Why the magic number seven plus or minus two

Article *in* Mathematical and Computer Modelling · August 2003

DOI: 10.1016/S0895-7177(03)90083-5

Mujgan Sagir Ozdemir
Eskisehir Osmangazi University

PERGAMON

Mathematical and Computer Modelling 38 (2003) 233–244

MATHEMATICAL AND COMPUTER MODELLING

www.elsevier.com/locate/mcm

Why the Magic Number Seven Plus or Minus Two

T. L. SAATY
University of Pittsburgh
Pittsburgh, PA 15260, U.S.A.
Saaty@katz.pitt.edu

M. S. OZDEMIR
Osmangazi University
Bademlik Eskisehir 26030, Turkey
mujgano@ogu.edu.tr

(Received and accepted July 2003)

Abstract—In 1956, Miller [1] conjectured that there is an upper limit on our capacity to process information on simultaneously interacting elements with reliable accuracy and with validity. This limit is seven plus or minus two elements. He noted that the number 7 occurs in many aspects of life, from the seven wonders of the world to the seven seas and seven deadly sins. We *demonstrate* in this paper that in making preference judgments on pairs of elements in a group, as we do in the analytic hierarchy process (AHP), the number of elements in the group should be no more than seven. The reason is founded in the consistency of information derived from relations among the elements. When the number of elements increases past seven, the resulting increase in inconsistency is too small for the mind to single out the element that causes the greatest inconsistency to scrutinize and correct its relation to the other elements, and the result is confusion to the mind from the existing information. The AHP as a theory of measurement has a basic way to obtain a measure of inconsistency for any such set of pairwise judgments. When the number of elements is seven or less the inconsistency measurement is relatively large with respect to the number of elements involved; when the number is more it is relatively small. The most inconsistent judgment is easily determined in the first case and the individual providing the judgments can change it in an effort to improve the overall inconsistency. In the second case, as the inconsistency measurement is relatively small, improving inconsistency requires only small perturbations and the judge would be hard put to determine what that change should be, and how such a small change could be justified for improving the validity of the outcome. The mind is sufficiently sensitive to improve large inconsistencies but not small ones. And the implication of this is that the number of elements in a set should be limited to seven plus or minus two.

Keywords—Paired comparisons, Consistency, Decision making, Analytic hierarchy process, Information processing, Mental limits.

1. INTRODUCTION

In his book "Number, the Language of Science", Dantzig [2] relates the story of a squire determined to shoot a crow that made its nest in the watchtower of his estate. He successively sent one, two, three, four, and finally five men to shoot the bird. In each case, the crow flew away and

0895-7177/03/$ - see front matter Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX
doi: 10.1016/S0895-7177(03)00216-4

watched until all the men left. It lost count when five went in and four left. It returned to its nest only to be shot by the fifth man left behind. The crow's instinct could not cope with the number five or more, do our own instincts place a limit on our ability to tell the difference among groups of different numbers of objects, and why? Dantzig observed that the human mind has a sense for numbers that is primitive and predates true counting; namely the ability to recognize that a small collection of objects has increased or decreased when things are added to it or subtracted from it. This is an intuitive talent that is not the same as counting. He also speculates on whether the concept is born of experience or whether experience merely serves to render explicit what is already latent in the mind.

Quantity and number play a fundamental role in our lives not only because we need to measure and count, but also because our brains and memory are structured in a way that depends on how many responses to stimuli we can deal with at one time, thus limiting the quantity of information we can process efficiently and consistently. Even a casual observer of the ladder on which we belong in the animal kingdom would conclude that there must be limits on our ability to process information. What size limit, and why such a limit and not a greater or smaller one is necessary is what we explore in this paper.

In his famous paper of March 1956, that appeared in psychological review, Miller [1] wrote the following.

> "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 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... what about the magical number seven... seven wonders of the world, seven seas, seven deadly sins, seven daughters of Pleiades, seven ages of man,.... Perhaps there is something deep and profound behind all these sevens, something just calling out for us to discover.... For the present, I propose to withhold judgment."

In a recent seminal paper with 39 commentaries of which the article relates, 15 accept, seven strongly oppose and 17 are neutral, Cowan [3] argues that Miller's number should be reduced from seven down to four. The debate indicates that it is not clear what is exactly being considered as short term memory, or capacity limit, and what role the time needed to remember plays in these considerations. To us, working memory capacity to identify relations in the form of "consistent" comparisons among remembered "chunks" limits the number of chunks we can handle simultaneously. The number 7 is perhaps a measure of short term memory capacity for processing cognition and is a different limit than the number 4, perhaps a measure of attentional capacity. We are grateful to Cowan for sending us a copy of his work and for graciously reading this paper and encouraging us to publish it.

