CHAPTER 3 The direction of associations

Introduction

In this chapter all numbers 1-100 will be put to a so called "discrete" association test.[1] This means that subjects must respond to a number, as quickly as possible, with just one other number. The resulting measures are comparable but not identical to those obtained by continued association. The most important similarity between the two types of experiments is in the actual procedure, which is specifically designed to obtain an "abstract", that is context-free, representation of the concepts involved. In both free and continued association subjects are therefore presented with isolated stimuli, and instructed to follow their minds' direction in the choice of a response. However, there are also important differences.

In the previous experiment people were instructed to give as many associations as they could think of during thirty seconds. Thus, response time is fixed. In discrete association, it is one of the response variables. RTs can be used to distinguish between stimuli. In word association, RTs average around a second and a half (1500 milliseconds), but concrete words have shorter RTs than abstract words.

Another important difference is the limitation of responses to one single association per subject per number. It guarantees that leading associations will be determined by the "one man one vote" principle. In the previous experiment, individual subjects did not have to choose between, say, 5 and 11 when associating to 55, but could name both, adding 50, 60, 56, 10, 110 and whatever number might further come to mind. Response commonality as measured in continued and discrete association are in some respects different measures. The first reason has just been given: when subjects give more associations, there is also more chance of overlap. Take the example of the stimulus number 10, with 100 as its most frequent response, which in the previous experiment had a commonality score of hundred percent. This perfect commonality does not mean that each subject named 100 as a first associate to 10; it only indicates that each of the 18 subjects ultimately named 100 as an association, perhaps after considerable thinking. In discrete association one single response number must be selected, and it must be done quickly. As a result, commonality will be lower. In word association the proportion of subjects giving the most popular response averages around thirty percent. In a study of Lauteslager, Schaap and Schievels (1986), relatively unfamiliar and abstract words, such as attention, thought and belief, had commonality scores as low as seventeen percent. Well known concepts on the other hand, such as names of popular colours (e.g., red) and fruit (e.g., apple) may have commonality percentages as high as forty (see also Schievels, 1988). There is a second, more technical, reason why the two measures of commonality differ. Because in continued association response commonality is positively influenced by a number's m-score, numbers that give rise to fewer associations are at a double disadvantage. This can be illustrated by a simple example. Consider the following set of responses, produced by three fictitious subjects, to the stimulus number 10.

	1	***2***	***3***	***4***	***5***
Subject A:	1	0	100	11	1000
Subject B:	100	1000	10.000	25	0
Subject C:	20	50	5	1	100

First suppose that these people produced no more than three associations each, being the first three numbers in each row. This would give 10 an m-score of 3. The leading response is the number 100, with a commonality of sixty-seven percent. Next assume that each subject produces a fourth response, being the numbers printed in the fourth column. M is now 4 and no single leading response is obtained, because 100 and 1 both score sixty-seven percent. The addition of the fifth column raises the m-score to 5 and the commonality of the leading response, which is now 100 again, to hundred percent. The example illustrates that commonality may be raised as a function of m. Such a confounding of measures is avoided in discrete association, where each number obtains an equal number of responses.

Numbers are all abstract, but not equally familiar. The category of tabled numbers (occurring in the multiplication tables 1-12, see Chapter 2, Table 2.1), which contains the frequently used single-digit numbers as well as the tens, is without doubt more familiar than the category of non-tabled numbers, which contains such difficult exemplars as 67 and 79. In fact, the prime defining attribute of the non-tabled numbers is their relative unrelatedness. These are the numbers that remain after the semantically more privileged groups have been taken care of. However, this is only the objective position. The question is if, and to what degree, this potential advantage of tabled numbers is reflected in people's mental representations.

To investigate this matter tabled numbers and non-tabled numbers will be compared on three measures. The first is leading response commonality, known in the literature as AFP (associative frequency of the primary response). The second is percentage of omissions. Sometimes subjects fail to produce an association. This happens more frequently with abstract than with concrete words and it seems reasonable to expect that it will happen more often with non-tabled than with tabled numbers. The third measure is reaction time (RT). On the one hand it might be expected that tabled numbers, being well-connected semantically, should be responded to more quickly than non-tabled numbers. The question then becomes whether such an RT-advantage is caused by the semantic content of these numbers, or by their frequency.

Frequently encountered concepts are, in general, more easy to handle than infrequent ones. A possible explanation is that frequent concepts, having stronger links to other concepts, are more easy to respond to. However, in the experiments of De Groot (1989) stimulus frequency had no influence whatsoever on associative RT after another factor, concreteness or imageability, had been controlled for. Should these results be replicated here, an alternative account of the representation of frequency is needed. In the previous chapter it was suggested that the stronger position of frequent numbers may primarily become manifest through response selection. Frequent numbers may “fit” themselves to many other numbers, while infrequent numbers do not. In that context I referred to Rosch (1975), who introduced the distinction between reference concepts and non-reference concepts to explain asymmetries in semantic relationships. In the experiments of Rosch, the reference numbers were tens, or powers of ten. The investigated asymmetries were those between neighbours, such as 79 and 80. However, additional forms of conceptual asymmetry may well exist. The results of the previous experiment in fact suggest that single-digit numbers are reference concepts to many other numbers. This notion will be tested by investigating the associative reference structure within the set of numbers 1-100. It is possible that concepts with referent status (such as single-digit numbers and tens) form distinct associative subcategories, in the sense that members of these subcategories are predominantly associated to members of the same group and have few associations outside it. Concepts with no referent status, on the other hand, could be associated with reference concepts as well as with non-reference concepts. If this should be indeed the case, associative asymmetries could be explained from a hierarchical ordering of semantic knowledge.

