ACCESSING NUMBER MEANING
IN ADULTS AND CHILDREN

A thesis submitted to the
Faculty of Science
of the University of London
for the degree of Doctor of Philosophy

Luisa Girelli

Department of Psychology
University College London

September 1998
ABSTRACT

This thesis aims to clarify different aspects of the mechanisms underlying number processing. In particular, the research has two main purposes: first, to explore the conditions under which numerals may autonomously access semantics; second, to investigate whether transcoding from one number format to another can be accomplished bypassing the semantic system. These theoretical issues were addressed in two series of experimental studies and in a single-case study.

In the first series of experiments, a number-Stroop paradigm was used to explore similarities and differences between intentional and unintentional processing of numerical information with the specific purpose of tracing the developmental changes in the autonomous processing of Arabic numerals. Pre-schoolers, primary-school children of different ages and university students compared either numerical size or physical size of Arabic numerals varying along both dimensions. Task- and age-related differences in the size-congruity effect strongly suggest that the autonomous processing of numerical information arises gradually as numerical skills progress.

In the second series of experiments, the hypotheses of a modality-independent automatic access to magnitude representation and of an asemantic mapping between Arabic and verbal numerals were investigated by means of two matching tasks differing in the level of processing required. The results challenge the assumption of a completely autonomous semantic access to magnitude information and suggest that this phenomenon is modulated by both stimuli selection and task demands.

Finally, a neuropsychological single-case study relevant to the addressed issues is reported. The study investigates the effect of different notations and the effect of different tasks on the production of verbal numerals in a phonological dyslexic patient. The results are consistent with the assumption of semantic and asemantic pathways in number transcoding and point to the importance of a more refined distinction between different type of number meanings.

The results are critically evaluated and possible implications for future research are discussed.
Acknowledgements

First and foremost, I wish to express my gratitude to my supervisor, Professor Brian Butterworth, for his invaluable help, scientific insight and continuous support.

I would like to thank Daniela Lucangeli who gave me the opportunity to test preschoolers and primary-school children. A sincere thanks to Allistar McClelland for his statistical advice.

I would like to thank all my friends and in particular Gina Alexandratou, Alessia Grana’ and Manuela Piazza for their warm and genuine support.

Finally, I am indebted to Margarete Delazer who gave me many cogent comments, suggestions and continuous encouragement.
TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION ................................................................. 11

CHAPTER 2
LITERATURE REVIEW ......................................................... 14

OUTLINE ............................................................................. 14
2.1 MODELS OF NUMBER PROCESSING .................................. 14
   2.1.1 McCloskey’s semantic model ........................................ 16
   2.1.2 Seron and Deloche’s asemantic transcoding algorithms .... 21
   2.1.3 Campbell and Clark’s encoding complex view ............... 22
   2.1.4 Noel and Seron’s preferred entry code hypothesis and the Intermediate semantic representation hypothesis ......................................................... 25
   2.1.5 Dehaene’s triple-code model ......................................... 27
   2.1.6 Cipolotti and Butterworth’s multiple-route model .......... 31
   2.1.7 Single and multiple routes: further considerations ......... 33
2.2 AUTONOMOUS PROCESSING OF NUMBER MAGNITUDE ...... 38
   2.2.1 SNARC effect in non-magnitude based tasks ............... 41
   2.2.2 Numerical versions of the STROOP-task ..................... 42
   2.2.3 Number-matching tasks .............................................. 50
   2.2.4 Additional studies ..................................................... 51
   2.2.5 Format effect ............................................................. 52
   2.2.6 Further considerations .............................................. 52
2.3 NUMBER PROCESSING IN CHILDREN ............................. 54
   2.3.1 The acquisition of number words ................................. 56
   2.3.2 The acquisition of Arabic numerals ............................... 58
   2.3.3 The acquisition of number comparison skills ............... 61
   2.3.4 Autonomous semantic processing of Arabic numerals .... 63

CHAPTER 3
THE RISE OF AUTOMATICITY IN ACCESSING NUMBER MAGNITUDE .......... 66

INTRODUCTION ................................................................. 66
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>EXPERIMENT 1</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>3.1.1 Method</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>3.1.2 Results</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>3.1.3 Discussion</td>
<td>82</td>
</tr>
<tr>
<td>3.2</td>
<td>EXPERIMENT 2</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>3.2.1 Screening procedure</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>3.2.2 Method</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>3.2.3 Results</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>3.2.4 Discussion</td>
<td>98</td>
</tr>
<tr>
<td>3.3</td>
<td>EXPERIMENT 3</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>3.3.1 Method</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>3.3.2 Results</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>3.4.3 Discussion</td>
<td>110</td>
</tr>
<tr>
<td>3.4</td>
<td>SUMMARY AND CONCLUSIONS</td>
<td>112</td>
</tr>
</tbody>
</table>

**CHAPTER 4**

**THE AUTONOMOUS PROCESSING OF MAGNITUDE INFORMATION:**
**INFLUENCE OF STIMULI SELECTION, NOTATION AND TASK DEMANDS.**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>EXPERIMENT 4N</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>4.2.1 Method</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>4.2.2 Results</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>4.2.3 Discussion</td>
<td>129</td>
</tr>
<tr>
<td>4.2</td>
<td>EXPERIMENT 4P</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>4.2.1 Method</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>4.2.2 Results</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>4.2.3 Discussion</td>
<td>133</td>
</tr>
<tr>
<td>4.3</td>
<td>EXPERIMENTS 5N AND 5P</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>4.3.1 Method</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>4.3.2 Results</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>4.3.3 Discussion</td>
<td>143</td>
</tr>
<tr>
<td>4.4</td>
<td>EXPERIMENTS 6N AND 6P</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>4.4.1 Method</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>4.4.2 Results</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>4.4.3 Discussion</td>
<td>155</td>
</tr>
<tr>
<td>4.5</td>
<td>EXPERIMENTS 7N AND 7P</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>4.5.1 Method</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>4.5.2 Results</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td>4.5.3 Discussion</td>
<td>163</td>
</tr>
<tr>
<td>4.6</td>
<td>SUMMARY AND CONCLUSIONS</td>
<td>165</td>
</tr>
</tbody>
</table>
TABLE OF TABLES

Chapter 3

Table 3.1 Examples of the stimuli used in the Experiment 1........................................... 78
Table 3.2 Experiment 1: Mean Reaction times, Standard deviations and Error rates as a function of task, congruity and numerical distance.............................................................. 79
Table 3.3 Preliminary task for Experiment 2: Percentage of correct responses in the preliminary tasks for the different groups of children ........................................................................... 79
Table 3.4 Preliminary task, number comparison: Mean Reaction times, Standard deviations and Error rates as a function of numerical distance and age.............................................. 80
Table 3.5 Experiment 2: Harmonic means and Error rates of 4 to 5 year-old and 5 to 6 year-olds as a function of task, congruity and numerical distance ......................................................... 89
Table 3.6 Experiment 3: Mean Reaction times, Standard deviations and Error rates for the different grades as a function of task, congruity and numerical distance ........................................... 91
Table 3.7 Experiment 3: Mean interference and facilitation ratios for numerical and physical comparisons as a function of grade level .............................................................................. 103

Chapter 4

Table 4.1. Examples of the stimuli used in Experiments 4-7.............................................. 125
Table 4.2. Experiment 4N: Mean Reaction times, Standard deviations and Error rates as a function of notation and numerical disparity ................................................................. 126
Table 4.3. Experiment 4P: Mean Reaction times, Standard deviations and Error rates as a function of notation and numerical disparity ................................................................. 131
Table 4.4. Experiment 5N: Mean Reaction times, Standard deviations and Error rates as a function of notation and numerical disparity ................................................................. 137
Table 4.5. Experiment 5P: Mean Reaction times, Standard deviations and Error rates as a function of notation and numerical disparity ................................................................. 140
Table 4.6. Experiment 6N: Mean Reaction times, Standard deviations and Error rates as a function of notation and numerical disparity ................................................................. 150
Table 4.7. Experiment 6P: Mean Reaction times, Standard deviations and Error rates as a function of notation and numerical disparity ................................................................. 152
Table 4.8. Preliminary numerical comparison task: Reaction times, Standard deviations and error rates as a function of grade and of numerical distance ............................................. 157
Table 4.9. Experiment 7N: Mean Reaction times, Standard deviations and Error rates as a function of grade, notation and numerical disparity ............................................................. 158
Table 4.10. Experiment 7P: Mean Reaction times, Standard deviations and Error rates as a function of grade, notation and numerical disparity ............................................................. 161
Table 4.11. Distance effect in numerical and physical matching tasks as a function of notation and Experiment ................................................................. 167
Chapter 5

Table 5.1. Non-numerical tasks, number comparison, calculation and recognition of arithmetic signs in 1992

Table 5.2. Transcoding tasks in 1992

Table 5.3. Percentage of correct answers in reading high and low frequency words in different grammatical classes (1992).

Table 5.4. Reading verbal numerals and Arabic numerals in 1992. Percentage of correct answers.

Table 5.5. Percentage of correct answers in reading high and low frequency words in different grammatical classes (1996).

Table 5.6. Number of errors in reading Arabic numerals and distribution of errors across classes.

Table 5.7. Number of errors in reading written verbal numerals and distribution of answers across different classes.

Table 5.8. Verbal errors to simple multiplications presented verbally and visually.
# TABLE OF FIGURES

## Chapter 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Schematic representation of the McCloskey's model</td>
<td>16</td>
</tr>
<tr>
<td>2.2</td>
<td>Schematic representation of Dehaene's triple code model</td>
<td>28</td>
</tr>
<tr>
<td>2.3</td>
<td>Schematic representation of the Cipolotti and Butterworth's multi-rote model</td>
<td>32</td>
</tr>
<tr>
<td>2.4</td>
<td>Schematic representation of the processes underlying intentional and unintentional number processing as proposed by Tzelgov et al. (1992)</td>
<td>48</td>
</tr>
</tbody>
</table>

## Chapter 3

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Experiment 1: Mean Reaction times as a function of task and congruity</td>
<td>81</td>
</tr>
<tr>
<td>3.2</td>
<td>Experiment 2: Mean Error rates for 4 to 5 year-olds and 5 to 6 year-olds as a function of task and congruity</td>
<td>94</td>
</tr>
<tr>
<td>3.3</td>
<td>Experiment 2: Mean Error rates in the numerical size comparison task as a function of age, congruity and numerical distance</td>
<td>96</td>
</tr>
<tr>
<td>3.4</td>
<td>Experiment 3: Mean Reaction times in numerical and physical tasks as a function of grade</td>
<td>105</td>
</tr>
<tr>
<td>3.5</td>
<td>Experiment 3: Mean Reaction times as a function of task, congruity and grade level</td>
<td>107</td>
</tr>
</tbody>
</table>

## Chapter 4

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Schematic representation of the putative processing pathways in a number matching task</td>
<td>117</td>
</tr>
<tr>
<td>4.2</td>
<td>Experiment 4N: Mean Reaction times and Error rates as a function of notation and numerical disparity</td>
<td>127</td>
</tr>
<tr>
<td>4.3</td>
<td>Experiment 4P: Mean Reaction times and Error rates as a function notation and numerical disparity</td>
<td>131</td>
</tr>
<tr>
<td>4.4</td>
<td>Experiment 5N: Mean Reaction times and Error rates as a function of notation and numerical disparity</td>
<td>138</td>
</tr>
<tr>
<td>4.5</td>
<td>Experiment 5P: Mean Reaction times and Error rates as a function of notation and numerical disparity</td>
<td>141</td>
</tr>
<tr>
<td>4.6</td>
<td>Experiment 6N: Mean Reaction times and Error rates as a function of notation and numerical disparity</td>
<td>151</td>
</tr>
<tr>
<td>4.7</td>
<td>Experiment 6P: Mean Reaction times and Error rates as a function of notation and numerical disparity</td>
<td>153</td>
</tr>
<tr>
<td>4.8</td>
<td>Experiment 7N: Mean Reaction times and Error rates as a function of grade, notation and numerical disparity</td>
<td>159</td>
</tr>
<tr>
<td>4.9</td>
<td>Experiment 7P: Mean Reaction times and Error rates as a function of grade, notation and numerical disparity</td>
<td>159</td>
</tr>
</tbody>
</table>
numerical disparity............................................................................................................. 162

Chapter 5

Figure 5.1. Schematic representation of the McCloskey's model............................................. 171
Figure 5.2. Schematic representation of the Cohen and Dehaene's model of number reading....... 172
Figure 5.3 Schematic representation of the Cipolotti and Butterworth's model of number reading. 174
CHAPTER 1

INTRODUCTION

Ever since scientists started to investigate the way in which the human mind operates, it was revealed that apparently simple and effortless performances imply complex and elaborate mechanisms. The ability to comprehend, manipulate and produce numbers\(^1\) is no exception to this rule.

Yet, very little attention has been paid to the mental processes involved in numerical abilities and, till relatively recently, numbers were mainly incidentally used as stimuli, in studies whose primary object of investigation was anything but numbers, as for example, studies on memory, attention or visual perception.

This reductive approach has been recently abandoned and in the last decades the study of numerical skills has become an extremely active area of research. The increasing interest in the study of the cognitive mechanisms underlying number processing abilities is certainly justified by the paramount importance of numbers in our cognitive milieu.

Indeed, numbers are used in countless every-day activities and in several different contexts and, accordingly, they convey different meanings (Fuson, 1992). Numbers stand for prices, height, weight and speed; they indicate the time, the age and sizes. Numbers are also used to ordering things in our environment, as the houses in a street, the rooms in an office or the floors in a building. Moreover, they constitute the far most preferred code system to identify different exemplars in the same category, as car models, telephone users or bus lines. Numbers are also the medium for calculation; in this context they refer to specific numerosities and they may be transformed through arithmetical

\(^1\) In the thesis, the term *number* is used to refer to any symbols, whatever its notational code, representing a number. The term *Arabic numerals* refer to numbers in digit form (e.g. 4), verbal numerals to numbers in word form (e.g., four), whether spoken (*spoken verbal numerals*) or written (*written verbal numerals*).
operations (e.g., 5+4=9). Thus, the ability to comprehend and manipulate numbers is of obvious practical importance and the study of the cognitive processes involved in it deserves concerted effort.

Our understanding of the multiple mechanisms involved in number processing is based on three different sources of evidence: performance of normal adults (e.g., Banks, 1977; Dehaene, Bossini & Giroux, 1993; Moyer & Landauer, 1969); developmental investigations (e.g., Gelman & Gallistel, 1978; Fuson, 1988) and brain-damaged subjects’ performance (e.g., Cipolotti & Butterworth, 1995, Cohen, Dehaene & Verstichel, 1994; Deloche & Seron, 1987; McCloskey, Sokol & Goodman, 1986).

These distinct areas of research investigate different aspects of the same cognitive processes: how numerical knowledge is acquired, normally used once established, and occasionally lost after cerebral lesion. Each of these disciplines, i.e., experimental psychology, developmental psychology and clinical neuropsychology, use different methods of investigation and draw inferences on different sources of data. Yet, it is becoming more and more evident that these approaches can greatly benefit from each other and that favouring connections between them may be extremely fruitful and be the engine of major progress (e.g., Dehaene, 1992). For example, studies of number processing in adults may benefit from understanding the developmental sequence characterising the acquisition of numerical competence during childhood. On the other hand, the detailed analysis of patterns of impaired performance in brain-lesioned patients may be informative about the cognitive architecture and functioning of normal numerical abilities (Caramazza, 1986; McCloskey & Caramazza, 1988). At the same time, the development of explicit cognitive models may greatly facilitate the interpretation of numerical disorders. In fact, only a theoretically-driven evaluation of a patient’s performance, based on specific hypotheses about the mechanisms involved in number processing and calculation, may lead to formulate a precise functional diagnosis of the deficit. Finally, though caution must be exercised in drawing parallels between the learning process in normal development and the re-acquisition of lost abilities, the rehabilitation of numerical disorders may contribute to what we know about the strategies and procedures that facilitate the acquisition of numerical skills over development.
In the present work we bring together these different approaches (developmental, experimental and neuropsychological research) in order to investigate the effects of semantic information in number processing. In particular, the role of magnitude is probed by investigating whether magnitude information may influence the performance in tasks that do not require semantic processing. It has been recently suggested that the mere presentation of an Arabic numeral may determine the activation of its magnitude representation (e.g., Dehaene et al., 1993; Henik & Tzelgov, 1982). However, the extent to which the access to number meaning is autonomously initiated and modulated by notational constrains are still open questions.

Though Arabic and verbal numerals are symbols that are arbitrarily associated to the meaning they convey, the evidence suggests that magnitude is a salient and distinctive attribute of a number. Yet, it is plausible to assume that the association between the written symbols and the magnitude they represent is gradually strengthened as numerical skills progress. We address this issue by exploring the rise of autonomy in accessing number magnitude and by comparing children's automatic and intentional processing of numerical information over the course of development.

All current models postulate notation-specific input and output systems for the comprehension and production of Arabic and verbal numerals. However, the way in which these two peripheral systems are connected is a rather controversial issue. Some authors postulate that input and output systems only communicate via a central semantic system (McCloskey, 1992), other authors posit additional asemantic pathways through which numerical stimuli may be transcoded (Cipolotti & Butterworth, 1995; Dehaene, 1992; Deloche & Seron, 1982a, b). This issue is specifically addressed in a single-case study where the effect of different notations, different number meanings and different task demands on the production of spoken verbal numerals were investigated.
CHAPTER 2

LITERATURE REVIEW

OUTLINE

The literature review has three components. The first outlines the proposed componental models of number processing; both neuropsychological and experimental evidence supporting the different theoretical accounts are discussed. The second examines experimental research focused on autonomy in semantic processing of numbers. The different paradigms and effects used to investigate the extent to which access to number magnitude representation is autonomously activated are considered. The third part reviews the studies which contributed to clarifying how these mechanisms develop over skills acquisition. In particular, attention is directed to research that investigate the processes by which numbers acquired symbolic meaning in the course of development.

