CHAPTER 6 Size and difficulty in subtraction

Introduction

In the vocabulary of arithmetic all sums, however simple, are called problems. There are easy problems, such as 2 + 2, and more difficult problems, such as 78 x 35. Many problems change from difficult to easy during primary school. Sums involving combinations of single digit numbers (the basic combinations) are first solved by counting and later by what is often called "direct retrieval" from memory. This notion of direct retrieval is based on its outside manifestations. A problem such as 4 + 5 is solved by adults in what seems to be one step. No discrete in-between steps are manifest even in the experience of the subject, who will state that the correct solutions just "came to mind" automatically.

Dividing thought into discrete steps is sensible when the focus is on complex problem solving, which is a multi-step activity by agreement or definition (cf. Van Lehn, 1989). The psychologist watches people as they struggle to bridge the gap between their initial knowledge and the solution prescribed by the task. Building up an adequate representation can take minutes or even hours (cf. Elshout, 1976; Simon, 1969, 1989). Subjects will mumble to themselves or talk aloud when asked to. The words they use are closely related to the mental steps taken to solve the problem. These words give a name and a content to what happens "inside", and producing them is not a serious problem in language-like tasks (see Elshout, 1976; Ericsson & Simon 1984; Hamel, 1990). Words thus have the appearance of being discrete mental entities, and at a certain level of analysis they certainly are. This does not mean, however, that words and numbers are really stored in some definite form, as terms like direct access and direct retrieval suggest. The difficulties inherent to such definite-form representations are very apparent in the size effects which are characteristically found in number processing tasks.

That number size is a good general predictor of problem difficulty is an undisputed fact (Ashcraft, 1992; Dehaene, 1992). Response times lengthen and errors increase as a function of the magnitude of the numbers involved. Though the correlation between problem size and problem difficulty is always imperfect (Campbell & Graham, 1985), it is usually strong and significant, for all types of problems (Ashcraft, 1992).

Such size effects are easy to understand when people calculate by counting. Given a counting procedure, 6 + 2 will be less difficult than 6 + 3 as a matter of course. But it is not so clear why 6 + 3 should also be more difficult for people who "directly retrieve" its solution from memory. When the thing to be produced is a symbol to start with, its processing load can only be represented external to its meaning. Larger numbers must then be made more heavy somehow, or stored in remote places, or inhibited by other pieces of knowledge.

Several such explanations have been proposed for number size effects. An early network model, by Ashcraft, suggested that people look up sums in mental tables (Ashcraft, 1982). Though such a representation (mental fingers moving downward and to the right from the relevant source numbers to the target number) gives a reasonable fit with the data, it was considered psychologically implausible. A different type of explanation was offered by Siegler, in his distribution-of-associations model (Siegler, 1988, 1989). It is a learning model, which starts by counting sums out. In the course of experience the associative connection between the problem and its solution becomes stronger. When the associative strength exceeds a certain threshold, direct retrieval replaces counting. The size effect is primarily caused by errors in counting, which are more frequent in the case of larger numbers. Such errors become part of the network, and interfere with the correct associations. The model handles different problems separately and does not intend to present a coherent picture of long-term knowledge of numbers. Such a coherent view is offered by Campbell, in his "encoding-complex perspective" on number processing (Campbell, 1987; Campbell & Clark, 1992). Campbell states that long-term representations of numbers consist of visual and auditory mental codes, which are more or less strongly interconnected. Mental processing differs with problem-format (e.g., words or digits) and may be facilitated or interfered with in various manners. Campbell does not believe in an "abstract mental representation" of numbers in terms of their formal properties, such as size. But according to McCloskey and his colleagues numbers must have an abstract representation in the mind, since calculation would otherwise be impossible (e.g., McCloskey, 1992; McCloskey, Macaruso & Whetstone, 1992; Sokol, McCloskey & Goodman-Schulman, 1992; McCloskey, Sokol & Goodman, 1986). The solution McCloskey offers is clear and simple. At the core of his "abstract modular" model is a module containing prescriptions sufficient to attain complete magnitude information. This "abstract semantic representation", which in McCloskey's view must be always accessed during calculation, has been the focus of much criticism. How can McCloskey assume that the mind contains an infinite number of formal specifications such as {5}10EXP3, {3}10EXP1, which in his model stands for 5030?

It certainly looks implausible. On the other hand, the difficulty this part of the model tries to solve is a real one. People do respond to calculating problems in terms of their formal, "abstract" demands. The representations that are used in calculation must incorporate certain formal properties to be correct.