The association experiment to be presented here falls into two parts. The only difference between the two settings is in the instruction. In the first experiment, subjects were instructed to respond to the presented stimulus with “any other number, the first that came to mind”. This is in fact the standard instruction for such experiments. However, many subjects found the task difficult to perform. This was apparent from their comments, and also indicated by some of the results. This led us to perform a second experiment, with a different instruction. This time, subjects were explicitly told to think of connections between numbers, and respond along such lines. Separate analyses of the two experiments will first be given. Afterwards their results will be compared.

EXPERIMENT 1

Method

Stimuli

The stimuli consisted of all numbers 1-100.

Subjects

Subjects were 54 psychology students, fulfilling a course requirement. They were tested in individual sessions. All subjects were presented with a complete set of stimuli.

Apparatus

The experiment was run on a Macintosh Plus ED computer. The presentation of the stimuli, in a random order which was newly created for each session, was controlled by a program written in Course of Action®. Stimuli were presented on the monitor screen in Arabic digit format. Before the appearance of a stimulus, a warning beep sounded. Responses of subjects were typed in on the computer's keyboard, out of the subject's view. Response times (RTs) were recorded by means of a voice-key. When a response was registered by the voice-key's microphone, the stimulus-number automatically vanished. The next stimulus appeared when the experimenter pressed a key.

Procedure

After the subject sat down in front of the computer screen, the instruction was read out, stating that her or his task in this experiment was to respond to a number appearing with any other number that came to mind, as quickly as possible. If no number response came to mind, the subject could say so, after which an omission would be recorded. An omission was also automatically recorded when no response was given within five seconds.
Subjects were then presented with ten practice stimuli, which had been drawn from the set of numbers between 101 and 200. During these practice trials, the voice key was also tested. If necessary, its position or sensitivity was adapted.

Results

Response frequency correlates .90 with the frequency distribution obtained in the continued association task (see Chapter 2). Within the set of 28 numbers for which general language frequency scores are available, the present measure correlates .65 with the Dutch frequency count (numerical instances only). This is considerably lower than the correlation between continued association frequency and the Dutch count, which is .89.

Tabled vs non-tabled numbers

For all numbers three scores were calculated: 1. AFP (proportion of subjects giving the most frequent response), 2. Omissions (percentage of subjects failing to respond), and 3.

RT (response time). Average RTs were calculated for each stimulus after removing the trials leading to omissions. Scores of tabled and non-tabled numbers on these measures were compared by Anova. The category means are given in Table 3.1.

Table 3.1

Mean scores of tabled and non-tabled numbers on three response measures.

	Tabled (n=53)	***Non-tabled (n=47)***	***Overall***
AFP	18.1	14.2	16.2
Omissions	11.8	13.4	12.6
RT	1269	1334	1300

Note. AFP and Omissions are given in percentage, RT in msecs.

Tabled numbers have higher AFP-scores (relative frequency of the primary response) than non-tabled numbers. The category means are 18.1 and 14.2 respectively. The difference is significant, $F(1, 98) = 17.22$, $p < .0001$. Six numbers had AFP-scores higher than 25 percent. These numbers, with their leading responses, are 99 : 100; 81 : 9; 18 : 9; 77 : 7; 98 : 100 and 100 : 10. Ten numbers had AFP-scores lower than 10 percent. These numbers are 57, 46, 58, 85, 13, 68, 87, 42, 34 and 56. Most of these have no single leading response.

Tabled numbers are also responded to more quickly than non-tabled numbers (1269 vs 1334 milliseconds, $F = 18.21$, $p < .0001$). Twelve stimulus numbers had RTs under 1200 milliseconds. These numbers are (from fast to slow): 99, 44, 22, 14, 45, 15, 72, 40, 18, 30, 71 and 80. Nine numbers had response times over 1400 milliseconds. These are (in the same order) 97, 96, 95, 46, 23, 93, 54, 10 and 57.

A marginal effect of stimulus category was obtained for omissions. Tabled numbers tended to have less omissions than non-tabled numbers (11.8 percent vs 13.4 percent, $F = 2.82$, $p = .09$). Ten stimulus numbers had omission scores lower than 6 percent. These are (starting with the lowest): 33, 18, 25, 48, 45, 16, 68, 42, 27 and 81. Twelve numbers had omission scores higher than 15 percent. These are (in the same order), 41, 43, 53, 82, 4, 59, 7, 3, 40, 85, 87, 5 and 1.

The influence of stimulus frequency

Tabled numbers have higher overall frequency, as measured by continued association (see Chapter 2, Table 2.8) than non-tabled numbers. The average tabled number was mentioned almost four times as often as the average non-tabled number. (The mean frequencies were 10 and 2.8, respectively.) The effects of tabledness might therefore to some degree be effects of response frequency as measured by continued association. Multiple regression tests were performed to determine the combined and independent effects of the variables tabledness and frequency. The results are presented in Table 3.2.

