CHAPTER 5 A memory test for numbers

Introduction

The present chapter involves a memory test for all numbers 1-100.[1] The purpose of the experiment is to determine each number's chance of being correctly recalled or recognized. We will also determine its chance of being "recalled" or "recognized" by mistake. Numbers which are easier to memorize are presumed to have stronger individual footholds in people's minds. A number will only be adequately remembered if it evokes a mental pattern that is sufficiently distinct to start with and sufficiently stable to be reproduced again. Memorization scores will first be analyzed in terms of the non-mental characteristics of the numbers involved, such as size and tabledness. Subsequently, we shall look at the relationship between memorization scores and other psychological measures.

To the uninformed observer all trees, all chess positions and all babies seem to be very much alike. The categories are recognized as such, but no sufficient knowledge exists to reliably discriminate one instance from the next. Such a confusion of data is amazing to the more expert observer. Who could possibly fail to recognize the distinctness of this unique person, object, or idea? But of course the expert's ability to discriminate was not there from the beginning. By practice and thinking long term mental patterns have been established which are now available for instant cognition in various contexts.

One such context is memorizing. Experts have better memories for information related to their subject. This phenomenon, which was first discovered by De Groot (1946, 1965) in his famous chess study, is now common knowledge in psychology (Van Lehn, 1989). Of course, people may be specifically trained to memorize certain things (e.g., Chase & Ericsson, 1981; Ericsson & Faivre, 1988; Ericsson & Polson, 1988a, 1988b). Some people may be naturally endowed with extraordinary memories for certain materials (Hunter, 1977; Luria, 1968; Sacks, 1985). But as a rule, good memory performance seems to spring from well developed knowledge (Chi, 1978; Ericsson & Smith, 1991).

In the course of cognitive development people learn to categorize and to distinguish. Small children will first use such words as "dog" for all seemingly similar animals (Bowerman, 1984, gives a series of examples). Later on, only dogs will be called dog and some dogs may be instantly perceived as a pitbull, or a poodle. That specialized knowledge influences instant categorization has been demonstrated by Tanaka and Tailor (1991). These authors compared subjects with a special interest in either birds or dogs with ordinary lay subjects. Various category naming tasks were used to compare the detailedness of the initial identifications, which, following Rosch, was assumed to reflect the "basic level" of categorization. They found that the acquisition of expertise produced a shift in the level of semantic differentiation. Presented with any type of bird, the lay person would respond with bird, while the bird-lover instantly responded on a more specific level, saying blue jay, or crow. The conclusion of the authors was that the "basic level" of categorization is a function of the detailedness of people's long term knowledge. Martin (1992) gives a similar explanation for the loss of semantic distinction in patients suffering from Alzheimer's disease.

Detailed knowledge of numbers must be learned. Geary (1995) distinguishes between numerical abilities which are biologically primary, and those which are biologically secondary. Primary abilities have been shaped in the course of evolution, and are part of people's natural endowment. Secondary abilities are culturally induced extensions of the primary ones; they do not develop spontaneously in any society, but must be specifically taught and trained. While some basic understanding of numerosity seems to be part of our biological heritage, its development into specific knowledge of many numbers is a product of formal education.

In the process of acquisition of such knowledge, which is necessary to distinguish them, small numbers may have a natural advantage. Very young children can discriminate between one and two objects, and between two and three (Starkey, Spelke, and Gelman, 1990). The symbols which denote such small quantities can therefore be connected to distinct mental patterns, with strong informal foundations. This is not the case with numbers such as 46 and 47, whose individual meaning is not so immediately obvious. Knowledge about the properties of larger numbers is, of course, acquired by education. Even so, the psychological advantage of smaller over bigger numbers may be a permanent one. Numbers may also have acquired individual distinction or salience for other reasons. An example is 13, which has a special affective significance. Other examples are numbers such as 10, 12, 24, 25, 50, and 100, which are often encountered and easy to work with. Such numbers may also have a clearer than average identity.

In the previous chapters numbers were shown to differ on several psychological measures, such as associative frequency, response commonality, meaningfulness and affect. Differences between numbers on these measures could often be traced to non-mental distinctions such as size, tabledness and being odd or even. Different aspects of long-term knowledge were tapped in these experiments, reflecting different aspects of meaning. The mental patterns that are activated in the context of memorization may include several aspects of long term knowledge.