These contradictory demands are difficult to meet within a "direct retrieval" framework. When content is conceived of as permanently present, everything that hinders its appearance must also be explicitly represented. We then see numbers fighting each other, numbers living in remote regions of the mind, numbers being either inhibited or activated. The less a person knows, the more complex the representation of that knowledge will become.

There are other possible viewpoints, however. Elshout and Milikowski have suggested that the size effects in number processing are, in reality, effects of "well-knownness" (see Elshout & Milikowski, 1995a, 1995b). Well-knownness, which for numbers is negatively correlated with size, has a qualitative and a quantitative aspect. The qualitative aspect is concerned with the "what". Thus, a memory node for "6" may potentially activate associations such as 2 and 3 (divisors), 36 (square), six (numberword), or even 666 (beast, Bible, Apocalypse). Whether these or other associations are in fact produced depends on two quantitative parameters. The first of these represents the long term strength of a memory item. Thus, "6" will have a greater permanent strength than has, for instance, 86 (see also Chapter 5). The second quantitative parameter, degree of activation, represents the momentary state of accessibility of a certain part of the network. Degree of activation, or accessibility, will be determined by the specific interactions of short term demands and purposes, with long term strengths and possibilities (see also Elshout, 1978). From this viewpoint, symbolic properties of numbers or words need not be thought of as permanently inhabiting the mind in a specified form to be adequately represented by it. The variability of mental concepts, which is stressed by Campbell as well as by Barsalou (1987, 1989, 1993), can be explained by it, without giving up on the endeavour to represent long term knowledge in terms of its content. Symbolic content is represented as a potential rather than as a given, but as such it has permanence. Nodes may exist for some numbers, but not for others. Which associations are in fact produced depends on inner as well as outer conditions.

Such a viewpoint distinguishes between long term memory structures, objective (outside) information, and the process of attaining representations which are adequate to the task. Tasks may thus be differentially sensitive to numbersize, as an example will illustrate. If the task demands counting, the size effect must be large. RTs can be expected to be a linear function of objective magnitude. If the task can be performed by recognition, however, size effects should be much smaller. The explanatory status of frequency will also be different in these two tasks. The influence of number-frequency on speed of number recognition is direct. But the time consumed by counting is determined by a number's serial position rather than by its frequency. The correlation between counting RTs and the target number's frequency is only indirect. Both variables can ultimately be traced to the serial position of that number.

When the representation of a number is constructed to fit a specific task, magnitude information need not always be at its core. A very different representation may be used for the assessment of a number's pleasantness than for a judgement about its size or oddness. When the task is calculation, the focus of a representation will presumably be on magnitudes. Even so, it may not be necessary for each number occurring in a problem to be fully elaborated, or represented, mentally. When the task is to solve the subtraction 44 - 43 a superficial assessment of the magnitudes involved may suffice to produce a correct answer. If this should be the case, 88 - 87 will not be significantly more difficult, even though the numbers 44 and 88 differ considerably in size. Changing the problem to 44 - 41, on the other hand, might make more of a difference, since the change now concerns the very number - an absent one - whose magnitude must be discovered.

The aim of the experiment to be reported in this chapter is to investigate the effects of three different number size manipulations within a set of subtraction problems. Most of the research into size effects in calculation has its focus on the basic combinations. These single-digit additions and multiplications are the problems most familiar to us. At the same time the size effects within this set suggest that they are not all equally familiar. Ashcraft has argued that the size effects observed may be largely due to differences in exposure and practice (Ashcraft, 1992). Text books used in primary school over-represent small-sized problems (Hamann & Ashcraft, 1986). This makes it difficult to separate two possible influences on RTs and errors. Exceptionally frequent problems have a better chance of being recognized as a single pattern. This might partly obscure the role of the individual numbers. For that reason, we have not used these basic combinations here. Instead, all combinations in the present problem set involve one one-digit number, and two two-digit numbers.

The reason to study subtraction rather than addition problems is that subtraction offers more attractive opportunities for the manipulation of magnitude. I shall briefly illustrate the advantage, referring to the Method section for a more extensive account.