2.1 MODELS OF NUMBER PROCESSING

In recent years, the cognitive mechanisms underlying numerical processing and calculation abilities have attracted the attention of a substantial number of psychologists and neuroscientists. The growing interest in this area of research is well reflected by the impressive increment of systematic studies that appeared in the literature. This cumulative evidence has provided the basis for the development and the evaluation of cognitive models of number processing and calculation (e.g., Campbell & Clark, 1988; Cipolotti & Butterworth, 1995; Dehaene, 1992; McCloskey, Caramazza & Basili, 1985; McCloskey,
1992). While some of these models attempt to provide a general architecture of numerical processing (Campbell & Clark, 1992; McCloskey et al., 1985; Dehaene, 1992), other contributions elucidate specific aspects of it (Cipolotti, 1995; Cipolotti & Butterworth, 1995; Cohen, Dehaene & Verstichel, 1994; Deloche & Seron, 1982a, b; 1987; Noel & Seron, 1993, 1997). The evidence in support to the different theories comes from both neuropsychological single-case studies and experimental investigations with normal subjects. Overall, the former have played a major role in it, but the importance of combining both approaches has been recently acknowledged (e.g., Dehaene, 1992; Noel & Seron, 1997).

At present, there is little disagreement about basic observations grounded on solid and highly replicable empirical findings. However, maybe not surprisingly, fundamental issues, such as the nature and the role of mental number representations, are still highly controversial. In particular, current theories differ in their assumptions about 1) the format in which numbers are mentally represented and 2) the extent to which a semantic representation is central to any numerical processing.

Numbers may be represented by different notations, among which the most familiar are the Arabic (e.g., 4) and the verbal format (e.g., four). Thus, the ability to recognise and produce numbers in either code implies the existence of notation-specific representations as both input and output modalities. However, the way in which these two peripheral systems are connected is, at present, a matter of debate. Some authors postulate that input and output systems only communicate via a central semantic system (McCloskey, 1992), other authors posit additional asemantic pathways through which numerical stimuli may be transcoded (Cipolotti & Butterworth, 1995; Dehaene, 1992; Seron & Deloche, 1982a, b).

A second point of debate concerns the nature of the numerical representation at the level of the semantic system. Within current theories, these representations are assumed to convey magnitude information; though other aspects of numerical knowledge are not totally neglected (e.g., Dehaene et al., 1993; Delazer & Butterworth, 1997) they are not explicitly incorporated in any models. Representations which are abstract and precise (McCloskey, 1992), analogue and approximate (Dehaene & Cohen, 1995) as well as tied to specific input code (Campbell & Clark, 1992) have been proposed.
In the following section, the current models of number processing will be described and emphasis will be given to their specific assumptions related to the access and the nature of numerical representations.

### 2.1.1 McCloskey’s semantic model

A comprehensive and influential theory of number processing has been developed by McCloskey and his colleagues. This model was first proposed in 1985 (McCloskey, Caramazza & Basili, 1985) and successively refined in several works (Sokol, Goodman-Shulman and McCloskey, 1989; Sokol et al., 1991; Sokol & McCloskey, 1991; McCloskey et al., 1991; Macaruso et al., 1993). The model is characterised by a highly modular architecture and by the pivotal role of the semantic system (see Figure 2.1).

---

**Figure 2.1 Schematic representation of the McCloskey's model**

![Diagram illustrating McCloskey's semantic model](image)
The model posits a first basic distinction between the number-processing system and the calculation system.

The number-processing system is composed of comprehension and production subsystems, which comprise distinct components for processing number in different modalities (spoken verbal numerals, written verbal numerals, Arabic digits). The comprehension mechanisms convert the surface forms of the numbers (e.g., 6, six or /siks/) into a single, modality-independent abstract code while the number-production mechanisms are assumed to translate these abstract internal representations into the appropriate form of output, either verbal (written or spoken) or Arabic.

The abstract internal representations are modality independent and semantic in nature: they are assumed to specify the basic quantities of the number and their associated power of ten. For example, the internal abstract representation of 5493 would be as follows: \( \{5\}10^{3}, \{4\}10^{2}, \{9\}10^{1}, \{3\}10^{0} \). The digits in brackets (e.g., \{5\}) stand for the semantic representation of the basic quantities of the number and \( 10^{n} \) specifies the power of ten associated with each quantity (e.g., \( 10^{3} \) specifies 10 to the third power, or thousand). The abstract internal representation does not include representation of the quantity \{0\}; thus, for example, the abstract internal representation of the number 2500 would look like \{2\} \( 10^{3} \) \{5\} \( 10^{2} \).

A further distinction is drawn within the comprehension and production mechanisms between lexical and syntactic processing mechanisms. Lexical processing involves comprehension or production of the individual element in a number (e.g., the digit 6 or the word /siks/) whereas syntactic processing involves processing of relation among the elements (e.g., order of the words within a verbal numeral) in order to comprehend or produce a numeral as a whole. For example, the translation of the verbal numeral *six hundred fifty-two* into the semantic representation \{6\} \( 10^{2} \), \{5\} \( 10^{1} \), \{2\} \( 10^{0} \) would require lexical processing to generate internal representations for the words, six-hundred-fifty-two as well as the syntactic processing to determine that, since "hundred" followed "six" in the stimulus, the quantity \{6\} should be associated with the second power of ten represented by the marker \( 10^{2} \), in the final semantic representation. Finally, within the lexical processing components for the verbal numeral comprehension and
production a distinction is drawn between phonological processing mechanisms for processing spoken verbal numerals and graphemic processing mechanism for processing written verbal numerals. For example, spoken production of the word 'six' would require retrieval of a phonological representation from a phonological output lexicon, whereas written production of the same word would require retrieval of a graphemic representation from a graphemic output lexicon. The phonological/graphemic distinction does not apply to syntactic processing: the same syntactic mechanisms are assumed to mediate the processing of both spoken and written verbal numerals.

The comprehension mechanisms subserving the translation of a numerical input into the correspondent amodal semantic representation have not been fully specified; on the other hand, the production process that converts these internal representation in the appropriate form of output has been the object of great attention (e.g., McCloskey et al., 1986; McCloskey, Sokol, Caramazza & Goodman-Schulman, 1990).

Further to number comprehension and production mechanisms the McCloskey’s model postulates three functionally independent components specific to calculation: a) mechanisms devoted to the comprehension and production of arithmetical signs ("+", "plus"), b) an arithmetical facts store, including all operation the solution of which is directly retrieved from memory (e.g., 4x5=20), and c) the knowledge of arithmetical procedures that specify the sequence of steps required for the solution of written multi-digit calculation (e.g., use of carry). One of the fundamental assumptions of the model is that the calculation system operates only on internal abstract representations: both input to and output from the calculation system are in the form of abstract semantic representations.

Single-case studies of cerebral-lesioned patients presenting selective difficulties in some aspects of number processing and/or in calculation abilities provide support for many of the functional distinctions depicted in the McCloskey’s model (see for a review, McCloskey, 1992). For example, patients showing a dissociation between intact comprehension and impaired production processes have been reported (Benson & Denckla, 1969; McCloskey et al., 1986; Singer & Low, 1933) as well as patients with a selective deficit in the comprehension of verbal numerals and intact comprehension of Arabic numerals.
(McCloskey et al., 1985; McCloskey & Caramazza, 1987; McCloskey et al., 1986). The inverse pattern of performance (intact comprehension of verbal numerals and impaired comprehension of Arabic numerals) was also reported (patient KK, McCloskey et al., 1985). The Arabic-verbal distinction has been also observed at the production level; in particular, a patient with error free production of Arabic numerals and impaired production of verbal numerals has been described (McCloskey, Sokol, Goodman-Shumann & Caramazza, 1990). Lexical mechanisms, responsible for the comprehension and production of the single elements of a numeral (e.g., the digits 2, 5 in the numeral 25), and syntactic mechanisms, devoted to the processing of the relations among these elements (e.g., the relative order of the single digits within a numeral) appear to be clearly dissociable as documented by the selective occurrence of lexical or syntactic errors (e.g., Cipolotti et al., 1994; Delazer & Denes, 1998; Deloche & Seron, 1982a, b; McCloskey et al., 1985; McCloskey et al., 1990; McCloskey et al, 1986; Noel & Seron, 1993; 1995; Singer & Low, 1933).

Neuropsychological data also support the hypothesis that the distinctive cognitive mechanisms specific to calculation constitute selectively vulnerable components of the calculation systems (Benson & Weir, 1972; Dagenbach & McCloskey, 1992; Delazer & Benke, 1997; Ferro & Botehlo, 1980; Girelli & Delazer, 1996; Lucchelli & DeRenzi, 1992; McCloskey et al., 1991; Hittmair-Delazer, Semenza & Denes, 1994; McNeil & Warrington, 1994; Pesenti, Seron & Van der Linden, 1994; Warrington, 1982).

To summarise, the McCloskey’s model explicitly postulates that the performance of any transcoding task is accomplished via obligatory activation of an internal semantic representation. Reading aloud a number or writing a number should always initially involve the generation of a number internal semantic representation. These representations are amodal and constitute the input to and output from the calculation system. This assumption implies that calculation, as well as other numerical tasks, such as number comparison, are the same irrespective of the encoding format. This hypothesis predicts, that for the same task (e.g. multiplication), format-related differences in the performance, such as different types of errors, would result only from differences in the encoding stage.
Support for this assumption comes, for example, from PS, a patient described by Sokol et al. (1991). PS showed a selective deficit in the retrieval of multiplication facts in absence of any peripheral problem at the comprehension or production levels. Indeed, PS’s error rate was very similar across different stimulus and response formats providing evidence for a unique mental representation underlying calculation. However, this evidence does not support unequivocally the assumptions that the internal representations are abstract and provide the basis for other numerical tasks.

Further evidence for lack of format-effects in numerical tasks comes from several experimental studies. LeFevre et al. (1988) developed a task where participants had to decide whether a target number (e.g., 3) match or not one of the two numbers (e.g., 2+3 or 2 3 ) presented immediately before the target one. Results indicated that it takes longer to reject a probe when it corresponds to the sum of the initial pairs, a finding that has been attributed to the automatic activation of addition facts (see 2.3.4 for further discussion). Moreover, LeFevre et al. showed that the interference effect was equally significant whether the initial pair was presented in Arabic or verbal format. This result would favour the hypothesis that retrieval of arithmetical fact relies upon a single amodal numerical representation. In a recent study, Blankenberger and Vorberg (1997) have drawn a similar conclusion. They investigated format-effects in simple addition (e.g., 1+2= ?, answer 3) and addition-plus-one tasks (e.g., 1+2= ?, answer 4). They found that the cost of the addition-plus-one operation was the same across problems in digit, verbal and dice format (Blankenberger & Vorberg, 1997, Experiment 1).

In line with the assumption of a unique, modality-independent semantic representation, no format effects have been so far reported in numerical tasks, such as number comparison. Indeed, standard effects in number comparison (i.e., distance effect, magnitude effect and serial-position effect) have been replicated in studies with Arabic numerals (Aiken & Williams, 1968; Banks, Fuji & Kayra-Stuart, 1976; Buckley & Gillman, 1974; Dehaene, 1989; Duncan & McFarland, 1980; Moyer & Landauer, 1967; Parkman, 1971; Sekuler & Mierkiewicz, 1977), written verbal numerals (Foltz, Poltrock & Potts, 1984; Takahashi & Green, 1983) and patterns of dots (Buckley & Gillman, 1974).
Nonetheless, the assumption that number production mechanisms could only be accessed through abstract internal representations yields several strong predictions that have been recently challenged in several neuropsychological case-studies (Cohen et al., 1994; Cipolotti, 1993, 1995; Cipolotti & Butterworth, 1995; Noel & Seron, 1995). Similarly, the claim that calculation is only mediated by abstract internal representations has been questioned and the hypothesis that number processing and calculation rely on modality specific representations has provided the basis for the development of alternative models (Campbell & Clark, 1988; Dehaene, 1992).

Overall, in the latest years the McCloskey's model has received extensive critiques. It has been also pointed out that the model is clearly not complete since it does not account for some numerical skills such as counting or conceptually-based arithmetic (Hittmair-Delazer, et al. 1994; Seron & Noel, 1992). However, this model has provided a useful taxonomy for number processing disorders and its impact on the development of this area of research is unquestionable. Its simple and constrained architecture has allowed to generate specific and straightforward predictions, against which the model has been tested giving impulse to a fruitful proliferation of neuropsychological and experimental studies.

2.1.2 Seron and Deloche’s asemantic transcoding algorithms

Even before McCloskey’s model was proposed, Deloche and Seron were interested in examining numerical transcoding in aphasic patients (Deloche & Seron, 1982a, b; 1987; Seron & Deloche, 1983, 1984). They suggested that the translation from one numerical format to a different one may be achieved by "asemantic" transcoding algorithms without computing intermediate semantic representations. In particular, they developed a detailed model of the processes underlying transcoding of written verbal numerals to Arabic numerals and vice versa. The production model proposed, involves the serial application of a set of transcoding rules that operate from left to right on each element of the numerical string. The model postulates four functionally independent and interactive components: 1) the left-to-right parsing process, that segments the lexical primitives in the number string; 2) the lexical categorisation/identification component that extracts from the lexical
primitives information about the lexical class (e.g., units, tens, teens..) and position within the class (e.g., first, second, third...); 3) the transcoding process itself, which consists of a set of rewriting rules, and finally 4) the output encoding process which consists of the production of the actual digit or verbal form. The transcoding errors in brain-damaged patients were interpreted in terms of the disruption to one or more of these components (Deloche & Seron, 1987; Seron & Deloche, 1983, 1984; Deloche, Seron & Ferrand, 1989).

Unfortunately, the authors did not provide specific evidence that a semantic representation is not activated in the transcoding process, undermining the strength of their proposal. However, the merit of Deloche and Seron’s contribution consists of having clearly formulated, for the first time, the hypothesis that transcoding tasks could be achieved without involving semantics. It does not constitute a comprehensive model of number processing since the authors focused their attention on specific transcoding mechanisms, but they did not reject the existence of comprehension and production mechanisms, clearly indispensable in many numerical tasks such as magnitude comparison or calculation.

2.1.3 Campbell and Clark’s encoding complex view

The most extreme non-modular model of number processing has been proposed by Campbell and Clark. Their interactive architecture strongly rejects the assumption of a central abstract semantic representation (Campbell & Clark 1988, 1992; Campbell, 1994; Clark & Campbell, 1991). According to the authors, number comprehension, production and calculation are mediated by multiple format- and modality-specific representations. They assume both verbal (e.g., auditory, orthographic, articulatory) and non-verbal codes (e.g. visual and analogue representations). These multiple representations are interconnected in an associative network so that any code could potentially activate other representations at any time. Indeed, multiple codes are assumed to be implicated at any point in a given task depending on excitatory and inhibitory mechanisms (but also depending on individual differences related to the learning history, use of specific strategies
and so forth). The model also postulates that all different numerical tasks involve common, rather than independent, cognitive representations and resources.

However, as specified by Campbell (1992), the denial of a central amodal representation does not entail the rejection of a numerical semantic representation, though the role of semantic representation(s) in number processing was not explicitly formulated. Within this framework, magnitude is not a unitary concept, "rather the representation of magnitude corresponds to a set of specific learned relations and processes (e.g., labelling of perceptual groups or intensities, use of counting and other basic relations to present changes in quantity)" (Campbell, 1992, p.553). Moreover, these semantic multiple representations are assumed to be modality-specific, as for example, visuo-spatial (in the form of a number line) or verbal (in the form of well-learned counting sequence), but certainly not activated automatically. In this respect, even if not neatly formulated, the model seems to assume that different transcoding routines and different representations may be activated depending on the task demands. Furthermore, "different number formats can differentially activate internal number representations and associations that are functional in performance" (Campbell & Clark, 1992, p.461). For example, they assume that Arabic numerals may be preferentially associated to magnitude code, given their higher frequency in magnitude-based tasks (e.g., calculation and number comparison), while verbal numerals may have privileged access to phonological representations. Thus, in principle, the encoding complex view postulates both semantic and asemantic transcodings but is neither explicit nor specified when and how these pathways contribute to the performance.

The encoding complex view was first proposed on the basis of a re-analysis of neuropsychological data (Campbell & Clark, 1988). Recently, support for this model came from experimental studies with normal subjects showing an effect of the presentation format on arithmetical performance^2 (Campbell & Clark, 1992; Campbell, 1994). More

---

^2 It is worth noticing that, in principle, format-specific effects in arithmetical performance may be considered evidence against a central amodal representation but little they said about the role of semantic information in this task if not explicit hypotheses are formulated.
specifically, these studies indicated that in production tasks, the presentation format (i.e., Arabic and verbal) interacted with both types of arithmetic operation and problem difficulty. Moreover, the qualitative analysis of the errors led the authors to claim that the observed differences may not simply be attributed to the encoding stage but more likely to the retrieval stage. Overall these results are difficult to accommodate within a strictly modular architecture, nevertheless, McCloskey and colleagues (McCloskey, Macaruso & Whetstone, 1992) argued that, once specific assumptions about the encoding stage are considered, these results are not clear evidence against the modular model. Specifically, one has to assume that subjects adopt a response deadline that limited the time required to respond. Since it takes longer to encode pairs of verbal numerals compared to pairs of digits, and this difference increased with number magnitude (Dehaene & Mehler, 1992), the time available to retrieve an answer is shorter for word problems than for digit problem. This would make word problems more prone to specific errors (see for a discussion Campbell, 1992 and McCloskey et al., 1992). Recent empirical evidence is in line with the McCloskey et al.’s alternative interpretation. Noel, Fias and Brysbaert (1997) have replicated and extended Campbell and Clark’s study on format-effect in arithmetic. They reported an interaction between format of presentation (Arabic versus verbal) and size of the numbers not only in a multiplication task, but also in a number matching where the same number pairs had simply to be read aloud. Furthermore, Campbell and Clark attributed the occurrence of operand intrusion errors (one of the operand appears in the error response: e.g., 5 x 4 = 24) to an interference between reading and retrieval processes: in the Noel et al.’s study this interpretation was challenged by the fact that similar error patterns were observed in English, French and Dutch speaking subjects, though the latter reverse the order of the number names (e.g., vier-en-twintig [four-and-twenty] instead of twenty-four). On the basis of this result the authors suggested that the interference effect is more likely to occur at the verbal output stage rather than at the retrieval stage.

The encoding complex view, at present, does not constitute a well structured model of number processing, on the basis of which specific and testable hypotheses may be generated (McCloskey et al., 1992). The legitimacy of this criticism has been acknowledged by the authors themselves, who however favoured a still unspecified theory that, according to
them, deals with the plasticity and complexity of numerical cognition, rather than "a relatively simple [but reductive] taxonomy of number processing subsystems" (Campbell, 1992, p.486).