The experimental section of this chapter consists of two parts. In the first part, the numbers are tested and their scores are analyzed in terms of their non-mental attributes. Because numbers have many different non-mental properties of potential interest, a stepwise regression test is performed to select the best predictors. These are then used to form distinct categories of numbers, whose memorization scores will be compared by Anova. In the second part, memory scores will be related to other psychological measures. Correlations with individual variables such as frequency and meaningfulness will be determined. Nineteen previously determined variables will be factor analyzed, looking for a simpler set of dimensions on which a number may be mentally represented. As a last step, the obtained dimensions are correlated with memorization scores.

PART 1 : MEMORY TEST

Method

Subjects

Subjects were 597 first year psychology students. They were tested collectively, in two separate sessions. Of these subjects, 289 participated in the recall condition and 308 in the recognition condition.

Materials

All numbers 1 through 100 were used as stimuli. Five lists were composed, of twenty numbers each, presented as Arabic numerals. (See Appendix 5). Three rules were followed: small numbers and tens had to be distributed evenly; no more than three numbers from a decade could appear in the same list, and lists should be balanced as to the proportion of odd and even numbers. Each list was then ordered so that obvious clusters were avoided. Of each of the five lists, three versions were composed, to control for differences in the serial position of numbers. The position of the numbers was varied by dividing the lists into parts a, b, and c, consisting of 7, 6, and 7 numbers respectively, and making the following combinations: abc, bca, and cab. Lists were printed in one column on sheets of A4 paper.
As testing was to take place in two conditions, two versions of the response sheets were prepared. For the recall condition, a response-sheet of paper with twenty open lines was printed. For the recognition condition, all hundred numbers were printed on the response-sheet in order of magnitude.

Procedure

The numbers-test was part of a large battery of tests and questionnaires. Five experimenters were present to assist. Subjects sat behind desks, as in an exam. They were advised to follow the printed instruction carefully, which was read out by one of the experimenters. The instructions informed subjects they would be given 90 seconds to study the numbers printed at the back of their sheet, and that they would be tested later. Subjects were not allowed to write anything down. The experimenters kept time with a stop-watch. After 90 seconds, subjects were instructed to put the numbers sheet into a folder together with other completed tests. The interval between studying time and testing time was over an hour in both conditions. During this interval, six questionnaires on various subjects were completed, and a short rest was taken. Subjects were then given 90 seconds to fill out their response sheet. In the recall condition, this meant writing down numbers on twenty blank lines. In the recognition condition, it meant marking the target numbers in the presented hundred numbers. In both conditions, guessing was discouraged by the announcement that incorrect responses would be subtracted from their scores.

Variables

Seven attributes of numbers were selected for participation in a stepwise regression analysis. Some are related to size and some are not.

1. Size. Standing for the numbers' actual value and varying from 1 to 100.
2. Syllables. Standing for the number of syllables in the number-word. (Number of syllables varies between one and six in Dutch).
3. Even/odd. A dichotomous variable, distinguishing between 50 even and 50 odd numbers.
4. Singles. A dichotomous variable, distinguishing between the single digit numbers (n = 9) and all others.
5. Doubles. A dichotomous variable, distinguishing between 11, 22, 33, 44, 55, 66, 77, 88, 99 (n = 9) and all others.
6. Teens. A dichotomous variable, distinguishing between 10-19 (n = 10) and all others.
7. Tabled. A dichotomous variable, distinguishing between the numbers used in the multiplication tables 1-12 (n = 53) and all others. This category overlaps with some of the other categories. In addition to the numbers 1-12, the tabled numbers are 14, 15, 16, 18, 20, 21, 22, 24, 25, 27, 28, 30, 32, 33, 35, 36, 40, 42, 44, 45, 48, 49, 50, 54, 55, 56, 60, 63, 64, 66, 70, 72, 77, 80, 81, 84, 88, 90, 96, 99, 100.

Data treatment

For each number two individual scores were determined. These scores consisted of 1. the proportion of subjects assigned to the relevant list who had correctly recalled the number, and 2. the proportion of subjects assigned to the relevant list who had correctly recognized the number. Additionally, each number was scored for its appearance as an incorrect number in both conditions. For subjects absolute recall and recognition scores were also determined and pooled over list-groups to check for list-effects.

Results

In both conditions six answering sheets were judged to be invalid, in eleven cases because it could not be determined which list had been studied, in the twelfth case because the response-sheet said "I have cheated in this test". A total of 585 answering sheets remained (283 recall and 302 recognition). Distribution of the five different lists of numbers had not been completely equal, but the differences were not significant statistically, X^2 (4, N = 580) = 3, p = .45.