A problem such as 72 - 9 can be contrasted with problems using different-sized numbers in three independent ways. One manipulation concerns the first and largest number. For example, 72 - 9 may be compared with 32 - 9, to determine the influence of the size of the first number. The second contrast concerns the smallest number involved in a problem. In both of the above examples the smallest number is 9. Sums involving different-sized smallest numbers may be compared, leaving first numbers as they are. Examples are 72 - 9, versus 72 - 5. The third contrast concerns response size. A problem involving the same three numbers can be presented in two different versions, e.g., 72 - 9 (= 63) and 72 - 63 (= 9). This manipulation leaves the first number unchanged. It also leaves the smallest number unchanged. However, in the first version this smallest number is given, while in the second it must be construed and named. Consequently, the size of the response number is systematically different for the two versions. It can be either small or large. Addition problems do not afford this third opportunity for independent manipulation. Compare the problems 63 + 9 (= 72) and 63 + 72 (= 135, not 9), which are different on more than one dimension.

Method

Design and variables

Number size was manipulated independently for three problem components, in a set of

144 subtraction problems. The range of these problems varied between 20 - ... to 99 - ...
The first variable was response size. For each of 72 problems, two versions were constructed. In the first version, the correct response was a small (one-digit) number. In the second it was a larger (two-digit) number. Examples of this manipulation are:
43 - 36 = ? (Answer: 7)
versus
43 - 7 = ? (Answer: 36)
The average size of a correct small response is 5. The average size of a correct large response is 55.
The two versions will be compared to determine whether problem difficulty is a function of response size.

The second variable was first number. In simple subtraction problems the size of the first number is half the sum of the three numbers in the problem. Of the set of 144 problems, half had first numbers between 20 and 59, while the other half had first numbers between 60 and 99. Thus, each set covered four decades. Each of these decades was represented by nine problems. The two sets were balanced for response size (large or small). Examples of this manipulation are
71 - 3 = ? and 71 - 68 = ?
versus
31 - 3 = ? and 31 - 28 = ?
The average size of the smaller first numbers was 40. The average size of the larger first numbers was 80.

The third variable is smallest number. The smallest number covers the difference between the two largest numbers involved in a problem. In all 144 problems the smallest number is a number between 1 and 9. It can either be the response number (as in 71 - 68), or the second given number (as in 71 - 3). All one-digit numbers 1 to 9 were used as smallest numbers equally often, that is once in each first number decade group. Thus, the variable was balanced both for first number and for response size. Three size levels of smallest number were compared, containing 48 problems each. The first level contains problems using 1, 2, and 3 as smallest numbers. The second level contains all problems using 4, 5, and 6. Problems with 7, 8, and 9 constitute the third level.

The fourth factor, which is unrelated to size, is borrowing. This factor was also manipulated independently. Thus, for each level of each size factor, half of the problems involve borrowing.

Subjects

Subjects were 37 first year psychology students fulfilling a course requirement.

Apparatus and stimuli

Problems were presented on the monitor screen of a Macintosh LC computer. The order of presentation was controlled by a randomizing program. Prior to the presentation of the problem, an asterisk appeared at the target spot. Response times were registered by means of a voice key connected to the computer.

Procedure

Subjects were instructed to respond both quickly and accurately. They were also informed about the time pressure built into the experiment. One second after its appearance a problem would disappear from the screen. Subjects were given twenty problems to practice. These problems were unrelated to the experimental set. Responses were typed in by the experimenter, who then made the next asterisk appear. No mechanical limit was set on the available time. After a block of ten problems a five second pause was introduced automatically.

Data treatment

For each problem, a mean RT (including those of non-responses and errors) was determined by averaging over subjects. Errors (incorrect answers) and omissions (indicated by subjects' statements such as "I don't know") were collapsed, for each number, into a single Error score.

Results

The mean response time of all 144 problems, averaged over subjects, was 594 msecs, with a standard deviation of 381 msecs. Mean RTs of individual problems range between 155 and 1739 msecs. The problem 94 - 4 had the shortest RT, and the problem 72 - 9 had the longest. Table 6.1 presents the ten slowest and the ten quickest solved problems. RTs and errors of all problems are given in Appendix 10.

Not all problems were correctly solved by all subjects. On one problem (72 - 9) fourteen subjects (38 percent) failed, either by mistake or by omission. In Table 6.1 the problems with the highest error scores are presented. Forty-five problems were correctly solved by all subjects. Thirty problems produced one failure only.

The correlation between RT and errors is .78.

Table 6.1

Ten problems with shortest RTs, ten with longest RTs and eleven problems with the highest number of errors.

