sidney Smith and was reported by him before the Eastern Psychological Association in 1954.

Begin with the observed fact that people can repeat back eight decimal digits, but only nine binary digits. Since there is a large discrepancy in the amount of information recalled in these two cases, we suspect at once that a recoding procedure could be used to increase the span of immediate memory for binary digits. In Table 1 a method for grouping and renaming is illustrated. Along the top is a sequence of 18 binary digits, far more than any subject was able to recall after a single presentation. In the next line these same binary digits are grouped by pairs. Four possible pairs can occur: 00 is renamed 0, 01 is renamed 1, 10 is renamed 2, and 11 is renamed 3. That is to say, we recode from a base-two arithmetic to a base-four arithmetic. In the recoded sequence there are now just nine digits to remember, and this is almost within the span of immediate memory. In the next line the same sequence of binary digits is regrouped into chunks of three. There are eight possible sequences of three, so we give each sequence a new name between 0 and 7. Now we have recoded from a sequence of 18 binary digits into a sequence of 6 octal digits, and this is well within the span of immediate memory. In the last two lines the binary digits are grouped by fours and by fives and are given decimal-digit names from 0 to 15 and from 0 to 31.

Table I. Ways of Recoding Sequences of Binary Digits

Binary Digits (Bits)	1	0	1	0	0	0	1	0	0	1	1	1	0	0	1	1	1	0
2:1	Chunks	10	10	00	10	01	11	00	11	10								
	Recoding	2	2	0	2	1	3	0	3	2								
3:1	Chunks	101		000		100		111		001		110						
	Recoding	5		0		4		7		1		6						
4:1	Chunks	1010		0010		0111		0011		10								
	Recoding	10		2		7		3										
5:1	Chunks	10100		01001		11001		110										
	Recoding	20		9		25												

It is reasonably obvious that this kind of recoding increases the bits per chunk, and packages the binary sequence into a form that can be retained within the span of immediate memory. So Smith assembled 20 subjects and measured their spans for binary and octal digits. The spans were 9 for binaries and 7 for octals. Then he gave each recoding scheme to five of the subjects. They studied the recoding until they said they understood it -- for about 5 or 10 minutes. Then he tested their span for binary digits again while they tried to use the recoding schemes they had studied.

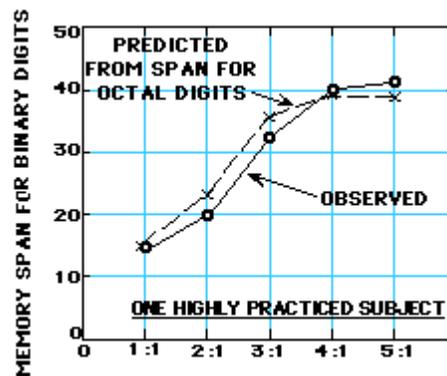
The recoding schemes increased their span for binary digits in every case. But the increase was not as large as we had expected on the basis of their span for octal digits. Since the discrepancy increased as the recoding ratio increased, we reasoned that the few minutes the subjects had spent learning the recoding schemes had not been sufficient. Apparently the translation from one code to the other must be almost automatic or the

subject will lose part of the next group while he is trying to remember the translation of the last group.

Since the 4:1 and 5:1 ratios require considerable study, Smith decided to imitate Ebbinghaus and do the experiment on himself. With Germanic patience he drilled himself on each recoding successively, and obtained the results shown in Fig. 9. Here the data follow along rather nicely with the results you would predict on the basis of his span for octal digits. He could remember 12 octal digits. With the 2:1 recoding, these 12 chunks were worth 24 binary digits. With the 3:1 recoding they were worth 36 binary digits. With the 4:1 and 5:1 recodings, they were worth about 40 binary digits.

It is a little dramatic to watch a person get 40 binary digits in a row and then repeat them back without error. However, if you think of this merely as a mnemonic trick for extending the memory span, you will miss the more important point that is implicit in nearly all such mnemonic devices. The point is that recoding is an extremely powerful weapon for increasing the amount of information that we can deal with. In one form or another we use recoding constantly in our daily behavior.

Figure 9. The span of immediate memory for binary digits is plotted as a function of the recoding procedure used. The predicted function is obtained by multiplying the span for octals by 2, 3, and 3.3 for recoding into base 4, base 8, and base 10, respectively.



In my opinion the most customary kind of recoding that we do all the time is to translate into a verbal code. When there is a story or an argument or an idea that we want to remember, we usually try to rephrase it "in our own words." When we witness some event we want to remember, we make a verbal description of the event and then remember our verbalization. Upon recall we recreate by secondary elaboration the details that seem consistent with the particular verbal recoding we happen to have made. The well-known experiment by Carmichael, Hogan, and Walter [3] on the influence that names have on the recall of visual figures is one demonstration of the process.

The inaccuracy of the testimony of eyewitnesses is well known in legal psychology, but the distortions of testimony are not random -- they follow naturally from the particular recoding that the witness used, and the particular recoding he used depends upon his whole life history. Our language is tremendously useful for repackaging material into a few chunks rich in information. I suspect that imagery is a form of recoding, too, but images seem much harder to get at operationally and to study experimentally than the more symbolic kinds of recoding.

It seems probable that even memorization can be studied in these terms. The process of memorizing may be simply the formation of chunks, or groups of items that go together, until there are few enough chunks so that we can recall all the items. The work by Bousfield and Cohen [2] on the occurrence of clustering in the recall of words is especially interesting in this respect.