Lists

Subjects assigned to different lists did not perform significantly different in either respect: F(4, 580) = 37.8, p = .83 for correct scores; F (4, 580) = 1.61, p = .17 for intrusions. All lists were equally difficult to remember.

Numbers

Numbers differed greatly in difficulty, as measured by proportions correct recall and recognition. The best recalled numbers (scoring 89 percent correct) were 7 and 8. The best recognized numbers were 1 and 7 (scoring 96 and 94 percent correct). In the recall condition 79 scored worst, followed by 41 (12 and 15 percent correct respectively). For recognition the lowest scoring numbers are 59 and 62 (with 17 and 18 percent correct, respectively). Table 5.1 presents the best and worst remembered numbers in both conditions.

Table 5.1.

Best and worst remembered numbers (named easy and difficult) in two conditions. For each condition, the ten highest and the ten lowest scoring numbers are presented, ordered by the magnitude of their scores.

Easy numbers		***Difficult numbers***	
Recall	Recognition	Recall	Recognition
8	7	82	39
1	2	56	61
100	5	61	87
2	100	94	83
17	8	85	79
5	12	45	41
9	3	83	82
10	99	59	84
99	17	41	62
11	44	79	59

Stepwise regression

A stepwise regression analysis (F to enter = 4) revealed that of the seven variables that were introduced (see methods section), four made a significant contribution to the explanation of numbers' memorization scores. This was the case both for recall-scores and for recognition scores. These four variables are, ordered by the magnitude of their contribution: 1. Singles, 2. Teens, 3. Doubles, 4. Tabled. The multiple correlation of the four variables with retention scores was quite high in both conditions. For recall an R of .81 was obtained, F (4, 95) = 45.98. For recognition the obtained R was .77, F (4, 95) = 35.47. Both analyses were performed in six steps, with identical patterns for recall and recognition. Size was the first variable selected, its correlation with retention scores of numbers being .58 and .57 respectively. Size was removed again as a fourth step, after singles and teens had been selected. After extraction of the four variables mentioned, size was left with non-significant F's to Enter of 1.27 for recall and 2.57 for recognition. The contributions of Evenness and Syllables were not significant at any moment in the procedure.

Categories

Because we were interested in a comparison of category means, the variables found to be significant were used to divide all numbers into five distinct (non- overlapping) categories. A full specification is given in Appendix 2. The category structure is built up by a stepwise removal of overlaps. Thus, the first category is Singles (nrs 1-9). The second is Teens (nrs 10-19). The third is Doubles (starting with 22). The fourth is Large Tabled, being the tabled numbers (see methods section) minus the singles, the teens and the doubles, leaving 29 numbers. The last category is Others, consisting of all numbers that do not belong to any of the categories mentioned.

Table 5.2

Mean percentages of correct reproduction of five categories of numbers, in two experimental .Standard deviations are given within parentheses.

Category	***% Recalled***	***% Recognized***	***% overall***
Singles (n = 9)	82 (7)	84 (11)	83 (9)
Teens (n = 10)	72 (9)	72 (9)	72 (6)
Doubles (n = 8)	59 (9)	66 (14)	62 (11)
Large tabled(n = 29)	42 (14)	49 (14)	45 (13)
Others (n = 44)	34 (11)	38 (12)	36 (11)
Overall (n = 100)	47·% (20)	51·% (20)	49·% (19)

Comparison of categories

As could be expected, the overall difference between the category means is highly significant for both testing conditions, F (4, 95) = 46.95, $p < .0001$ for recall, and F (4, 95) = 38.94, $p < .001$ for recognition.

Comparison of conditions.

Recall and recognition scores correlate .90. A repeated measures Anova shows that performance is somewhat worse when measured by recall than by recognition, F (1, 95) =

48.88, $p < .0001$. Mean percentages correct retrieval are 47 and 51 respectively. No interaction was found between numbers categories and testing condition, $F(4, 95) = 1.41$, $p = .24$.

Category scores

A post hoc comparison revealed that for recall-scores, the comparison between singles and teens is not significant. For recognition-scores, the comparison of teens and doubles is not significant. All other comparisons are significant at an alpha level of .05.

