Storage and retrieval of addition facts: The role of number comparison

Brian Butterworth

University College London, London, UK

Marco Zorzi

University of San Raffaele, Milan, Italy

Luisa Girelli

University of Trieste, Trieste, Italy

A.R. Jonckheere

University College London, London, UK

It is proposed that arithmetical facts are organized in memory in terms of a principle that is unique to numbers—the cardinal magnitudes of the addends. This implies that sums such as 4 + 2 and 2 + 4 are represented, and searched for, in terms of the maximum and minimum addends. This in turn implies that a critical stage in solving an addition problem is deciding which addend is the larger. The COMP model of addition fact retrieval incorporates a comparison stage, as well as a retrieval stage and a pronunciation stage. Three tasks, using the same subjects, were designed to assess the contribution of these three stages to retrieving the answers to single-digit addition problems. Task 3 was the addition task, which examined whether reaction times (RTs) were explained by the model; Task 1 was a number naming task to assess the contribution of the pronunciation stage; Task 2 was a magnitude comparison task to assess the contribution, if any, of the comparison stage. A regression equation that included just expressions of these three stages was found to account for 71% of the variance. It is argued that the COMP model fits not only the adult RT data better than do alternatives, but also the evidence from development of additional skills.

The basic phenomena involved in single-digit addition performance are robust, widely replicated and well known, yet there has been much controversy as to the psychological processes involved. It is generally agreed that competent adults use some mixture of memory retrieval and procedures, but there is little agreement as to how the addition facts are represented and

---

---

organized in memory or on the related issue of how retrieval plans are set up. One important issue is whether our memory contains both members of a commuted pair. That is, does memory contain both the fact that $5 + 3 = 8$ and the fact that $3 + 5 = 8$, or is one of these forms privileged in some way?

Groen and Parkman (1972) proposed that children generally, and adults at least sometimes, count on from the larger of the two addends: They represent the larger number (henceforth, the "max") and count the smaller number, the "min". This model treats commuted pairs as equivalent, as to solve both $3 + 5$ and $5 + 3$ the count is 3 on to the 5. Adults are held to use retrieval from memory for most solutions to single-digit problems, and Groen and Parkman suggested that retrieval times were more or less the same whatever the problem. However, for a proportion of the problems, the solution will not be retrievable in an acceptable time, and the counting on procedure is applied. It is this proportion that yields the differences in RTs among problems.

The observation that difficulty of simple addition increases with numerical size is "one of the most extensively researched phenomena of cognitive arithmetic" (Campbell, 1995, p. 123). This "problem-size effect" was first explained by Groen and Parkman (1972). For them, it was a by-product of increasing the min—for example, $5 + 3 = 8, 5 + 4 = 9$. So addition reaction times (henceforth, ARTs) will differ among problems with same sum but different mins, but will be the same for different sums with the same min ($5 + 3 = 8, 6 + 3 = 9, 7 + 3 = 10$, etc.). One crucial exception to the problem-size effect, and one of which they were well aware, was the ties effect. In this, the answers to additions such as $3 + 3$ or $9 + 9$ are found more quickly and more accurately than other problems with the same sum. For Groen and Parkman this effect arises because tie problems are almost always retrieved from memory, and so min counting is rarely used. The 0-problems were tested but not analysed separately. Presumably, 0-problems are fast because the min requires no counting at all.

In another set of proposals, addition facts are held to be stored in a two-dimensional matrix with integer coordinates $0 \rightarrow 9 \times 0 \rightarrow 9$. Here magnitude is implicit, though only well-ordering is strictly entailed (Ashcraft & Battaglia, 1978; Widaman, Geary, Cormier, & Little, 1989).

Retrieval in these proposals consists of activating the first and the second addends, regarded as the x and y coordinates of a matrix, which jointly activate their sum, and thus both of the commuted pair are stored. In these accounts, retrieval time itself depends on the size of both addends. Typically they postulate a table with addends from 0 to 9 as coordinates ("entry nodes"), and their sums in the cells. In one version (Ashcraft & Battaglia, 1978), to find the sum of $5 + 3$ entails moving two pointers from the origin to the 5 and 3 coordinates, such that the time will be a function of the sum of the pointer movements. Ashcraft and Battaglia (1978) found that the best fit to the ARTs was not a linear function of the addends, but the square of their sum, ART = 906 ms + 1.15 (sum squared), which accounted for 48% of the variance. Notice, however, that this means that 75% to more than 90% of the ART consists of the unexplained constant, 906 ms. (A further problem is that the regression equation excludes ties.) Sum squared suggested that "the entry nodes are not evenly spaced, but rather become more widely separated as the numbers increase" (Ashcraft & Battaglia, p. 537). Widaman et al. (1989) point out that strictly this suggestion should be supported not by a fit to sum squared, that is $(a+b)^2 = (5+3)^2 = 64$, but by a fit to the sum of the squares of the addends, that is, $a^2 + b^2 = 5^2 + 3^2 = 34$, by no means the same thing.

However, Widaman et al. (1989) found that the best predictor of addition verification times turned out to be the *product* of the addends, which led them to propose a table look-up model were equally spaced coordinates defined a search space of 5×3 cells, and ART reflected exhaustive search of this space, starting at the 0,0 origin. However, ties do not fit their model, as the ARTs for ties, and for the product of the tie addends, do not show a problem-size effect (their Table 2). In their general regression equation for the whole data set, they arbitrarily set the value of the product of ties to zero. They offer no explanation of how tie facts are stored or retrieved, and 0-problems are not explicitly discussed. From the regression equation in their Table 2, the addition verification time $= 926 + 7.91$ (product of addends). As the product of N and 0 would be 0, this reduces for all 0-problems to 926 ms.

A somewhat different approach has been taken by Siegler, who argued that the memory representation of arithmetic facts contains both correct and incorrect answers. Correct answers will tend to have stronger associations with problems, and the relative strength of connections to the correct answer as compared with those to incorrect answers will increase with experience. ARTs will be a function of strategy selection and number of memory searches for the retrieval strategy until a pre-set criterion of confidence in the solution is reached. These will depend on how "peaked" the distribution is about the correct answer. Flat distributions will fail to reach criterion, and further searches will be needed (Siegler & Shrager, 1984)

One issue is whether there is something special about arithmetical facts. Certainly, they appear to be more orderly than facts in other domains. However, do their storage and retrieval simply reflect general laws of associative learning? Ashcraft (1992) writes, "The memory representation is organized and structured in terms of the strength off individual connections, and reflects varying degrees of relatedness among the elements; and the strength with which the elements are stored, hence the probability or speed of retrieving information, *depends critically on experience, especially acquisition, rather than on numerical characteristics inherent in the information itself* [italics added]" (p. 85).

In Ashcraft's (1992) network retrieval model, retrieval of single-digit addition answers is achieved by spread of activation through previously established connections. "The strength with which nodes were stored and interconnected was a function of frequency of occurrence and practice . . . associative strength between problems and correct answers . . . was corroborated by data taken from elementary school textbooks, which showed problem size and difficulty to be a direct function of both order and frequency of presentation" (p. 86). In support, Ashcraft and Christy (1995) note specifically that frequency and problem-size are confounded in that "Large facts are practised less frequently than small facts" (p. 414) and conclude that they are therefore "probably learned to a lower degree of mastery" (p. 414). By college age, something like 15–25% of the variance in ARTs was accounted for by their frequency in school textbooks (Ashcraft & Christy, 1995, Table 2). It is worth noting that for a given sum, the fastest ARTs are for $N + 0$ and $N + N$ problems, which are by no means the most frequent (Ashcraft & Christy, 1995; Hamann & Ashcraft, 1986).

One version of fact retrieval accounts claims that arithmetical facts are stored only in verbal form (Dehaene & Cohen, 1995). This model does not exclude the possibility that problems may be manipulated—"semantically elaborated"—prior to a retrieval plan being formulated. Presentation of problems in arabic numerals would thus require recoding into verbal codes so that they can be matched to the storage code of the facts.