2. OBSERVATIONS ON THE LIMITS TO HUMAN CAPACITY TO PROCESS INFORMATION

General limitations on human performance are very familiar in the literature of psychology [1,4] and are often classed together as cognitive spans. Such limits are widely known as "memory span", "attention span", "apprehension span", "perceptual span", "span of absolute judgment", "central computing space", and "channel capacity". They are limits on the number of sensations, impressions, or distinctions that can be held in mind briefly and grasped at once, or used as a basis for making judgments.

Absolute judgment is the identification of the magnitude of some simple stimulus—for instance, the brightness of a light, the loudness of a tone, or the curvature of a line—in terms of standards in memory about similar stimuli. This is in contrast with relative judgment that is the identification of some relation between two stimuli both present to the observer.

In experiments on absolute judgment, an observer is considered to be a communication channel. There is an amount of information in the stimuli, and there is another amount of information in the observer's responses. The overlap is the stimulus-response correlation as measured by the amount of transmitted information. One increases the amount of information received and then measures the amount transmitted after processing. If the observer's absolute judgments are sufficiently "accurate" or consistent, the response would transmit most of the input information. However, with increased information there is also the likelihood of more errors. The goal then is to test the limits of the observer's absolute judgments when the information is increased. One would also expect that when the amount of input information is increased, the transmitted information would increase at first and then level off at some value. This value is the channel *capacity of the observer*. Channel capacity is the greatest amount of information about the stimulus that can be transmitted as an absolute judgment by the observer. It is an *upper limit* on the observer's ability to respond to the stimuli received.

3. COMPARISONS AND PRIORITIES [5]

When a subject responds to an event involving several sensations, these sensations must be all related in some way in the mind to make it possible to distinguish among them in a *consistent* way that correctly relates each of them to the entire set of sensations. For example, such a relation is often needed to find the most dominant stimulus among several, and then again the next most dominant one, and so on, a process that needs a total ordering of the sensations. The simplest way to order n sensations is to choose one and compare it with another, retaining the more dominant of the pair, and in turn compare that with another, again retaining the dominant one, continuing until the entire collection has been ordered. This process requires making $n-1$ comparisons to find the most dominant member, $n-2$ comparisons to find the next most dominant member, continuing in this manner by making a total of $n(n-1)/2$ comparisons. The result is an ordering of the sensations according to dominance without knowledge of their numerical values.

But there is a better way to compare and order n sensations that also involves making a total of $n(n-1)/2$ comparisons yet obtaining an ordering according to estimated relative magnitudes among them. It relies on relative judgment about the degree or intensity of dominance of one stimulus of a pair over the other with respect to a given property present to the observer. Such comparison is made by first identifying the smaller or lesser stimulus as the unit and then estimating how many times the greater stimulus is a multiple of that unit. Numerous experiments [5] confirm that people can do this and an example is given in Section 4 to illustrate how it is done. When all the comparisons are made, a scale of priorities is derived from them that represents the relative dominance of the stimuli. We learn from this approach that not only must the sensations be homogeneous or close in order for the comparisons to be meaningful (otherwise, we place them in different homogeneous groups of elements of descending order, with a common pivot from a group to an adjacent group to link the measurements), but also that there must be a limit to the number that we can process at one time while at the same time maintaining consistency in our judgments.

Assume that there are n stimuli present to an observer. The goal of that observer is to

(1) provide judgments on the relative intensity of these stimuli;

(2) ensure that the judgments are quantified to an extent that also permits quantitative interpretation of the judgments among all the stimuli.

Clearly, goal (2) will require appropriate technical assistance.

We describe a method of deriving, from the observer's quantified judgments (i.e., from the relative values associated with *pairs* of stimuli), a set of weights to be associated with *individual* stimuli. These weights should reflect the individual's quantified judgments. What this approach achieves is to put the information resulting from (1) and (2) into usable form without deleting information residing in the qualitative judgments.

Let $A_1, A_2, \ldots, A_n$, be the set of stimuli. The quantified judgments on pairs of stimuli A_i, A_j, are represented by an n-by-n matrix $A = (a_{ij}),\ ij = 1, 2, \ldots, n$. The entries a_{ij} are defined by the following entry rules.

RULE 1. If $a_{ij} = a$, then $a_{ji} = 1/a,\ a \neq 0$.

RULE 2. If A_i is judged to be of equal relative intensity to A_j, then $a_{ij} = 1,\ a_{ji} = 1$; in particular, $a_{ii} = 1$ for all i. Thus, the matrix A has the form

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \dfrac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{1}{a_{1n}} & \dfrac{1}{a_{2n}} & \cdots & 1 \end{bmatrix}.$$

Having recorded the quantified judgments on pairs (A_i, A_j) as numerical entries a_{ij} in the matrix A, the problem now is to assign to the n stimuli $A_1, A_2, \ldots, A_n$ a set of numerical weights $w_1, \ldots, w_n$ that would "reflect the recorded judgments". In order to do so, the vaguely formulated problem must first be transformed into a precise mathematical one. This essential, and apparently harmless, step is the most crucial one in any problem that requires the representation of a real-life situation in terms of an abstract mathematical structure. It is particularly crucial in the present problem where the representation involves a number of transitions that are not immediately discernible. It appears, therefore, desirable in the present problem to identify the major steps in the process of representation and to make each step as explicit as possible to enable the potential user to form his own judgment on the *meaning and value* of the method in relation to *his* problem and *his* goal.