Table 5.2 presents the category means obtained in both conditions. The table shows that single-digit numbers are most easy to remember, being correctly reproduced in 82 percent of all cases in the recall condition, and in 84 percent of the cases in the recognition condition. Next come teens, with a percentage correct reproduction of 72, in both conditions. Doubles are third, with percentages of 59 and 66, followed by the category of large tabled numbers, with percentages of 42 and 49. The numbers that do not belong to any specific category are remembered worst, being correctly reproduced in 34 and 38 percent of the cases only.

Incorrect numbers

Subjects produced 2.01 (recall) and 2.78 (recognition) incorrect numbers on the average. These "incorrects" are numbers that did not occur in the studied list. Larger numbers tend to appear somewhat more frequently as intrusions, $r = .36$, $p < .001$. The scores of best remembered categories (smaller numbers) do not appear to have been contaminated by mistaken reproduction.

Discussion

The differences found in this experiment are quite large. Small numbers are remembered well. They have a chance of over eighty percent of being correctly recalled or recognized. In contrast, numbers belonging to no specific subcategory have a poor chance of being remembered, with average percentages of less than forty. Teens, doubles and tabled numbers take intermediate positions.

Size has a strong influence. For the most part, this influence can be captured within a simpler structure of subcategories of numbers. These subcategories organize numbers largely on the basis of their size. The prime distinction in this respect is between singles, teens, and all other categories. However, even the subcategory of large tabled numbers, which in this comparison does not contain the tabled numbers under twenty and the doubles, is of a smaller average size than the unspecified category called "others". Mean sizes are 52 and 63 respectively. Moreover, memorization scores within each subcategory have a negative correlation with size. These correlations vary between -.18 for tabled numbers and -.26 for singles. None of them are significant, but all are in the expected direction.

A special case are the doubles, which are the only high scoring subcategory. Its average size

(60) is greater than that of the whole set (50). The advantage of the doubles may have been partly perceptual, standing out as they do by the repetition of two digits. Singles, equally, may well have obtained some extra profit from their visual distinctiveness.

Though the negative influence of size is large and pervasive, it seems to be indirect rather than direct. By this I mean that it is not to be explained by the specific contents of a number's representation but, rather, by the presence or absence of any content. The problem with some larger numbers is not that their representation is so large, but that it is insufficiently specified. It does not contain the information needed for correct identification. For that reason, they are sometimes confused with other numbers that are equally underspecified, and may share some of their characteristics.

It is quite illuminating to compare the ten best remembered numbers in both conditions with those scoring worst. These numbers are printed in Table 5.1. As can be seen, all well remembered numbers are either singles, teens or doubles, with the exception of 100, which is the sole representative of the large tabled numbers. Most of the lowest scoring numbers, in contrast, belong to the nameless category of Other numbers.

People do not like such numbers. It will be remembered that, at one point in our research, we had subjects rate all numbers 1-100 on a series of scales (See Chapter 4). Some of the findings reported in that chapter are pertinent to the present discussion. One scale we used was a good-bad scale. This scale proved to be quite reliable, with an alpha of .81 for all numbers, and one of .91 for a selection consisting of the twelve top and the twelve bottom numbers. It is interesting to compare this selection, which is given in Table 5.3, with the ratings obtained in the present experiment. The twelve "good" numbers are, in order of the magnitude of their goodness ratings: 10, 100, 36, 8, 24, 66, 16, 4, 1, 88, 21, and 12. As can be seen, they all belong to one of the privileged categories. The twelve "bad" numbers are, in the same order: 37, 93, 41, 51, 39, 17, 13, 59, 29, 43, 53, and 67. Remarkably, all belong to the category of numbers which is also difficult to remember. In a small follow-up experiment, we had 21 subjects memorize the lists of "good" and "bad" numbers, which were presented to them in Arabic format, on a sheet of paper. These students were given a minute to study a list of twelve numbers, after which they had to reproduce it immediately. Ten subjects first memorized and reproduced the list of good numbers, and tackled the bad list after a short pause. Eleven subjects took the tests in the reverse order. Of particular interest here is the nature of the intrusions that were produced. For the "good" list, which was perfectly reproduced by fifteen out of twenty-one subjects, only one incorrect response was obtained, being 64. Subjects had more problems with the "bad" list, which was correctly reproduced by only one of them. Moreover, there were fifteen intrusions. We have listed them in Table 5.3, because they seem to shed some light on the kind of things people do remember about such "bad" and difficult numbers.