Summary

I have come to the end of the data that I wanted to present, so I would like now to make some summarizing remarks.

First, the span of absolute judgment and the span of immediate memory impose severe limitations on the amount of information that we are able to receive, process, and remember. By organizing the stimulus input simultaneously into several dimensions and successively into a sequence or chunks, we manage to break (or at least stretch) this informational bottleneck.

Second, the process of recoding is a very important one in human psychology and deserves much more explicit attention than it has received. In particular, the kind of linguistic recoding that people do seems to me to be the very lifeblood of the thought processes. Recoding procedures are a constant concern to clinicians, social psychologists, linguists, and anthropologists and yet, probably because recoding is less accessible to experimental manipulation than nonsense syllables or T mazes, the traditional experimental psychologist has contributed little or nothing to their analysis. Nevertheless, experimental techniques can be used, methods of recoding can be specified, behavioral indicants can be found. And I anticipate that we will find a very orderly set of relations describing what now seems an uncharted wilderness of individual differences.

Third, the concepts and measures provided by the theory of information provide a quantitative way of getting at some of these questions. The theory provides us with a yardstick for calibrating our stimulus materials and for measuring the performance of our subjects. In the interests of communication I have suppressed the technical details of information measurement and have tried to express the ideas in more familiar terms; I hope this paraphrase will not lead you to think they are not useful in research. Informational concepts have already proved valuable in the study of discrimination and of language; they promise a great deal in the study of learning and memory; and it has even been proposed that they can be useful in the study of concept formation. A lot of

questions that seemed fruitless twenty or thirty years ago may now be worth another look. In fact, I feel that my story here must stop just as it begins to get really interesting.

And finally, what about the magical number seven? What about the seven wonders of the world, the seven seas, the seven deadly sins, the seven daughters of Atlas in the Pleiades, the seven ages of man, the seven levels of hell, the seven primary colors, the seven notes of the musical scale, and the seven days of the week? What about the seven-point rating scale, the seven categories for absolute judgment, the seven objects in the span of attention, and the seven digits in the span of immediate memory? For the present I propose to withhold judgment. Perhaps there is something deep and profound behind all these sevens, something just calling out for us to discover it. But I suspect that it is only a pernicious, Pythagorean coincidence.

References

1. Beebe-Center, J.G., Rogers, M.S., and O'Connell, D.N. Transmission of information about sucrose and saline solutions through the sense of taste. *J. Psychol.*, 1955, **39**, 157-160.
2. Bousfield, W.A., and Cohen, B.H. The occurrence of clustering in the recall of randomly arranged words of different frequencies-of-usage. *J. Gen. Psychol.*, 1955, **52**, 83-95.
3. Carmichael, L., Hogan, H.P., and Walter, A.A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. Exp. Psychol.*, 1932, **15**, 73-86.
4. Chapman, D.W. Relative effects of determinate and indeterminate *Aufgaben*. *Amer. J. Psychol.*, 1932, **44**, 163-174.
5. Eriksen, C.W. Multidimensional stimulus differences and accuracy of discrimination. *USAF, WADC Tech. Rep.*, 1954, No. 54-165.
6. Eriksen, C.W., and Hake, H.W. Absolute judgments as a function of the stimulus range and the number of stimulus and response categories. *J. Exp. Psychol.*, 1955, **49**, 323-332.
7. Garner, W.R. An informational analysis of absolute judgments of loudness. *J. Exp. Psychol.*, 1953, **46**, 373-380.
8. Hake, H.W., and Garner, W.R. The effect of presenting various numbers of discrete steps on scale reading accuracy. *J. Exp. Psychol.*, 1951, **42**, 358-366.
9. Halsey, R.M., and Chapanis, A. Chromaticity-confusion contours in a complex viewing situation. *J. Opt. Soc. Amer.*, 1954, **44**, 442-454.

10. Hayes, J.R.M. Memory span for several vocabularies as a function of vocabulary size. In *Quarterly Progress Report*, Cambridge, Mass.: Acoustics Laboratory, Massachusetts Institute of Technology, Jan.-June, 1952.
11. Jakobson, R., Fant, C.G.M., and Halle, M. *Preliminaries to speech analysis*. Cambridge, Mass.: Acoustics Laboratory, Massachusetts Institute of Technology, 1952. (Tech. Rep. No. 13.).
12. Kaufman, E.L., Lord, M.W., Reese, T.W., and Volkman, J. The discrimination of visual number. *Amer. J. Psychol.*, 1949, **62**, 498-525.
13. Klemmer, E.T., and Frick, F.C. Assimilation of information from dot and matrix patterns. *J. Exp. Psychol.*, 1953, **45**, 15-19.
14. Külpe, O. Versuche über Abstraktion. *Ber. ü. d. I Kongr. f. Exper. Psychol.*, 1904, 56-68.
15. Miller, G.A., and Nicely, P.E. An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Amer.*, 1955, **27**, 338-352.
16. Pollack, I. The assimilation of sequentially encoded information. *Amer. J. Psychol.*, 1953, **66**, 421-435.
17. Pollack, I. The information of elementary auditory displays. *J. Acoust. Soc. Amer.*, 1952, **24**, 745-749.
18. Pollack, I. The information of elementary auditory displays. II. *J. Acoust. Soc. Amer.*, 1953, **25**, 765-769.
19. Pollack, I., and Ficks, L. Information of elementary multidimensional auditory displays. *J. Acoust. Soc. Amer.*, 1954, **26**, 155-158.
20. Woodworth, R.S. *Experimental psychology*. New York: Holt, 1938.