2.1.4 Noel and Seron's preferred entry code hypothesis and the Intermediate semantic representation hypothesis

Recently, asemantic transcoding algorithms of the type described by Seron and Deloche have been included in the preferred entry code hypothesis proposed by Noel and Seron (1993). According to this approach, the access to number semantic representations may be accomplished either from the verbal or the Arabic code, according to the individual's idiosyncratic preference to maintaining the information in working memory in an auditory or a visual code. Thus, before any numerical task may be performed, transcoding between the input format to the preferred format would be realised through asemantic mechanisms. This hypothesis, originally proposed to account for a specific pattern of errors disclosed by a patient across different numerical tasks (Noel & Seron, 1993), is still waiting for solid empirical validation; however, the assumption that the preferred numerical format to perform numerical tasks may vary across individuals has been claimed by many authors (e.g., Campbell & Clark, 1992; Gonzales & Kolers, 1982; Marsh & Maki, 1976; Smith, 1988).

The debate about the nature of the semantic numerical information has mainly concerned the format of these representations, with views in favour of abstract amodal representations (McCloskey, 1992) and views supporting the existence of format-dependent internal representations (Campbell & Clark, 1992; Dehaene, 1992). Recently, Noel and Seron (1995, 1997) addressed the issue of the content of these representations by investigating to what extent the presentation format of a numeral may influence post-encoding processes, such as magnitude comparison or arithmetic facts retrieval. In particular, they proposed that it is the lexico-syntactic structure of the numerical input rather than the notational system per se (e.g. Arabic or verbal) to play a role beyond the peripheral stages of numerical processing. They hypothesised that, during the encoding
stage, an intermediate representation (IR) bound to the lexical primitives and the syntactic relationships between them, is activated. For example, encoding twenty-four would require the identification of the lexical primitives ‘twenty’ and ‘four’ as well as the sum relationship that connect them. The corresponding IR would thus keep track of both information in the form of “<20>+<4>” (where <20> stands for the semantic representation of the number 20); similarly, the encoding of the Roman numerals XXIV would activated the following IR: <10>+<10>+<5>-<1>. It follows that different input formats may well activate the same IR, as long as they share the same lexico-syntactic structure as, for example, the verbal numerals twenty-four and the Arabic numerals 24. Moreover, the authors did not argue that these intermediate representations differ in their format neither did they favour an amodal or modality-dependent format. In support of their hypothesis the authors reported a series of experiments that indicate that the lexico-syntactic structure of the numerals influence the performance on tasks such as verification of simple arithmetic, magnitude comparison and calculation; these tasks are generally assumed to be based on abstract (McCloskey et al., 1985) or modality specific (Dehaene, 1992; Campbell & Clark, 1992) semantic representations. Furthermore, the evidence suggests that this intermediate representation may be semantic in nature, i.e., it does not only reflect the lexical structure of the stimulus (e.g. VI as <5> and <1>) but conveys information about the quantity represented and the specific arithmetical relationships that connect them (e.g. VI as <5> + <1>). It is worth noticing that the IR assumption is not an alternative to the central amodal representation of number magnitude, rather it is assumed to be interposed between the encoding stage and this central representation.

The hypothesis of the intermediate semantic representation has been put forward on the basis of solid, even if still limited, empirical evidence. By definition, it may only elucidate the mechanisms underlying the manipulation of numerical stimuli with a specific lexico-syntactic structure, as Roman numerals and complex verbal numerals. This limitation, however, is recognised by the authors themselves who suggest that the IR is only one of the multiple semantic attributes associated with a number representation (parity status, magnitude information, power relation being among the others). It is plausible that,
depending on the task requirements, the relevant attributes of the presented numeral are all activated exerting in combination an influence on any post-encoding processing.

2.1.5 Dehaene’s triple-code model

In an attempt to reconcile Campbell and Clark’s multiple-code hypothesis and the modularity assumption postulated by McCloskey, Dehaene proposed a triple-code theory (1992). Within this framework three types of mental numerical representations are postulated: visual-Arabic number form, in which numbers are represented as string of digits, auditory-verbal number form in which numbers are represented as syntactically organised sequence of words and analogue magnitude code in which magnitude information is represented as distribution of activation on an oriented number line (Restle, 1970). Each of these representations is connected to the others through internal transcoding routes and is tied to notation-specific input and output procedures, similar to those included in the McCloskey’s model (see Figure 2.2). For example, reading of Arabic numerals would rely on a dedicated procedure that categorises them into string of digits in the visual Arabic number form (Cohen & Dehaene, 1991). Similarly, writing of Arabic numerals would employ output routines that convert the internal Arabic code into the appropriate motor programme for writing them down.

Critically, the model assumes that each representation mediates specific numerical activities. The visual-arabic representation is postulated to support multi-digit operations as well as to have exclusive access to parity information (e.g., Dehaene et al., 1993). The auditory-verbal code is supposed to mediate verbal input and output, some counting processes and also to provide the medium of representation for memorised addition and multiplication facts (Dehaene & Cohen, 1997). The analogue representation is assumed to be used in any task based on quantity information as, for example, in number comparison or approximate calculation (e.g., Cohen & Dehaene, 1991; Dehaene, 1989; Dehaene, Dupoux & Mehler, 1990). This latter is the only representation that conveys semantic information in the form of an approximate magnitude code with no syntactic sophistication.
In other words, at this level, precise information about the elements composing a given numeral, such as ‘212’, is lost and this latter is simply associated to a round (and more familiar) number such as 200 (Dehaene & Cohen, 1995, p.87). This representation is thus inadequate for transcoding tasks, such as reading or writing, but is supposed to be sufficient for estimation, magnitude comparison or approximate calculation.

Thus, the very first stage of any given task, is the translation of the input stimulus into the correspondent notation-appropriate representation. Then, via internal transcoding routines this representation would be transcoded into the appropriate internal code required by the specific task. In this respect, the triple-code model does not predict format-effects to modulate the performance given that, regardless of the stimuli format, any given numerical operation operates upon a single internal code. However, it has been suggested that some numerical tasks, such as simple arithmetic, may simultaneously activate more than one
internal representation (see for further discussion, Dehaene & Cohen, 1997). Given that each representation is linked to the others, the triple-code model posits both semantic and asemantic transcoding routes. As depicted in Figure 2.2, Arabic and verbal codes are connected by direct asemantic pathways as well as by way of routines converging onto the analogue magnitude representation. Finally, the model holds also specific hypothesis about the neuroanatomical substrates of the different numerical representations and the functional pathways interconnecting them (Dehaene & Cohen, 1995).

In support to his model, Dehaene provided multiple evidence from single-case studies as well as from experimental data from normal subjects' performance. The assumption of direct and indirect transcoding routes has been documented mainly by neuropsychological data (Cohen & Dehaene, 1991; 1995; Cohen et al., 1994). For example, Cohen et al. (1994) reported a deep dyslexic patient whose difficulties in reading numbers were constrained to unfamiliar numerals while meaningful numbers, such as famous dates, were better mastered. On the basis of this observation, the authors proposed a dual-route model, which parallels models of alphabetic reading, where familiar numerals are supposed to be processed via a lexical route and non familiar numerals via a non lexical surface route (see Chapter 5 for further discussion of this case).

In support of the distinction between precise number processing and approximation abilities, Dehaene and Cohen (1991) reported a patient who could no longer answer simple arithmetic problems neither could he classified numbers as odd or even. On the other hand, he was able to compare the relative magnitude of 1- and 2-digits numbers as well as to reject grossly incorrect solution to arithmetical problems (see also Warrington, 1982). The authors interpreted this profile as the result of a disturbed symbolic processing of numbers, which would support exact manipulation of them, and preserved analogue magnitude processing which would allow approximate computations. It is important to note that this case could also be explained within the single semantic representation framework (McCloskey), assuming a deficit at the semantic level responsible for incorrect but plausible answer in calculation. Recently Dehaene and Cohen (1997) reported two acalculic patients who, besides preserved transcoding abilities, showed somehow opposite pattern of spared
and impaired numerical knowledge. One patient showed an inability to deal with rote verbal knowledge (e.g., reciting the alphabet) including retrieval of multiplication tables, but he had no difficulties in a series of tasks based on quantitative numerical knowledge (e.g., number comparison, approximate calculation). The other patient was impaired in tasks which required manipulation of numerical quantities even if his rote arithmetical knowledge was relatively better preserved. This evidence seems in line with the triple-code assumption that approximate quantitative representations and verbal representations subserve distinct and dissociable numerical competencies. However, at a closer sight, the presumed dissociation was less clear-cut than had been predicted by the authors’ interpretation (e.g., the patient’s performance in some tasks was not in agreement with the postulated functional deficit\(^3\)). Also, at a more general level, it seems that the proposed distinction between approximate and rote numerical knowledge is, at most, insufficient to account for skilled arithmetical performance.

The assumption of an analogue magnitude representation pictured as an oriented and compressed number line, originally proposed to account for the distance effect in number comparison (i.e., the time to compare to digits is logarithmic function of the distance between them: Aiken & Williams, 1968; Moyer & Landauer, 1967; Restle, 1970) has been tested in experimental studies with normal subjects. Data from 2-digit numbers comparison indicated that, even in cases where the comparison of the decades would unable the judgement, the units had an influence on the reaction times. These results support an holistic model of number comparison, according to which decades and units are first combined in a unique analogue magnitude representation rather than processed serially at the digit level (Dehaene, 1989; Dehaene et al., 1990). Furthermore, Dehaene, Bossini and Giraux (1993) observed that in a parity judgement task, subjects were consistently faster to respond to larger numbers with their right-hand and to small number with their left-hand. In subsequent experiments the SNARC effect (Spatial-Numerical-Association of Response Codes) was shown to be unrelated to handedness or hemispheric dominance but to depend on the direction of writing. Overall, this small-left/large-right association was interpreted as

\(^{3}\) For example, patient MAR whose functional deficit was located at the level of quantitative number processing, performed well above chance in a number comparison task and in estimating quantities.
evidence for an automatic recoding of numerals as analog representations onto a left-to-right oriented number line (for further discussion see Section 2.2).

Overall, the triple-code model may be considered successful in its attempt to provide a structured and testable functional architecture of number processing. However, some of the proposed components and the procedures operating upon them are still largely unspecified. Further, if the model has the merit of explicitly referring to numerical competencies that were previously neglected (e.g., approximation, subitizing), it is far from being exhaustive. In this respect, the postulated number semantics seems particularly unsatisfactory: it is hard to conceive that numbers may only convey approximate values and indeed very little can be done with fuzzy magnitude (for a discussion, Delazer & Butterworth, 1997).

2.1.6 Cipolotti and Butterworth’s multiple-route model

Cipolotti et al. (1993; 1995; Cipolotti & Butterworth, 1995; Cipolotti, Butterworth & Warrington, 1995) proposed a modified version of the McCloskey’s model in order to account for the patients’ performance that would be problematic to interpret within a single-semantic route model. In particular, patient SF, described by Cipolotti (1995), showed a selective deficit in reading aloud Arabic numerals in absence of comprehension and production problems. Patient SAM (Cipolotti & Butterworth, 1995) presented a dissociation between impaired verbal and Arabic numeral production in transcoding tasks and preserved spoken and Arabic numerals production in calculation tasks (for further discussion see Chapter 5). In both cases, the observed difficulties to produce a particular number code was task specific. On the basis of these results the authors suggested a multiple routes model for number processing. They proposed that numeral output systems can be accessed not only via abstract semantic representations, but also through asemantic routes that bypass the abstract internal representation (Figure 2.3).

The "Arabic numeral to spoken number name route" allows the direct translation of the numerical input into spoken verbal numerals; this asemantic translation requires the activation of a fixed set of transcoding rules.
Similarly, by analogy to what has been suggested for word processing (e.g., Coltheart, Curtis, Atkins & Haller, 1993; Shallice, Warrington & McCarthy 1983) reading aloud written verbal numerals could presumably be accomplished through the print-to-sound procedures; as the writing to dictation of Arabic numerals through the verbal numerals-to-Arabic conversion rules. Finally, an asemantic phonological transcoding process is supposed to allow the repetition of spoken number words. The existence of these additional asemantic pathways, however, it is not by itself sufficient to interpret the observed results. If, in both cases, the asemantic transcoding routes were impaired or operating inefficiently, one may question why the intact semantic route was not engaged to accomplish the task. The authors proposed an explanation in terms of task demands. They postulated dedicated control mechanisms that, on the basis of task requirements, would select one specific route and inhibit the others. Thus, for example, the task of reading aloud Arabic numerals would
preferentially activate the direct route from Arabic to spoken verbal numerals and simultaneously inhibit the semantic one (see also, Cohen & Dehaene, 1995). Though the plausibility of this interpretation is not disputed, it will remain a tentative suggestion until empirically validated.

2.1.7 Single and multiple routes: further considerations

All in all, from the overview of current models of number processing, it might be concluded that the existence of both asemantic and semantic pathways has received much more consensus, at least at the theoretical level, than the single semantic route assumption. McCloskey himself (1992) acknowledged that the hypothesis of asemantic transcoding deserves further attention. In this respect, he pointed out that the transcoding of verbal numerals, whether in reading, writing or repetition is not truly a matter of contention given that all these tasks may be reasonably accomplished through the same asemantic processes postulated for word processing in general (see also, Seron and Noel, 1995). He further suggested “that analogous [direct] routes might carry out mappings between digit representations and phonological or graphemic number-word representations” (p.121) possibly through the application of an acquired set of transcoding rules (as suggested by Seron & Deloche, 1987). However, as argued by McCloskey, the plausibility of additional asemantic routes is not per se evidence of their existence. In other words, only empirical data that can unequivocally indicate the need of asemantic mechanisms may resolve the debate.

This specific issue has been thoughtfully discussed by Seron and Noel (1995) in a critical review of the neuropsychological evidence that has been so far reported in favour of the asemantic/semantic distinction. The authors pinpoint a number of methodological and theoretical problems that may undermine the legitimacy of this distinction in the realm of numbers. A major point of concern is that no double dissociation between asemantic and semantic processing has been observed until now. In fact, all reported cases performed better in task relying on semantic elaboration and no one showed the reverse dissociation (e.g., intact transcoding of Arabic numerals and impaired understanding of them). Till a
patient with selective deficit in semantic processing and preserved asemantic transcoding is described, alternative interpretations based on differential impairment in the semantic route may always be provided (see Seron & Noel, 1995 for an extensive discussion). Even more questionable is the distinction between semantic and asemantic tasks; given that all well-formed numbers represent a numerosity, any so-called asemantic task, such as reading, can be turned into a semantic one by addressing the represented numerosity. Similarly, the repeatedly claimed parallels between the semantic and asemantic reading routes employed in alphabetical material and the reading routes employed in reading Arabic notation are also questionable. As Seron and Noel (1995) state, every Arabic numeral can be transcoded through a non-lexical route, since no irregular numerals exist. Thus, even an asemantic surface route should transcode all possible numerals correctly. On the other hand, if the asemantic surface route is damaged, the semantic route should still be able to interpret all numerals, since they all represent a numerosity. All possible combinations of single digits represent legal multi-digit numerals, denoting a numerosity, while the same is not true for alphabetical material, where randomly generated combinations of letters can result in nonwords. Overall, these critical questions do not make the semantic/asemantic debate less worthwhile; on the contrary, they emphasise the need of additional data, of more caution in drawing conclusions from them and, above all, of more refined theoretical assumptions about number representation and mechanisms operating upon them.

The paramount contribution of neuropsychological data in elucidating the mechanisms involved in number processing is clear. The investigation of cerebral-lesioned patients may indeed allow observations that would be hardly possible to detect in normal subjects’ performance. In particular, it is very likely that the postulated alternative routines and representations are normally activated simultaneously, contributing, to a different extent, to a given task: this could make the investigation of these issues in normal subjects extremely challenging. On the other hand, this latter offers undeniable advantages over the studies of cerebral-lesioned patients (e.g., patients may develop strategies to compensate for their deficit) and provides a powerful method for testing theoretical assumptions derived from neuropsychological observations (e.g., the IR hypothesis in Noel & Seron, 1993, 1997).
As already pointed out, the issue of single or multiple routes in number processing has been indirectly addressed by investigating format-effects in arithmetic (Campbell & Clark, 1992; Campbell, 1994; Noel & Seron, 1992; Noel, Fias & Brysbaert, 1997). The overall evidence favour the hypothesis that arithmetical-fact retrieval is mediated by a single internal representation, whether amodal, auditory-verbal or of other nature still remains to be clarified. Regardless of how advantageous this approach might be, patterns of results and their interpretation are very much complicated by the complexity of the task itself (i.e., calculation).

Indeed, a number of alternative experimental paradigms have been recently developed with the aim of directly tackling the issue of semantic/asemantic transcoding in number processing. In line with the suggestion that the alternative transcoding routes may be recruited depending on the task demands (Cipolotti, 1994; Cohen & Dehaene, 1995), in studies with normal subjects task instructions are usually manipulated to favour one or the other processing type.

This specific criterion was adopted by Dehaene and Akhavein (1995) in a series of experiments where participants performed same-different judgement tasks. Pairs of numbers written in the same (pure trials, e.g., 2 2, ONE NINE) or different notations (mixed trials, e.g., 2 TWO, 1 NINE), had to be matched at the semantic level (numerical matching task) or at a physical level (physical matching task). In the numerical matching task, subjects were required to judge if the two stimuli represented the same number, ignoring variation in their notation (e.g., '2 TWO' had to be answered same); in the physical matching task they had to take their decision on the basis of the physical characteristics of the stimuli (e.g., '2 TWO' had to be answered different). The symbolic distance effect (i.e., the time to compare two numbers is an inverse function of the numerical distance between them; Moyer & Landauer, 1967) was employed to measure semantic processing. In particular, if subjects were accessing number magnitude, they should have answered slower to numerically close pairs (e.g., 1 2) than to far pairs (e.g., 1 9) pairs, though both pairs had to be responded as different. The hypothesis of asemantic transcoding was tested by looking at the response profile in mixed trials (e.g., 1 ONE) in the physical matching task and in particular, at whether participants could match an Arabic
numeral to its corresponding verbal form without showing any evidence of semantic processing. It was assumed that the absence of a distance effect together with a slow and/or inaccurate performance in answering different to Arabic-verbal pairs representing the same quantity (e.g., 2 TWO) would be evidence for asemantic transcoding. The reaction times data failed to show any effect, but a trend for responding same to mixed pairs representing the same number was observed. Even though this result is rather weak, Dehaene and Akhavein study seems to favour the hypothesis of an addition direct asemantic mapping between Arabic and verbal numerals (the methodological drawbacks of this study are thoroughly discussed in Chapter 4).