Table 5.3

List of twelve "good" and twelve "bad" numbers, with the intrusions produced by 21 subjects

"Good" list

Members	10, 100, 36, 8, 24, 66, 16, 4, 1, 88, 21, 12
Intrusions (n=1)	64

"Bad" list

Members	37, 93, 41, 51, 39, 17, 13, 59, 29, 43, 53, 67
Intrusions (n=15)	63 (3 x), 19 (3 x), 69 (2 x), 83, 79, 57, 49, 33, 31, 23

As the table shows, all intrusions of the bad list are odd numbers, consisting of two digits, the second of which is either 1, 3, 7, or 9. In these respects the intrusions are similar to the stimulus numbers.

Of equal interest in the context of this discussion is a comparison with the numbers' Meaningfulness (m), as determined in the association experiment. Following Noble (1963), m was defined as number of associations produced during a fixed amount of time, which in this case was thirty seconds. For the complete set of numbers, m has a moderately strong correlation with memorization scores (.45 and .42 for recall and recognition, respectively). However, when we look at the extremes, it is even more evident that the categories found to distinguish numbers in the present experiment have relevance for their meaningfulness too. The ten numbers with the highest m-scores are, in descending order, 12, 50, 10, 60, 15, 16, 100, 11, 34, and 22. With the exception of 34, these are all numbers belonging to one of the privileged categories. The ten numbers with the lowest m-scores are, in the same order, 74, 87, 78, 98, 43, 59, 68, 85, 76, and 67. All these numbers belong to the category of numbers that are so difficult to remember.

In the next part we shall take a closer look at the relation between memory scores and the psychological measures obtained in the previous chapters.

PART 2: RELATING MEMORY SCORES TO OTHER EXPERIMENTAL VARIABLES

Method

Recall and recognition scores were collapsed into one variable, called "memorization". This variable denotes the overall proportion of correct responses for each number. Erroneous citations were similarly collapsed over recall and recognition conditions. The variable "errors" thus sums up all instances when a number was selected while it had not been part of that subject's list.

From the experiments reported in the previous chapters nineteen variables were selected for a comparison with positive and negative memory scores. These variables were previously shown to tap different aspects of the numbers representation. Appendix 6 gives their means and standard deviations. I will only briefly describe the variables here, referring to the relevant chapters of this book for further information.

1. Frequency general. This variable measures the overall Response frequency of each number in continued association, see Chapter 2 and Appendix 1;

2. Frequency difficult. This variable indicates the Response frequency obtained for each number in discrete association in the "difficult" condition. See Chapter 3; Experiment 1.

3. Frequency easy. The same, for the "easy" condition. See Chapter 3, Experiment 2;

4. Pleasant score. This variable indicates the frequency of a number's being named as pleasant. See Chapter 4, Experiment 2;

5. Unpleasant score. This variable indicates the frequency of a number's being named as unpleasant. See Chapter 4 , Experiment 2;

6. Meaningfulness (m). This variable measures the number of responses generated by each stimulus number. See Chapter 2 and Appendix 1;

7. Commonality general. This variable measures response commonality (proportion of subjects giving the most frequent response), for each stimulus number, as assessed by continued association. See Chapter 2 and Appendix 2;

8. Commonality 1st. This variable measures the commonality of the first named response in continued association. See Chapter 2 and Appendix 2;

9. Commonality difficult. This variable measures response commonality obtained in discrete association in the "difficult" condition. See Chapter 3, Experiment 1 and Appendix 2;

10. Commonality easy. This same, for the "easy" condition. See Chapter 3, Experiment 2, and Appendix 2;

11. Response diversity difficult. This variable measures the number of different responses generated by each stimulus number in discrete association. See Chapter 3, Experiment 1;

12. Response diversity easy. The same, for the "easy" condition. See Chapter 3, Experiment 2;

13. Omissions difficult. This variable measures the proportion of failed responses to each stimulus number in discrete association, "difficult" condition. See Chapter 3, Experiment1;

14. Omissions easy . The same, for the "easy" condition. see Chapter 3, Experiment 2;

15. RT difficult. This variable measures average times (discounting omissions) for each stimulus number in discrete association, "difficult" condition. See Chapter 3, Experiment 1;

16. RT easy. The same, for the "easy" condition. See Chapter 3, Experiment 2;

17. Good. This variable measuresthe "good" ratings of all numbers obtained by means of the Semantic Differential. See Chapter 4, Experiment 1;

18. Heavy. The same, for "heavy" ratings. See Chapter 4, Experiment 1;

19. Excitable. The same, for "excitable" ratings. See Chapter 4, Experiment 1.

Results

The intercorrelations between the variables are given in Appendix 7. Table 5.4 presents their correlations with the memorization and error scores obtained in the memory test.