Campbell (1995) has modelled fact retrieval in a network of associations among the various "codes" in which numbers can be represented—spoken words, written words, numerals, and magnitudes. As in the previous models, answers are retrieved through spread of activation through the network. There are connections from the operand nodes and also from internal representations of whole problems to number nodes. A problem activates not only the correct answer but many other answers that could interfere with its production—and do so to produce characteristic errors. "Retrieval in the [simulation of the] model involves a series of processing cycles. . . . On each cycle, each node receives excitatory input determined both by physical-code similarity and magnitude similarity to the presented problem. The excitatory input is modulated by inhibitory input that is proportional to the total activation associated with all currently activated nodes. Across cycles, the strength of the excitatory and counteracting inhibitory inputs gradually approach equilibrium, and a response is generated when one of the nodes in the network reaches a critical threshold level of activation. The excitatory input to the correct node is generally the same for all problems, so the differences in retrieval difficulty arise primarily from differences in inhibition due to the activation of incorrect nodes" (p. 122).

Campbell (1995) argues that "the problem-size effect arises . . . because larger-number problems are more similar in magnitude to their neighbours than are smaller-number problems (i.e., because the magnitude scale is relatively compressed for larger compared to smaller numbers). This causes larger problems to activate neighbours more strongly: Consequently, larger-number problems encounter more interference by way of inhibition from neighbours than do smaller-number problems. This slows the rate of activation of the correct for larger-number problems and makes them more susceptible to retrieval errors" (p. 125).

In this model ties form a subcategory of problems (errors likely to be answers to other ties), and within-category activation is stronger, and hence there will be less interference to slow ARTs. In the simulation (Campbell, 1995), however, the ties effect is achieved by manipulating the "similarity" variable, so that ties are 5% less similar to nontie problems than they would have been on the basis of their physical and magnitude features.

Overall, Campbell's (1995) model RTs for the 64 additions (over digits 2–9) are sensitive to sum and to ties, and they give a qualitative fit to experimental ARTs, but he does not present a regression equation fitting the model RTs to the experimental results. A limitation of the simulation as presented in Campbell is that it models only addition involving the digits 2–9. Campbell takes the view that 0-problems and 1-problems are probably solved by rule and therefore fall outside the scope of the network interference model of fact retrieval.

The domain of arithmetical facts is unique in that these facts can all be derived from calculation as well as being retrieved from memory. (One cannot work out the name of the capital of Guinea Bissau if you do not know it.) The mixture of retrieval and other calculation processes has been advanced by LeFevre, Sadesky, and Bisanz (1996) as an explanation of the problem-size effect. In their account, both the problem-size and ties effects are largely a matter of the proportion of slower non-retrieval methods used to find the answers: Large problems are more likely to elicit transformations (e.g., turning $9 + 6$ into $10 + 5$), whereas ties are almost always retrieved from memory.

Although there is agreement on the phenomena of skilled addition, there is little consensus as to the best way to model the underlying processes as the competing models all fit the data reasonably well. Another way to approach this problem is to start by considering how children lay down these facts and to see if this can help resolve the theoretical impasse.

Children above the age of 6 years seem to understand the commutativity of addition. At roughly the same time, children begin *counting on* from the larger number, rather than *counting all* from the first number presented (see Baroody & Ginsburg, 1986). The change to counting on may be based on an understanding of the commutativity principle (Groen & Resnick, 1977; Resnick & Ford, 1981), or it may be based on the more limited idea that the order in which the sets are taken is irrelevant (perhaps in the same way that children come to understand that the order in which the members of a set is counted does not matter (Gelman & Gallistel's, 1978, "order-irrelevancy principle"). Baroody and his colleagues have argued that counting on from the larger number develops as a short-cut or trick to reduce cognitive load: There will be fewer counting steps to keep track of if they start with the larger number (Baroody & Gannon, 1984). They have found that children have invented the practice of counting on from the larger number without apparently an understanding of commutativity; and those who do seem to understand it may still count all (see Baroody & Gannon, 1984, Table 4). More recently, he has argued that counting on may develop through generalizing from the more limited idea of counting 1 from N, in N+1 and 1+N problems (Baroody, 1995). However, Siegler and his colleagues (Siegler & Jenkins, 1989; Siegler & Shrager, 1984) have found that counting on from the larger number (his "min strategy") can first be shown for sums where neither number is 1. Children can also learn counting from explicit teaching (Secada, Fuson, & Hall, 1983). Counting on presupposes that the child first compares the two addends and selects the larger to count on from. This suggests that comparing the magnitudes of two numbers is part of the earliest stages of learning to add.

Children who are counting on to solve a particular problem do not, at that developmental moment, achieve the answer by reference to facts stored in memory. A number of studies suggest that counting to achieve the sum comes to be replaced as the strategy of choice by memory retrieval (Siegler, 1987). Siegler and his colleagues (Siegler & Jenkins, 1989; Siegler & Shrager, 1984) have found that both retrieval and counting strategies will co-exist for a period in the child's development, usually with retrieval being used for the smallest sums and counting for the larger ones, with counting the larger sums eventually being replaced by retrieval. This transition from a counting procedure to memory retrieval has been simulated in adults using "alphabet arithmetic". In these studies, subjects are presented with problems such as A + 2 = ?, G + 4 = ?. Initially, they count from the letter to the solution (C, K), but after some training they are able to recall the solution without counting (Logan & Klapp, 1991).

As experience of addition increases, counting on from the larger addend could serve as the basis of the organization of facts in memory. If this is so, then the privileged form is likely to be 5 + 3 = 8 rather than 3 + 5 = 8. The organization of addition facts then would have the following two properties at least: Organization by cardinal magnitude and storage of just one of the commuted pairs. It also implies that the process of solving a simple addition problem involves comparing the two addends to determine the larger and transforming, if necessary, a smaller–larger sum into a larger–smaller sum. (In what follows, the larger number in a sum will be called the "max" and the smaller as "min", with ties, by convention, as "max + max").

It has been known since Moyer and Landauer's (1967) seminal paper, that the time taken to select the larger is closely related to the difference between the numbers. This has been explicitly related to other kinds of "symbolic distance" effects, and the "split effect", as it has come to be known, is highly reliable across a wide range of paradigms. If the process of comparing

addends contributes to ARTs, it would fit well with what is known about the time taken to make the comparison.

In fact, a comparison stage fits well with two other well-known addition time phenomena. The first is "ties effect", which is another exception to the problem-size effect. Problems such as $7 + 7$ and $4 + 4$ are solved more quickly than non-ties with the same sum ($9 + 5, 5 + 3$) and in general do not show a significant problem-size effect (see, for example, data in Ashcraft & Battaglia, 1978; Groen & Parkman, 1972; Miller, Perlmutter, & Keating, 1984, Figure 1; Widaman et al., 1989, Table 2). Many explanations have been offered for this effect. Groen and Parkman (1972) as well as LeFevre et al. (1996) suggest that answers to tie problems are almost always retrieved from memory, whereas other sums are often computed using counting or some other strategy. Our suggestion is that the comparison stage is particularly quick as both numbers are the same.

The second are 0-problems ($N + 0, 0 + N$), which are solved more quickly that other problems with sum N (see Results, Table 3). In a verification task, Groen and Parkman (1972) found that these problems were solved fastest in children (their Figure 2) and adults (their Figure 4). Ashcraft and Battaglia (1978), Miller et al. (1984), and Widaman et al. (1989) used an experimental design incorporating all the 100 problems from combing the digits 0–9, but did not report the data from 0-problems separately. It has been argued that 0-problems are solved by using a "rule" (Baroody, 1983). However, when subjects are asked to say how they solved 0-problems after they had produced the answer, 12% said they used a "zero-rule", whereas the remaining 88% said they just knew the answer and were presumed to have retrieved them (LeFevre et al., 1996). For these problems, a comparison stage could be especially fast as the presence of the zero allows a short-cut to selecting the larger.

It is possible that 1-problems may also form a special subset, as Baroody (1983, 1995) has argued from the developmental data, but this still has to be established that it is the case for adults.

THE COMP MODEL

We wish to propose here a new model of the retrieval of single-digit addition facts—the COMP model. It is based on the hypothesis that children build addition fact memory on a max + min organization, with no separate representation for the min + max commuted version.