Adopting an analogue of the standard Stroop-task in which an irrelevant stimulus is presented adjacent to the target one (flanker task; Dallas & Merikle, 1976; Glaser & Glaser, 1989; see MacLeod 1991 for a review), Fias, Brysbaert and d'Ydewalle (1996) have investigated the mutual influence between Arabic and verbal numerals in both semantic and asemantic tasks. Subjects were presented simultaneously an Arabic and a Verbal numeral (range 0 to 9) that could represent the same quantity (congruent condition) or a different quantity (incongruent condition). For half of the subjects the target stimulus was the Arabic numeral, for the other half the verbal numeral. When the task consisted of a parity judgement, a task that presumably requires access to the semantic system, the distractor influenced the processing of the target stimuli whatever its notation and the interference effect was indeed of the same size in both notations. Moreover, since there was no effect of the response congruency within incongruent trials (i.e., whether target and distractor were of the same, e.g. 2 FOUR or different parity, e.g., 2 FIVE, made no difference) the source of interference was located at an early stage of processing. Critically, when the task simply consisted in reading aloud the target numerals, the interference effect was highly asymmetrical: only verbal distractors exerted an influence on Arabic digits slowing down the reading times.

These results were interpreted with reference to the dominance rule theory (Glaser & Glaser, 1989), according to which Stroop interference occurs when the irrelevant dimension has privileged access to the system at the level of which the appropriate response is selected. Thus, results from the parity judgement task would suggest that both Arabic
and verbal numerals have equal access to the semantic system (at the level of which the odd/even classification may take place); on the other hand, the uni-directional interference from verbal to Arabic numerals in the naming task would be evidence for a privileged access of those latter to the spoken verbal output system. These conclusions were further corroborated by the fact that in a phoneme monitoring task (i.e. is there a /e/-sound in the name of a numerical stimulus?), Fias et al. found a SNARC effect, considered as evidence for semantic processing (see also Section 2.2), only with Arabic numerals but not with verbal numerals: thus, while Arabic to verbal (phonological) numerals evidenced access to semantic information, for verbal numerals the task seemed to be successfully mediated by asemantic transcoding.

To account for the divergence of results between Arabic and verbal numerals, Fias et al. (1996) proposed an hybrid model of number processing that postulates the existence of both asemantic and semantic routes from written verbal numerals to spoken verbal numerals but a single semantic route for transcoding of Arabic numerals. Given that the existence of multiple routes for processing of verbal numerals has been acknowledged by all authors (though McCloskey has not provided any formalisation of it in his model), what Fias et al. emphasised is the pivotal role of semantic access for the processing of Arabic numerals.

Overall, Fias et al.’s results clearly indicated, maybe not so surprisingly, that verbal numerals have privileged access to phonological information. They also show, together with several other studies (e.g., Dehaene & Akhavein, 1995, Henik & Tzelgov, 1982; Tzelgov et al., 1992), that Arabic numerals are processed semantically, even when magnitude information is irrelevant to the task. It is still possible, however, that semantic effects in processing of Arabic numerals are determined by the dominance of the indirect route over the direct asemantic one, rather than by the existence of a single transcoding pathway. This would be even more the case if the semantic access for Arabic numerals is largely automatic as repeatedly suggested. An autonomous and fast semantic access may thus mask the role of a putative asemantic route (Dehaene & Akhavein, 1995). This might be an important confounding factor in the evaluation of asemantic transcoding and certainly

37
deserves further attention. In the next session the experimental research focused on autonomy in number semantic processing is examined.

2.2 AUTONOMOUS PROCESSING OF NUMBER MAGNITUDE

As already pointed out, numbers may be used in several different contexts and accordingly, convey different meanings, only some of which are truly numerical (e.g., cardinal meaning). Nevertheless it is undeniable that the magnitude is the most distinctive attribute of a number. Indeed, both Arabic and verbal numerals are symbols that are arbitrarily associated to the meaning they convey, i.e., the relation between the Arabic numeral 2 and the quantity <2> must be learned and it will be later on discussed how lengthy this acquisition process may be (see Section 2.3). However, once this association is established it seems hardly possible to ignore it. Namely, the mere presentation of an Arabic numeral seems to determine the activation of its magnitude representation. Recent findings suggest that this property extends to written verbal numerals (Dehaene & Akhavein, 1995; but see Fias et al., 1996). Ground for this claim comes from the fact that magnitude-related effects have been found in several tasks in which number meaning was task-irrelevant (e.g., Duncan & McFarland 1980; Henik & Tzelgov, 1982; Dehaene et al., 1993).

A clarification of terms needs to be introduced. Within these studies, the terms automatic and autonomous have both been used to refer to a process that takes place even when irrelevant to the task. However, when clearly stated, the access to numerical information has been said to be automatic by virtue of being autonomous, i.e., it begins and runs to completion without intention (Zbrodoff & Logan, 1986). Autonomy is thus a property of automatic processes (i.e., processes that are fast, effortless, autonomous and

---

4 The term magnitude is here used to refer to the numerosity a number stands for and does not imply any reference to the analog magnitude representation has depicted in the Dehaene's model.
unconscious; e.g., Logan, 1980; Posner & Snyder, 1975; Shiffrin & Schneider, 1977) and is the only one relevant to the present discussion.

Thus, the evidence for the autonomous processing of magnitude information comes from studies that reported semantic effects in tasks where magnitude information was irrelevant to the performance.

So far, two different effects have been used as markers of semantic number processing, i.e., the distance effect and the SNARC effect.

The distance effect refers to the inverse relation between the time required to compare two numbers and the numerical distance between them (e.g., Moyer & Landauer, 1967). For example, it takes longer to select the larger (or smaller) between 2 and 3 than between 2 and 9, and this effect has been also reproduced with 2-digit numbers (e.g., Dehaene, 1989; Dehaene et al., 1990; Hinrichs, Yurko & Hu, 1981). Interestingly, the effect seems to resist extensive practice (Dehaene et al., 1990; Fairbank, 1969). Moreover, the distance effect is not restricted to number comparison (e.g., Banks et al., 1976; Buckley & Gillman, 1974; Duncan & McFarland, 1980; Foltz et al., 1984; Parkman, 1971; Sekuler, Rubin & Amstrong, 1971) but occurs in any comparative judgement of physical parameters (e.g., length, weight), in comparison of semantic linear orderings (e.g., letters of the alphabet, Lovelace & Snodgrass, 1971; Parkman, 1971; months of the years, Friedman, 1974; time, quality and temperature, Holyoak & Walker, 1976), as well as in memorial comparison (e.g. size comparison of named animals, e.g. Paivio, 1975). In all these cases, the response time is a logarithmic function of the distance (numerical or physical) between the two stimuli to be compared. On the basis of this evidence, it has been suggested that numbers are not compared at the symbolic level but converted to analogue magnitude before the comparison takes place (Moyer & Landauer, 1967; Restle, 1970; Gallistel & Gelman, 1992; Dehaene, 1989). In this perspective, the distance effect should be a reliable index of access to magnitude information; however its semantic origin has been recently called into question and tested against an alternative hypothesis. As suggested by Dehaene and Akhhavein (1995), one cannot exclude the possibility that the distance effect may partially arise from the intralexical association between consecutive or close numbers in the counting sequence. According to this hypothesis, the effect would be a lexical distance rather than a
purely semantic distance effect. However, the authors showed that the distance effect was still significant when the lexical distance between the to be compared numbers was kept constant (Dehaene & Akhavein, 1995; Experiment 3). These findings favour a semantic interpretation of the distance effect.

The SNARC effect (see also 2.2.5) refers to a numerical analog of the Simon effect according to which large numbers preferentially elicited a rightward response and small numbers a leftward response. This Spatial-Numerical Association of Response Codes has been interpreted as a spatial-congruency between the response side (left and right side of the extracorporal space) and the relative position of the represented numerical quantity on a mental oriented number-line (Dehaene et al., 1993). This effect has been firstly identified and extensively investigated by Dehaene et al. (1993), in a series of experiments where subjects had simply to performed a parity judgement task. Regardless of their parity status, large numbers were responded faster with the right-hand key and small numbers were responded faster with the left-hand key. This effect was equally significant for Arabic and verbal numerals (Dehaene et al., 1993; see also, Fias et al., 1996) but tends to disappear with numbers larger than 10, whatever the notation. This suggests that small numbers, i.e., single digits, have privileged access to magnitude representation, either because they are more frequent or more easy to represent mentally (Dehaene & Mehler, 1992). Critically, this effect was shown to depend on the relative magnitude of the number with respect to the tested range, rather than to the absolute characteristic of the numbers (e.g., 4 elicited a faster rightward response when presented within the 0-4 interval and a leftward response within the 4-9 interval). The semantic nature of the SNARC effect was verified by looking whether or not it was still significant in a consonant-vowel letters classification task. Letters, though ordered in a overlearned sequence, did not yield any response-side preference, suggesting that the SNARC effect was indeed related to the cardinal rather than

---

5 In principle, letters constitute the ideal control stimuli to verify the ordinal vs cardinal source of the SNARC effect. However, in order to balance the proportion of consonant and vowel, the actual set of stimuli used by Dehaene et al. did not included a sequence of consecutive letters. Given that the alphabet is overall less familiar than the number sequence 1 to 9, we doubt that the set of letters and set of number used might be adequately compared.
to the ordinal representation of numbers. However, the extent to which the properties of this representation are invariant and ubiquitous is still an open question. Indeed, it has been recently shown that the response-side x magnitude interaction may be reversed simply by asking the subjects to conceive of numbers as hours on a clock' face (Bachtold, Baumuller & Brugger, 1998). On the basis of these findings, the authors suggested that the response-side congruency depends more on what side of the representational space a number is associated with rather than on its relative magnitude.

At this point, we need to evaluate the adequacy of the specific experimental paradigms where the aforementioned semantic effects were claimed to indicate autonomous access to magnitude information. Numerous studies addressed this issue, but when carefully evaluated the evidence they provide in favour of autonomous semantic processing is not always definitive. The major concern lays on the nature of the tasks adopted and, more specifically, on the extent to which numerical information was truly irrelevant to it. The paradigms employed vary widely along several critical dimensions and a detailed analysis will consent to draw more careful conclusions.

2.2.1 SNARC effect in non-magnitude based tasks

When a SNARC effect was first reported in a parity judgement task, Dehaene and colleagues (1993) claimed that it was a clear index of automatic access to magnitude information, given that "numerical magnitude is mathematically irrelevant when judging the parity of a number" (p. 380). However, both magnitude and parity are semantic dimensions, thus requiring a higher processing; moreover, since knowledge of magnitude precedes knowledge of parity in development (Miller & Gelman, 1983) and since magnitude is more readily available than parity (Sudevan & Taylor, 1987), its effect on a parity task is less than surprising. It is nevertheless true that the SNARC effect emerges with impressive regularity in odd/even classification tasks (Fias et al., 1996; Huha, Berch & Krikoran, 1995; Iversen & Willmes, 1998).

More convincing evidence comes from the occurrence of the SNARC effect in tasks that require a purely phonological or perceptual processing of Arabic numerals. Thus, large
numbers still yielded a faster rightward response and small numbers a faster leftward response, even if the task required simply to detect the presence of a specific phoneme in the corresponding number name (Fias et al., 1996) or to classify the first letter of the corresponding number name as vowel or consonant (Huha et al., 1995). Similar results were obtained in a perceptual task that required to classify Arabic numerals as symmetrical (e.g., 8) or asymmetrical (e.g., 9) (Huha et al., 1995). However, in this case, the SNARC effect was much less pronounced and it was yielded by only half of the numerical stimuli, suggesting caution in drawing firm conclusions.

2.2.2 Numerical versions of the STROOP-task

A few studies have adopted a Stroop-like procedure to investigate the autonomous property of number processing (e.g., Henik & Tzelgov, 1982; Tzelgov et al., 1992). Interestingly, numerical versions of the Stroop-task appeared in the literature as early as the 60s and since then, various modifications of the task have been introduced. However, in these early studies, numbers were only incidentally used as stimuli but the primarily objects of investigation were the mechanisms of selective attention and automaticity. Nevertheless, some of these studies may well shed some light on the way in which magnitude information is processed when irrelevant to the task as it will be later on discussed.

It is only recently that the numerical-Stroop paradigms have been used to directly investigate different issues in number processing, as for example notational effects in number processing (Besner & Coltheart, 1979; Foltz et al., 1984; Vaid & Corina, 1989) and, more critically to the present review, the autonomous access to number magnitude (Henik & Tzelgov, 1982; Tzelgov, et al., 1992; Algom et al., 1996).

Two different numerical-Stroop paradigms have been used so far, which differ in the nature of the conflicting dimensions adopted: in the numerosity-Stroop task, the numerical value represented by a digit could match or mismatch the number of numerical symbols displayed (e.g., 2 2 2 versus 2 2; the two dimensions are not integrated); in the number-Stroop task, the physical and numerical size of the digit are put in competition (e.g., 1 5, the two dimensions are integrated). Overall, in both cases, magnitude information is never
totally irrelevant to the tasks, thus the paradigms do not offer a strict test for the autonomy of numerical semantic access; however, in some cases the evidence provided is extremely compelling (e.g., Tzelgov et al., 1992).

**Numerosity-Stroop task**

In the numerosity-Stroop task, participants are presented with a set of numbers, in which the numerical value represented by the digit may or may not correspond to the numerosity of the set. The task requirement may be either to name the numeral or to name the numerosity of the set. The critical condition for evaluating the autonomous processing of number magnitude is the enumeration task. In particular, if the enumeration is slowed down by the incongruity of the number value (e.g., it takes longer to enumerate three Twos, e.g., "2 2 2" than three Threes, e.g., "3 3 3") this would suggest that the digit has been recognised, even if not strictly relevant to the task. This result has been consistently found across different studies (Morton, 1969a, b; Flowers, Werner & Polansky, 1979; Shor 1971; Fox, Shor & Steinman, 1971; Hock & Petrasek, 1973; Windes, 1968). However, little can be said from these results, since it is conceivable that the interference may simply be determined by the mismatch between the name of the numeral and the named numerosity, with no need to assume any semantic access to the magnitude represented by the digit. This issue has been addressed in Hock and Petrasek's (1973) study. The authors’ attention was directed to establishing the extent to which the amount of interference depends on the 'semantic' relation between relevant and irrelevant dimensions. In a first experiment, the relevant dimension was the size of a displayed ellipse (size range, 1 to 6) and the irrelevant dimension was the numerical value of a number printed in the centre of the ellipse (range, 1 to 6). The interference effect was found to be inversely correlated with the numerical distance between the displayed digit and the ellipse size. Interestingly, the same distance effect was also reported in a numerosities-Stroop task, indicating that the interference caused by the irrelevant numerals was indeed of a semantic nature.
**Number-Stroop task**

In the number-Stroop task, subjects are requested to compare either the numerical size or the physical size of numbers varying along both dimensions. In the numerical comparison, they have to select the larger (or smaller) in magnitude between the two numbers; in the physical comparison they have to select the larger (or smaller) digit in physical size, ignoring semantic information. Trials may be congruent, when the numerically larger number is physically larger (e.g., 2  6); incongruent, when the numerically larger number is physically smaller (e.g., 6 2); a neutral condition may also be included (numbers are displayed in the same size; e.g., 2  6 or the same numbers are displayed in different size e.g., 2  2, for numerical and physical tasks respectively). The size congruity effect, originally reported by Paivio (1975) in a picture comparison task, refers to the slower RTs yielded by incongruent trials compared to neutral and congruent ones. More importantly, it occurs in both numerical and physical comparisons: not only the incongruent physical size interferes with comparison of the numerical size (e.g., Besner & Coltheart, 1979; Foltz, Poltrock & Potts, 1984; Hatta, 1983; Peereman & Holender, 1984; Takahashi & Green 1983; Vaid, 1985; Vaid & Corina, 1989) but also the numerical size interferes with comparison of physical size (Henik & Tzelgov, 1982; Tzelgov et al., 1992).

In the majority of these studies, the physical size of the stimuli was manipulated as irrelevant dimension and the actual comparison was based on their numerical (semantic) size. The magnitude of the Stroop effect, thus, was primarily used to measure speed of semantic access, and in particular to evaluate the effect of number notation in accessing magnitude information (e.g., Besner & Coltheart, 1979; Vaid, 1985; Vaid & Corina, 1989; Takahashi & Green 1983).

More critical to the issue of the autonomy of numerical processing are those studies where the physical size is the relevant dimension to the task. This specific task offers a relatively stringent condition for testing the above issue given that, even assuming that physical and numerical sizes are processed in parallel, they do so at different rate; the physical size, being a perceptual dimension, is likely to be extracted prior to the numerical (semantic) size. Accordingly with several studies on Stroop-like interference (e.g.,
MacLeod & Dunbar, 1988), the congruity effect in physical size comparison is not simply accounted for in terms of relative difference in the speed of processing. Physical comparison is clearly faster than numerical comparison, yet the congruity effect is rather symmetrical. A model that invokes both the degree of automaticity and the relative strength of activation of the competing processes as factors modulating interference seems, at present, more adequate (e.g., Cohen, Dunbar & McClelland, 1990).

Overall, only three studies adopted both numerical and physical sizes as comparative dimensions and the findings on the physical task are worth a careful examination. Henik and Tzelgov (1982) were the first to report a size congruency effect in a physical size comparison task. In a first experiment they found that incongruent trials yielded slower reaction times than congruent ones when the comparison required simply to attend to the physical size of the numerals. This evidence indicated that the symbolic value of the digits was accessed and interfered with the perceptual comparison, regardless of the evidence that processing of the numerical size took longer than processing of the physical size. These results were replicated in a second experiment where the authors also explored the nature (fine-grained or coarse) of the numerical information accessed in autonomous condition. To this end, numerically close (i.e., 2 units apart) and numerically far (i.e., 4 units apart) pairs were presented in both congruent and incongruent conditions, the hypothesis being that an effect of the distance on the performance would suggest a refined processing of the numerical information. In this condition, however, the distance effect was expected to be in the inverse direction: the incongruity between physical and numerical dimensions is indeed larger when the digits are numerically far apart. As expected, far pairs yielded a larger congruency effect and overall they were answered slower than numerically close pairs. On the basis of these findings, the authors suggested that physical and numerical information were extracted in parallel and that the latter did not simply consist of a crude small/large categorisation but in a relatively detailed semantic elaboration.

---

6 Tought all three studies included both numerical and physical size comparisons, only the latter is strictly relevant to the present discussion and will be here considered.
Similar results were obtained by Peereman and Holender (1984) in a study primarily concerned with the differences in processing alphabetic (verbal numerals) and ideographic (Arabic numerals) stimuli. They also found that the irrelevant numerical dimension interfered with the comparison of the physical size of the stimuli, even though this effect was limited to the processing of Arabic numerals; with regards to the effect of numerical distance, the analysis did not disclose any regular pattern.