The major question is the one concerned with the meaning of the vaguely formulated condition in the statement of our goal: "these weights should reflect the individual's quantified judgments". This presents the need to describe in precise, arithmetic terms, how the weights w_i should relate to the judgments a_{ij}; or, in other words, the problem of specifying the conditions we wish to impose on the weights we seek in relation to the judgments obtained. The desired description is developed in three steps, proceeding from the simplest special case to the general one.

STEP 1. Assume first that the "judgments" are merely the result of precise physical measurements. Say the judge or judges are given a set of stones $A_1, A_2, \ldots, A_n$, and a precision scale. To compare A_1 with A_2, they put A_1 on a scale and read off its weight—say, $w_l = 305$ grams. They weigh A_2 and find $w_2 = 244$ grams. They divide w_l by w_2, and get 1.25. They pronounce their judgment, "A_1 is 1.25 times as heavy as A_2" and record it as $a_{12} = 1.25$. Thus, in this ideal case of exact measurement, the relations between the weights w_i and the judgments a_{ij} are simply given by

$$\frac{w_i}{w_j} = a_{ij} \quad \text{or} \quad w_i = w_j a_{ij} \qquad (i, j = 1, \ldots, n) \tag{1}$$

and we have

$$A = \begin{array}{c} \\ A_1 \\ \vdots \\ A_n \end{array} \begin{array}{c} \begin{array}{ccc} A_1 & \cdots & A_n \end{array} \\ \begin{bmatrix} \dfrac{w_1}{w_1} & \cdots & \dfrac{w_1}{w_n} \\ \vdots & \cdots & \vdots \\ \dfrac{w_n}{w_1} & \cdots & \dfrac{w_n}{w_n} \end{bmatrix} \end{array}.$$

However, it would be unrealistic to require these relations to hold in the general case. Imposing these stringent relations would, in most practical cases, make the problem of finding the w_i (when a_{ij} are given) unsolvable. First, even physical measurements are never exact in a mathematical sense; and hence, allowance must be made for deviations; and second, in human judgments, these deviations are considerably larger.

STEP 2. In order to see how to make allowance for deviations, consider the i^{th} row in the matrix A. The entries in that row are $a_{i1}, a_{i2}, \ldots, a_{ij}, \ldots, a_{in}$. In the ideal (exact) case these values are the same as the ratios $w_i/w_1,\ w_i/w_2, \ldots, w_i/w_j, \ldots, w_i/w_n$. Thus, in the ideal case, if we multiply the first entry in that row by w_1 the second entry by w_2, and so on, we would obtain

$$\frac{w_i}{w_1}w_1 = w_i, \frac{w_i}{w_2}w_2 = w_i, \ldots, \frac{w_i}{w_j}w_j = w_i, \ldots, \frac{w_i}{w_n}w_n = w_i.$$

The result is a row of identical entries $w_i, w_i, \ldots, w_i$ whereas in the general case, we would obtain a row of entries that represent a statistical scattering of values around w_i. It appears, therefore, reasonable to require that w_i should equal the average of these values. Consequently, instead of the ideal case relations (1)

$$w_i = a_{ij}w_j \qquad (i, j = 1, 2, \ldots, n),$$

the more realistic relations for the general case take the form (for each i)

$$w_i = \text{the average of } (a_{i1}w_1, a_{i2}w_2, \ldots, a_{in}w_n).$$

More explicitly, we have

$$w_i = \frac{1}{n}\sum_{j=1}^{n} a_{ij}w_j \quad \text{or} \quad nw_i = \sum_{j=1}^{n} a_{ij}w_j \qquad (i = 1, 2, \ldots, n). \tag{2}$$

The relations in (2) represent a substantial relaxation of the more stringent relations (1) and indicate proportionally between each weight w_i and the weighted sum of its corresponding judgments a_{ij}. The proportionality is turned to equality with the constant n that is the same as the order of the matrix A. There still remains the question: is the relaxation *sufficient* to ensure the existence of solutions; that is, to ensure that the problem of finding unique weights w_i when the a_{ij} are given is a solvable one?