Table 3.2.

Multiple correlations and standard regression coefficients (beta weights) of the stimulus variables tabledness and frequency with five response measures.

	Mult. Corr.	***Tabledness***	***Frequency***
AFP	.40	.33**	.10
Omissions	.44	-.44**	.49**
RT	.40	-.43**	.06

* = p < .05 ** = p < .001

The multiple R's are statistically significant for all three measures, p < .0001. Tabledness has an independent positive influence on AFP (r = .33). It has an independent negative influence on percentage of omissions (r = -.44) and RT (r = -. 43).

Frequency, on the other hand, has no significant independent influence on RT (r = .06) and AFP (r = .10). However, it has a considerable independent positive influence (r = .49) on omissions.

The differential effects of frequency and tabledness may be explored further, by comparing the scores of six subcategories of numbers. For this comparison we have selected the following subcategories: 1. Single-digit numbers (1 - 9) ; 2. Tens (10, 20 etc., up to 100); 3. Teens (11 through 19); 4. Double digit numbers (22, 33, 44 etc.,); 5. Other Tabled numbers from 21 up (21, 24, 25, etc.,), and 6. the Remaining numbers, which do not belong to any of the mentioned subcategories (23, 26, 29, etc., n = 44). The mean frequencies of these categories as determined by continued association (see Chapter 2) are 413, 191, 197, 131, 75, and 44, respectively. In fact, the subcategory structure that is given here captures most of the variance in the frequency variable (r = .89). The first three subcategories are those for which general language frequency scores are available, which seems to be a further indication of their relatively privileged position.

Table 3.3.

A comparison of six subcategories of numbers on three response measures.

	Singles (n = 9)	***Tens*** (n = 10)	***Teens*** (n = 9)	***Doubles*** (n = 8)	***OthTab*** (n = 20)	***NonTab*** (n = 44)
AFP	17.9	19.1	17.7	23.6	15.7	14.0
Omissions	18.1	13.0	10.3	8.8	10.3	13.5
RT	1296	1278	1249	1216	1289	1336

Note. AFP and Omissions are given in percentage, RT in msecs. The column OthTab contains the tabled numbers which are neither singles, nor tens, or teens. The column NonTab contains all numbers that do not occur in the multiplication tables 1-12.

Of particular interest is the very high proportion of omissions (18.1 percent) produced by the most frequent subcategory, the one containing single-digit numbers. For these single-digit numbers the proportion omissions is even higher than is AFP, the proportion of subjects who gave the primary response (17.9).

The category of double digit numbers (11, 22, etc.,), which is not particularly frequent, has an AFP-score of 23.6, the highest of all. It has, moreover, the lowest percentage omissions (8.8 percent), and the shortest RT (1216 milliseconds).

The distributions of responses

A stimulus-response matrix was produced (see Table 3.4) to inspect associative relationships within and between subcategories of numbers. Stimuli were divided into four subcategories: 1. Single-digit numbers (n = 9), 2. Tens (n = 10), 3. Teens (n = 9) and 4. Other numbers between 1 and 1200(n = 72). Responses were divided into six response categories: 1. Single-digit numbers, 2. Tens, 3. Teens, 4. Other numbers, 5. Omissions, and 6. Numbers not represented in the stimulus set, named outsiders. Responses obtained in all 5400 experimental trials (100 numbers times 54 subjects) could thus be scored. The resulting matrix is given in Table 3.4.

Table 3.4.

Absolute and proportional distributions of the responses in Experiment 1. Frequencies are given for four groups of numbers: single-digit numbers, tens, Teens and Other numbers. The columns are the stimulus groups. The rows are the response groups.

Experiment 1	***Stimulus numbers***				
	Singles (n = 9)	Tens (n = 10)	Teens (n = 9)	Others (n = 72)	All (n = 100)
Response numbers					
Singles (9)	252	128	203	1306	1889
%	51.9	23.7	41.8	33.6	35.0
Tens (10)	51	204	65	575	895
%	10.5	37.8	13.4	14.8	16.6
Teens (9)	42	28	86	558	714

%	8.6	5.2	17.7	14.4	13.2
Others (72)	41	70	81	967	1159
%	8.4	13.0	16.7	24.9	21.5
Outsiders (~)	12	18	1	29	60
%	2.5	3.3	.2	.7	1.1
Omissions	88	92	50	453	683
%	18.1	17.0	10.3	11.7	12.6
Totals	486	540	486	3888	5400
%	100	100	100	100	100

Table 3.4 can be read in two directions. The columns represent the distributions of all responses, including omissions, given to the four categories of stimulus numbers. Let us first consider the column of single-digit stimulus numbers. The totals at the bottom of the column give the number of experimental trials involving this stimulus category. Nine single-digit numbers times 54 subjects gives a Total of 486. Of these 486 trials, 88 (18.1 percent) failed to produce a response. These are the omissions. Twelve responses (2.5 percent) involved numbers not represented in the stimulus set (outsiders). The largest response category (252 responses, or 51.9 percent) consists of single-digit numbers. Tens (10, 20, etc) were given as a response in 51 cases (10.5 percent) and teens (11, 12, etc.,) in 42 cases (8.6 percent). In 41 trials (8.4 percent) a member of the large category of other numbers was named as a response.