More recently, Tzelgov et al. (Tzelgov, Mayer & Henik, 1992), in a thorough study based on the numerical-Stroop procedure, expanded the issue of autonomous number processing addressing the question whether the same numerical information is accessed under intentional and autonomous conditions. In particular, on the basis of what they called the ‘independent encoding postulate’ according to which the comparison process follows the recoding of two objects into an internal scale, they predicted that numbers are first, simply classified as small (1, 2, 3 and 4) or large (6, 7, 8 and 9) and only at a second stage, more precisely associated to their specific magnitude. This dichotomous classification would depend on a basic property of the decimal system, and specifically that among numbers from 1 to 9, 5 has a special status of midpoint. One of the reasons for attributing to five such a status would be, for example, that once all digits from 1 to 9 are paired with each other, 5 is the only one to be, with equal frequency, the smaller and the larger of the pair. On the basis of this assumption they suggested that size congruity has to be considered an intradigit, rather than an interdigit, property; thus, its effect should equally characterise number comparison in both selection (i.e., a pair of numerals are presented to be compared) and classification paradigms (i.e., a single digit is presented to be classified as larger of smaller than a fixed standard). Furthermore, Tzelgov et al. put forward the hypothesis that when autonomously accessed, the numerical information readily available mainly consists of a coarse small-large classification of the stimuli and this would prevent a standard numerical distance effect emerging. For example, when compared to a fixed

---

7 Tzelgov et al. seem to suggest that in comparison with a fixed standard, the small/large classification does not involve any comparison stage. I agree with the authors that the nature of the tasks, i.e. selection and classification comparison, are substantially different; however the presence of distance effect strongly suggest that, whether necessary or not, a comparison stage is invoked also in the classification with fixed standard. In this respect, Experiment 2 certainly provide a more appropriate test for their hypothesis.
standard, e.g. 5, the numbers 1 and 3 would yield similar responses being both classified as 'small', regardless of their relative numerical distance from 5 (i.e., a distance of four and two respectively). These predictions were confirmed in two experiments where digits, varying along both numerical and physical sizes, had to be classified as physically smaller or larger than a fixed internal standard. Interestingly, similar results occurred when the internal standard was a 5 of intermediate physical size (Tzelgov et al., Experiment 1) as well as an intermediate-size rectangle (Tzelgov et al., Experiment 2). In this latter condition, the internal standard does not contain information on the irrelevant dimension (i.e., numerical size) thus providing further support for the intradigit origin of the congruency effect. Finally, the aforementioned hypotheses were further tested in a selection task, in which two digits were simultaneously presented to be compared for their physical size (Experiment 3). According to the independent encoding postulate, pairs of numbers may be classified as unilateral, when both digits belong to the same (small or large) category (e.g., 2 3; 7 8) or as bilateral, when the two digits belong to different categories (e.g., 2 8); a third set includes pairs with 5. In line with previous results, a size congruency effect was found, but it was of a larger size in bilateral pairs than in unilateral pairs, albeit in the latter it was still significant. Thus, even though within unilateral and bilateral pairs no effect of numerical distance was detected, suggesting that refined magnitude representations were not readily available, the presence of the congruency effect in unilateral pairs indicated that the access to those representations was somehow initiated. Interestingly, pairs with the digit 5 yielded responses similar to bilateral pairs, supporting the hypothesis that the number 5, even when semantic access is autonomously activated, may equally belong to the small and large categories depending on the context.

Overall, Tzelgov et al.'s results corroborate the hypothesis that number meaning may be autonomously accessed; however, they also point to the qualitative difference between intentional and autonomously activated numerical information. In particular, following Logan's (1988) two-process theory of skill performance, they proposed that autonomous activation of numerical knowledge results in a 'memory-based' (i.e., retrieved from memory in a single-step) small/large categorisation process; on the other hand, intentional activation would rely on a precise 'algorithm-based' (i.e., computed on line) mapping of the
stimulus on a number scale. Tzelgov et al. suggested that, though both processes acquired autonomous property in the course of skill acquisition (Logan, 1988), the 'memory-based' component would dominate over 'algorithm-based' component when numerical knowledge is irrelevant to the task; this latter would still be activated but too slow to modulate the performance (Figure 2.4). This last assumption would account for the size congruency effect emerged in the unilateral pairs in absence of a distance effect.

Figure 2.4. Schematic representation of the processes underlying intentional and unintentional number processing as proposed by Tzelgov et al. (1992).

Tzelgov et al.'s study provides an intriguing interpretation for the multiple mechanisms involved in number recognition; however some of their suggestions are still tentative and waiting for further empirical validation. For example, the way in which the 'algorithm-based' component operates in autonomous conditions, determining size congruity effect to occur in both unilateral and bilateral pairs but preventing any distance effect, still remains to be specified. Moreover, the absolute property of the 'memory-based' classification component seems, at least, questionable (but see Algom et al., 1996, Experiment 4). It is reasonable to conceive that the small and large status of a number depends, to some extent.
upon the considered numerical interval, as also indicated by some empirical evidence (Dehaene et al., 1993). The authors seem to assume that in condition of autonomous processing, this ‘scaling’ does not occur and we acknowledge that such an evidence has not been yet reported. However, it is not clear how the proposed dichotomous categorisation may efficiently operate according to the task demands.

Overall, the studies which adopted the number-Stroop paradigm provide support for the hypothesis of autonomous processing of magnitude information. However, there are several reasons that suggest caution in drawing conclusions. First, as already pointed out, in a physical size comparison task, the numerical dimension is not merely irrelevant to the task, i.e., it is not needed for an accurate performance, but is also semantically related to the relevant dimension; specifically, it conveys meaning that conflicts with the meaning of the intended dimension (similarly to the standard word-colour Stroop task). In principle, this latter provides a more stringent test for mechanisms of selective attention compared to a dimension that is only irrelevant to the task (e.g., ignoring numerical size when the relevant dimension is the colour of the stimuli).

Second, even if physical size, similarly to other perceptual dimension, is likely to be processed early in visual processing, it has been shown that its analysis does not prevent access to semantic information (Boucart & Humphreys, 1994). In particular, studies on object recognition indicate that automatic access to semantic information occurs when the physical size of the object has to be attended; similar results have been reported when orientation or global shape are the critical dimensions, suggesting that semantic access is unavoidable when global analysis of the stimulus is required (Boucart & Humphreys, 1994, 1992; Boucart, Humphreys & Lorenceau, 1995).

Third, the studies so far reported differ along the main factors of numerical size (i.e., range of numbers) and physical size (relative dimension of the digits presented). It is very likely that the amount of interference is partially related to the distinctiveness of the two dimensions: thus for example, if the difference between physically small and physically large stimuli is minimal, the selection process would take longer, increasing the probability for semantic information to be accessed. This critical point has been addressed by Algom et al.
(Algom, Dekel & Pansky, 1996) in a study where the interaction between physical and numerical sizes of numerals has been extensively tested under a variety of experimental conditions. The main concern of Algom et al. was to control for the level of discriminability of the numerical and physical dimensions, given that evidence from the Stroop literature (Melara and Mounts, 1993) indicates that interference effect critically depends upon this factor. Discriminability is defined as an intradimensional index, which specifies the psychological difference separating two stimulus values along a dimension (Melara & Mounts, 1993). So far, multiple numerical values (e.g., numbers from 2 to 8) were always contrasted with dichotomic physical values (i.e., small and large) and never was controlled that numerical and physical differences were equally discriminable. In a series of four experiments using a speeded classification paradigm (Garner, 1974), Algom et al. showed that when numerical and physical sizes are adequately matched for discriminability, they constitute separable dimensions, namely they can be selectively attended with no interference arising from the irrelevant dimension. According to the authors, these findings undermine the assumption of a mandatory access to magnitude information, and more generally suggest that Arabic numerals, when advantageous for the task, may be successfully treated as mere graphic stimuli. Nevertheless, it seems plausible to assume that, given the frequency with which numbers are used in meaningful contexts, semantic processing is indeed the preferred and dominant mode.

2.2.3 Number-matching tasks

In this task, subjects are simply required to decide whether two numbers are identical or different. Albeit both numerically close and far pairs have to be answered different a significant distance effect has been found (Duncan & McFarland, 1980; Dehaene & Akhavein, 1995). Thus, even though the task may be performed on the basis of perceptual information of the stimuli, the evidence suggests that numbers are semantically processed. However, in this paradigm, the symbolic value of a number and its physical shape are confounded dimensions, thus, attending to the semantic meaning may only increase the accuracy of the task. This factor calls into question the strength of the evidence for
autonomous semantic access. However, the selection of the dimension along which same-
different judgements are performed may be critical to this issue. For example, Dehaene and
Akhavein’s physical matching task, maximising the role of physical similarity and reducing
the facilitation effect from semantic processing (e.g., “2 two” required a different
response), certainly provides a more stringent test for autonomous processing.

It is worth noticing that exception to the obligatory activation of numerical magnitude
in same/different tasks has been reported. In particular, Garner et al. (Garner, Podgrony &
Frasca, 1982) failed to observed any distance effect when a set of number pairs had to be
compared for physical identity, but also for same odd/even status and same curvilinearity.
One possible explanation for these results may rely on the methodological differences
across the studies, and specifically to the range of number pairs employed. This issue will
be expanded and further discussed in Chapter 4.

2.2.4 Additional studies

Additional evidence for the primacy of magnitude information in numbers processing
comes from priming and short-term memory studies. In a short-term memory study Morin,
Derosa and Stultz (1967) observed that participants, after memorising a set of numbers
(e.g., 8, 6, 7, 5), were slower to reject a probe number when it was close in magnitude to
the range of the others (e.g., 4) than when it was far (e.g., 1). In a digit naming task,
Marcel and Forrin (1974) reported that the priming effect from one trial to the next one
varied systematically with the distance between the two digits. In this study, the possibility
that the distance effect arose from a lexical rather than from a semantic level seems,
however, hard to exclude. Similarly, in a letter-digit classification task, den Heyer and
Briand (1986) reported that the amount of facilitation varied inversely with the numerical
distance between the prime and the target number. Moreover, LeFevre et al. (1988, 1991)
reported a semantic-related effect in a number matching tasks where participants had to
verify the presence of a target number (e.g., 3) in a previously presented pair (e.g., 2 3).
Adult participants, consistently showed longer reaction times to reject a probe that
corresponds to the sum of the initial pairs compared to an unrelated probe. Moreover,
longer reaction times were also reported for target digits that were close in magnitude to the initial pair (e.g., 3+1=5) than for target digits that were far apart (e.g., 3+1=7). This pattern of results supports not only the semantic processing of the presented numbers but also the automatic activation of their sum. More recently, Brysbaert (1995) reported both number magnitude effect and number priming effect in naming Arabic numerals from 1 to 99.

2.2.5 Format effect

Little attention has been so far paid to exploring whether or not autonomous access to magnitude information is bound to a specific stimuli format. A minority of studies have used both Arabic and verbal numerals to investigate this issue and the evidence so far is not straightforward. Among the numerical-Stroop studies, physical size comparison of written verbal numerals has been performed only once, and a null effect of congruency was reported (Peereman & Holender, 1984). Additional evidence of this sort was provided by Fias et al.'s study (1996, see also Section 2.2.1). In a phoneme monitoring task based on the transcoding from written verbal to spoken verbal numerals no SNARC effect was observed. These results seem to suggest that autonomous access to semantic information does not occur for verbal numerals, or if it does so, it may be less efficient than in the case of Arabic numerals. This last alternative would be supported by the findings of Dehaene and Akhhavein (1995) that indicate a distance effect for pairs of written verbal numerals in physical matching task. The reliability of these results will be further discussed in Chapter 4; however, it seems cautious to conclude that the evidence in favour of automatic activation of magnitude information from verbal input is rather scanty.

2.2.6 Further considerations

To summarise, several studies have so far addressed the issue of autonomous processing of numerical information. Not all the evidence is equally strong, but overall the data favour the hypothesis that Arabic numerals do initiate semantic processing in an autonomous fashion. Magnitude information seems a salient and distinctive attribute of an
As suggested by Zbrodoff and Logan (1986), a process may be autonomous to different degree: a completely autonomous process should initiate and proceed similarly under intentional and unintentional conditions; on the other hand when qualitative differences emerged they indicate that the criteria for autonomy are only partially met. In this latter case, they claim, semantic activation, for instance, may occur but being less effective or extensive. With regard to numerical processing, it has been suggested that intentional and unintentional semantic processing may indeed differ: the former would activate a refined and distinctive numerical representation on a number scale (or number line), the latter would be dominated by a coarse dichotomous classification of numbers in small and large categories (Tzelgov et al., 1992). Support for this hypothesis comes from physical-size comparison of Arabic numerals; however, contrasting evidence arises from paradigms employing the SNARC effect as a marker of semantic access. In this latter, the magnitude information available under unintentional conditions, seems more than a mere small/large classification, given that the size of the effect is function of the magnitude of the number (Fias et al., 1996). Yet, we must be cautious in comparing directly different experimental paradigms and semantic effects. For example, the paradigms measuring the SNARC effects are all classification tasks, namely, numbers are presented one a time to be classified according to a given criterion and all single-digit numerals are presented. The paradigms where the distance effect is used as an index of semantic processing, require the subject to select between or compare two digits simultaneously presented and, often, only a repeated set of number pairs is employed. Moreover, both brief and unlimited presentation of the stimuli has been used (e.g., Dehaene & Akhavein, 1995; Tzelgov et al., 1992). These methodological differences may critically modulate the type of processing and, in turn, the size of the effects measured.
2.3 Number Processing in Children

The acquisition of the notational systems in the number domain is one facet of the more general process of learning about symbolic representations (e.g., words) and their underlying meanings. However, in the number domain this process may be particularly challenging for at least two reasons. First, numbers (i.e., cardinals) do not refer to individual items, neither to a specific attribute, but to a property of a set of items. As clearly stated by Resnick (1992), in the number domain there is no denotable object that a child may refer to in order to abstract the concept of it, e.g., one can point to a set of three elements but not to a three. Second, the same number may be represented by more than one notational code, i.e., the Arabic and the verbal codes. Thus, children have to learn two distinct notational systems (i.e., Arabic and verbal), governed by their own set of composition rules, and map the one onto the other (e.g., 24 <----> twenty-four).

Tracing the development stages that characterise the acquisition of the number concept is beyond the scope of the present review. However, to some extent, understanding how children learn about number representations is strictly related to understanding of their acquisition of number knowledge. Number words are the first numerical notation that children acquire. Indeed, the mastery of the verbal number sequence has received a great deal of attention from developmental psychologists, attention further motivated by the critical link between the acquisition of number words and counting skills. Much less investigated is the later stage in which children face with written numerical notations and learn about Arabic numerals, in particular how these latter are linked to the number words they already know.

For the purpose of the present work, the attention is focused on the acquisition and manipulation of single-digit Arabic numerals and their corresponding number words. This is only the very first stage of a longer and laborious learning process, during which children have to deal with the many irregularities of the number-word sequence (at least in most western countries, where the number-words from 11 to 99 do not map directly onto the underlying base-ten structure of the number system) and with the complex rules governing...
the mapping of these words into the corresponding Arabic numerals. The lack of transparency between linguistic properties of the numeral system and the quantitative aspects of it, has been indeed identified as the source of many problems in the acquisition of both verbal numerals (Fuson & Kwon, 1991; Miller & Stigler, 1987) and arithmetic skills (Fuson, 1990; Fuson & Kwon, 1992). Similarly, the specific difficulties children are confronted with in learning to transcode numerals from one code to another have motivated targeted investigations (Power & Dal Martello, 1990; Seron, Deloche & Noel, 1992; Seron & Fayol, 1994). We refer to the above studies for an exhaustive analysis of the developmental stages in the acquisition of complex Arabic and verbal numerals.

Long before entering school, children familiarise with the number words through several number-related activities in the context of social play (e.g., counting fingers; counting toys; singing song with numbers). Similarly, they become acquainted with the number symbols that they encounter regularly in their environment, as in shops, cars or buses as well as in printed material or electrical appliances (e.g., TV). Moreover, since a very early age, both number words and number symbols constitute quite distinctive categories within the child's conception of language and written material in general (e.g., Fuson et al., 1982: Gelman & Meck, 1983; Sinclair & Sinclair, 1986). Critically, from their every-day experiences, children must also acknowledge that numbers are used in several different contexts and that, accordingly, they convey different meanings. Fuson (1988, 1992) indicated as much as seven different contexts where a child may encountered a number and only in one of these, the cardinal context, “the number word refers to a whole set of entities (a discrete quantity) and describes the manyness of the set” (Fuson, 1992, p. 127). For example, a number may indicate the relative position of an element within and ordered set (ordinal context), or simply refer to its specific position within the number sequence (sequence or recitation context). In this latter, the number term bears the only meaning of being part of the number sequence with no reference to any other entity. Finally, number may also have a nominal value when used in non-numerical contexts, such as in telephone numbers or zip codes. Thus, the child's understanding of number implies the appreciation of the multiple ways and contexts in which numbers are used; by
differentiating these distinct meanings and by learning the relationship between them (in particular between counting, sequence and cardinal meanings) a child may develop an adequate numerical knowledge. This process is a rather slow and difficult one, taking several years to a child to accomplish it (Fuson, 1988).