Table 5.4.

Correlations of 19 experimental variables with memorization scores and error scores of 100 numbers obtained in the memory test (recall and recognition combined).

	Memorization	***Errors***
Frequency general	.78	-.41
Frequency difficult	.69	-.35
Frequency easy	.70	-.31
Pleasant	.67	-.23
Unpleasant	.59	-.21
M-score	.44	-.32
Commonality general	.39	-.29
Commonality 1st	.27	-.24
Commonality difficult	.39	-.24
Commonality easy	.38	-.31
Resp. diversity diff.	-.54	.50
Resp. diversity. easy	-.65	.46
Omissions. difficult	.05	-.03
Omissions easy	-.46	.21
RT difficult	-.25	.19
RT easy	-.39	.31
Good	.43	-.22
Heavy	-.47	.20
Excitable	.25	.03

Note. Al full description of the variables is given in the Method section.

As Table 5.4 shows, most of the nineteen variables have some relation with memorability. The only exception are the omissions in Experiment 1 of Chapter 3 (Omiss. Diff.), which, it may be remembered, gave rise to doubts even at its first appearance. Entering the 19 variables in a multiple regression analysis gives multiple R's with memorization and error scores of .89 and .62, respectively. These are respectable values. Using the four objective predictors from Part I of this chapter gives lower R's, of .81 and .37, with both memorization and errors. However, the present pattern is difficult to interpret. The multiple regression test works by spreading the explanatory credit over many variables. As a result, only three variables obtain an individual p-value smaller than .05 by this test. These are Commonality General, Omissions Easy, and RT Easy. This outcome is not really satisfactory, considering the correlations given in Table 5.4. A stepwise regression does not solve the problem either. By this test only those variables are selected which make a significant contribution on their own. It starts with the most powerful one, which may use as much of the variance as it can, and then looks for a next one, which may help itself to the remains. The procedure goes on for as long as new variables can be found to make significant additional contributions. In the present case, one single variable takes almost all of the predictive credit for memorization scores. It is the first variable, Frequency General, which measured the frequency of a number's being used as a response in the continued association test. It has a correlation of .78 with memorization scores. The next to enter is Omissions Easy, which raises the multiple R to .80. As a third step Commonality General is introduced, (R = .81) and the last variable to be entered is Commonality Difficult (R = .82). The multiple R of .54 with error scores obtained by the stepwise procedure rests on the contribution of two variables only, being the two Response Diversity measures (difficult and easy). Many potentially interesting variables are thus left out.

An alternative approach is to search for a set of underlying dimensions used in subjects' representation of the numbers. One such dimension may presumed to be frequency or general availability. A second dimension might underlie positive appreciation, and may also be related to richness of connections with other numbers. Some numbers may be prominent on one dimension only; others may manifest themselves on both. Both dimensions, and possibly others too, may be relevant in defining a number's memorability. To investigate these possibilities we have first performed a Principal Components analysis on the 19 variables. As a next step, we have correlated the obtained factors with the memorization scores.

Factor analysis

On the 19 variables described above (see also Appendix 6) a principal components analysis was performed. Five components were extracted by default procedure (extracting all components with eigenvalues greater than 1, plus the next one in size) and then rotated to the Varimax criterion. Together, these factors explain 70 percent of the original variance in the numbers' scores. Four components have eigenvalues greater than one. Their magnitudes are 7.47, 3.18, 1.41 and 1.35, respectively. The fifth has an eigenvalue of 0.86. The proportions of the variance explained by the factors are 39, 17, 7, 7, and 5 percent respectively. The dimensional structure obtained after orthogonal rotation is given in Appendix 8. I will briefly describe its main features here. Information about these features comes from three sources. The first source is the factor matrix (see Appendix 8), which gives the correlations between the obtained factors and the original variables. The second is the correlation matrix printed in Table 5.4. It gives the correlations between the factors and five objective characteristics: 1. Size, 2. Even/odd, 3. Tabledness, 4. Number of Syllables, and 5. Number of Digits. The third source of information are the scores of individual numbers on the five factors (see Appendix 9).

Table 5.5

Correlations between the obtained factors and five objective attributes of numbers .