Longest Rts		*Shortest RTs*		*Most errors*	
Problems	***msecs***	***Problems***	***msecs***	***Problem***	***%Errors***

72 - 9	1739	50 - 49	236	72 - 9	38%
72 - 63	1648	89 - 80	234	73 - 4	35%
97 - 88	1640	49 - 40	231	97 - 9	30%
82 - 74	1588	50 - 1	230	33 - 7	27%
91 - 5	1492	76 - 6	225	43 - 37	27%
97 - 9	1424	36 - 6	224	66 - 8	24%
33 - 7	1420	24 - 23	216	97 - 88	24%
53 - 7	1377	49 - 9	210	53 - 46	24%
73 - 66	1319	29 - 9	207	46 - 38	24%
43 - 37	1291	94 - 4	155	54 - 9	24%
				53 - 7	24%

The correlations between parallel problems with either large or small response numbers (e.g. 72 - 9 and 72 - 63) are r = .88 for RT and r = .55 for errors (including omissions).
Two four factor Anova's were performed, the first on RT and the second on errors (including omissions). The factors have been described in the Method section. They are response size (two levels), first number (two levels), smallest number (three levels), and borrowing (two levels). Mean values are given in Table 6.2.

Table 6.2

RTs and proportions of errors and omissions in 72 Non-borrowing and 72 Borrowing problems using different sized numbers.

	Non-Borrowing		***Borrowing***		***All problems***	
	RT	***% Errors***	***RT***	***% Errors***	***RT***	***% Errors***
Small Response	356	3%	801	10%	578	6%
Large Response	349	3%	871	11%	609	7%
1st number 20-59	332	3%	763	10%	547	6%
1st number 60-99	372	3%	909	11%	641	7%
Smallest number 1-3	320	3%	493	4%	407	3%
Smallest number 4-6	385	3%	841	9%	613	6%
Smallest number 7-9	351	3%	1174	19%	762	11%

Response times

For RT, three significant main effects were obtained. RT was influenced by borrowing, smallest number, and first number. Mean values are given in Table 6.2. The largest effect was produced by borrowing, $F(1, 143) = 177.06$, $p < .0001$. The mean RT for the 72 borrowing problems was 836 msecs. The mean RT for the 72 non-borrowing problems was 352 msecs.

The effect of smallest number was also highly significant, F (2, 143) = 32.22, p < .0001. The 48 problems with 1, 2, or 3 as smallest numbers had a mean RT of 407 msecs. The problems with 4, 5, or 6 had a mean RT of 613 msecs. Those with 7, 8, or 9 had a mean RT of 762 msecs. Each of these comparisons was statistically significant (p < .05).

The third significant main effect was obtained for first number, F (1,143) = 6.58, p < .05. Problems with first numbers between 60 and 99 took significantly longer to solve than problems with first numbers between 20 and 59. The mean RT for the group of 72 problems with large first numbers was 641 msecs. The mean RT of the other group was 547 msecs.

Response size had no significant influence on RT, F (1, 143) = 0.25, p > .10. Problems demanding large responses took 610 msecs on the average. Parallel problems demanding small responses had an average RT of 578 msecs.

A significant interaction was observed between borrowing and smallest number, F (2, 143) = 26.94, p < .0001. RT-differences associated with this latter variable were much larger for borrowing problems than for non-borrowing problems. For the borrowing problems, mean RTs were 493, 841, and 1174 msecs for small (1-3), medium (4-6), and large (7-9) numbers, respectively. In a separate analysis of the 72 borrowing problems the effect of smallest number is significant F (2, 71) = 32.78, p < .001. For the 72 non-borrowing problems the mean RTs are 320, 385, and 350 msecs, respectively. In this group the effect of smallest number is non-significant, F (2, 71) = 0.21, p > .10.

Errors and omissions.

The proportion of errors (including omissions) was significantly influenced by borrowing (F (1, 143) = 61.02, p < .0001), and smallest number (F (2, 143) = 20.08, P < .0001). Borrowing problems were associated with more mistakes and omissions than non-borrowing problems. The proportions were 11 percent and 3 percent, respectively.

Errors were also influenced by the variable smallest number. The proportions are 11 percent for large sized (7-9), 6 percent for medium sized (4-6) and 3 percent for small sized (1-3) smallest numbers. Only the two extremes differ significantly in a post hoc comparison (p < .05).

First Number Size and response size had no significant effects on Errors, F (1, 143) = 0.9, p > .10 and F (1, 143) = 0.59, p > .10, respectively.