There is no independent evidence for this hypothesis. Indeed, the a priori presumption must be the opposite: namely that the child will store in memory every addition problem with its solution, or at least the child's best attempt at a solution. (A variant of the hypothesis, with possibly the same effect on the long-term representation of facts in memory, is that members of the commuted pair are stored according to the child's experience, but the max + min version comes to be privileged, due to greater use, or for some other reason.)

The hypothesis entails that the organizing principle of arithmetical fact memory is the domain-specific property of numerical magnitude. It further implies at a processing level that the magnitude of the addends must be compared. The model will be tested against the performance of adults, in particular the distribution of ARTs.

Figure 1 shows the proposed stages in the addition process. We assume that each stage makes an independent contribution to ARTs. In the first stage, the numbers are identified. Generally, this should apply whatever the input format of the numbers is—arabic numerals visually presented, number words visually presented, or number words auditorily presented. In this study, only visually presented arabic numerals will be considered. We assume without argument that identification means that the identity of the number is established, abstracting away from the font, size, colour, and so on of the stimulus, in a manner analogous to the
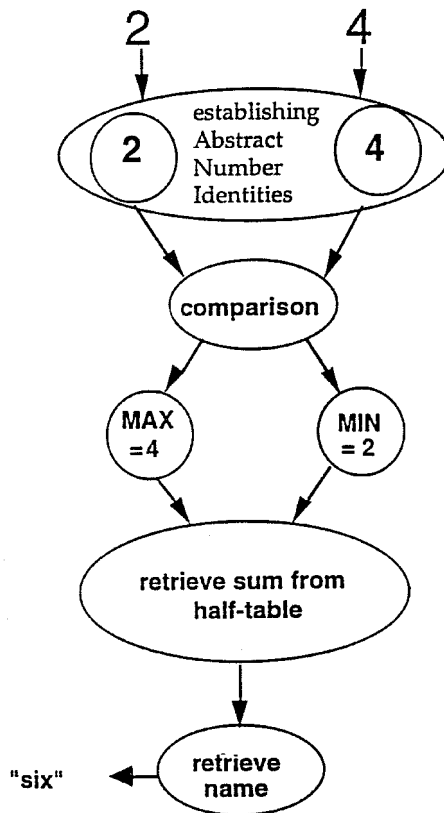

Figure 1.   The COMP model. The organization of the processes is given by the ovals. The first stage establishes the identity of the numbers in the problem. It is assumed that this process takes a small fixed amount of time. The second, comparison stage finds the max (larger) and the min (smaller) of the two numbers (ties are treated, by convention, as max + max). The retrieval stage uses the max and min to access stored addition facts. The pronunciation stages takes an abstract representation of the retrieved sum and retrieves the task-appropriate output form, in this case the spoken number name.

   Addition facts are stored in a network that connects a sum with representations of the magnitudes of max and min in terms of the number of input units activated in separate sets of units for max and for min (see text for further details).

establishment of abstract letter identities (ALIs) in reading (McClelland & Rumelhart, 1981), to yield abstract number identities (ANIs). Unlike ALIs each number can have a meaning.[1]

There is evidence that cardinal magnitude information is automatically activated even when it is irrelevant to the task. Physical identity between numerals (Duncan & McFarland, 1980), physical size comparisons (Foltz, Poltrock, & Potts, 1984), and parity judgements (Dehaene, Bossini, & Giraux, 1993), all show the effects of the cardinal magnitude of the stimulus numbers. This suggests that cardinal magnitude will be activated automatically. We assume therefore that the outputs from the identification stage are cardinal magnitudes.

There are no published estimates of the time taken to establish the identity of the numbers—ANIs. In the studies described later, we simply assume that this process takes the same small fixed amount of time for all the single-digit numbers for each of the three tasks—naming, comparison, and addition—even though task requirements are obviously rather different. We also assume that the identity of the arithmetical operation is established on the basis of the stimulus "+".

The comparison stage takes the magnitudes of the addends as input and compares them.[2] The larger addend is assigned the designation max(n) and the smaller the designation min(n). For example, for the sum "3 + 5" the comparison stage outputs max(5) + min(3) and will produce the same output for "5 + 3". For the sum "8 + 5", it will produce the output max(8) + min(5). It is assumed that ties such as "5 + 5" produce an output of the form max(5) + max(5). The time it takes to carry out the comparison as part of the addition task should reflect the time it takes in a pure comparison task. That is to say, it should reflect the numerical difference (Moyer & Landauer, 1967). For ties, we assume that subjects can make something like a physical match judgement (Posner, 1978) and thereby circumvent the magnitude comparison process. However, for 0-problems, we do not know whether subjects will be making a genuine, albeit quick and easy, comparison, or whether they will use some other strategy, in advance of the experiment revealing a split effect or not.

One potential complication arises because explicit comparison tasks, such as those requiring subjects to name the larger of two numbers, may not be a completely satisfactory model of the implicit comparison process we assume is involved in addition.

The sum retrieval stage will use max and min to locate the relevant sum. As only one of the commuted pairs is stored in memory—max + min version—there will be a half table rather than a whole table.

Once the sum has been retrieved, its output form, appropriate to the task in hand, will be produced. In the case of the experiments here, this will be the spoken number name. This

---

[1] Indeed several meanings. For example, the number 5 can have a cardinal meaning when it refers to the number of things in a collection, it can have an ordinal meaning when it refers to the position of an entity in a sequence, it can have nominal meaning as part of a telephone number or a product designation (as in "Renault 5"), and so on (Delazer & Butterworth, 1997; Fuson, 1988). In the context of arithmetic, numbers are usually taken to have cardinal meanings (Giaquinto, 1995). McCloskey and his colleagues (McCloskey, 1992; McCloskey, Caramazza, & Basili, 1985) have argued on the basis of studies of acalculic patients that the activation of numerical meaning, presumably cardinal meaning, is a necessary stage from input to calculation in adults.

[2] An alternative to the assumption that magnitudes, automatically activated by identification, are the input to comparison is to assume that ANIs are the input and that the comparison process itself intentionally activates the magnitudes of the numbers. The difference between these positions is not critical to the claims being made in this paper, though clearly it is an important issue for further research.

pronunciation stage is assumed to take an abstract form of the solution as input and to produce the spoken number as output. Factors that may contribute to the time taken to do this include number magnitude, the frequency of the number name in the language (Dehaene & Mehler, 1992), and the acoustic–phonetic properties of the name, especially the initial sounds (first consonant and vowel), as these are known to affect word-naming latencies in general (Treiman, Mullennix, Bjeljac-Babic, & Richmond-Welty, 1995). It should be noted that variance due to initial sounds may not be entirely psychological in origin, but may also reflect the response characteristics of the voice key, which will trigger when a threshold amplitude has been reached. Some speech sounds take longer to reach the maximum than others. In particular, vowels contain most energy, and the longer the vowel is delayed the longer the key will take to trigger. The high-pitched noise bursts for the sibilant "s" in "six" and "seven" contain less energy than the vowel that is delayed by them.

We make the strong, testable, hypothesis that, in the COMP model, each stage will make an independent contribution to the total ART.

Outline of the study

The three tasks are designed to test the COMP model: The third task yields the ARTs for all single-digit additions from 0 through 9. This production task in which subjects uttered the answer was preferred to a verification task, as recent studies indicate that arithmetic verification can be accomplished by a familiarity or plausibility strategy that does not require retrieval of specific facts (Campbell & Tarling, 1996; Lemaire & Fayol, 1995; Zbrodoff & Logan, 1990). The second task yields comparison times for all these pairs of numbers, presented in identical form to the additions and requiring the name of the larger number. The first task is a simple number-naming task for all numbers 0 through 18. The same subjects participate in the three experiments, which has the advantage of enabling us to make better estimates of the contribution of the independent naming and comparison estimates to ARTs. This design is similar to that employed by Miller et al. (1984), though we put the data to rather different use.