2.3.1 The acquisition of number words

Several studies have explored the way in which young children acquired number words, and more importantly, how they learn to map number words into numerosities (e.g., Fuson, 1988; Gelman & Meck, 1983; Wynn, 1990, 1992). Verbal counting is one of the earliest number-related activities and much can be inferred from it about the child's conception of numbers. Counting proficiency entails distinct basic skills, such as learning the correct sequence of number words, creating a one-to-one correspondence between number words and counted items and coming to understand that the last number word denotes the numerosity of the counted set (e.g., Fuson, 1988; Gelman & Gallistel, 1978). Children begin to use number words when counting from the age of 2, but often the sequence they produce is not in the standard order (e.g., one, two, four) though, at least in older children (above 3 y.o.), this nonconventional sequence does not include repetition and is stable across the counted sets (Gelman & Gallistel, 1978; Fuson et al., 1982; Baroody & Price, 1983). This evidence has favoured the hypothesis that children implicitly know that different number words denote different quantities (e.g., one-to-one principle) and that the order of the sequence is a relevant factor (i.e., stable-order principle) (Gelman & Gallistel, 1978). However, there is evidence indicating that children may well be able to count, also knowing the meaning of some number words, before learning that counting determines the numerosity of a set (e.g., Fuson, 1988; Wynn, 1990). This knowledge is not acquired before the age of 3 or 4, despite the fact that younger children may know the number sequence. Longitudinal studies indicate that learning the cardinal number words is a lengthy process; children take about an year to map precisely number words into numerosities despite their implicit and general knowledge that number words denote numerosities (Wynn, 1992). For example, children may be able to point to a numerosity of 'four' when
presented with two cards with respectively 4 and 1 balloons, but they could not point to the same numerosity when confronted with a 5 balloons card. By the age when they know that ‘one’ denotes a single element, and that ‘four’ denotes a numerosity they still do not associate number words to their corresponding quantities. By the age of 3/4, children have acquired the cardinal meaning of all number words within their counting range, but this learning process seems to proceed sequentially from smaller to larger number words (Wynn, 1990). The children’s ability to accurately name small numerosities seems strictly related to their abilities to recognise them by subitization (e.g., Gelman & Tucker, 1975; Silverman & Rose, 1980), even though this does not imply that the cardinal meaning of the number words in the subitizing range are acquired all at once (Wynn, 1992). On the other hand, the evidence suggests that, once children come to understand that counting determines numerosities (i.e., the ‘cardinal word principle’; Gelman & Gallistel, 1978; Fuson, 1990) they acquired the cardinal meaning of larger number words simultaneously and this occurs roughly by the age of 4 (Wynn, 1990).

Although most researchers acknowledge that both an innate sensitivity to numerosities and the child’s experience with numbers play a role in the acquisition of counting skills, there is much debate about the relative contribution of these two factors. Gelman and colleagues (e.g., Gelman & Gallistel, 1978; Gelman & Meck, 1983; Gelman & Greeno, 1989) claim that children’s counting behaviour and the learning of number words are greatly facilitated by the non-linguistic and innate knowledge of a set of “how-to-count” principles that serve as guidelines for counting. The “one-to-one” principle states that the to-be-counted items must be put in one-to-one correspondence with the counting tags; the stable-order principle states that counting tags have to be used consistently in a fixed order and finally, the cardinality principle states that the last counting tag indicates the numerosity of the counted set. Thus, children’s overt and verbally driven counting should be supported by the pre-verbal principles and specifically, by putting in correspondence these mental number tags with the newly acquired number words. The advocates of the alternative view acknowledge the primacy of the child’s experience with counting and numbers in the development of principled knowledge. The inborn sensitivity to numerosities may well provide the foundation for later development, as for example, in learning to map number
words to numerosities in the subitizing range. However, the principles governing counting as well as the multiple meanings of number words are assumed to be learned through experience (Briars & Siegler, 1984, Fuson, 1988; Wynn, 1992).

2.3.2 The acquisition of Arabic numerals

In marked contrast with the learning of number words, relatively little attention has been devoted to the child’s acquisition of Arabic numerals. However, Arabic numerals possibly constitute the preferred code for representing numbers, at least in specific contexts as in arithmetic or in more ecological situations as for example, representing time and prices. Similarly to number words, Arabic numerals are widely used in our environment for a variety of different purposes and long before they enter school children become highly familiar with Arabic numerals. In fact, they are soon engaged in the complex task of understanding the different meanings an Arabic numeral may stand for. Sometimes numbers represent cardinalities (e.g., the quantity of objects in a box), sometimes ordinal properties (e.g., houses in the street). Arabic numerals could also stand for measures of various kinds (e.g., sizes, dates, scores) or simply identify elements in a specific class (e.g., buses, car number-plate). The interesting issue is to understand the way in which pre-schooler’s knowledge of written numerals interact and develop with their acquisition of number concepts. To this end, two approaches have been so far adopted: the first one, is to investigate the children’s interpretation or ideas of the written numerals they encountered in every-day situations (Sinclair & Sinclair, 1984; Ewers-Rogers & Cowan, 1996); the second one, is to observed their own production of early numerical notations (Ewers-Rogers & Cowan, 1996; Hughes, 1982; 1986; Sinclair, Siegrist & Sinclair, 1983).

Sinclair and Sinclair (1984) presented 4 to 6 year-old children with pictures where numbers were used in familiar contexts, such as birthday cakes, buses, prizes, houses and so forth. Children were explicitly questioned about the meaning or the function of the numerical symbols. Not surprisingly, a clear-cut developmental trend for answering in a more adequate and functional way emerged. Younger children were more likely to answer by describing or naming the numeral itself (e.g., “it is a number”, “it is a five”), but also by
attributing to it a general function not strictly associated to the specific context (e.g., the number on the bus is .."for the people...”). However, by the age of 5, a large proportion of explanations related to the specific function of the numeral reflected the emerging distinction between numerical (e.g., the number on the birthday cake stands for the age of someone) and non numerical context (e.g., the number on the bus indicates where it goes). Interestingly, children’s knowledge about the number function seemed relatively independent from their knowledge of the number names.

More recently, Ewers-Rogers and Cowan (1996) presented 3 and 4 year-old children with pictures of familiar objects or situations (e.g., a telephone, a coin) where numerals were removed to test whether children do notice numbers in their environment. Results indicated that children rarely noticed the missing numbers, but when they did so, they were always able to explain their purpose. In the same study, the authors investigated children’s methods of representing number in a variety of familiar situations as, for example, completing a party invitation or labelling tins containing either 1, 2, 3 or no toy bricks. Following Hughes (1982, 1986), children’s written representations were classified as idiosyncratic, when no obvious relation with the number was observed, pictographic, when the representations bearded some aspect of the appearance of the object, iconic when the representations were based on a one-to-one correspondence with the number and symbolic, when conventional Arabic numerals were used. None of the children used all symbolic representations but there was some consistency across the tasks; moreover, production of numerals was more frequent in children who could adequately explain their function.

In Hughes’s study (1982, 1986) 3 to 8 year-old children were simply requested to represent how many objects were displayed in front of them (range 1-5). The proportion of symbolic representation increased consistently with age, being 11% in the younger children, 32% in 5 year-old children and 50% in both 6 and 7 year-old children. Similar results were reported by Sinclair et al. (1983). In their study only few 5 year-old but all 6 year-old children represented the set of objects by the corresponding written numeral (range 1 to 8), though other notations were also used. Some of the children who mastered the cardinals, sometimes represented them as written number words. From 3 years old, children produced successfully iconic representations, using a one-to-one correspondence between the
elements and some written marks (e.g., XX, OO, AIP, I I I). Interestingly, children at this stage were often able of naming Arabic numerals as well as writing them from dictation, but still favour iconic representations. Sometimes children tried to represent their counting procedures, as for example by writing down the numerals themselves (e.g., 12345) or repeating the cardinal value (e.g., 55555). These transcriptions clearly indicated the transition between the iconic and the symbolic stage, consisting in the attempt to integrate their knowledge of the written notation with their perception of the reality. As Sinclair and Sinclair suggested (1984), however, a clear understanding of cardinality emerged only when a single 5 will be used to denote five objects.

Additional evidence supports children's early appreciation of magnitude as central attribute to numerals. Miller and Gelman (1983) investigated children's number representation throughout judgements of similarity between triads of numbers. They results indicated that since the age of three, numbers were mainly associated on the basis of their closeness in the counting sequence, a dimension that gradually acquired a magnitude value by the age of 5. Over the course of the primary-school years, a set of numerical relations (e.g., parity status, calculation relations) based on a more and more complex understanding of numerical meanings emerged. This study corroborates the hypothesis that children are able to make consistent judgements about written numbers since a very early age, and certainly by the age of 5, they recognise their cardinal value as a central numerical property.

In conclusion, these studies indicate that young children posses a remarkable range of knowledge of numerals used in different contexts and they reveal an appreciation of the quantitative and non quantitative use of them. In particular, the production of spontaneous notational systems reflects their developing conceptual understanding of numbers. Clearly, the ability to recognise, name and write Arabic numerals, that emerges from the age of 3, precedes and does not imply their use as cardinals (Sinclair et al., 1983; Bialystok, 1992). In other words, at this stage, numbers are distinctive and familiar to children, without being fully understood as symbolic representations. It is only by the age of 5-6 that children come to map consistently Arabic numerals to their numerosities, as for example, by accurately representing with bricks their cardinality (Hughes, 1986). However, even at this age, they do not always prioritise the use of Arabic numeral to represent numerositites (Hughes,
Children's development through a symbolic understanding of the written numerals is an extremely important accomplishment; only then, can children use them engaged in the acquisition of arithmetic skills.

2.3.3 The acquisition of number comparison skills

Once children come to grasp the symbolic function of Arabic numerals, they should successfully compare them along their magnitude. Indeed, a simple method for testing understanding of Arabic numerals is to present children with two digits and ask them to indicate which is the larger (or smaller) of the two. Accurate performance of the task requires access to the meaning, i.e., to the symbolic internal representation, of the digits. Moreover, the qualitative profile of the performance may reveal the nature of the underlying mechanisms operating upon these representations.

The literature on number comparison is abundant and the standard effects have been repeatedly replicated (Moyer & Landauer; 1967, Banks, 1977; Dehaene, 1989; Parkman, 1971). Among all, the most robust result in numerical comparison is that reaction times decrease as the numerical distance between the to be compared digits increased. Similarly to what has been observed in the comparison of physical parameters (e.g., length, weight, brightness), the response time is a logarithmic function of the distance between the two items. On the basis of the distance effect it has been suggested that numerals are not compared at the symbolic level but converted to analogue magnitudes before the comparison takes place.

Fewer studies have investigated numerical comparison in children at different ages (Duncan & McFarland, 1980; Schaffer, Eggleston & Scott, 1974; Sekuler & Mierkiewicz, 1977; Siegler & Robinson, 1982). Overall, these studies suggest that number comparison skills are mastered well before school-age and that by the age of five, both reaction times and errors distributions mimic the adults' performance. Because of the observed distance effect (Sekuler & Mierkiewicz, 1977; Duncan & McFarland, 1980), it has been suggested that children as young as five years have an internal number representation similar to the adult's oriented number line (e.g., Resnick, 1982). However, Sekuler and Mierkiewicz
(1977) observed that the relation between reaction times and numerical distance was steeper in younger subjects (5 and 6 year-old). They interpreted this finding by suggesting that younger children's internal analogue number line representation a) might be more compressed, and/or b) presents more discriminative dispersion; in either cases, their subjective distance among number representations would be smaller than in older children and adults.

A similar pattern of results was reported by Duncan and McFarland (1980), testing children from 5 to 11 years of age. A distance effect was shown by all age-groups, even if younger children presented a steeper relation between reaction times and numerical distance. Since this relationship between age and distance effect was not modified by slowing down the encoding stage (e.g., altering the quality of the stimuli, Experiment 1), this result was attributed to age-related differences in the comparison process rather than in the encoding stage. Yet, they argued that developmental changes in number comparison are quantitative (e.g., speed of processing) rather than qualitative; the same basic process operating of mental analogues would underlie number comparison at all ages.

However, studies investigating number comparison by younger children suggest that the number line representation is only the final result of a long and gradual process of differentiation (Siegler & Robinson, 1982; Schaffer, Eggleston & Scott, 1974). Multidimensional scaling and clusters analyses indicate that children, first categorise numbers in small and large classes, then they start to differentiate between intuitive numbers (1 to 5) and gradually extend this process to larger numbers (6 to 9). For example, in Siegler and Robinson's analysis of 4 year-old children's data, numbers seem to fall in four clusters: (1), (2,3), (4,5), (6,7,8,9). The authors suggested that comparison of numbers within the same cluster is more difficult than comparison between clusters. Additional evidence suggested that this size-categorisation of numbers in preschoolers results from a long-term memory representation: contrary to adults, young children would label 5 as a medium-size number and 9 as a large-size number, regardless to whether the considered interval is 1 to 9 or 1 to 1 trillion (Siegler & Robinson, 1981, reported by Siegler & Robinson 1982). Moreover, even though 3 year-old children were reasonably accurate in

---

8 Age-related decline of symbolic distance effect has been also reported in magnitude comparison of concrete objects (both pictures and spoken names) (Wright and Berch; 1992).

62
classifying all single-digits numbers according to their magnitude, they did not seem to use this knowledge to compare numbers, a task that they performed at chance level. However, by the age of 4, children’s accuracy increased significantly (80% correct) reflecting the gradual refinement of their number knowledge.

To summarise, the ability to compare numerals clearly entails the comprehension of their cardinal meanings; young children (3 year-old) may be not able to accomplish this task even though they could recognise and name Arabic numerals. Mapping numerals onto their verbal representations may be a developmental pre-cursor to number comparison skills, but obviously not a unique pre-requisite. Interestingly, by the same age, they might also possess an early knowledge about the absolute size of numbers (small, medium and large), but still failed to use it to judge their relative magnitude in comparative conditions. This categorical and absolute representation of numerals, seems to dominate the children’s early successful performances in comparison tasks (4 year-old). However, as their acquaintance with numbers improves, children’s number representation undergoes a continual refinement and differentiation process. By the time they enter school (5-6 years of age), children can compare single-digit numerals with relative accuracy (error rates ranging from 10% to 12%) and speed (mean RTs ranging from 1400-1700 ms). More importantly, their performance is qualitative similar to the adult’s one, and they consistently show a standard distance effect.

2.3.4 Autonomous semantic processing of Arabic numerals

In section 2.2, we have seen that adults’ performance suggests, not only that a magnitude-related representation is likely to be central to a variety of numerical tasks, but also that it is autonomously accessed whatever the task requirements. One may assume that, by the time children attribute to Arabic numerals a symbolic value and they successfully operate upon their cardinalities, they may exhibit the ubiquitous primacy of magnitude information, as adults do. As a matter of fact, this issue has been greatly overlooked by both developmental and cognitive psychologists; the attempt to locate any relevant studies points to only three contributions.
In a thorough study on numerical comparison, Duncan & McFarland (1980) presented both adults and children of different ages (from 5 to 11) with a same-different judgement task (see also Section 2.2.3). Subjects had simply to decide whether two digits were identical (e.g., 4 4) or different (e.g., 4 8). Despite the fact that both close and far pairs had to be answered different, numerically distant pairs (e.g., 2 8) yielded faster responses than numerically close pairs (e.g., 2 3). The authors interpreted these results as evidence that subjects used ‘symbolic’ information even if the decision could have been made simply on the basis of perceptual information. Surprisingly, this pattern of results characterised both adult’s and children’s performances to the same extent. On the light of theories which claim a primacy of perceptual over ‘symbolic’ information in cognitive development (Bruner, Olver & Greenfield, 1996; Piaget, 1952), these results were unexpected. It seems that children as young as 6 years old, were accessing number magnitude to facilitate their decision and, more critically, this information was readily available to them. These results would suggest that, also for relatively young children, the cardinal value is a highly distinctive dimension of a written numeral, which may be hardly possible to disregard.

Two additional studies were indirectly concerned with the autonomous property of numerical processing, even though both focused on mastery of simple arithmetic. LeFevre, Kulak and Bisanz (1991) tested children in Grades 3 through 5 with the number-matching task previously used with adults to evaluate the relative strength of sum activation (see also Section 2.1.1). We may recall that this task requires participants to indicate whether a subsequent probe matched one of the numbers in a initial pair. The results from the adult’s performance indicated that probes that correspond to the sum of the initial pair took longer to be rejected than neutral probes. In line with the authors’ expectations, however, 8 to 10 year-old children showed minimal effects of sum activation (e.g., initial pair 2+7, probe 9) but larger effects of probes close in magnitude to the initial pair (e.g., initial pair 2+7, probe 6). This distance effect was assumed to reflect the strength of number-line associations resulting from the extensive use of counting strategies in the solution of simple addition (e.g., Siegler & Shrager, 1984). Interestingly, the distance effect was more pronounced in low-skilled children, who are more likely to rely on non-retrieval back-up strategies in the solution of simple additions. This study was replicated and expanded by Lemaire et al.
An analogous number matching task was employed, where the autonomous activation of addition facts was tested under more stringent conditions, i.e., the addition sign between the initial pair was omitted and the range of numbers included large integers (e.g., 06 08, probe 14). Results indicated that primary-school children did show an interference effect although it varied with the size of the problems (i.e., the interference effect was magnified in smaller problems) and the grade-level (i.e., older children showed larger interference). The hypothesis that mental arithmetic is a partially autonomous process since the early stages of acquisition, was corroborated by the observation that children exhibited associative confusion effects between addition and multiplication facts. Since Grade 3 (8 year-old), children took longer to reject equations such as 3+4=12 or 3x5=8, where the proposed result corresponds to the correct solution to a different operation between the same factors; the interference effect was initially constrained to easy problems and gradually generalised to all problems as a function of experience and mastery of arithmetic.

In conclusion, these studies provide convincing evidence that in primary-school children the associations between a number pair and its sum or product are strong enough to produce interference in a variety of tasks. The effect is substantially stable across development, even though the range of problems to which it applies, increases gradually as skills progress. This pattern well reflects the gradual building up of an associative network in memory where problem and answer associations become more and more specific over development (Ashcraft, 1992; Siegler, 1988; Siegler & Shrager, 1984). Of particular value is the evidence provided by the number-matching task. Here, the interference effect presupposes strong numerical associations, the activation of which is clearly initiated autonomously by the mere presentation of numbers to be held in memory. These associations, reflecting both sum relations and counting-on relations, initially involved small integers (1 to 5), and only later on, extend to larger numbers.
INTRODUCTION

The purpose of the study reported in the present Chapter was to examine how and when children come to map Arabic numerals to numerosities and to trace the rise of automaticity of this association. Thus, this work does not primarily concern with the development of the number concept itself, but with the process through which this concept is extracted from its symbolic representation, i.e., Arabic numerals.

To our knowledge, no studies have explored the developmental trend of automatic number processing. The only exception is the study by Duncan and McFarland (1980) where they presented adults and children of different ages with a same-different judgement task (Experiment 2). Subjects had simply to decide whether two digits were identical or different. Despite the fact that both close and far pairs had to be classified as different, all subjects showed a significant distance effect. Thus, children as young as 6 years old were processing the numerals all the way to semantics despite a simple visual matching would have been sufficient to answer the task. This result seems to favour the hypothesis that the automatic access to magnitude representation is quite and early achievement; however, this will remain a tentative suggestion till further evidence proves its legitimacy.

To pursue the aforementioned objectives, a number Stroop-like paradigm was adopted consisting of two comparison tasks where the relevant dimension was conceptual
numerical comparison) and physical (physical comparison) size respectively. As illustrated in the literature review, this paradigm has been extensively used in experimental psychology; however, the combination of both numerical and physical comparisons has been used in a minority of studies (Henik & Tzelgov, 1982; Peereman & Holender, 1984; Tzelgov et al., 1992) and, to our knowledge, neither of this task has ever been used with children. Stroop-like tasks proved to be useful in identifying changes in word recognition ability determined by learning experience or development. There is no reason why this should not apply to the numerical domain as well.