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*
Size	-.64	-.17	.09	-.27	.37
Tabled	.28	.62	.20	-.02	-.31
Even/odd	-.04	.38	-.20	-.45	-.11
n-Syll	-.75	-.35	-.07	.07	.13
n-Digits	-.79	.04	-.13	-.07	.07

Note. The variable named n-syll represents the number of syllables a number contains when considered as a numberword. The variable n-Digits represents its number of digits.

The largest factor, Factor 1, has strong negative correlations with size, number of syllables and number of digits (see Table 5.5). It has strong positive correlations (between .84 and .94) with all five response frequency measures (see Appendix 8). It has negative correlations with response diversity and heaviness. It has a positive association with omissions in Experiment 1 of Chapter 3. In that experiment, it may be remembered, omissions were very frequent for single digit numbers. An inspection of the factor scores confirms that the Factor 1 is in fact a frequency factor. If we look at its extremes, as represented by the numbers scoring 1.5 standard deviation above or below the mean of 0 (see Appendix 9), we find that numbers 3, 7, 5, 1, 2, 4, 8, 10, 11, 9, 6, and 13 all stand out at the positive side. No numbers score below –1.5. This factor evidently separates the numbers 1 - 13 from all other numbers.

Factor 2, which is also large, is concerned with tabledness (.62), evenness (.38) and number of syllables (negatively, -.35) as objective attributes. It is strongly associated with the experimental measures of meaningfulness (.68) and goodness (.74). It has positive correlations with commonality measures and negative ones with RTs, omissions and response diversity. The variable pleasant has a positive correlation of .21 with this factor, while unpleasant has a negative one of -.16. Again, the factor scores well illustrate what Factor 2 stands for. The highest scoring numbers on the positive side are 10, 16, and 100, followed by 8 and 24. The numbers standing out on the negative side are (starting from the most negative one): 59, 53, 58, 87, 67, 41, and 13.

Factor 3 has no clear objective referents, as is demonstrated by Table 5.4. However, its nature is not so difficult to determine. The factor has positive correlations between .60 and .77 with response commonality in the association experiments, and negative ones (-.47 and -.68) with response diversity. It is also negatively correlated with RT. The factor scores confirm the impression that this component separates numbers with one dominant associate from numbers that lack such manifest connections. High scoring numbers on this factor are, on the positive side, 2, 99, 9, 1, 100, 59, 77, 98, 81, and 55. Examples of such dominant associations are 100 to 99, and 7 to 77. Its representatives on the negative side are 56, 34, 42, 68, and 63. These numbers have no single most prominent association.

Factor 4 has a negative correlation of -.45 with evenness, which means that it tends to favour odd numbers. As to the experimental measures, its largest correlation is with excitability (.86). It also has a positive correlation of .31 with unpleasantness, and a negative one of -.35 with goodness. Other variables come in beside these affective measures. The factor is negatively associated with omissions and heaviness. The highest scoring number on the positive side of this fourth factor is 13, followed by 81, 9, 33, 28, and 27. All numbers with culturally established affective connotations have positive standard scores on this factor. The numbers 11, 7, and 3 have scores of 1.22, 1.04, and .61 respectively. Numbers standing out on the negative side are 80, 82, 87, 36, 40, and 85. Though the factor is primarily an affective one, some different sources of residual variance have been used in its actual composition.

Factor 5 has a positive correlation of .37 with size and a negative one of -.31 with tabledness. Its prime loading is on RT in Experiment 1 of chapter 3, the difficult version of the discrete association experiment (.88). Its correlation with RT in the easier version of Experiment 2 in Chapter 3 is lower (.44). It is also positively correlated with heaviness (.50). In this case, the factor scores offer no ready clues to the factor's interpretation. It distinguishes the numbers 10, 100, 93, 54, 95, 57, 96, and 85 on the positive side from 44, 62, 72, 22, 26, 32, 35, and 15 on the negative side. Some residual variance in these numbers' heaviness and/or slowness of being responded to has evidently been used to construct this factor, which, it should be noted, is a very small one to begin with.

Correlations with memorization scores

Table 5.6 presents the correlations of the five factors with the scores obtained in the memory experiment. To the prediction of correct memorization, all five factors make a significant contribution ($p < .01$) with positive beta coefficients. Only the first three factors make a significant contribution to the prediction of the error scores ($p < .05$), with negative standard coefficients.

Table 5.6.

Individual and multiple correlations of the five factors with memorization and error scores of 100 numbers.