STEP 3. To seek the answer to the above essentially mathematical question, it is necessary to express the relations in (2) in still another, more familiar form. For this purpose, we need to summarize the line of reasoning to this point. In seeking a set of conditions to describe how the weight vector w should relate to the quantified judgments, we first considered the ideal (exact) case in Step 1, which suggested relations (1). Next, realizing that the real case will require allowances for deviations, we provided for such allowances in Step 2, leading to the formulation (2). Now, this is still not realistic enough; that is, that (2) which works for the ideal case is still too stringent to secure the existence of a weight vector w that should satisfy (2). We note that for good estimates, a_{ij} tends to be close to w_i/w_j, and hence, is a small perturbation of this ratio. Now as a_{ij} changes there might be a corresponding solution of (2), (i.e., w_i and w_j can change to accommodate this change in a_{ij} from the ideal case), if the proportionally constant n were to also change. We denote this value of n by c and we have

$$cw_i = \sum_{j=1}^{n} a_{ij}w_j \qquad (i = 1, \ldots, n). \tag{3}$$

This problem involves solving a system of homogeneous linear equations for which we know that there is a unique solution if c has a particularly chosen value. What we have done so far has been to give an intuitive justification of our approach. It is useful to say all we said above in terms of elementary matrix algebra.

We begin by formulating the condition for a solution in the ideal case

$$Aw = \begin{array}{c} \\ A_1 \\ \vdots \\ A_n \end{array} \overset{A_1 \dots A_n}{\begin{bmatrix} \frac{w_1}{w_1} & \dots & \frac{w_1}{w_n} \\ \vdots & \dots & \vdots \\ \frac{w_n}{w_1} & \dots & \frac{w_n}{w_n} \end{bmatrix}} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = n \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = nw,$$

where A has been multiplied on the right by the column form of the vector of weights $w = (w_1, \dots, w_n)$. The result of this multiplication is nw. Thus, to recover the scale from the matrix of ratios, one must solve the problem $Aw = nw$ or $(A - nI)w = 0$ where I is the identity matrix. This is a system of homogeneous linear equations. It has a nonzero solution if and only if the determinant of $A - nI$, a polynomial of degree n in n (it has a highest degree term of the form n^n, and thus, by the fundamental theorem of algebra has n roots or eigenvalues), is equal to zero, yielding an n^{th} degree equation known as the characteristic equation of A. This equation has a solution if n is one of its roots (eigenvalues of A). But A has a very simple structure because every row is a constant multiple of the first row (or any other row). Thus, all n eigenvalues of A, except one, are equal to zero. The sum of the eigenvalues of a matrix is equal to the sum of its diagonal elements (its trace). In this case, the diagonal elements are each equal to one, and thus, their sum is equal to n, from which it follows that n must be an eigenvalue of A and it is the largest or *principal* eigenvalue, and we have a nonzero solution. The solution is known to consist of positive entries and is unique to within a multiplicative (positive) constant, and thus, belongs to a ratio scale.

When $a_{ij}a_{jk} = a_{ik}$, the matrix $A = (a_{ij})$ is said to be consistent and has the ideal form $a_{ij} = w_i/w_j$ and as we have seen, its principal eigenvalue is equal to n. Otherwise, it is simply reciprocal with $a_{ji} = 1/a_{ij}$ and its principal eigenvalue is the perturbed value c of n that we denote by $\lambda_{\max}$, to indicate the largest or principal eigenvalue of A. Near consistency is essential for response to stimuli because when it is used to compare stimuli that are intangible, human judgment is approximate and mostly inconsistent. If with new information one is able to improve inconsistency to near-consistency, that could improve the validity of the priorities derived from the judgments. To derive priorities from an inconsistent matrix $A = (a_{ij})$, it is necessary to obtain the principal right eigenvector w to represent these priorities.

In simplest terms, a priority vector w can be used to weight the columns of its matrix and sum the elements in each row to obtain a new priority vector and repeat the process, thus obtaining an infinite set of priority vectors. The question is which is the real priority vector? Such ambiguity is eliminated if we require that a priority vector satisfies the condition $Aw = cw$, $c > 0$. In other words, ratios of priorities in the new vector coincide with the same ratios in the old vector. It should now be transparently clear why c and w must be the principal eigenvalue and corresponding eigenvector of A and we will not prove it here.

4. NUMERICAL JUDGMENTS [5]

In the judgment matrix A, instead of assigning two numbers w_i and w_j and forming the ratio w_i/w_j, we assign a single number drawn from a fundamental scale of absolute numbers to represent the ratio $(w_i/w_j)/1$. It is a nearest integer approximation to the ratio w_i/w_j. The derived scale will reveal what w_i and w_j are. This is a central fact about the relative measurement approach. It needs a fundamental scale to express numerically the relative dominance relationship. A person may not be schooled in the use of numbers but still have feelings and understanding that enable him or her to make accurate comparisons. Such judgments can be applied successfully to compare stimuli that are not too disparate in magnitude. If they are far apart, they are grouped together through a filtering process into clusters each of which includes homogeneous stimuli. By homogeneous we mean fall within specified bounds. The clusters can be appropriately linked through their elements by using a pivot stimulus from a cluster to an adjacent cluster.