We hypothesize that the ARTs will be a simple function of the two numbers plus the time it takes to compare them and the time it takes to name the sum. If 0-problems are not solved by reference to memorized sums, but by a procedure in which the subject simply reads N, for example, then our naming experiment will provide data for comparing naming times with ARTs for these problems. If 1-problems are not solved by sum retrieval, but some other procedure, then their ARTs should not be affected by the factors that affect other problems, such as problem-size or comparison time.

Method

Subjects

A total of 20 undergraduate and postgraduate students (10 female and 10 male) from University College London participated in this experiment. The students' median age was 26.5 years with a range of 20 to 34 years. Each subject was paid a small fee for participating. All participants reported normal or corrected-to-normal vision. The same 20 subjects completed the three tasks.

Procedure

The three experiments were designed and carried out on a Macintosh IIsi running PsychLab 0.85 (Gum & Bub, 1988). For all three experiments, each trial began with a start fixation point lasting 500 ms followed 100 ms later by a stimulus. The subject's vocal response activated a voice key connected to the computer so that latencies were automatically recorded while the vocal responses were tape-recorded. The stimulus stayed on the screen until the subject's response was given. The interval between the response and the beginning of the next trial was 1500 ms.

Each task began with ten practice trials; these problems were randomly selected from the stimuli included in the second half of the list to be shown. A short break was allowed after 50 trials.

In all three tasks, where subjects made an irrelevant noise (e.g., "um") or a self-correction, the trial was disregarded in the RT analysis. Analyses of RTs were carried out on correct trials, though errors were noted.

Number naming task

The purpose of the number-naming task was to assess the contribution made by the act of uttering the number names to the comparison reaction times (CRTs) in Task 2 and the ARTs in Task 3. Of course, reading a numeral aloud will not elicit exactly the same processes as generating the number from a comparison or from a mental calculation. Nevertheless, both of these last two tasks will share the processes of reading numbers and uttering number names. Systematic differences, if any, in the time taken to name numbers must be taken into account in interpreting patterns of CRTs and ARTs. It should be noted that these differences might not be due to differences in the psychological and physiological processes involved in number name utterance, but simply to the sensitivities of the apparatus to the phonetic/acoustic properties of the names.

The stimuli consisted of four blocks of 19 trials, each containing the arabic numerals ranging from 0 to 18. Stimuli were displayed using black characters (font: Geneva; 25 point) in the centre of a white window of 5 × 5 cm, with subjects seated approximately 75 cm from the screen. Subjects were asked to name the numbers as quickly and accurately as they could. They were asked to utter a single name and to avoid filled hesitations (e.g., "um") and self-corrections.

Number magnitude comparison task

The purpose of this task was to assess whether a comparison process was involved in finding addition sums by seeing whether CRTs were reliable predictors of ARTs. For this reason, the comparison task used exactly the same stimulus format as the addition task. It is possible that this will prompt subjects to add the two numbers unconsciously, as well as to compare them as the task demands. However, even without the addition sign, subjects may still add two numbers in a task-irrelevant way, as shown by LeFevre, Bisanz, and Mrkonjic (1988).

Each subject was twice presented a list of the 100 problems in two different pseudo-random orders (A and B). Problems were ordered with the constraints that no addend or sum was repeated on consecutive trials; moreover problem complements (e.g., 2 + 4 and 4 + 2) did not appear in the same half of the list. Problems were displayed horizontally using black characters (font: Geneva; 25 point) centred on a 7 × 7-cm white window on a computer screen. The addends were next to the sign without an intervening space.

Subjects were asked to disregard the addition sign and simply to name the larger of the two numbers as quickly and accurately as they could. For ties (e.g., 4 + 4), they were asked to name the repeated number (e.g., "four"). Subjects were asked to give a single answer and to avoid filled hesitations (e.g., "um") and self-corrections.

RT analyses were carried out on the mean of each subject's two correct responses to all 100 problems. Three factors were predicted to determine RTs: the absolute difference between the numbers ("split"); the time taken to identify the larger (i.e., its position in the pair); and the time taken to name the (larger) number (derived from Task 1).

Addition task

ARTs were the dependent variable of primary theoretical interest. The stimuli were identical to those used in the number comparison task and were shown in the same order. Thus, each subject was presented with each problem twice. Subjects were asked to produce the sum of the two numbers as quickly and as accurately as they could.

Half of the subjects received the arithmetical task first whereas the other half received the magnitude comparison task first. Instructions in the tasks emphasized both speed and accuracy. Subjects were asked to give a single answer and to avoid filled hesitations (e.g., "um") and self-corrections.

Analyses were carried out on the mean of each subject's two responses to all 100 problems.

Statistical analysis

For the comparison and the addition tasks, the basic datum was the mean RT of each subject's two correct responses. If the subject failed to produce a correct response, then the remaining RT was taken as the mean. All the statistical analyses are based on the medians over 20 subjects of these basic data, unless otherwise noted in the text.

Results

Task 1. Naming

Figure 2 shows the naming RTs for the numbers from 1 to 18. Note that the RT for 0 was high because of the ambiguity and low frequencies of its names (nought or zero). It has been excluded from further analyses.
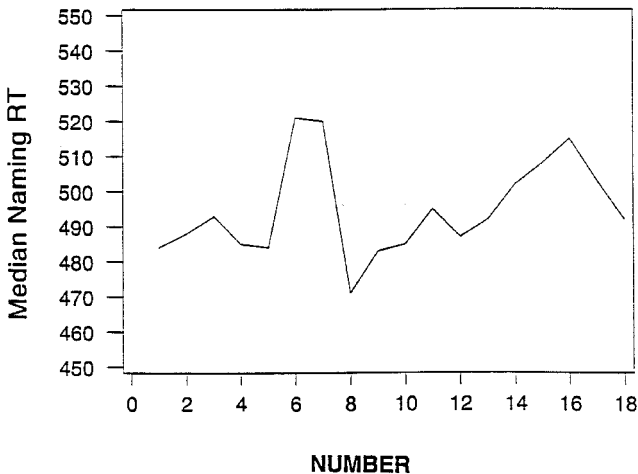

Figure 2. The median naming RTs of 20 subjects.

Analysis of variance revealed that there was a significant difference among naming RTs, $F(17, 323) = 4.596, p < .0001, MSE = 4289.1$.

Figure 2 suggested that the slowest numbers apart from 0 were 6, 7, 15, 16, and 17. The initial sound of these numbers is a fricative, followed by a tense vowel. Multiple regression on the naming RTs (excluding 0) using magnitude, fricative (+/−) and tense vowel (+/−) as predictors, showed that significant contributions were made by fricative ($p = .01$) and tense vowel ($p = .03$), but not by magnitude ($p > .8$). (The best fitting regression equation for naming RTs was: naming RT = 480 + 13.8 fricative + 12.5 tense. This accounted for 60% of the variance.)

Discussion of the naming task

The purpose of the naming task was to use the naming RT to estimate the contribution to CRTs, and especially ARTs, of utterance generation time and/or differential apparatus sensitivity. The values of the naming times are entered into the regression equations for CRTs and ARTs below.

The two variables that made a measurable difference to naming times were the initial sound and the tenseness of the first vowel of the number name. Word-naming studies also have found significant effects of the initial sound, especially where it is a fricative (Treiman et al., 1995). No effect of numerical magnitude was observed. Notice that magnitude is highly correlated with frequency of the number names in the language (Dehaene & Mehler, 1992).

Task 2. Magnitude comparison

The error rate in the 4000 trials was very low ($n = 26$; <1%); 33 responses (<1%) were invalid due to irrelevant noises, self-corrections, or equipment failures. Table 1 presents the median correct CRTs, pooled over subjects, for each pair of numbers to be compared.