It is evident from this table that single-digit numbers are predominantly responded to with numbers belonging to the same subcategory. "Other numbers" are infrequently given as a response to single-digit numbers. The differences become even more salient when the inequality between categories is taken into account. While the 252 single-digit responses were gathered by a group of nine stimulus numbers (giving an average of 28 per number) , the 41 "other number"- responses must be shared by a group of 72 stimulus numbers (an average of 0.57). The remarkable fact that single-digit numbers, as a stimulus category, are responsible for so many omissions has already become manifest from Table 3.3.

Tens, also, seems to invite responses from the same category. Of the 540 trials involving a ten as a stimulus number, 204 (37 percent) lead to a ten as a response. Tens are the most frequent response category for tens. For the stimulus group of teens, single-digit numbers are the preferred responses (203 cases, or 41 percent). Single-digit numbers are also the largest response category for "other numbers" (1306 cases, or 33.6 percent). Teen-responses take second place for teen-stimuli, and "other number"-responses take second place for "other number"-stimuli.

The rows represent the scores obtained by each response category. These are summated in the last column. The largest category is that of single-digit numbers, which are given as a response in 1885 (35 percent) of all cases. The second largest is that of "other numbers" (1159 cases, or 21.5 percent). Tens come third (895, or 16.6 percent), then Teens (714, or 13.2 percent). The total number of omissions is 683 (12.6 percent). Only 60 responses (1.1 percent) involve numbers from outside the stimulus set (outsiders). For a proper understanding of the unevenness of this distribution it should again be considered that the categories have an unequal number of members. Thus, each of the "other numbers" occurs as a response 1159 : 72 = 16.1 times on the average. These averages are 209.8 for single-digit numbers, 89.5 for tens, 79.3 for teens and infinitely small for outsiders.

Primary associative relationships

Leading stimulus-response combinations (see Appendix 2) were scored by the method used in Chapter 2. This time, eight categories were needed to score all leading stimulus-response relationships. Several of these have already been described (see Chapter 2, Table 2.5). Other categories are neighbours (such as the stimulus-response pair 19 : 20), complements (consisting of a stimulus number and its complement to the next multiple of ten, e.g., 97 : 3), and next ten (e.g., 98 :100). The formula category contains associations which seem to represent a combination of the two digits of the stimulus (e.g., 67 : 13). An eighth category (absent) contains stimuli with no single leading associate. Table 3.5 presents all categories, with their frequencies and AFP-scores.

Table 3.5

Prime response categories as obtained in both experiments, with mean commonality scores for each category.

Category	*N*	*AFP*	*Examples*
Divisors (23) & multiples (1)	24	18.8%	(14 : 7)
Ambiguous	24	18.4%	(25: 5)
Neighbours	13	18.0%	(59 : 60)
Digits	8	12.7%	(83 : 8)
Next Ten	7	15.4%	(95 : 100)
Formula	7	12.7%	(67 : 13)
Complement	5	13.0%	(97 : 3)
Absent	12	11.2%	
Total	100	16.2%	

Table 3.5 shows that the three largest categories, those containing multiple-divisor pairs (n = 24), ambiguously related pairs (n = 24), and pairs that are neighbours on the numberline

(n = 13) have higher AFP-scores (18 to 18.8 percent) than the less frequent categories (12.7 to 15.4 percent).

Discussion

It seems clear from these results that subjects did not find the numbers association task easy to perform. It is true that tabled numbers score better than non-tabled numbers: 18.1 versus 14.2 percent on AFP (response commonality), and 11.2 versus 13.4 percent on omissions. However, the commonality scores are in fact rather low compared to those obtained for words, and the omission scores are relatively high. De Groot (1989) obtained AFP-scores of 28 and 40 percent for words with low and high imageability. Lauteslager and his colleagues obtained an average AFP-score of 30 percent for 540 words (Lauteslager et al., 1986; see also Schievels, 1989). Percentages omissions in these experiments with words were well below ten. It might be argued that numbers are highly abstract stimuli, and should therefore be compared with highly abstract categories of words. A possible example is the category "mental concepts" in the Lauteslager study, which has a average response commonality of only 16 percent, and an average omission score of 20 percent. The problem with this argument is that the highly familiar group of single-digit stimuli produces even more omissions (18.1 percent) than the less familiar group of non-tabled numbers (13 percent). Of all stimulus numbers, 1 has the most omissions, closely followed by 5. It is difficult to see why these numbers should be more abstract than, say, 14 or 62. The reverse would make more intuitive sense.

Possibly people just did not understand the task very well. We decided to test a new group of subjects in the same experiment, using an instruction in which the concept of relations between numbers is explicitly referred to.

EXPERIMENT 2

Method

Subjects and apparatus

Subjects were 50 first year psychology students fulfilling a course requirement.

Apparatus and stimulus material were the same as in Experiment 1.