From logarithmic stimulus-response theory [5], we learn that a stimulus compared with itself is always assigned the value 1 so the main diagonal entries of the pairwise comparison matrix are all 1. We also learn that we must use integer values for the comparisons. The numbers 3, 5, 7, and 9 correspond to the verbal judgments "*moderately more dominant*", "*strongly more dominant*", "*very strongly more dominant*", and "*extremely more dominant*" (with 2, 4, 6, and 8 for compromise between the previous values). Reciprocal values are automatically entered in the transpose position. We are permitted to interpolate values between the integers, if desired. If two stimuli are much closer, they are compared with other contrasting stimuli to obtain the small difference between them by favoring one of them over the other slightly in comparing them with the other stimuli.

Here is an example developed by a group of 30 people that shows that the scale works well on homogeneous elements of a real life problem. A matrix of paired comparison judgments is used to estimate relative drink consumption in the United States. To make the comparisons, the types of drinks are listed on the left and at the top, and judgment is made as to how strongly the consumption of a drink on the left dominates that of a drink at the top. For example, when coffee on the left is compared with wine at the top, it is thought that coffee is consumed extremely more and a 9 is entered in the first row and second column position. A 1/9 is automatically entered in the second row and first column position. If the consumption of a drink on the left does not dominate that of a drink at the top, the reciprocal value is entered. For example, in comparing coffee and water in the first row and eighth column position, water is consumed more than coffee slightly and a 1/2 is entered. Correspondingly, a value of 2 is entered in the eighth row and first

Table 1. Which drink is consumed more in the U.S.? An example of estimation using judgments.

Drink Consumption in the U.S.	Coffee	Wine	Tea	Beer	Sodas	Milk	Water
Coffee	1	9	3	1	$\frac{1}{2}$	1	$\frac{1}{2}$
Wine	$\frac{1}{9}$	1	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
Tea	$\frac{1}{3}$	3	1	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{5}$
Beer	1	9	4	1	$\frac{1}{2}$	1	1
Sodas	2	9	5	2	1	2	1
Milk	1	9	4	1	$\frac{1}{2}$	1	$\frac{1}{2}$
Water	2	9	5	1	1	2	1

The derived scale based on the judgments in the matrix is:

Coffee	Wine	Tea	Beer	Sodas	Milk	Water
.142	.019	.046	.164	.252	.148	.228

with a consistency ratio of .01.

The actual consumption (from statistical abstract of the United States, 2001, for the year 1998) is:

.133	.014	.040	.173	.267	.129	.240

column position. At the bottom of the Table 1, we see that the derived values and the actual values obtained from various pages of statistical abstract of the United States are close.

The numbers are small, and thus, deviations tend to be accentuated. In this case the average absolute percent deviation from the actual is 12.6%. The theory itself provides us with a compatibility index. We denote by $x = (x_i)$ and $y = (y_i)$, respectively, the derived and actual scale vectors, and by $C = (c_{ij})$ where c_{ij} is obtained as the Hadamard or element-wise product $c = (x_i/x_j)(y_j/y_i)$ of one matrix of ratios of the two scales and the transpose of the other matrix of ratios. We then sum the elements of C and divide by n^2 to obtain 1.016 or .016 for deviation from the perfect consistency of the two ratios. This number is much less than the bound of 0.1 on inconsistency and incompatibility.

5. WHEN IS A POSITIVE RECIPROCAL MATRIX CONSISTENT?

In light of the foregoing, for the validity of the vector of priorities to describe response, we need greater redundancy and, therefore, also a large number of comparisons. We now show that, for consistency, we need to make a small number of comparisons. So where is the optimum number?

We now relate the psychological idea of the consistency of judgments and its measurement, to a central concept in matrix theory and also to the size of our channel capacity to process information. It is the principal eigenvalue of a matrix of paired comparisons.

Let $A = [a_{ij}]$ be an n-by-n positive reciprocal matrix, so all $a_{ii} = 1$ and $a_{ij} = 1/a_{ji}$ for all $i, j = 1, \dots, n$. Let $w = [w_i]$ be the principal right eigenvector of A, let $D = \mathrm{diag}(w_1, \dots, w_n)$ be the n-by-n diagonal matrix whose main diagonal entries are the entries of w, and set $E/D^{-1}AD = [a_{ij}w_j/w_i] = [\gamma_{ij}]$. Then E is similar to A and is a positive reciprocal matrix since $\gamma_{ji} = a_{ji}w_i/w_j = (a_{ij}w_j/w_i)^{-1} = 1/\gamma_{ij}$. Moreover, all the row sums of E are equal to the principal eigenvalue of A

$$\sum_{j=1}^{n} \varepsilon_{ij} = \frac{\sum_j a_{ij}w_j}{w_i} = \frac{[Aw]_i}{w_i} = \frac{\lambda_{\max} w_i}{w_i} = \lambda_{\max}.$$