In the number-Stroop paradigm subjects are presented with pairs of Arabic numerals to be compared. In the numerical comparison task, they have to select the larger in magnitude between the two numbers; in the physical comparison task they have to select the larger digit in physical size ignoring semantic information. Physical and numerical sizes are orthogonally combined to define three different congruity conditions: a) congruent, when the numerically larger number is physically larger (e.g., 2 6); b) incongruent, when the numerically larger number is physically smaller (e.g., 6 2); c) neutral, when the numbers are displayed in the same size (numerical comparison; e.g., 2 6) or the same numbers are displayed in different sizes (physical comparison; e.g., 2 2). Clearly, the time and efficiency in processing the relevant and irrelevant dimensions play a critical role in determining the size congruity effect, i.e., incongruent trials yield slower reaction times compared to both neutral and congruent trials. Given that processing of physical and numerical sizes is assumed to be differentially modulated by developmental changes, this paradigm offers the possibility to elucidate the interplay of perceptual and symbolic information in cognitive development as well as to establish to what extent the semantics of Arabic numerals is autonomously accessed by children at different ages.

When first reported, the effect of the perceptual properties of the numbers on the symbolic comparison was considered consistent with the assumption that memorial and perceptual comparisons involve similar processes (e.g., Moyer, 1973; Paivio, 1975). The locus of the interference, however, has not yet been unequivocally clarified. Banks and his

---

10 We use this term to refer to the 'magnitude' or numerosities a number stands for. This aspect is not the only semantic dimension of numbers (e.g., parity status) but certainly is the most central.
his colleagues (e.g., Banks, 1997; Banks & Flora, 1977; Banks et al., 1976) proposed a serial comparison model which includes a first encoding stage, where large and small codes are assigned to the-be-compared stimuli. When numbers are close in magnitude this coding does not discriminate between them and an additional comparison stage is required. Within this framework, the sources of the size congruity effect and of the distance effect are located respectively at the encoding stage and at the comparison stage. This assumption is supported by the evidence that congruity and distance have additive effects on reaction times (e.g., Banks, 1977; Henik & Tzelgov, 1982). However, this last result has not been consistently reported (e.g., Foltz et al., 1984; Tzelgov et al., 1992) and it seems to partially depend on some methodological differences across the studies. In principle, the congruity effect should be maximised when the numerical comparison takes longer, i.e., in numerically close pairs compared to numerically far pairs; this relation is assumed to be modulated by the relative distance between close and far pairs. Indeed, the levels of numerical distance vary in the different studies, and when a comparison was attempted, we noticed that an interaction between congruity and numerical distance emerged only when the latter had more than two levels (Foltz et al., 1984; Tzelgov et al., 1992) and the distance within far pairs was maximised.

Several studies have suggested that numerical comparison skills are mastered well before school-age and that, by the age of five, both reaction times and error distributions mimic the adult’s performance (e.g., Sekuler & Mierkiewicz, 1977; Duncan & McFarland, 1980; Siegler & Robinson, 1982; Schaffer et al., 1974). Thus, the task requirements in both numerical and physical comparisons, should not be beyond children’s skills. Nevertheless, the mismatch between numerical and physical sizes is likely to affect children’s performance, as it does for older subjects. The critical question is how the gradual acquisition and refinement in numerical knowledge affects the pattern of interference-facilitation across development.

Despite the fundamental differences between the numerical version of the Stroop-paradigm and the standard word-colour task, some key results in the developmental Stroop studies may be relevant to the present investigation.
There is general consensus that developmental changes in the Stroop word-colour task are the result of the combination of several factors. On the one hand, immature inhibitory mechanisms and a general low control over cognitive processes are regarded as responsible for the large interference effects observed in young children (e.g., Tipper, Bourque, Anderson & Brehaut, 1989). In fact, the ability to disregard interfering stimuli (or dimension of the stimuli) or, possibly, to process interfering and primary stimuli independently, is acquired later on in development. On the other hand, a prerequisite for interference to appear is for the irrelevant dimension to be automatically processed. For example, the name of the printed word ‘RED’ would interfere with the naming of the colour ink (e.g., green) if, and only if, a child can read (or recognise) the word. Thus, reading proficiency is obviously a critical factor for the interference effect to emerge. In fact, it has been shown that interference increases as reading comprehension increases, from non-readers to Grade 3-4, and then subsequently decreases through the adulthood (Comalli, Wapner & Werner, 1962; Schiller, 1966; Schadler & Thissen, 1981).

Similar predictions would apply to children’s performance in the picture-word version of the Stroop task. Rosinsky et al. (1975) showed that after only 1 year of reading experience (mean age = 7.7 y.o.), the presence of an incongruent word increased significantly the time needed for naming a picture and this effect was modulated by the semantic content of the word. Moreover, the decrease in interference with age was not determined by developmental changes in the semantic component of the effect but in attentional or responses processes efficiency (Rosinski et al., 1975). These results suggest that the automaticity of both reading processes and semantic access occurs early in the acquisition of reading skills (see also Tipper et al., 1989, Experiment 2). On the other hand, the lack of interference in the performance of young poor readers seems to suggest that automatic decoding may be a prerequisite for automatic access to word meaning (Ehri, 1976; but see Golinkoff & Rosinsky, 1976).

Overall, whatever the source of interference, the data support the general conclusion that interference emerges as early as reading skills begin to develop, and then gradually declines as development progresses. What is less clear is the extent to which this pattern applies to facilitation as well. The facilitation effect has been investigated less
systematically than interference, but overall it seems to be of a smaller size and highly dependent on the choice of the control condition (MacLeod, 1991; for a discussion see, Lindsay & Jacoby, 1994). However, the few studies that addressed this issue seem to indicate that facilitation remains relatively constant across development (Ehri, 1976; Shadler & Thissen, 1981).

As mentioned before, the experimental paradigm adopted in the present study consists of two tasks: the relevant dimension in one task (numerical comparison) acts as irrelevant dimension in the other one (physical comparison) and vice versa. Still adopting the standard terminology, we would measure, not only the ‘Stroop-effect’ (i.e., interference and/or facilitation of the irrelevant word reading on naming colour ink) but also the ‘reverse Stroop-effect’ (i.e., interference and/or facilitation of the irrelevant colour ink on word naming). Again, studies investigating this issue indicate that interference may indeed occur in both directions even if the ‘reverse Stroop effect’ appears as a weaker and less clear phenomenon (MacLeod, 1991). With regard to developmental studies, the evidence is scanty and not conclusive (e.g., Rosinsky et al., 1975).

Taking into account these considerations a set of predictions for the present study was put forward.

In the number-Stroop paradigm the reading requirement is very low since single-digit Arabic numerals only are presented. Recognition of Arabic numerals precedes alphabetic reading and it is not difficult to find preschoolers who can identify single digits despite their inability to read words. Thus, we did not expect reading proficiency to modulate children's performance. Moreover, more critically, the number-Stroop paradigm, requiring the comparison of two stimuli along a conceptual dimension (i.e., numerical size), concerns the comprehension of Arabic numerals rather than the ability to name them. The evidence suggests that children as young as 5 years old are able to compare single-digit numerals and since then, they show a normal distance effect (Sekuler & Mierkiewicz, 1977; Duncan & McFarland, 1980). Thus, we assumed that they may access and operate upon the conceptual dimension as required by the task. It follows that, in a number-Stroop task, children may show a pattern of interference and facilitation
similar to the adult’s one. However, we hypothesised that the size of these effects was very likely to be modulated by age-related differences in both attentional efficiency and number processing mechanisms.

A larger size congruity effect would be expected in young children given their immature ability to resist to interference, i.e. inefficient selective attention mechanisms (Dempster, 1992; Tipper et al., 1989). Moreover, it has been long suggested that perceptual information predominates symbolic information in children. Perceptual cues (e.g., appearance) can lead children to disregard cardinality, as in the classic piagetan task where children fail to recognise that two arrays of dots have the same numerosity if they are in different spatial arrangements. Thus, it might be particularly difficult for children to ignore the irrelevant dimension of a stimulus if this is a perceptual one, as in the case of the physical size of a digit; though, as a child grows older, the irrelevant physical dimension should become less distracting (Bisanz, Danner & Resnick, 1979).

However, the fact that children can compare the magnitude of two Arabic numerals does not imply that this dimension is autonomously activated by the mere presentation of the stimuli. It is plausible to assume that the association between an Arabic numeral and the meaning it conveys (i.e., the numerosity it represents), develops gradually over the course of learning. In other words, for younger children physical and numerical sizes of a numeral may be much more separable dimensions than for older children. If the size of the interference effect is largely determined by the “integrality” of the conflicting dimensions\(^\text{11}\) (e.g., Garner et al., 1982; Algom et al., 1996), we expected the interference effect to increase with age. That is to say that, after all, the task may be more compelling for older children and adults than for younger participants.

The physical comparison task may provide further support to this hypothesis. In fact, the size congruity effect in this task is said to result from an inability to ignore the numerical information conveyed by the Arabic numerals (Henik & Tzelgov, 1982; Tzelgov et al., 1992). In other words, the interference effect in this task reflects the autonomous access to number semantic representations.

\(^{11}\) In Garner’s terms, if subjects can selectively attend to one dimension while another varies irrelevantly, then the dimensions are considered separable; if this is not the case the dimensions are integral (Garner, 1974).
The evidence suggests that recognition of single letters becomes automatic early on in the acquisition of reading skills (e.g., Schadler & Thissen, 1986) and, very likely, this applies to Arabic numerals as well. However, automatic recognition does not imply automatic semantic access. If the meaning of an Arabic numeral becomes the core feature of it over the course of skill acquisition, well beyond the stage in which children may attribute to it a symbolic value and a phonological form, we could expect young children to perform the physical comparison task undisturbed by the irrelevant semantic dimension of the stimuli. However, as skills develop and Arabic numerals become more and more salient symbols that stand for specific numerosities, this semantic dimension may gradually become accessed automatically. If this is the case, when older children come to automatically access magnitude information, we may wonder if they do so as adults seem to do, i.e., by activating a less refined small/large categorisation of numbers (Tzelgov et al., 1992).

Tzelgov et al. suggested that when numerical size is irrelevant to the task, the available numerical information comes mainly from the primary encoding stage where the stimuli are simply classified as small (i.e., 1, 2, 3 and 4) or large (i.e., 6, 7, 8 and 9). Though a more refined recoding of the numbers onto an internal magnitude representation is also initiated, its activity would be too slow to exert an influence on the performance. Their hypothesis was corroborated by the fact that in the physical comparison task, the size congruity effect was maximised in pairs including numbers with different initial encoding (i.e., small-large, defined as "bilateral" pairs) compared to pairs including numbers with the same primary encoding (i.e., small-small or large-large, defined as "unilateral" pairs). Moreover, within bilateral pairs, the effect was not further modulated by numerical distance (e.g., ‘4 6’ and ‘4 8’ yielded the same RTs). This last result would be not easily explained within an alternative approach, according to which intentional and unintentional semantic activations would differ only quantitatively. In fact, it might be suggested that in condition of autonomous processing, the activation of the magnitude representations would be simply less efficient and, in turn, less refined

12 It follows that unilateral pairs are by definition numerically closer than bilateral pairs, thus laterality and numerical distance are partially overlapping effects.
preventing the discrimination between close numbers but not between distant numbers. However, in this case, the size congruity effect should be simply a function of the numerical distance between the numbers, and this effect did not emerged.

In summary, in Tzelgov et al.'s account, two different processes are invoked: first, a memory-based classification that simply discriminates between small and large numbers; second, an algorithm-based process through which numbers are mapped on their corresponding magnitude representations. Critically, skill acquisition would determine the development of the first process and cause the second one to acquire “ballistic” properties; however, the temporal aspects of this acquisition process remain to be specified. It might be suggested that children’s performance in number comparison, and in particular the occurrence of a distance effect, reflects their efficiency in the algorithm-based process. Children’s RTs and accuracy profiles in the physical comparison task, where the automatic processing of numerical information is evaluated, may shed light on the qualitative changes that occur in the course of skill acquisition. The observation of the size congruity effect in the physical comparison task would imply that children access numerical information in an autonomous way. If similarly to adult’s performance, this information results from a dichotomous classification process, the effect should be limited to bilateral-far pairs. However, if the primacy of the small/large classification process is established later on in development, and the rise of automaticity in number processing is determined by the automatic mapping of the numbers onto their magnitude representations, we would expect children to show a size congruity effect in all number pairs.

Our predictions are consistent with the theory according to which automatization is achieved gradually as the acquisition of a domain-specific knowledge progresses (e.g., Logan, 1988). One of the key assumption of this view is that automaticity is not an “all or none” phenomenon, but a continuum along which processes may differ (e.g., Logan, 1985; Schiffrin, 1989). Most important, practice would play a critical role in producing automaticity, suggesting that automaticity may be the product of learning. This issue is
not new to the Stroop-task literature, given that the very first study on practice effect was run by Stroop himself in his seminal paper (Stroop, 1935; Experiment 3). Most recently, this account has inspired a number of studies intended to clarify and quantify practice effects in Stroop-like tasks (e.g., MacLeod & Dunbar, 1988). The underlying assumption is that practising one dimension should lead to increase its automaticity, that, in turn, lead to increase its interference effect when the dimension is task-irrelevant. For example, MacLeod and Dunbar (1988) varied the extent of practice in an analogue of the Stroop colour-word task across different experiments. As the continuum of automaticity account would predict, the differential practice with one or the other dimension (e.g., naming colour and naming shapes) of a two-dimensional task, modulated the pattern of interference and facilitation.

In order to control for practice effect, these studies often involved unfamiliar (e.g., unfamiliar shapes) or arbitrary (association of nonsense syllable to colours or words) dimensions. In the present study, practice and familiarity with the dimensions involved (physical size and numerical size of Arabic numerals) could not be directly controlled. However, we assumed that both these factors increase with development, and that age could thus provide a reliable index for our purpose. In particular, we assumed that by virtue of their different nature (e.g., perceptual versus conceptual), the degree of automaticity of physical and numerical sizes processing differ significantly. If physical size may well be automatically processed at a very early age and be less sensitive to developmental changes, numerical size would become automatic through extensive learning experience.

Overall, only three studies using the number-Stroop paradigm adopted both numerical and physical sizes as comparative dimensions and they all differ along critical methodological factors (see Chapter 2). Numerical distance (or laterality) was not always evaluated (Henik & Tzelgov, 1982, Experiment 1) or the neutral condition not included, preventing the possibility of distinguishing facilitation and interference effects (Henik & Tzelgov, 1982, Experiment 2; Tzelgov et al., 1992, Experiment 3). When these two factors were both taken into account, presentation mode was varied (central versus
lateralised) for studying visual field asymmetry in numerical processing (Peereman & Holender, 1984).

In the present study, both numerical distance (i.e., 1 and 5, unilateral and bilateral pairs respectively) and level of congruity (congruent, incongruent and neutral) were varied in a standard selection paradigm. First, to evaluate the feasibility of the experimental paradigm and to partially replicate Tzelgov et al. results (1992), a group of university students participated in the experiment (Experiment 1). A shortened version of the tasks has been then adapted to test pre-school children (Experiment 2). Finally, to detect the developmental changes over the course of learning, first-grade, third-grade and fifth-grade children took part in the experiment (Experiment 3).

3.1 Experiment 1

In Experiment 1, university students took part in two experimental sessions where they were requested to compare pairs of numbers according to either their physical or numerical size. In both tasks, the factorial combination of numerical and physical sizes determined three congruity conditions: congruent, incongruent and neutral. The numerical distance defined two types of trials: numerically close or unilateral pairs, consisting of consecutive numbers (e.g., 1 2) and numerically far or bilateral pairs, including a small (1, 2, 3 or 4) and a large (6, 7, 8 or 9) number (e.g., 2 7).

In the present paradigm distance and laterality effects are completely confounded; yet, adopting as a criterion for selecting the stimuli the exclusion of the number 5, given its special status of medium number (Tzelgov et al., 1992), we favour the hypothesis that laterality plays a major role in determining the occurrence of the size congruity effect (Tzelgov et al., 1992). However, as a matter of terminology, we favour the term distance effect (i.e., difference in the processing time between unilateral and bilateral pairs) and the terms close (instead of unilateral) and far (instead of bilateral) pairs, being the ones conventionally used in the literature.

75
In the numerical comparison task, we expected the congruity between physical and numerical dimensions to yield both facilitation and interference effects, as observed in similar studies (e.g., Foltz et al., 1984; Henik & Tzelgov, 1982; Peereman & Holender, 1984). On the other hand, in the physical comparison task, interference effect proved to be more reliable than facilitation, possibly due to the already fast processing of the neutral condition. Moreover, the numerical distance was assumed to differentially modulate the congruity effect in the two tasks. In the numerical comparison task we predicted the size congruity effect to emerge in all pairs, while in the physical comparison task this effect should be constrained to or maximised in far pairs. In fact, in conditions of autonomous processing, as in the physical comparison task, it has been suggested that numerical information mainly consists of a small/large classification, on the basis of which only numerically distant numbers will be discriminated (Tzelgov et al., 1992). This would limit the incongruity between numerical and physical dimensions to far pairs.

3.1.1 Method

Subjects

Twenty-four undergraduates and postgraduates students (17 female, 7 male) from University of Padova participated in the experiment. The mean age was 23 (range from 20 to 29). All participants had normal or corrected to normal vision and were unpaid volunteers.

Stimuli

Two numerical distances, 1 and 5, defined the following experimental pairs: 1 2, 2 1, (2 3, 3 2), 3 4, 4 3, 6 7, 7 6, (7 8, 8 7), 8 9, 9 8 as close pairs (numbers in the pairs are both small, 1<x<4, or large, 6<x<9, thus unilateral); 1 6, 6 1, 2 7, 7 2, 3 8, 8 3, 4 9, 9 4 as far pairs (one number is small, 1<x<4, and the other one is large, 6<x<9, thus bilateral). 5 is not included given is special status of medium number (Tzelgov et al, 1992). The neutral trials in the physical comparison were the following: 1 1, 2 2, 3 3, 4 4, 6 6, 7 7, 8 8, 9 9. Pairs ‘2 3’ and ‘7 8’ were included so that every number from 2 to 8
appeared both as the smaller and as the larger digit, however, they were removed from further analysis in order to balance the number of observations in close and far pairs. Extreme values, i.e. 1 and 9, appeared equally often in close and far pairs (as all other numbers, once 23, 78 are removed).