Procedure

The procedure was also identical to the previous experiment, with the exception of the instruction. Whereas subjects in the previous experiment were instructed to respond to a stimulus-number with any other number that came to mind, they were now advised to respond to each stimulus-number "with its most likely partner, that is, the first that occurs to you."[2]

Results

Response frequency correlates .97 with frequency as obtained in Experiment 1, and .92

with response frequency in continued association. Within the set of 28 numbers studied by Dehaene and Mehler (1992), the present measure correlates .72 with the Dutch frequency count (numerical instances only). This is a better value than was obtained in Experiment 1 of this chapter (.65). It is still considerable lower, however, than the correlation of .89 which was obtained between continued-association frequency and the Dutch language count (see Chapter 2, Table 2.6).

Stimulus categories

Scores of numbers were calculated as in Experiment 1. The effects of stimulus category (tabled versus non-tabled numbers) are highly significant for each of the three measures. Category means of both experiments are given in Table 3.6.

Table 3.6

Mean scores of tabled and non-tabled numbers on five measures. Standard deviations are given within parentheses.

	Tabled (n=53)	***Non-tabled (n=47)***	***Overall***
Experiment 1			
AFP	18.1	14.2	16.2
Omissions	11.8	13.4	12.6
RT	1269	1334	1300
Experiment 2			
AFP	29.3	18.6	24.3
Omissions	3.9	8.1	5.9
RT	1330	1472	1397

Note. AFP and omissions are given in percentage, RT in msecs.

Tabled numbers have higher response commonality (AFP) and fewer omissions than non-tabled numbers. They are also responded to more quickly. The respective means of tabled and non-tabled numbers are 29.3 versus 18.6 percent for AFP , 3.9 versus 8.1 percent for omissions, and 1248 versus 1437 milliseconds for RT. Each of the comparisons is significant at an alpha level of .0001.

Experiments compared

On each response measure, differences between the two experiments were found (compare Table 3.6). AFP was higher in Experiment 2 than in Experiment 1 (24.3 percent versus 16.2 percent). Fewer omissions were registered (5.9 versus 12.6 percent). But responses also came more slowly (1397 versus 1300 milliseconds). All comparisons are significant , p

< .0001.

Interactions of Stimulus category and Experiments

The means given in Table 3.6 suggest that a comparison between the two experiments will give different results for tabled and non-tabled numbers. This is confirmed by an 2 x 2 (Experiment x Category) Anova . Significant interaction effects were obtained for all three measures, $F(1,196) = 11.9$, $p < .0001$ for AFP, $F(1, 196) = 5.0$, $p < .05$ for Omissions and $F(1,196) = 9.9$, $p < .01$ for RT. AFP-scores of tabled numbers are more strongly enhanced by the new instruction than those of non-tabled numbers. Omission-scores of tabled numbers are more strongly reduced. For response times the reverse is true: RTs of tabled numbers are less strongly affected by the change in condition.

Tabledness and frequency

Multiple regression analyses were again performed to determine the combined and independent effects of the stimulus variables tabledness and frequency. The results are presented in Table 3.7. The multiple R's are again statistically significant for each of the three measures, $p < .0001$. Tabledness has an independent positive influence on AFP (r = .43). It has an independent negative influence on percentage of omissions (r = -.48) and RT (r = -.62). Frequency has no significant independent influence on any of the three measures in this experiment.

Table 3.7.

Multiple correlations and standard regression weight (beta coefficients) of the stimulus variables tabledness and frequency obtained for three response measures.

	Mult. Corr.	*Tabledness*	*Frequency*
Experiment 1			
AFP	.40	.33**	.10
Omissions	.44	-.44**	.49**
RT	.40	-.43**	.06
Experiment 2			
AFP	.55	.43**	.18
Omissions	.55	-.48**	-.10
RT	.62	-.62**	.00

* = p < .05 ** = p < .001

Interestingly, the positive influence of stimulus frequency on percentage of omissions has completely disappeared (see Table 3.7). It is now even slightly negative, though the coefficient is not significant. For AFP and RT the patterns of coefficients are left essentially unchanged. In both experiments tabledness has a strong positive influence on AFP, while the influence of frequency is slight. And in both experiments frequency completely fails to influence RT when the effect of tabledness is left out.

Six subcategories of stimuli

The changes in the pattern that is manifest from the changed standard coefficients can be further illustrated by a comparison of the scores of the subcategories of numbers with high and low frequencies, which are given in Table 3.8. In Experiment 2, the subcategories of frequent numbers have manifestly higher commonality, as measured by AFP, than the less frequent subcategories. Of particular interest is the drop in the omissions of single-digit stimulus numbers, from a very high 18.1 percent to a moderately low 4.0 percent.

Table 3.8.

A comparison of six subcategories of numbers on three response measures.

	Singles (n = 9)	***Tens*** (n = 10)	***Teens*** (n = 9)	***Doubles*** (n = 8)	***OthTab*** (n = 20)	***NonTab*** (n = 44)
Experiment 1						
AFP	17.9	19.1	17.7	23.6	15.7	14.0
Omissions	18.1	13.0	10.3	8.8	10.3	13.5
RT	1296	1278	1249	1216	1289	1336
Experiment 2						
AFP	31.1	33.8	27.3	30.2	24.8	18.7
Omissions	4.0	3.2	4.2	2.0	5.1	8.3
RT	1261	1213	1362	1131	1293	1433

Note. AFP and Omissions are given in percentage, RT in msecs.