The one significant interaction is, again, between borrowing and smallest number, F(2, 143) = 19.09, p < .0001. The pattern is similar to that obtained for RT. The size of the smallest number had a strong influence within the category of borrowing problems. Proportions errors and omissions are 19 percent, 9 percent, and 4 percent for large, medium, and small numbers, respectively. Each of the three comparisons is significant. For the non-borrowing problems, on the other hand, the size of the smallest number made no difference whatsoever. The proportion of omissions and errors is 3 percent in each of the three groups.

Size versus frequency

What is the better predictor of problem difficulty: size or frequency?

To answer this question two multiple regression analyses were performed on problem RT. In the first analysis the predictor variables were: 1. Borrowing; 2. Size of first number; 3. Size of smallest number.

In the second analysis the predictors were 1. Borrowing; 2. Frequency of first number; 3. Frequency of smallest number.

The results are given in Table 6.3. In the first analysis, using Size, the combined effect of the three variables is R = .78. The beta's are .65 for borrowing, .16 for size of the first number, and .41 for size of the smallest number. In the second analysis, using frequency, the combined effect of the three variables is R = .80. The beta weights are .72 for borrowing, –.32 for frequency of the first number, and –.33 for frequency of the smallest number.

By these analyses, the frequency of the first number is somewhat more relevant to the prediction of RT than its size. For smallest numbers the reverse is true. The size of the smallest number is a better predictor of performance than its frequency. The highest multiple correlation is attained by a combination of borrowing, frequency of the first number, and size of the smallest number. This combination gives an R of .81. (see Table 6.3).

Table 6.3.

Multiple R's and beta weights of different combinations of the variables Borrowing, Size and Frequency.

	R	***Borrowing***	***1. First number***	***2. Smallest number***
RT				
Both Size	.78	.65**	.16*	.41**
Both Frequency	.80	.72**	-.32**	-.33**
1. Frequency, 2. size	.81	.72**	-.30**	.35**
Errors				
Both Size	.62	.48**	.07 (n.s)	.39**
Both Frequency	.61	.53**	-.22*	-.28**
1. Frequency, 2. Size	.64	.52**	-.19*	.36**

* $p < .01$

** $p < .0001$

This particular mixture also gives the best description of the pattern of errors (including omissions). Its multiple R is .64, with beta's of .52, –.19, and .36 for borrowing, frequency of the first number, and size of the smallest number, respectively. A combination of borrowing with the two size variables, on the other hand, gives an R of .62. In this combination, the contribution of the first number is small (a beta coefficient of .07) and non-significant. Th beta weights while those of borrowing and size of the smallest number are .48 and .39, respectively. When frequency is introduced to replace size, the picture changes considerably. Frequency of the first number makes a significant contribution (–.22), while size of the first number did not). However, the frequency of the smallest number is less of a predictor than its size (-.28 vs .39), though both are significant.

Magnitude of RT-differences.

It is interesting to compare the three size manipulations for their effects on RT. The values are given in Table 6.2. The first comparison concerns response size. This variable was manipulated by changing the place of the smallest number (as in 72 - 9 = ? versus 72 - 63 = ?). The effects of this manipulation were insignificant, and they were also very small in terms of RT-differences. The category of problems with large response numbers had a mean RT of 610. The category with small response numbers had a mean RT of 578. The average size of the response numbers used in these two groups were 55 and 5 respectively, which gives an average RT-difference of less than one millisecond (32 / 50 = 0.64) per unit.

The largest size effect is produced by the variable smallest number. However, this effect is only present in the category of borrowing problems. Within this set, the mean RTs are 493, 841, and 1174 msecs for small(1-3), medium (4-6), and large (7-9) smallest numbers. Adding 3 to the smallest number in these problems thus produces a rise of over 300 milliseconds on RT.

The third manipulation of size concerned the variable first number. The average difference between the two levels (first numbers 20 - 59 versus first numbers 60 -99) is 40. The RTs are 547 and 641 msecs respectively, a difference of 94. However, borrowing makes a considerable difference on this variable also. For the non-borrowing problems the average RT-difference between larger and smaller first numbers is only 40 msecs, while for the borrowing problems it is 145 msecs.

Discussion

The results confirm that number size is not a unitary psychological variable even in calculation, where size is of evident importance. One interesting result is the absence of an effect of response size. The answers to problems such as 72 - 63 are not produced more quickly, or more accurately, than the answers to problems such as 72 - 9, which call for a larger-sized response number. That these problems are about equally difficult is also confirmed by the correlation of .88 between the RTs of the two sets of parallel problems. This finding is not so self-evident as its intuitive plausibility suggests. That 63 is not more difficult to retrieve when 72 - 9 is given than is 9 when 72 - 63 is given, at least suggests that the difficulty of problems within the present range does not critically depend on the magnitude of their outcome.