TABLE 1
Median reaction times over 20 subjects for the magnitude comparison task
(Experiment 2)

Left side numbers	Right side numbers									
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	654	564	580	551	546	546	603	555	532	536
1	529	529	586	591	587	548	575	573	570	562
2	552	577	563	605	642	585	617	581	557	567
3	549	592	563	573	661	599	616	622	564	538
4	542	564	593	609	572	616	640	634	595	552
5	538	537	573	539	603	557	637	643	582	582
6	560	576	566	602	613	627	607	643	596	590
7	573	576	579	598	572	612	712	589	614	577
8	515	549	514	551	534	542	614	652	572	563
9	514	499	519	541	552	561	574	646	607	543

The following analyses excluded the 0 + 0 problem, as the naming experiment showed that naming 0 was an outlier in the distribution of naming RTs because of the ambiguity in its name ("nought" or "zero", which are both very uncommon words in English). Median latencies for correct responses only were submitted to a regression analysis where split (the absolute difference between the numbers) and naming RTs were used as predictors. Consistent with other studies, the split predicted a significant percentage of the variance ($r^2(97) = .248, p < .001$) for the complete data set without 0 + 0. As ties all have a split of 0, the data set without ties ($n = 9$) yielded $r^2(88) = .373, p < .001$. CRTs for non-ties ($n = 90$) were regressed onto the Welford function log [(larger−smaller)/(larger+smaller)]. (For ties, this function cannot be computed as larger−smaller = 0). The equation was CRT = 551 + 41.5 Welford; R-square was 45.3%, which is standard for comparison experiments.

Ties and 0-problems

For tie pairs (e.g., 5+5, $n = 9$), regression analysis. revealed that naming RTs were very good predictors of CRTs ($r^2(7) = .544, p = .02$) suggesting that subjects simply name the repeated number rather than comparing the two.

For 0-problems (N+0, 0+N) the split was not a good predictor ($r^2(16) = .08, p > .2$) whereas the naming RTs explained 50% of the variance ($r^2(16) = .504, p < .01$). This suggests that selecting the larger in these items is not achieved by comparing the magnitude of the two numbers, as the results do not show the characteristic signature of comparison—the split effect—but rather, selection is achieved by a process in which the presence of a 0 triggers a strategy of simply naming the non-zero number.

Order effect

There was a 13-ms trend for CRTs to be shorter where the larger digit was in the first position (e.g., 6 + 3) than where it was in the second position (e.g., 3 + 6); t test, mean difference 13.3 ms, $t(44) = 2.88, p < .01$, ties excluded.

Discussion of the magnitude comparison Task.

The task used in this experiment differs from other studies of magnitude comparison that have used key press responses (Duncan & McFarland, 1980; Moyer & Landauer, 1967; Parkman, 1971). The reason for the use of naming the larger number is that we wished to be sure not only that the subject had discriminated between the magnitudes, but had achieved the identity of the larger in the context of an addition problem. For this reason, another difference in design was employed: that of using exactly the same stimuli for both comparison and addition, that is, with the addition sign between the two numerals.

In spite of these differences, our results were comparable with those obtained in previous studies.[3]

One new finding is that split is not a predictor of pairs containing zero,-0-problems. Here, naming RT (from Experiment 1) is a good predictor (50% of variance), but split is not (8% of the variance). This suggests that subjects are not using a procedure that is sensitive to the numerical difference between the two numbers. A plausible interpretation is that where one of the numbers is zero, subjects simply read the other number. Similarly, naming RT turned out to be a good predictor for tie problems, again suggesting that subjects make no attempt to compare the two numbers, but simply name one of them.

Task 3. Addition

Accuracy

Of 4000 stimuli, 48 responses (1.2%) were invalid due to irrelevant noise or equipment failures, and 120 responses were incorrect (3.0%). Errors were not uniformly distributed across the tables but increased with the size of the problems. Figure 3 shows the percentage of errors plotted against Sum.

Reaction times

Problem-size effect.   Median RTs over 20 subjects for the 100 problems are presented in Table 2.

Our data confirm the common finding that solution latencies and error rate increase very roughly with problem size (sum). In Figure 4, ARTs for correct trials have been plotted as a function of the sum.

It can be seen that ARTs are not a simple function of the sum, but for each sum, the ARTs are widely spread. For example, for sum = 12, 6 + 6 takes 650 ms, 8 + 4 takes 745 ms, and 5 + 7 takes 928 ms. (Similar comparisons can be made by following the appropriate diagonals in Table 2).

Structural variables.   The standard "structural variables"—numbers and order—were used to predict median reaction times. Regression analyses were calculated using minimum addend (min, Groen & Parkman, 1972); correct sum, sum squared (Ashcraft & Battaglia,

---

[3]In his review of an earlier draft of this paper, Mark Ashcraft argued that by presenting the numbers to be compared in the same format as the addition task—namely as "5 + 3"—subjects would automatically add the two numbers together (as LeFevre et al.'s, 1988, study has shown), activating a nontarget number (8) that would interfere with the production of the target (5). Ashcraft suggests that this phenomenon could systematically affect comparison reaction times so that they indeed covaried with the sum reaction time (from the separate addition experiment). LeFevre et al. showed that the presentation of two numbers, whether with the addition sign or not, would induce the task-irrelevant activation of the sum, when it is below 10. If there is interference, Ashcraft suggests that it would increase as the sum increases, hence comparison times would predict addition times. However, consider the following stimuli: 6 + 1 and 4 + 3. Each has the sum 7, but which would suffer the most interference if the sum competed with the correct response, 6 or 4? In general, there is more interference from numbers that are closer in magnitude (Pavese & Umiltà, 1998). Thus, 6 + 1 should suffer more than 4 + 3. Hence, Ashcraft's presumption predicts that comparison times should be inversely related to split in this experiment. As we show, in line with previous findings, 6 + 1 is compared faster than 4 + 3, and Experiment 3 demonstrates that it is also added faster.
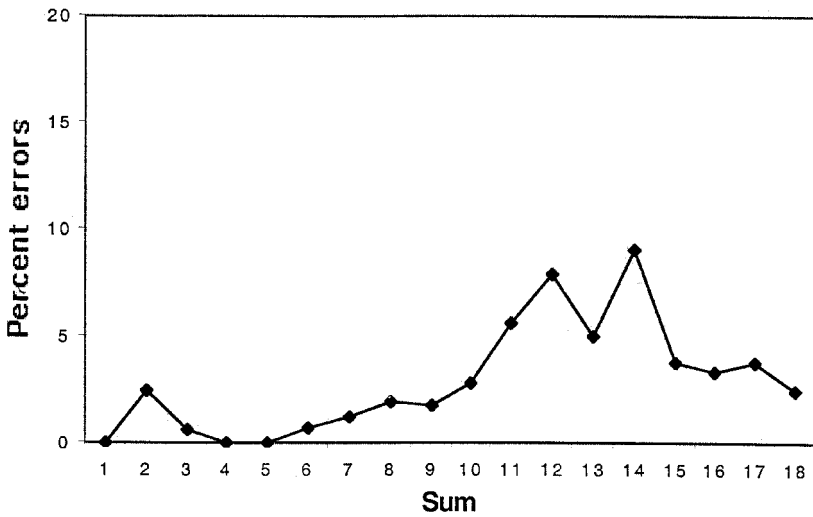
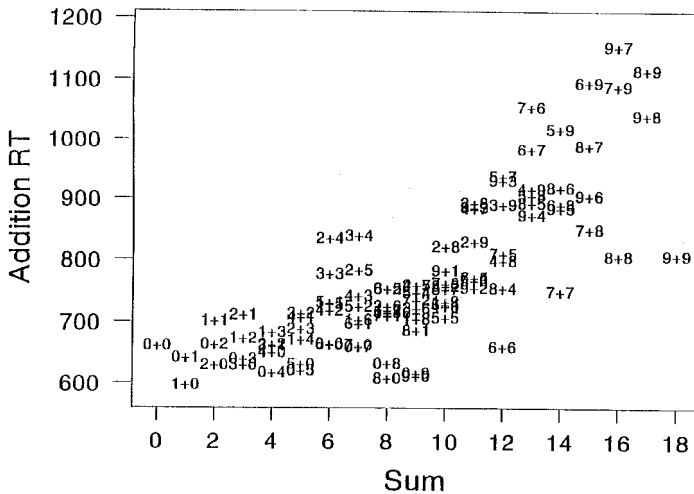**Figure 3.** Addition errors as function of sum.

Figure 4. Addition reaction times (ARTs) plotted by sum.