Response frequency

Response frequency in Experiment 2 correlated .97 with the response frequency distribution obtained in Experiment 1, and .92 with response frequency as measured by continued association (see Chapter 2). As in Experiment 1, frequency count were performed for four stimulus categories and six response categories. Absolute frequencies were transformed into percentages, to facilitate a comparison between the two experiments.

Table 3.9.

Absolute and proportional distributions of the responses in Experiment 2. Frequencies are given for four groups of numbers: single-digit numbers, tens, teens and *other numbers*. The columns are the stimulus groups. The rows are the response groups.

	Stimulus numbers				
	Singles (n = 9)	Tens (n = 10)	Teens (n = 9)	Other nrs (n = 72)	All nrs (n = 100)
Response numbers					
Singles (9)	357	145	250	1659	2411
	79.3	29.0	55.6	46.1	48.2
Tens (10)	32	265	40	382	719
	7.1	53.0	8.9	10.6	14.4
Teens (9)	19	19	92	576	706
	4.2	3.8	20.4	16.0	14.1
Others (72)	16	40	43	693	792
	3.6	8.0	9.6	19.2	15.8
Outsiders (~)	8	15	6	49	78
	1.8	3.0	1.3	1.4	1.6
Omissions	18	16	19	241	294
	4.0	3.2	4.2	6.7	5.9
Overall	450	500	450	3600	5000
	100	100	100	100	100

The response frequencies of Experiment 2 are given in Table 3.9. This table may be compared with Table 3.4, which contains the response frequencies of Experiment 1. The pattern revealed by Experiment 1 is even clearer in the present experiment. Now, 79.3 percent of all responses to single-digit numbers and 53.0 percent of the responses to tens are from the same category. These percentages were 51.9 and 23.7 in Experiment 1. The percentage single-digit numbers in the total response set has risen from 35.0 to 48.2.

Stimulus-response categories

All leading stimulus-response pairs could be traced to six categories. Table 3.10 presents the categories of leading associative relationships, with the frequency of their occurrence and their average commonalities.

Table 3.10
Prime response categories as obtained in both experiments, with average commonality scores for each category.

	Experiment 1		Experiment 2	
Category	***N***	***AFP***	***N***	***AFP***
Divisors (23) & Multiples (1)	24	18.8%	27	32.4%
Double meanings	24	18.4%	29	26.3%
Neighbours	13	18.0%	3	16.0%
Digits	8	12.7%	9	17.8%
Next ten	7	15.4%	-	-
Formula	7	12.7%	-	-
Complement	5	13.0%	20	20.0%
Absent	12	11.2%	12	15.0%
Total	100	16.2%	100	24.3%

As can be seen from Table 3.10, no "next ten" or "formula" responses were found among the leading responses in Experiment 2. In Experiment 1, these categories had been found to lead in seven cases each. The category of "complement" responses was much more popular in the second experiment (20 versus 5 cases). The remaining categories have similar frequencies in both experiments. As could be expected, all categories have higher AFP-scores in the second experiment. The difference is most pronounced, however, for the divisor responses, whose AFP-scores were raised from 18.8 percent to 32.4 percent.

General Discussion

Experiment 2 was performed because the results of the first were considered unsatisfactory. Our number stimuli scored much worse than words on prime response commonality (AFP) and proportion of omissions, a result that was both unexpected and difficult to explain considering the outcomes of the continued association test. Under the changed instruction subjects did indeed perform much better. Overall response commonality was raised from 16.2 tot 24.3 percent, and the proportion of omissions was reduced from 12.6 to 5.9. Moreover, the differences between the categories of tabled and non-tabled numbers were larger and more consistent in Experiment 2.

At first sight, the variable RT seems to be an exception to the general facilitating influence of the altered instruction. As can be seen by inspection of Table 3.2, RTs in the second experiment are longer. This may seem rather surprising, even though the difference in mean RT is exclusively caused by the group of non-tabled numbers (see Table 3.2). However, the RT advantage of Experiment 1 is probably an artefact. As stated in the Method section, the mean RT of each stimulus-number was calculated after removing the trials leading to omissions. An omission is a sign of response difficulty, as is a long RT. In Experiment 2, omissions were much fewer, indicating that many "difficult" numbers were now ultimately responded to. The effect was to raise the mean RT. This is confirmed by a comparison of RTs based on all trials, including omissions.3 Mean RT-values in the Experiments 1 and 2 are now 1457 milliseconds and 1396 milliseconds, respectively, revealing a main effect in the expected direction ($F (1, 98) = 11.00, p < .001$). The cell means are 1437 and 1482 milliseconds for tabled and non-tabled numbers in Experiment 1 and 1330 and 1472 milliseconds for tabled and non-tabled numbers in Experiment 2. Tabled numbers have a significantly shorter RT in both Experiments. There is also a significant interaction, indicating that tabled numbers are more strongly affected by the change in the instruction than non-tabled numbers.

It is an interesting question whether the changed instruction has led to a changed content of response patterns. The comparison offered in Table 3.5 suggests that the distributions of leading response types are rather similar for both experiments: multiple-divisor pairs (e.g., 21 : 7), and pairs categorized as ambiguous (e.g., 55 : 5), are most frequent in both cases. The difference is in the proportion of subjects backing the preferred response. There are some changes however. In Experiment 2, complement associations (97 : 3)gained in popularity, while next ten associations (e.g., 97 : 100) disappeared, together with the formula associations (e.g., 68 : 14). Neighbour responses (e.g., 19 : 20) were obtained as leading associations less often in Experiment 2 than in Experiment 1. A comparison of the leading responses printed in Appendix 2 confirms that these are different numbers in several cases.