The computation

$$n\lambda_{\max} = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \varepsilon_{ij} \right) = \sum_{i=1}^{n} \varepsilon_{ii} + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} (\varepsilon_{ij} + \varepsilon_{ji}) = n + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \left(\varepsilon_{ij} + \varepsilon_{ij}^{-1}\right) \geq n + \frac{n^2 - n}{2} = n^2$$

reveals that $\lambda_{\max} \geq n$. Moreover, since $x + 1/x \geq 2$ for all $x > 0$, with equality if and only if $x = 1$, we see that $\lambda_{\max} = n$ if and only if all $\gamma_{ij} = 1$, which is equivalent to having all $a_{ij} = w_i/w_j$.

The foregoing arguments show that a positive reciprocal matrix A has $\lambda_{\max} \geq n$, with equality if and only if A is consistent. As our measure of deviation of A from consistency, we choose the *consistency index*

$$\mu \equiv \frac{\lambda_{\max} - n}{n - 1}.$$

We have seen that $\mu \geq 0$ and $\mu = 0$ if and only if A is consistent. We can say that as $\mu \to 0$, $a_{ij} \to w_i/w_j$, or $\varepsilon_{ij} = a_{ij}w_j/w_i \to 1$. These two desirable properties explain the term "n" in the numerator of μ; what about the term "$n - 1$" in the denominator? Since trace $(A) = n$ is the sum of all the eigenvalues of A, if we denote the eigenvalues of A that are different from $\lambda_{\max}$ by $\lambda_2, \dots, \lambda_{n-1}$, we see that $n - \lambda_{\max} + \sum_{i=2}^{n} \lambda_i$, so $n - \lambda_{\max} = \sum_{i=2}^{n} \lambda_i$ and $\mu = (1/(n-1)) \sum_{i=1}^{n} \lambda_i$ is the average of the nonprincipal eigenvalues of A.

It is an easy, but instructive, computation to show that $\lambda_{\max} = 2$ for every 2-by-2 positive reciprocal matrix

$$\begin{bmatrix} 1 & \alpha \\ \alpha^{-1} & 1 \end{bmatrix} \begin{bmatrix} 1 + \alpha \\ (1 + \alpha)\alpha^{-1} \end{bmatrix} = 2 \begin{bmatrix} 1 + \alpha \\ (1 + \alpha)\alpha^{-1} \end{bmatrix}.$$

Thus, every 2-by-2 positive reciprocal matrix is consistent.

Not every 3-by-3 positive reciprocal matrix is consistent, but in this case we are fortunate to have again explicit formulas for the principal eigenvalue and eigenvector. For

$$A = \begin{bmatrix} 1 & a & b \\ \frac{1}{a} & 1 & c \\ \frac{1}{b} & \frac{1}{c} & 1 \end{bmatrix},$$

we have $\lambda_{\max} = 1 + d + d^{-1}$, $d = (ac/b)^{1/3}$, and

$$w_1 = \frac{bd}{(1 + bd + c/d)}, \qquad w_2 = \frac{c}{d(1 + bd + c/d)}, \qquad w_3 = \frac{1}{(1 + bd + c/d)}.$$

Note that $\lambda_{\max} = 3$ when $d = 1$ or $c = b/a$, which is true if and only if A is consistent.

In order to get some feel for what the consistency index might be telling us about a positive n-by-n reciprocal matrix A, consider the following simulation: choose the entries of A above the main diagonal at random from the 17 values $\{1/9, 1/8, \ldots, 1, 2, \ldots, 8, 9\}$. Then fill in the entries of A below the diagonal by taking reciprocals. Put ones down the main diagonal and compute the consistency index. Do this 50,000 times and take the average, which we call the *random index*. Table 1 shows the values obtained from one set of such simulations and also their first-order differences, for matrices of size $1, 2, \ldots, 15$.

Figure 1 is a plot of the first two rows of Table 2. It shows the asymptotic nature of random inconsistency.