1978); product of the addends (product, Widaman et al.,1989); and a binary variable when the sum was greater than 10 (greater 10) (cf., Siegler, 1987). Table 3 shows how each structural predictor accounted for the variance in ARTs. When all problems are analysed (ties and 0 included), each of the structural predictors accounted for a significant percentage of the variance. The best predictor was the product, followed by sum^2 and sum. When zero–problems were removed from the data set and the analysis applied again, the best predictors were the sum^2 followed by product. These results are broadly in line with ARTs found by other researchers (Miller et al., 1984; Widaman et al., 1989; etc).

TABLE 2
Median reaction times[a] and errors[b] over 20 subjects for the addition task (Experiment 3)

Left side numbers	Right side numbers																		
	0		1		2		3		4		5		6		7		8		9	
	RT	Error	RT	Error	RT	Error	RT	Error	RT	Error	RT	Error	RT	Error	RT	Error	RT	Error	RT	Error
0	650	0	631	0	653	0	629	0	607	0	610	0	655	0	649	0	622	0	606	0
1	586	0	691	7.5	663	0	673	0	661	0	719	0	693	0	709	2.5	695	0	723	0
2	619	0	701	0	653	0	678	0	825	0	773	2.5	715	0	749	0	812	7.5	820	0
3	619	2.5	650	0	703	0	767	5	829	2.5	744	5	712	5	743	5	883	7.5	881	7.5
4	640	0	697	0	708	0	732	0	705	0	749	0	716	0	874	10	789	5	907	0
5	621	0	721	0	715	5	706	7.5	737	2.5	697	0	758	5	928	10	896	5	1007	0
6	653	0	687	0	743	2.5	705	0	719	2.5	762	0	650	0	973	7.5	881	17.5	1084	0
7	652	0	700	0	725	10	752	7.5	762	20	801	17.5	1044	7.5	740	5	840	0	1078	7.5
8	597	0	676	0	748	0	876	5	745	0	883	7.5	906	7.5	979	7.5	797	0	1105	7.5
9	601	0	773	2.5	745	2.5	920	15	863	2.5	875	5	895	7.5	1143	2.5	1030	0	798	2.5

[a] In ms.
[b] In percentages.

TABLE 3
Structural predictors of addition reaction times (R^2 values)

	Stimulus set and number of stimuli				
	All 100	Min > 0 81	Min = 0 19	Min = 1 18	Ties 9
Min	.598**	.457**			.365*
Sum	.610**	.567**	.019	.307**	
Sum^2	.632**	.602**	.039	.366**	
Greater 10	.564**	.504**			
Product	.679**	.587**		.307**	.440*

* $.05 < p < .10$.
** $p < .01$.

Non-structural variables

Frequency effect.   It has been suggested that retrieval latencies depend significantly on the frequency of presentation of the different problems (e.g., Ashcraft & Christy, 1995; Hamann & Ashcraft, 1986). In particular, the higher frequency of small problems compared to larger problems has been considered a critical factor in determining the problem-size effect. However, new studies on frequency of arithmetical facts show poor correlation with the problem-size (Ashcraft & Christy, 1995) suggesting that the frequency of occurrence is only one of the numerous factors affecting mental arithmetic.

We investigated the contribution of frequency to the ARTs for the addition problems with reference to the index of frequency reported by Ashcraft and Christy (1995), who do not include $0 + 0$ in their tables of frequencies. Correlations between this frequency measure and ARTs fail reach significance ($n = 99$; $r = .022$; ns).

Retrieval probability.   LeFevre et al. (1996) have shown that subjects frequently use non-retrieval strategies to solve problems, especially where the problems are large (their Figure 1). These strategies include counting on and transforming the problem into one easier to solve. They found that the proportion of trials in which subjects reported using a non-retrieval strategy was a predictor of ART. Using the proportions they report in their Appendix C as a predictor of our own data, we also found it a better predictor, accounting for 64% of the variance (for $n = 99$).

0–problems.   A separate analysis for the 0-addend problems was conducted. Chronometric studies (e.g., Ashcraft & Stazyk, 1981; Miller et al., 1984) as well as neuropsychological data (e.g., Sokol, McCloskey, Cohen, & Aliminosa, 1991) strongly suggest that these problems are solved differently from others.

As already reported by LeFevre et al. (1996) no single structural variable accounted for a significant percentage of the variance of the solution times for 0-problems. However, a significant regression has been found for median ARTs on naming RTs ($r^2 = .486$, $p < .001$). This is strikingly different from the situation with the set of all problems, where the regression value is much lower ($r^2 = .093$, ns). These data are consistent with a strategy for 0-problems in which subjects simply name the non-zero number.

1-Problems.   According to Baroody (1983, 1984, 1995) problems including 1 as addend are likely to be solved via a so-called "number after" rule rather than by retrieving a stored answer. LeFevre et al.'s (1996) college subjects frequently reported using a counting strategy. The different status of the N+1 problems compared to problems with addends larger than 1 has implicitly been assumed by researchers who did not include these items in testing a model of arithmetical fact retrieval (e.g., Campbell, 1994, 1995; Zbrodoff, 1995).

We performed a separate analysis of the 1-problems in order to test for the hypothesis of a rule-based processing. We assumed that if a "number after" rule is applied, rather than fact retrieval, latencies should not be correlated with any of the structural variables that affect problems with min > 1. On the other hand, a significant correlation with the naming RTs for the sums might be found. As shown in Table 3, as with the outcome of the overall analysis, the structural variables were significant predictors of the ARTs.

Order effect.   There was a 9-ms trend for problems with max on the left to be faster than those with max on the right. However, this trend failed to reach significance: Max + Min vs. Min + Max; $t(44) = 1.22$, $p = 23$; ties excluded. It will be seen later that the order effect is modulated by the magnitude of the numbers. This may be why the trend fails to reach significance in a t test. It should be noted that the trend is in the same direction as the significant effect for the magnitude comparison task.

Ties.   The ARTs for ties of 2 + 2 or greater were found to be consistently shorter than those for problems with the same min (note that the sum for non-ties in this comparison is always one greater than that for ties); see Figure 5.

These results are consistent with all studies that have looked at ties since Groen and Parkman (1972).

As can be seen from Table 4, the other structural variables—min, sum, sum^2, and product—are also good predictors, suggesting that these factors are relevant. It is clear, however, that some part of the process is different from and quicker than that for non-ties.
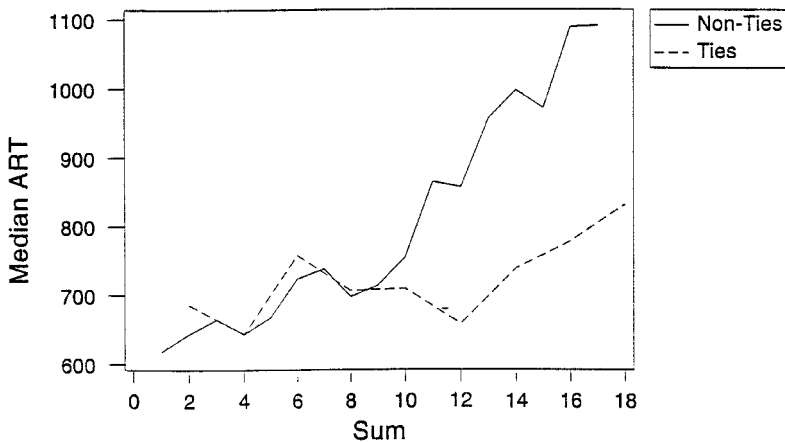

Figure 5.   Addition reaction times (ARTs) for ties and non-ties.

TABLE 4
Comparison RTs as predictors for addition reaction times (ARTs):
Values of R^2

	Stimulus set and number of stimuli		
	All 100	Split > 0 90	Min > 0 split > 0 72
Comparison RT* naming larger RT	.285*	.202*	.191*
Comparison RT (diff)			.152*

Note: Comparison RT from Experiment 2. Naming larger RT from Experiment 1.
Comparison RT (diff) = difference between comparison RT for M(>n) and naming
RT for M.
*p < .05 for all values.