However, the differences are not as large as they may seem. One indication is the high correlation between the two response frequency distributions. Another is the observation that the response sets obtained for each stimulus number show considerable overlap. Often, the difference between the two experiments is that first and second response numbers have changed places. For example, the two leading associates of 95 are 100 and 5 in both experiments. In Experiment 1, 100 was the first leading response, while 5 came second. In Experiment 2, these positions were reversed. A similar exchange between first and second response numbers has taken place in other cases: 98 (100, 2); 20 (2, 10); 30 (3, 10); 44 (4, 11), and 60 (6, 10). It is obvious in these cases that two response numbers which are in fact complementary to each other have changed ranks between experiments. But in fact, many of the leading combinations imply a third partner. Frequently this third partner also makes an appearance on the response list.

The notion that many associative combinations implicitly involve three numbers, opens a different perspective on the obtained commonalities. Consider the stimulus number 35 in response to which, in Experiment 1, eight subjects named 5, and eight other subjects 7. Measured by the commonality of either of these two responses, the number 35 has an AFP of 14.8. Should we, however, view the combination as a triangular one (35 : 5 implying 7 and 35 : 7 implying 5), AFP-scores would double and rise to a very respectable 29.6 percent.

I have inspected the response hierarchies of both experiments for such first and second placed "complementary" response numbers. These occurred in 23 cases in Experiment 1, and in 27 cases in Experiment 2. For the most part, the same stimulus-response triangles are involved in both experiments. Of the 23 triangular combinations in Experiment 1, six involve a stimulus from the category of tens (e.g., 20 : 2/10 and 90 : 10/9), six involve multiples of eleven (e.g., 33 : 3/11; 88 :8/11), seven involve other tabled numbers (e.g., 8 : 4/2; 27 : 9/3) and the remaining four involve stimulus numbers close to a hundred (e.g., 95 : 100/5; 97 : 3/100). Measured by their AFP-score, these 23 stimulus number have a mean commonality of 19.4 percent. Should we measure their commonality by the proportion of subjects giving either of the two leading responses, their mean commonality would be 32.8.

The triangular combinations in Experiment 2 involve the same subcategories of stimulus numbers (eight tens, eight multiples of eleven, eight other tabled numbers and three numbers close to 100). These 27 stimulus numbers have a mean AFP of 30.8, but a mean triangular commonality of 50.7.

These figures suggest that conceptual agreement between subjects is more extensive than the AFP-scores suggest. If this is so, our judgement of the results of Experiment 1 must perhaps be partly reconsidered. Commonality may have been less poor than we originally thought. However, the large proportion of omissions, our second source of dissatisfaction, still stands unexplained. In word association experiments a high proportion of omissions can generally be explained either by subjects' ignorance of a word's meaning, or by its abstractness. An example of the first case is the word "fennel" (venkel in Dutch) which in the Lauteslager study had an omission score of fifty percent (Lauteslager et al., 1986). Schievels' suggestion that most subjects - who were pupils from secondary schools - probably did not know its meaning seems a reasonable explanation (Schievels, 1988). An example of the second group is the stimulus word fact (feit in Dutch), used in a study of De Groot (1981). Though the word is a highly frequent one and presumably known well enough by the average subject, it is nonetheless difficult to associate to, as witnessed by its low commonality and high proportion of omissions.

In Experiment 1 of this chapter, single-digit numbers had the highest proportion of omissions. These numbers are neither more infrequent nor more abstract than other categories. They are, to the contrary, more frequent and also less abstract, in the sense that they are easier to imagine as quantities, or patterns, than larger-sized numbers. What, then, can have caused people's problems during Experiment 1 in responding to those numbers? The only explanation I can think of is that in Experiment 1 subjects had a greater tendency to respond to the more superficial attributes of a stimulus. Some suggestions to that effect come from the many "formula" responses in Experiment 1. Examples of these are 46 : 8 (a series), 31 : 4, 52 : 7, 53 : 8, 67 : 13, and 65 : 11 (all additions). This response type was much less prominent in Experiment 2, and completely absent in its set of leading associations. It is possible that a response strategy which concentrates on what is visually available rather than on "deeper" connections, leaves subjects powerless when presented with a single-digit number. Nothing in the immediate appearance of such a number can be used to construct a response with.

The question may be asked which experiment gives the best account of numerical knowledge. It seems to me that something can be said for both of them. The complaints of subjects in Experiment 1 (and the many omissions they produced) underlines that associating to numbers does not come naturally, as producing word associations sometimes seems to do. Experiment 2 shows, however, that a clearer instruction will help to overcome these difficulties. It might, of course, be argued that the new instruction gave too much away; that it artificially induced a semantic response set that could not be obtained by more natural means. This is a valid objection. By suggesting that subjects respond in terms of "relatedness", we put them on a track they might not have discovered otherwise.