Since it would be pointless to try to discern any priority ranking from a set of random comparison judgments, we should probably be uncomfortable about proceeding unless the consistency index of a pairwise comparison matrix is very much smaller than the corresponding random index value in Table 2. The *consistency ratio* (C.R.) of a pairwise comparison matrix is the ratio of its consistency index: to the corresponding random index value in Table 2. The notion of order of magnitude is essential in any mathematical consideration of changes in measurement. When one

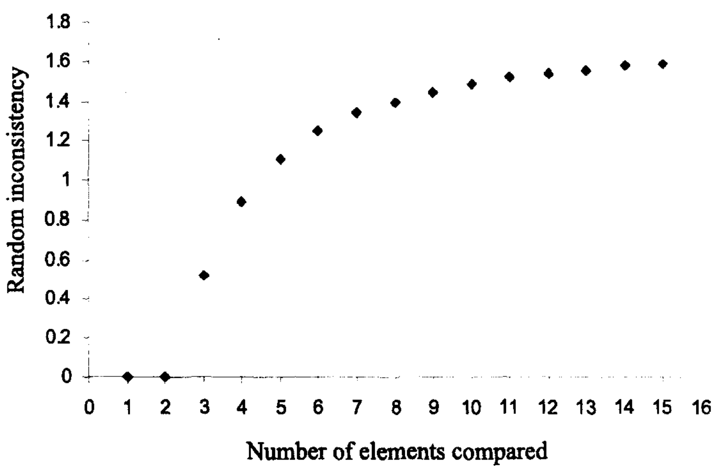

Figure 1. Plot of random inconsistency.

Table 2. Random index.

Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R.I.	0	0	0.52	0.89	1.11	1.25	1.35	1.40	1.45	1.49	1.52	1.54	1.56	1.58	1.59
First-Order Differences		0	0.52	0.37	0.22	0.14	0.10	0.05	0.05	0.04	0.03	0.02	0.02	0.02	0.01

has a numerical value say between 1 and 10 for some measurement and one wishes to determine whether change in this value is significant or not, one reasons as follows: a change of a whole integer value is critical because it changes the magnitude and identity of the original number significantly. If the change or perturbation in value is of the order of a percent or less, it would be so small (by two orders of magnitude) and would be considered negligible. However, if this perturbation is a decimal (one order of magnitude smaller), we are likely to pay attention to modify the original value by this decimal without losing the significance and identity of the original number as we first understood it to be. Thus, in synthesizing near consistent judgment values, changes that are too large can cause dramatic change in our understanding, and values that are too small cause no change in our understanding. We are left with only values of one order of magnitude smaller that we can deal with incrementally to change our understanding. It follows that our allowable consistency ratio should be not more than about .10. The requirement of 10% cannot be made smaller such as 1% or .1% without trivializing the impact of inconsistency. But inconsistency itself is important because without it, new knowledge that changes preference cannot be admitted. Assuming that all knowledge should be consistent contradicts experience that requires continued revision of understanding.

If the C.R. is larger than desired, we do three things.

(1) Find the most inconsistent judgment in the matrix (for example, that judgment for which $\varepsilon_{ij} = a_{ij}w_j/w_i$ is largest).
(2) Determine the range of values to which that judgment can be changed corresponding to which the inconsistency would be improved.
(3) Ask the judge to consider, whether he can change his judgment to a plausible value in that range. If he is unwilling, we try with the second most inconsistent judgment and so on. If no judgment is changed the decision is postponed until better understanding of the stimuli is obtained.

Judges who understand the theory are always willing to revise their judgments often not the full value but partially and then examine the second most inconsistent judgment and so on. It can happen that a judge's knowledge does not permit one to improve his or her consistency and more information is required to improve the consistency of judgments.

Before proceeding further, the following observations may be useful for a better understanding of the importance of the concept of a limit on our ability to process information and also change in information. The quality of response to stimuli is determined by three factors. Accuracy or validity, consistency, and efficiency or amount of information generated. Our judgment is much

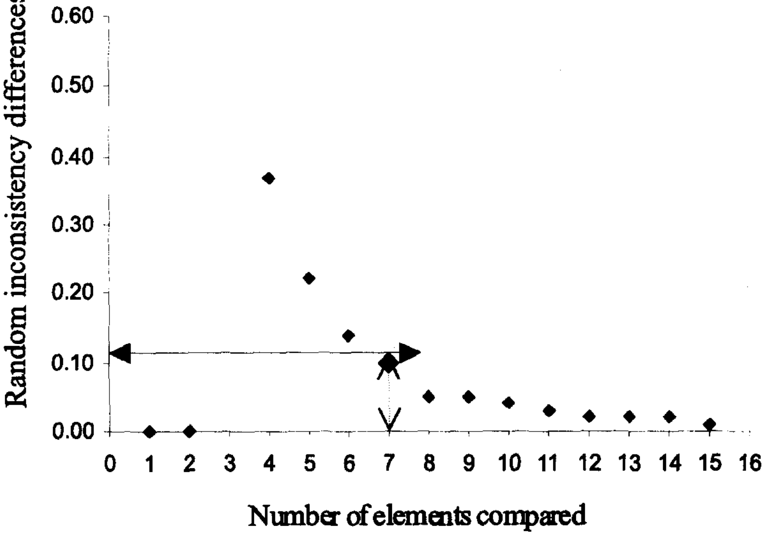

Figure 2. Plot of first differences in random inconsistency.

more sensitive and responsive to large perturbations. When we speak of perturbation, we have in mind numerical change from consistent ratios obtained from priorities. The larger the inconsistency, and hence, also the larger the perturbations in priorities, the greater is our sensitivity to make changes in the numerical values assigned. Conversely, the smaller the inconsistency, the more difficult it is for us to know where the best changes should be made to produce not only better consistency but also better validity of the outcome. Once near consistency is attained, it becomes uncertain which coefficients should be perturbed by small amounts to transform a near consistent matrix to a consistent one. If such perturbations were forced, they could be arbitrary, and thus, distort the validity of the derived priority vector in representing the underlying decision.