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*	*R*
Memorization	.62	.38	.25	.16	-.17	.80
Errors	-.27	-.21	-.29	.09	.13	.48

Comments

The reduction of the 19 variables to a simpler set of underlying dimensions has led to some loss in predictive values compared with the first multiple regression test, in which all variables were used. As will be remembered, that test gave multiple R's of .89 and .62 for correct memorization and errors respectively. These are now .80 and .48 (see Table 5.6). After cor-

rection for capitalization on chance (see Nunnally, 1967, p. 164) the multiple R's obtained for the 19 variables (.86 for memorization and .49 for errors) are still higher than those obtained for the five factors (.79 and .45 respectively). However, the gains are greater than the loss. From a statistical point of view, the factor structure is attractive because it presents us with a set of uncorrelated variables. As a result, the picture obtained by stepwise regression is almost identical to the one presented in Table 5.6.
But the principal gain is a conceptual one. A set of well interpretable psychological dimensions has been obtained. The first is a frequency dimension, which is primarily correlated with smallness. The second is a general evaluative dimension, which is correlated with the "roundness" of numbers (see Costermans, 1990b). The third factor is primarily associative and favours numbers with strong specific connections. The fourth factor singles out numbers which have culturally established affective connotations. The relevance of these dimensions is further confirmed by their significant correlations with positive and negative memorization scores.

General Discussion

Categorization of psychological data is almost always imperfect and every solution has its drawbacks. The categories used to describe the data in this chapter are no exception to these rules. In part I, numbers were subdivided according to their non-mental attributes. Though the structure gives a fair description of the data, it is by no means the only one that can do so. An alternative possibility, which we have used in other experiments (see Chapters 2 and 3), is-to distinguish between tabled and non-tabled numbers only, tabled numbers being defined as the numbers 1-12 and their products between 1 and 100. Such a subdivision contrasts the combined categories of singles, teens minus 13, 17 and 19, doubles and large tabled numbers with a slightly extended category of "others", which in that case also contains 13, 17 and 19. Separating numbers with different morphemic structures would also be an option. Such an approach would contrast the numbers with the simplest morphemic structure in Dutch (1 - 12), with those consisting of two morphemes (the remaining teens and the tens) and three morphemes (all others). All such subdivisions are defensible, and each will capture a good part of the variance in experiments such as the present one. The reason is that the numbers system itself discriminates between numbers. Some numbers participate in many substructures, while others participate in few. The category of "other numbers" is not particularly favoured by the numbers system. In general, its participants have relatively few intrinsic properties which make them candidates for individual attention.

In Part 2 we obtained a set of at least four independent psychological dimensions on which a number can be represented. These dimensions were obtained by a principal components analysis performed on nineteen experimental variables. This categorization does not primarily distinguish between numbers, as the first one does. Instead, it distinguishes between different aspects of their psychological representation, as measured by the experiments. One number may score high on more than one dimension; another may be inconspicuous on all dimensions. This structure gives a good description of the data, too. The dimensions that were thus obtained stand for the psychological meaning of numbers, which is made up of different components.

Of course, the obtained factors cannot be presumed to be a definite inventory of the psychological meaning of numbers. For one thing, real experts can distinguish many numbers which ordinary people are bound to confuse. Testing subjects with special knowledge of numbers might reveal different components in such numbers' representation. For another, variations of context and task demands, highlighting different features of numbers, can easily produce some changes in the hierarchies obtained, which may influence the factor structure. Introducing a new variable obtained in the same experiment may have similar effects. Moreover, a change in the statistical procedure may bring about a change in the output. We found, for example, that extracting and rotating the first four factors instead of the first five somewhat changed the composition of the fourth factor, though the first three remained unaffected.

Main findings

¶ The numbers in this study have very unequal chances of being remembered in a list learning task. Proportions correct memorization vary between 80 (for Single Digit numbers) and less than 40 (for numbers belonging to no specific subcategory).

¶ Correct memorization was well described by a combination of psychological measures obtained in other experiments. Associative frequency was the best single predictor.

¶ Nineteen experimental variables could be reduced to a simpler structure of five independent components, with no great loss of predictive value for memorization scores. Four of these general factors could be well interpreted. They stand for psychological frequency, general positive appreciation, exclusiveness of associations and specific affective meanings. These four factors were also found to make a significant individual contribution to the memorization of numbers.

1 Some of the results reported in Chapter 5 have also been published elsewhere. See Elshout and Milikowski (1995a, 1975b); Milikowski and Elshout (in press).