This result can only be understood in combination with the effect of smallest number. The effects of this variable are large, and do not depend on the place of the smallest number. The correlations between RT and smallest number size are almost equal for the two types of problems. For the problems demanding small number responses (e. g., 72 - 63) the correlation is .41. For the large number response problems it is .42. In the former set, the smallest number is indicated by the difference between two large numbers, which has yet to be named. In the latter, it is the second given number, which, of course, is named. It may therefore be concluded that the processes of construction, respectively decomposition, of a smallest number demand an equal amount of elaboration.

Another interesting point is that size differences in smallest numbers do not affect RT in problems that can be solved without borrowing. Solving such problems does not seem to demand much elaboration of the magnitudes involved. These problems are all solved very quickly and accurately. As can be seen from Table 6.2, first number makes hardly any difference, either, within this set. In fact, the distribution of errors is entirely flat, across all size distinctions. It is a further indication that number size is not a critical variable under all conditions. Size effects seem to depend on the demands set by a problem. Sometimes a superficial assessment of magnitudes suffices. At other times, these same magnitudes must be elaborated more extensively. This is when size comes in as a psychological factor. It is evident from these data that borrowing problems take considerable elaboration of the smallest number involved, while non-borrowing problems do not. In borrowing problems, it seems, the distance between the two largest numbers is bridged by a mental procedure which resembles counting.

Whether size effects are "really" effects of frequency (see Ashcraft, 1992; Dehaene,1992; Dehaene & Mehler, 1992), depends on the task as well as on the position of the number concerned. In this experiment, first number size is less of a predictor than is first number frequency. The effects produced by the smallest number, in contrast, are better described by its size. This, again, can be explained by a more superficial as opposed to a more elaborate representation of the number concerned.

The "abstract representation of magnitude", as it is realized in the model of McCloskey and his co-workers, cannot represent or explain these differential effects of size. In McCloskey's model, the magnitude information contained in a number's representation is always the same and unaffected by context. Even if such representations could exist, they have little to contribute, apparently, to the prediction of size-related processing times.

There is a considerable difference in RT and errors between the easiest and the most difficult problems in this set. The easiest problems are solved within 200 milliseconds, while the most difficult ones take over 1500 milliseconds to solve. Limiting exposure (subjects viewed the problem for one second only) may have contributed to this spread, as it was meant to do. It is interesting to compare the RTs of this experiment with the association RTs in Chapter 3. In the association experiments the average RT was approximately 1300 milliseconds. In this experiment the average RT is 600 milliseconds, approximately. While most calculating problems were solved within 1000 milliseconds, RTs in the association task were all longer than 1000 milliseconds. In fact, the shortest RT in Experiment 1 of Chapter 3 is 1007 msecs. (This is the RT of stimulus number 99, with 100 as its primary response.)

Similar RT-differences can be observed when word association is compared with word translation. De Groot (1989) reports average latencies of around 1500 msecs for discrete word association. Word translation is performed more quickly, on the average. De Groot, Dannenburg, and Van Hell (1994) obtained RTs between 850 and 1500 for different groups of words. Interestingly, each trial in the calculation task involved three numbers, whereas in the association task it involved two. Also, the combined magnitudes of the numbers in the calculation problems are much larger on the average than those in the association task. That calculation RTs are nonetheless considerably shorter is a further indication that different representations of the same numbers are used during the performance of different tasks. On the average, the association task seems to have called for a more elaborate representation of numbers than the calculation task.

Main findings

¶ The difficulty of subtraction problems is most strongly influenced by the size of the smallest number. It makes no difference if this smallest number was given in the problem, or must be retrieved as its solution. The size of the first number (which in subtraction problems is also the largest), affected RT and errors to a lesser degree, while response size completely failed to discriminate between problems involving the same three numbers.

¶ For first numbers, frequency is a stronger predictor of difficulty than size. RTs and errors were best described by a combination of smallest number size, and first number frequency.

¶ Large linear effects of objective size of the smallest number were obtained for borrowing problems only. In these problems, RTs and errors correspond to the distance between the two largest numbers in a problem. This indicates that solutions to these problems were constructed by some form of counting.