The third row of Table 2 gives the differences between successive numbers in the second row. Figure 2 is a plot of these differences and shows the importance of the number 7 as a cutoff point beyond which the differences are less than 0.10 where we are not sufficiently sensitive to make accurate changes in judgment on several elements simultaneously.

6. SECOND AFFIRMATION THROUGH THE EIGENVECTOR

Stability of the principal eigenvector also imposes a limit on channel capacity and also highlights the importance of homogeneity. To a first-order approximation, perturbation Δw_1 in the principal eigenvector w_1 due to a perturbation ΔA in the matrix A where A is consistent and is given by Wilkinson [6]

$$\Delta w_1 = \sum_{j=2}^{n} \left(\frac{v_j^\top \Delta A w_1}{(\lambda_1 - \lambda_j) v_j^\top w_j} \right) w_j.$$

Here $\top$ indicates transposition. The eigenvector w_1 is insensitive to perturbation in A, if

(1) the number of terms n is small;

(2) if the principal eigenvalue λ_1 is separated from the other eigenvalues λ_1, here assumed to be distinct (otherwise, a slightly more complicated argument given below can be made); and

(3) if none of the products $v_j^\top w_j$ of left and right eigenvectors is small but if one of them is small, they are all small.

However, $v_1^\top w_1$, the product of the normalized left and right principal eigenvectors of a consistent matrix is equal to n that as an integer is never very small. If n is relatively small and the elements being compared are homogeneous, none of the components of w_1 is arbitrarily small and correspondingly, none of the components of $v_1^\top$ is arbitrarily small. Their product cannot be arbitrarily small, and thus, w is insensitive to small perturbations of the consistent matrix A. The conclusion is that n *must be small*, and one must compare *homogeneous* elements.

When the eigenvalues have greater multiplicity than one, the corresponding left and right eigenvectors will not be unique. In that case, the cosine of the angle between them which is given by $v_i^\top w_i$ corresponds to a particular choice of w_i and v_i. Even when w_i and v_i correspond to a simple λ_i they are arbitrary to within a multiplicative complex constant of unit modulus, but in that case $|v_i^\top w_i|$ is fully determined. Because both vectors are normalized, we always have $|v_i^\top w_i| < 1$.

7. CONCLUSIONS

The consistency of judgments is necessary for us to cope effectively with experience but it is not sufficient. A mental patient can have a perfectly consistent picture of a nonexistent world. We need the redundancy of informed judgments to improve validity. Paired comparisons make redundancy possible. However, redundancy gives rise to inconsistency. For the sake of efficiency, we need to make a tradeoff between consistency and redundancy that implies validity. We know that when the inconsistency is small, whether contributed by one or several judgments, we are insensitive to making very small changes in judgment to improve consistency. We recall that when we make a judgment, we also automatically make the reciprocal judgment and our measure of inconsistency takes the inconsistency of both into consideration. Large inconsistency may be due to either one judgment that has considerable error in it such as using the reciprocal value instead of the value itself, or to incompatibility among several judgments. Our measure of random inconsistency reveals that as the number of elements being compared is increased the measure of inconsistency decreases so slowly that there is insufficient room for improving the judgments and, therefore, also consistency. From Figure 2, we conclude that to serve both consistency and redundancy, it is best to keep the number of elements seven or less. It appears that Miller's seven plus or minus two is indeed a limit, a channel capacity on our ability to process information.

REFERENCES

1. G.A. Miller, The magical number seven, plus or minus two: Some limits on our capacity for processing information **63** (2), 81–97, (1956).
2. T. Dantzig, *Number, The Language of Science*, The Macmillan Company, New York, (1954).
3. N. Cowan, The magical number 4 in short-term memory: A reconsideration of mental storage capacity, *Behavioral and Brain Sciences* **24**, 87–185, (2001).
4. A.L. Blumenthal, *The Process of Cognition*, Prentice-Hall, Englewood Cliffs, NJ, (1977).
5. T.L. Saaty, *Multicriteria Decision Making, The Analytic Hierarchy Process*, RWS Publications, Pittsburgh, PA, (1996).
6. J.H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, (1965).