Testing the COMP model

The COMP model consists of three identifiable stages, which we hypothesize will make independent variable contributions to ARTs (Table 5). It is assumed that the input to the comparison stage, which encodes the stimuli as ANIs and activates the magnitudes of the two numbers, takes a small fixed time. A first step in testing the COMP model is to evaluate whether or not RTs for comparing M and N to find the max and the RT for naming the sum of M and N tasks constitute good predictors of the ART for finding the sum of M + N

The median ART is expressed as follows:

$$\text{Median ART} = 23.6 \text{ left number} + 22.9 \text{ right number} +$$
$$(0.772 \text{ CRT} - 1.31 \text{ naming RT of max}) + 1.55 \text{ naming RT of sum} \quad (1)$$

In Equation 1, the expressions "x left number" and "y right number" mean that the numbers are multiplied by the coefficients; for example, for 5 + 3, the value of the left number expression is 23.6*5 (ms), and the value of the right number expression is 22.9*3 (ms). To get a better estimate of mental comparison times (by taking into account the process of selecting the name of the larger number), the naming latencies for the max were entered in the regression together with the CRTs; this is represented by the expression: (0.772 CRT − 1.31 naming RT of max). The effect of the retrieval time for the name of the sum is given by the expression: 1.55 naming RT of sum.

The regression equation obtained accounted for 71% of the variance for all sums, and each of the regression factors contributes significantly (p < .001) to the equation; t ratios: CRT, 3.24

TABLE 5
Stages of the COMP model

Stage	Predictors
Input to Comparison: Establishing Abstract Number Identities	
1. Comparison of M and N to identify Max	Comparison RT adjusted for Naming RT of Max
2. Retrieval of Sum of Max + Min from half-table	Size of Max and Min
3. Producing name of Sum	Naming RT of Sum

(5% of unique variance); naming RT of max, −3.44; left number, 9.03; right number, 8.18; naming RT of sum, 4.79). Equation 1 was calculated with no constant because the constant did not make a significant contribution. Separate analyses were conducted on ties and 0-problems.

Ties. The model in Equation 1 was a good fit to the ARTs for ties ($r^2 = .52$).

0-problems. On the other hand, neither the N + 0 nor the 0 + N data were predicted by the model ($r^2 = .175$, *ns*). This strongly suggests that neither comparison nor retrieval was implicated in solving these problems. As noted earlier, 0-problem ARTs were highly correlated with the naming times for N, suggesting that subjects simply read N when solving these problems.

Discussion of the addition task

In outline, the results replicated previous findings on generating answers to single digit additions. We found that the Sum was a significant predictor of ARTs: This is the problem-size effect. We also found an advantage for ties.

However, detailed analysis revealed that the problem-size effect is by no means simple. As can be seen from Figure 4, within a given sum, ARTs will differ by over 250 ms (more than 30% of the mean RTs) . Our analysis further revealed that the problem-size effect does not apply to N + 0 and 0 + N problems. Nor does the problem-size effect apply to ties in our study. (Even where previous studies found a statistically significant problem-size effect for ties, the effect was much smaller than that for non-ties; e.g., Ashcraft & Battaglia, 1978; LeFevre et al., 1996). On the other hand, N + 1 and 1 + N did show an effect of problem size.

By using the same subjects in the three tasks, it was possible to estimate the contribution of comparison processes and name retrieval processes to total ARTs. By relating naming times for the numbers of the sums (from Task 1) to ARTs, we could see that the ARTs for problems with zeros, unlike other problems, are predicted by the naming times. This suggests that these problems are not solved in the same way as other problems. So, if other problems are solved usually by retrieval of stored addition facts, then 0-problems seem to be solved in some other way.

Although the usual structural variables were found to be good predictors of ARTs (Table 3), better predictors were those involving explicit comparison of the two numbers. The COMP model explicitly claims that as solving one-digit addition involves comparison, comparison RT will predict ART. This is what was found in our regression analyses. Using data from explicit comparisons in Task 2 to predict addition times yielded a regression equation (Equation 1) that accounted for some 71% of the variance in ARTs. This is better than previous explanatory attempts, where the best results using structural variables, frequency of occurrence, or probability of a non-retrieval strategy was in the range 20–60%.

GENERAL DISCUSSION

Three tasks were carried out with the same 20 subjects to assess the contribution of three factors to the time to produce the sum of a two-digit addition. These factors were: (1) the time to retrieve and enunciate the name of the sum; (2) the time to compare the two numbers, which is

hypothesized to be part of the addition process; and (3) the time to find the sum of the two digits in memory.

In Task 1, subjects named (read aloud) all the numbers from 0 to 18 that are the sums of all single-digit additions. Excluding 0, which turned out to be ambiguous as between "nought" and "zero", average name times ranged from 470 to 520 ms. The main contributory factor, indeed the only identifiable contributory factor, in naming times was the nature of the initial sound of the number name. Number magnitude made no significant contribution.

In Task 2, subjects named the larger of two single digit numbers for all 100 pairs from 0–9. These were presented as if in an addition sum—that is, with a plus sign between them. For ties, subjects were instructed to name either number. The results of this experiment were consistent with those previously reported that used a button press methodology (Moyer & Landauer, 1967). Overall, the difference between the numbers (the split) was a good predictor of most comparisons. However, split failed to predict pairs containing 0. For these, the best predictor was the naming time for the non-zero number derived from Task 1. Ties too were well predicted by naming times, suggesting that subjects simply named the number rather than carrying out a comparison.

Task 3 required subjects to say the answer to the 100 single-digit addition sums. The results of this experiment were consistent with those previously reported that used a verification methodology (Ashcraft & Battaglia, 1978; Groen & Parkman, 1972; Widaman et al., 1989) and a production methodology (LeFevre et al., 1996; Miller et al., 1984). Subjects were accurate and fast, and the structural factors previously found to be relevant to ARTs were found to be relevant here: size of the sum, min, and ties, such that overall ARTs were longer for larger sums and longer for larger mins, and ARTs for ties were shorter than those for problems with the same sum or the same Min. Non-structural properties were also examined. An estimate of the frequency of occurrence of the additions in school textbooks (Ashcraft & Christy, 1995) turned out to make no significant contribution to ARTs. On the other hand, an estimate of the probability of simple sum retrieval as compared with other solution procedures (LeFevre et al., 1996) was a good predictor. However, none of the structural or non-structural factors was able to explain more than 64% of the variance. Further analysis revealed that ARTs for 0-problems were highly correlated with naming times found in Experiment 1. This makes sense in terms of a choice of a strategy but it is puzzling in terms of the findings of LeFevre et al. (1996) that subjects rarely reported having used a special (i.e., non-retrieval) strategy for these problems.

The COMP model

The basic idea of the COMP model is that addition information is stored in long-term memory in max + min form. This means that commuted versions have the same memory representation: $3 + 5 = 8$ and $5 + 3 = 8$ have the single representation max (5) + min (3) = 8. Of course, there is no general processing reason why arithmetical memory should be organized in this way. It is not clear, for example, that it is more efficient to store only half the possible sums: Memory savings may be more than offset by procedural costs.

It makes sense only if arithmetical facts are seen as being organized in a domain-specific way, in terms of their cardinal magnitude, rather than in some domain-independent way, for example in terms of frequency of occurrence or association strength, or in terms of another

domain, such as the phonological shape of the number words, and so on. Cardinal ordering, as was discussed in the Introduction, implies storing just one of the commuted pairs of addends.

As identifying the max and min implies comparing the magnitudes of the addends, the key predictive feature of the COMP model is a comparison stage that is an intrinsic part of the addition process. A more complete model of the process must also contain a stage or sub-process in which the sum of the addends is retrieved from memory, and a further stage in which the name of the sum is retrieved and uttered. The three tasks were designed to assess the contributions of comparison, naming, and sum retrieval.

The method was employed to assess the contribution of the postulated comparison stage to overall ART used comparison RTs from Task 2, modulated by an adjustment for the naming time for the larger of the two addends (as naming is the response in the comparison task). This expression was entered into regression Equation 1 along with the right and left addends and the naming time for the sum (also taken from Task 1). Each of these predictors made a highly significant contribution, with an r^2 of .71. Further analysis showed that this model gave a good account of ties, as the naming RT for ties in the comparison experiment was short, because subjects can solve this comparison problem without actually comparing magnitudes. However, 0-problems were not well predicted by the equation, as was evident from the fact that ARTs were highly correlated with the naming RT for the non-zero number (i.e., the sum of N + 0 or 0 + N).