It is not so clear, however, how serious this is. One curious feature of Experiment 1 concerned the relation between subjects' associative behaviour and their intelligence. When commonality was measured individually, by counting how often people responded with the dominant association, these individual scores correlated .38 with individual intelligence-scores.[4] In Experiment 2 this correlation was reduced to an insignificant and meaningless .05. Two components of the intelligence test were particularly relevant to response behaviour in Experiment 1. Those are Series and Conclusions, which correlate .34 and .36 with associative commonality. Both are tests of abstract reasoning. The correlations with other components (including Calculation Speed, r = .14, n.s.) were much lower.

These figures suggest that Experiment 1 was difficult in an unintended way. The purpose of an association experiment is to measure knowledge that is readily available. That means it should not present people with a puzzle. The instruction of Experiment 2 seems to have removed this unintended obstacle, by helping subjects to find the right track.

Stating that an association test measures what is readily available does not imply that the associations are completely prefabricated by earlier experience. The response times alone seem to contradict such a suggestion. In the two experiments reported here, RTs (excluding those of the omissions) averaged around 1300 milliseconds. Even longer latencies, of over 1500 milliseconds, have been reported for word association (De Groot, 1989). It is interesting to compare such associative RTs with those obtained in other tasks. In one of the experiments performed in the context of this dissertation (see Chapter 6) subjects were presented with 144 subtraction problems of the types 73 - 4 and 73 - 69. These problems were solved in 600 milliseconds on the average, which is less than half the time it takes to produce an association. A similar comparison can be made between word association latencies and RTs registered in translation tasks. The production of a response in an association experiment takes well over 1500 milliseconds on the average, while the translation of individual words takes considerably less time. De groot, Dannenburg, and Van Hell (1994) obtained RTs averaging around 1050 milliseconds. These comparisons suggest that the abstract representations which guide performance in association are not so much "accessed" as constructed during the task. A comparable viewpoint has been proposed by Barsalou (1987, 1989, 1990, 1993), who maintains that concepts are in fact produced in working memory. Barsalou did not perform association experiments, but studied how people use long term knowledge to produce semantic categories. Representations guiding performance in an association experiment may be analysed in a similar ways. These representations could also be viewed as ad hoc constructions, which use information from long term memory without being neatly determined in advance.

One strong determinant of associative content seems to be frequency. Frequency severely limits the set of numbers that can be used for the representation of a stimulus number's meaning. It is very clear from these experiments that the frequency dimension is not external to the content of semantic knowledge, but part of its very fabric. Small numbers, and to a lesser degree tens, are used to make sense of other numbers in many different ways. Table 3.4 gives evidence of a well structured asymmetry. Both tens and single-digit numbers behave as reference concepts in the sense of Rosch (1975). While single-digit numbers are essential to the representation of all other numbers, these other numbers do not contribute much to the meaning of the most frequent category. Single-digit numbers are also mutually related, in that most of the responses to such numbers (51.9 percent in Experiment 1 and 79.3 percent in Experiment 2) are from the same subcategory. Tens, also, shown signs of forming a special subcategory. Of all associations given in response to Tens 37.8 and 53 percent are other tens. That these effects are strongest in Experiment 2, where the response patterns were more semantic, is a further indication that frequency differences are built into the content of semantic knowledge.

Main findings

¶ The distribution of response frequencies obtained in discrete association are similar but not identical to those obtained in continued association. Of the three experimental frequency measures, response frequency in continued association gives the best approximation of general language frequencies.

¶ Stimulus-response combinations obtained by discrete number association fall into eight categories. The preferred associative response is a divisor of the stimulus number.

¶ Single-digit numbers make up the majority of all associative responses in discrete association, thus acting as reference concepts to many other numbers. This asymmetrical relationship can also explain why frequency is not effective as a stimulus variable, for instance, for the prediction of RT. Frequent numbers, though easy to respond with, are not particularly easy to respond to.

1 Some of the results reproted in Chapter 3 have also been published elswhere. See Milikowski and Elshout, 1994.

2 In Dutch: "Uw taak is om bij elk gegeven getal de meest voor de hand liggen de partner te

noemen, dat wil zeggen: de eerste die u te binnen schiet".

3 In a few cases, omissions were registered after five second. In most cases shorter Rt's were registered, since subjects would often state within five seconds that no association came to mind.

4 Intelligence scores (available for 44 out of 54 subjects in the first experiment and 44 out of 50 in the second) are based on a Guilford-type test, made up of six different components, which psychology students at the university of Amsterdam take during the first year of their study. The intelligence components (see also Elshout, 1976) are Vocabulary (representing CMU in the Guilford model), Conclusions (CMS), Series (CSS), Calculation Speed (NSI), Verbal Analogies (CMR) and Hidden Figures (NFT). The sums of the six compont-scores represent general intelligence. Individual commonality scores were determined by scoring, for each stimulus number, whether a subject did or did not produce the most frequent associative response. No scores were given for stimulus numbers that had no single winning associate.

The average componential intelligence of the 44 subjects in the first experiment was 20.9, with a standard deviation of 4.21. The average commonality score was 15.3, with a standard deviation of 9. In the second experiment average intelligence (of 44 subjects) was 20.2, with a standard deviation of 4.2. Average individual commonality was 23.2, with a standard deviation of 14.8.