It is by no means evident that having a special sub-process for 0-problems is necessarily more efficient than retrieving the solutions to these problems from memory. However, using a special sub-process is consistent with the subjects' domain-specific understanding that adding 0 to any number yields that number as the sum. Rote learning of the solutions for N + 0 and 0 + N problems, on the other hand, would appear to produce a set of associations qualitatively little different from any other single-digit addition. This modification has the advantage that 0 has a representation. In the original model (Figure 1), there was no distinct representation for the empty set and hence no obvious way of solving 0-problems.

In summary then, more of the variance in ARTs was explained by the regression equation derived from the COMP model than by any other previously published model using datasets collected in similar tasks (Ashcraft & Battaglia, 1978; Campbell, 1995; Groen & Parkman, 1972; LeFevre et al., 1996; Miller et al., 1984; Widaman et al., 1989). Application of alternative models to the dataset collected here was again less successful than application of the COMP model.

We now have a model for all single-digit additions that is consistent with the idea that skilled adults predominantly use simple fact retrieval, where the facts are organized in a domain-specific way. The max–min organization also reflects the way children acquire addition facts. The use of memory retrieval appears to follow the developmental stage where children are counting on from larger, though no doubt counting on can co-occur with retrieval, as it does to a lesser extent in adults (LeFevre et al., 1996). This account also appears to pose a problem for the influential model of fact retrieval advanced by Dehaene and Cohen (1995), in which facts are stored as verbal associations. On the other hand, it supports the general contention of Butterworth (1999) that numbers constitute a special cognitive domain with their own principles of organization. These have been embodied in a neural network model where the same numerical principles underlie both number comparison and addition (Zorzi & Butterworth, 1997, 1999).

REFERENCES

Ashcraft, M.H. (1992) Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75–106

Ashcraft, M.H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Memory and Learning*, *4*, 527–538.

Ashcraft, M.H., & Christy, K.S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in grades 1–6. *Journal for Research in Mathematics Education*, *26*(5), 396–421.

Ashcraft, M.H., & Stazyk, E.H. (1981). Mental addition: A test of three verification models. *Memory & Cognition*, *9*, 185–196.

Baroody, A.J. (1983). The development of procedural knowledge: An alternative explanation for chronometric trends of mental arithmetic. *Developmental Psychology*, *3*, 225–230.

Baroody, A.J. (1984). A reexamination of mental, arithmetic models and data: A reply to Ashcraft. *Developmental Review*, *4*, 148–156.

Baroody, A.J. (1995). The role of the Number-After rule in the invention of computational shortcuts. *Cognition and Instruction*, *13*(2), 189–219.

Baroody, A.J., & Gannon, K.E. (1984). The development of the commutativity principle and economical addition strategies. *Cognition and Instruction*, *1*(3), 321–339.

Baroody, A.J., & Ginsburg, H.P. (1986). The relationship between initial meaningful and mechanical knowledge of arithmetic. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Butterworth, B. (1999). *The mathematical brain*. London: Macmillan.

Campbell, J.I.D. (1994). Architectures for numerical cognition. *Cognition*, *53*, 1–44.

Campbell, J.I.D. (1995). Mechanisms of number-fact retrieval: A modified network-interference theory and simulation. *Mathematical Cognition*, *1*, 121–164.

Campbell, J.I.D., & Tarling, B. (1996). Production, verification and error priming in cognitive arithmetic. *Memory and Cognition*, *24*(2), 156–172.

Dehaene, S., Bossini, S., & Ciraux, P. (1993). The mental representation of parity and numerical magnitude. *Journal of Experimental Psychology: General*, *122*, 371–396.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*, 83–120.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*, 1–29.

Delazer, M., & Butterworth, B. (1997). A dissociation of number meanings. *Cognitive Neuropsychology*, *14*, 613–636

Duncan, E.M., & McFarland, C.E. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory and Cognition*, *8*, 612–622.

Foltz, G.S., Poltrock, S.E., & Potts, G.R. (1984). Mental comparison of size and magnitude: Size congruity effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 442–453.

Fuson, K. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.

Gelman, R., & Gallistel, C.R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Giaquinto, M. (1995). Concepts and calculation. *Mathematical Cognition*, *1*, 61–81.

Groen, G.J., & Parkman, J.M. (1972). A chronometric analysis of simple arithmetic. *Psychological Review*, *79*, 329–343.

Groen, G.J., & Resnick, L.B. (1977). Can preschool children invent addition algorithms. *Journal of Educational Psychology*, *69*, 645–652.

Gum, T., & Bub, D. (1988). *Psychlab v. 0.85*. [software].

Hamann, M.S., & Ashcraft, M.H. (1986). Textbook presentations of basic arithmetic facts. *Cognition & Instruction*, *3*, 173–192.

LeFevre, J.-A., Bisartz, J., & Mrkonjic, L. (1988). Cognitive arithmetic: Evidence for obligatory activation of arithmetic facts. *Memory & Cognition*, *16*, 45–53.

LeFevre, J.-A, Sadesky, G.S., & Bisartz, J. (1996). Selection of procedures in mental addition: Reassessing the problem-size effect in adults. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*(1), 216–230.

Lemaire, P., & Fayol, M. (1995). When plausibility judgements supersede fact retrieval: The example of the odd–even effect on product verification. *Memory and Cognition*, *23*(1), 34–48.

Logan, G.D., & Klapp, S.T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*, 179–195.

McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*, 107–157.

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition*, *4*, 171–196.

Miller, K., Perlmutter, M., & Keating, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 46–60.

Moyer, R.S., & Landauer, T.K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519–1520.

Parkman, J.M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, *91*(2), 191–205.

Pavese, A., & Umiltà, C. (1998). Symbolic distance between numerosity and identity modulates Stroop interference. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1535–1545.

Posner, M.H. (1979). *Chronometric explorations of mind*. New York: Oxford University Press.

Resnick, L.B., & Ford, W.W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Secada, W.G., Fuson, K.C., & Hall, J. (1983). The transition from counting-all to counting-on in addition. *Journal for Research in Mathematics Education*, *14*, 47–57.

Siegler, R.S. (1987). The perils of averaging over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250–264.

Siegler, R.S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Siegler, R.S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sokol, S.M., McCloskey, M., Cohen, N.J., & Aliminosa, D. (1991). Cognitive representations and processes in arithmetic: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*(3), 355–376.

Treiman, R., Mullennix, J., Bjeljac-Babic, R., & Richmond-Welty, E.D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*(2), 107–136.

Widaman, K.F., Geary, D.C., Cormier, P., & Little, T.D. (1989). A componential model for mental addition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 898–919.

Zbrodoff, N.J. (1995). Why is 9+7 harder than 2+3? Strength and interference as explanations of the problem size effect. *Memory & Cognition*, *23*, 689–700.

Zbrodoff, N.J., & Logan, G.D. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 83–97.

Zorzi, M., & Butterworth, B. (1997). On the representation of number concepts. *Proceedings of the Nineteenth Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Zorzi, M., & Butterworth, B. (1999). A computational model of number comparison. *Proceedings of the Twenty-First Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

COMP MODEL OF ADDITION

APPENDIX 1

Regression table for best-fitting model of ART including constant

Predictor	Coefficient	SD	t ratio	p
Comparison RT	1.0165	0.2246	4.53	0.000
Naming RT of max	−1.4528	0.4159	−3.49	0.001
Left number	21.056	2.372	8.88	0.000
Right number	20.685	2.556	8.09	0.000
Naming RT of sum	1.2038	0.4442	2.71	0.008
(Constant)	107.6	264.9	0.41	0.685

Median ART = 108 + 21.1 left number + 20.7 right number + 1.02 CRT − 1.45 naming RT of max + 1.20 naming RT of sum r^2 = .72.