There were three different congruity condition: a) congruent when the numerically larger number was physically larger, b) incongruent when the numerically larger number was physically smaller, c) neutral when the numbers were displayed in the same size (numerical comparison task) or the same numbers are displayed in different sizes (physical comparison task) (see Table 3.1). Each single pair was presented twice in each condition\(^{13}\), for a total of 120 stimuli (2 blocks of 60 stimuli).

Stimuli were presented in pseudorandom order according to the following criteria: 1) no same numbers in consecutive trials; 2) no same trials consecutively; 3) no more than three same (left or right) correct answer in a row; 4) no more than three same numerical distance in a row; 4) no more than two same congruity conditions in a row.

Before the experiment began, subjects were presented with 20 random training trials in order to practice with the task and the experimental setting.

**Procedure**

Subjects were presented with the numerical and physical comparison tasks in two sessions on different days (an average of seven days apart). This procedure was adopted to minimise the effect of the first task on the second one. In the numerical comparison task, they had to select the larger in magnitude between the two numbers; in the physical comparison task they had to select the larger number in physical size ignoring semantic information. The order of presentation of the two tasks was counterbalanced between subjects.

\(^{13}\) In the physical task, neutral pairs were presented 5 times (the frequency of the relative position of the physically larger digit was counterbalanced across the trials) for a total of twenty items (1/3 of the experimental trials).
Table 3.1 Examples of the stimuli used in the Experiment 1.

<table>
<thead>
<tr>
<th>DISTANCE = LATERALITY</th>
<th>CONGRUITY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>congruent</td>
</tr>
<tr>
<td>close (unilateral)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Numerical task)</td>
</tr>
<tr>
<td>far (bilateral)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>(Physical task)</td>
</tr>
</tbody>
</table>

Subjects were seated about 50 cm from the computer screen. Each trial began with a fixation point displayed at the centre of the screen for 500 ms and followed 500 ms later by a pair of stimuli to be compared. In the congruent and incongruent conditions the physically larger digit had a font of 48 point, the physically smaller had a font of 24 point. In the neutral condition they had an intermediate size (36 point). The stimuli appeared in a white centred window on a black background. Stimuli stayed on the screen until the subject responded. The interstimulus interval was 1500 ms and subjects had a few minutes rest in between the two blocks. Participants were instructed to press the left- or right-hand keys according to the position (left or right) of the target stimulus (larger number). Instructions emphasised both speed and accuracy. Each task lasted approximately 20 minutes.

**Apparatus**

A Macintosh Powerbook 175 was used to present the stimuli. Psychlab software (Version. 99, Gum, 1992) was used to display stimuli and record reaction times.
Design

Overall, the variables manipulated were: task (numerical comparison and physical comparison); congruity (neutral, congruent and incongruent) and numerical distance (close and far). Moreover, one between-subjects variable, order of the tasks (numerical first and physical first), was considered.

3.1.2 Results

Reaction times (means of medians) and error rates for each experimental condition are given in Table 3.2.

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>SD</th>
<th>% errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td>618.4</td>
<td>92.1</td>
<td>1.0</td>
</tr>
<tr>
<td>far</td>
<td>531.5</td>
<td>59.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Incongruent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td>676.6</td>
<td>79.2</td>
<td>4.4</td>
</tr>
<tr>
<td>far</td>
<td>619.0</td>
<td>75.7</td>
<td>1.8</td>
</tr>
<tr>
<td>Neutral</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td>650.2</td>
<td>96.8</td>
<td>1.6</td>
</tr>
<tr>
<td>far</td>
<td>573.1</td>
<td>65.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Physical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td>420.8</td>
<td>74.9</td>
<td>0.0</td>
</tr>
<tr>
<td>far</td>
<td>422.9</td>
<td>67.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Incongruent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td>429.7</td>
<td>75.2</td>
<td>0.5</td>
</tr>
<tr>
<td>far</td>
<td>449.6</td>
<td>77.7</td>
<td>2.1</td>
</tr>
<tr>
<td>Neutral</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>426.4</td>
<td>81.2</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 3.2 Experiment 1: Mean Reaction times, Standard deviations and Error rates as a function of task, congruity and numerical distance.
The overall error rate was very low (mean = 1%; the individual mean error rate never exceeded 3.6%) and it was not further analysed; however, error distribution fits the RTs data. In the numerical comparison task there were 0.5% errors in congruent trials, 3.1% in incongruent trials and 0.8% in the neutral ones. In the physical comparison task there were 0.15% errors in congruent trials, 1.3% in incongruent trials and 0.5% in the neutral ones.

For every subject median correct RTs for each condition were calculated and entered as dependent variable in a series of ANOVA analyses.

A first repeated measures ANOVA with task (numerical and physical), congruity condition (congruent, incongruent and neutral) as within-subjects factors and order of the tasks (numerical first, physical first) as between-subjects factor was carried out.

The main effect of order of the tasks was not significant and did not enter in any significant interaction; data were then collapsed over this factor.

Physical comparison was significantly faster than numerical comparison (608 ms and 428 ms, respectively; F[1, 23] = 195.3; p<.0001). The three levels of congruity differed significantly (F[2, 46] = 59.22, p<.0001) with means of 496 ms, 542.5 ms and 517 ms for respectively congruent, incongruent and neutral trials. Decomposition into contrasts indicated that each condition differed significantly from all the others (all, p<.0001).

However, the effect of congruity was highly modulated by task, (F[2, 46] = 35.6, p<.0001). While in numerical comparison all levels of congruity differed significantly from each others (all, p<.0001), in physical comparison incongruent trials differed significantly from both neutral (p<.05) and congruent ones (p<.01), but no difference was found between congruent and neutral trials (see Figure 3.1).

Data from the numerical comparison only were submitted to a repeated measures ANOVA with congruity (congruent, incongruent and neutral) and numerical distance (close and far) as within-subjects factors.

The main effect of congruity was significant (F[2, 46] = 47.7, p<.0001) replicating the previous analysis. Moreover, subjects were faster to answer to far (574.5 ms) than to
close pairs (648.4 ms) \( (F[1, 23] = 86.6; \ p < .0001) \); however, the interaction between numerical distance and congruity failed to reach significance \( (F[2, 46] = 2.40, \ p > .1) \).

To directly verify whether the congruity effect varied in close and far pairs and to what extent these effects depended on the judged dimension, a further ANOVA with task (numerical and physical), congruity (congruent and incongruent) and numerical distance (close and far) as within-subjects factors was carried out. The main effect of numerical distance was significant \( (F[1, 23] = 37.9, \ p < .0001) \), as well as its interaction with task \( (F[1,23] = 95.4, \ p < .0001) \) and with congruity \( (F[1, 23] = 8.1, \ p < .01) \). In fact, while in numerical comparison close pairs were answered slower than far pairs (647 ms and 575 ms, respectively; \( p < .0001 \)), the opposite pattern was found in physical comparison (425 ms and 436 ms, respectively; \( p < .05 \)). Moreover, while in numerical comparison congruity
effect was significant for all pairs (for both close and far pairs, \(p<.0001\)), in physical comparison congruity was modulated by numerical distance, being significant in far pairs only (\(F[1, 23] = 9.14, p<.01\) and \(F[1, 23] = 1.01, p>.3\), for far and close pairs, respectively). Yet, in physical comparison, the interaction between congruity and numerical distance failed to reach significance (\(F[1, 23] =2.86, p=.10\)).

Previous studies adopting an analogous within-subjects design (all participants performed both tasks), emphasised the critical influence that the experience of the first session may have on performing the second one (Henik & Tzelgov, 1982; Tzelgov et al., 1992). In particular, one may suggest that the size congruity effect could be maximised (e.g., in numerical comparison) or even induced (i.e., in physical comparison) by having to ignore a dimension that was previously relevant to the task. In the overall analysis, order of the tasks was not significant and did not modulate any of the critical effects; however, a further analysis was carried out to test separately the interaction between relevant dimension and congruity for each group of subjects. In both numerical and physical comparisons, the magnitude of the congruity effect did not differ in the two groups (both groups, \(p<.001\) and \(p<.005\) for the numerical and physical comparisons respectively) confirming that the autonomous activation of both physical and numerical sizes was independent from previous experience.

3.1.3 Discussion

Overall, the results of this experiment are in line with the main findings reported in previous research (Henik & Tzelgov, 1982; Peereman & Holender, 1984; Tzelgov et al. 1992). As already discussed, caution must be taken in comparing the effect sizes across different studies given that the main factors of numerical size (range of values and numerical distance) and physical size (relative dimension of the printed numerals) are not matched. The magnitude of both numerical and physical differences is very likely to affect
the magnitude of the size congruity effect. Nevertheless, the pattern of RTs and error rates are relatively similar across the studies.

Not surprisingly, physical comparison was always faster and more accurate than numerical comparison. The congruity conditions had an effect in both tasks, suggesting that the irrelevant dimension, either physical or numerical, was autonomously activated. However, the magnitude of the effect was modulated by the nature of the task.

The inspection of the results in the numerical comparison task indicates that the interaction of physical and numerical dimensions determined both interference and facilitation. The magnitude of these effects was indeed identical (36.5 ms; both, p<.0001). Moreover, the size congruity effect was not further modulated by numerical distance. Despite the standard result that close pairs took longer to be compared than far pairs, the irrelevant physical dimension affected all pairs to the same extent. A similar additive relation between numerical distance and congruity has been previously reported in numerical size comparison (e.g., Banks, 1977; Henik & Tzelgov, 1982; Peereman & Holender, 1984; but see, Foltz et al., 1984; Tzelgov et al., 1992). Yet, within the incongruent condition, errors tended to be more frequent in close than in far pairs, suggesting that when the comparison was more time demanding, the incongruent physical dimension had a greater impact on the response.

Even more interestingly, when subjects had to evaluate the physical dimension of the stimuli, the congruity effect was still significant. This result suggests that numerical size information was elaborated even when irrelevant to the task. In this case, the congruity was determined by an interference effect only. As already pointed out by Henik and Tzelgov (1982), this result may be partially attributed to the nature of the neutral trials in the physical comparison. These pairs, including identical digits (e.g., 3 3), may be particularly easy to be processed, thus reducing the chance for congruent trials to further speed up the decision. In this respect, we may suggest that the numerical comparison offers an ideal baseline condition, while this does not apply to the physical comparison task.
As predicted, the congruity effect was modulated by the numerical distance, being maximised in far pairs. The difference between congruent and incongruent conditions was not significant in close pairs, though it was in the expected direction. In the Tzelgov et al.’s study (1992, Experiment 3) close (unilateral) pairs yielded a smaller but still significant congruity effect (note that the significance level is not reported). However, while in their experiment close pairs differed by two units (i.e., 2 4; 6 8) in the present study, all close pairs included consecutive numbers (i.e., 1 2; 3 4; 6 7; 8 9). Tzelgov et al. (1992) accounted for their result suggesting that even though a simple small-large classification may dominate the autonomous processing of numerical size, some information about the relative location of the digits on the internal numerical scale is also available. Yet, it seems plausible that this information is not refined enough to discriminate between consecutive numbers. Alternatively, it might be suggested that numbers did always activated their internal representations, but that differentiation between close pairs, being more time-demanding, would not have been completed in time for interfering with the response.

In summary, the present results are consistent with the suggestion according to which a) numerical information is automatically activated when irrelevant to the task, but b) differs qualitatively and/or quantitatively, from the numerical information activated under intentional condition (numerical comparison).
3.2 EXPERIMENT 2

Overall, the results of Experiment 1 indicated that the experimental paradigm and the selected stimuli were adequate for the purpose of our study and we decided to adopt the same procedure with younger participants.

The investigation of the "size-congruity effect" in pre-school children was carried out taking into account that task requirements had not to be beyond children's numerical skills and attentional resources. To this aim, participants were selected according to specific proficiency criteria (see Screening procedure) and they were presented with a shortened paper and pencil version of the tasks.

In the present study, preschoolers participated in two experimental sessions where they were presented with the numerical comparison and the physical comparison tasks. Our predictions for the two tasks were substantially different. We expected variation in the physical dimension of the digits to have a significant impact in children's performance in the numerical comparison. It is well established that young children are rather poor in filtering out irrelevant information (e.g., Dempster, 1992), in particular when this information is of a perceptual nature (e.g., Bisanz et al., 1979). Most likely, the physical dimension would both facilitate and interfere with the processing of the numerical dimension. In any case, children should consistently show a distance effect, being faster and/or more accurate in comparing numbers far apart than numbers close in magnitude. On the other hand, we expected the physical comparison task to present minimal difficulty to children, in part because of the nature of the task (i.e., perceptual comparison being easier than symbolic comparison), in part because the irrelevant dimension is symbolic (i.e., numerical value) and not a perceptual property of the stimuli. Moreover, we assumed that, despite their familiarity with Arabic numerals, pre-school children do not access their numerical values in an autonomous fashion given their limited practice with these symbols in numerical contexts.

Whether the limited age-range tested (4 to 6 year-old children) allowed to capture not only quantitative but also qualitative changes in the performance was not certain. Overall,
we evaluated group differences in numerical comparison to be more likely than in physical comparison, assuming that qualitative changes in the latter would involve a rather lengthy and gradual refinement in children's numerical knowledge.

3.2.1 Screening procedure

Preliminary tasks

All subjects were recruited in kindergarten in northern Italy. 168 4 to 6 year-old children underwent a preliminary assessment where some basic numerical skills were evaluated. The following tasks were presented: counting forward from 1 to 10; reading aloud Arabic numerals from 1 to 9; matching Arabic numerals to numerosities (N=6)\(^{14}\); selecting the larger between two Arabic numerals (N=20). The tasks were presented in the stated order; selection criteria were a flawless performance in reading, counting and matching tasks and an error rate lower than 0.25 in numerical comparison. Although, possible dissociations between the different tasks may be of theoretical interest, this screening procedure only aimed to identify those children whose numerical knowledge was adequate for participating in the experimental study.

Results: qualitative analysis

Only 52 out of the 168 children tested satisfied the criteria for participating in the experimental study.

The remaining 116 children may be roughly divided into three different groups, called A, B and C, on the basis of their performance in the preliminary tasks (see Table 3.3). Group A includes thirty-one children of whom ten were 4 to 5 year-old and twenty-one were 5 to 6 year-old. They all could count from 1 to 9, but occasionally failed to read aloud Arabic numerals (5% error rate). More errors occurred in matching Arabic numerals to numerosities and their performance in the magnitude comparison task clearly indicated a limited understanding of cardinality, failing in over 20% of the trials. Means

\(^{14}\text{Stimuli consisted in individual card where a single-digit numeral (range 1-6) was printed above three dice-pattern of dots representing a) the correct numerosity, b) a close (+/- 1) numerosity and c) a distant (+/- 3) numerosity. The order of the alternative answers was counterbalanced across trials.}
error rates for the two age-group were 30% (range 4% to 11%) and 35% (range 5% to 10%) respectively. All children were more likely to err in close pairs (5-6 y.o. N=72; 4-5 y.o. N=40) than in far pairs (5-6 y.o. N=53; 4-5 y.o. N=29).

Group B includes 39 children of whom twenty-four were 4 to 5 year-old and fifteen were 5 to 6 year-old, whose ability to recognise Arabic numerals was still not completely mature. They all failed to name, at least, 2/3 digits out of the 9 presented, errors being mainly visually similar numbers (e.g., 6 for 9; N=64) but also omissions (N=62) and symbolic misrecognitions (e.g. “A” for 4; N=25). Few skipping errors appeared in the counting task. The remaining 46 children (Group C, twenty-eight 4 to 5 year-old and twenty-two 5 to 6 year-old) performed very poorly in reading Arabic numerals, most of the errors consisting of omissions (N=187). However, their ability to recite the number sequence was remarkable compared to their reading skills (37 out of 46 children were counting flawlessly). The overall performances in the two latter groups were similar; however the qualitative analysis of the reading errors disclosed an important difference. The former often attributed to Arabic numerals a symbolic value, the latter simply answer “Non lo so” [I don’t know] to the majority of the stimuli.

<table>
<thead>
<tr>
<th>4/5 y.o.</th>
<th>5/6 y.o.</th>
<th>Tot N</th>
<th>Counting</th>
<th>Reading</th>
<th>Matching</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>10</td>
<td>21</td>
<td>31</td>
<td>98.6%</td>
<td>95%</td>
<td>87.1%</td>
</tr>
<tr>
<td>Group B</td>
<td>24</td>
<td>15</td>
<td>39</td>
<td>86%</td>
<td>43%</td>
<td>nt*</td>
</tr>
<tr>
<td>Group C</td>
<td>28</td>
<td>18</td>
<td>46</td>
<td>92%</td>
<td>32%</td>
<td>nt</td>
</tr>
</tbody>
</table>

*Table 3.3. Preliminary task for Experiment 2: Percentage of correct responses in the preliminary tasks for the different groups of children (* not tested)*

Overall, these results confirm that the knowledge of the verbal number sequence precedes recognition of Arabic numerals (Bialystock, 1992). The performance in the comparison task was not as good as we expected from previous studies that tested children of the same age (e.g., Siegler & Robinson, 1982). The difficulty encountered in our task may be attributed to its greater level of abstractness: numerals were simply
visually presented and no verbal cues were given. This was also the modality of presentation of the experimental tasks and the ultimate motivation for adopting the specified selection criteria.

3.2.2 Method

Subjects

Forty-eight\textsuperscript{15} children contributed to the analysis of the experimental study: twenty-four (12 female, 12 male) had a mean age of 4 years, 7 months (range from 4.1 to 5.0), twenty-four (13 female, 11 male) had a mean age of 5 years, 8 months (range from 5.1 to 6.2).

Preliminary task: Magnitude comparison results

Stimuli consisted of pairs of Arabic numerals printed in single card. Five different numerical distances were used (1 to 5); four different pairs for each numerical distance were selected for a total of twenty stimuli. The position of the larger digit was counterbalanced across the trials. Subjects were required to indicate, as quickly and accurately as possible, the larger between the two numerals. Reaction times were hand-timed using a stop-watch.

Median correct RTs and arcsine-transformed error proportions for each condition were analysed in two separate 2x2 ANOVAs with age (4/5 y.o. and 5/6 y.o.) as between-subjects factor and numerical distance (five levels) as within-subjects factor.

Younger children were significantly slower (F[1, 46] = 38.02; p<.0001), and made more errors (F[1, 46] = 12.06; p<.005) than older ones. More importantly, both groups showed a normal distance effect: both reaction times (F[4, 184] = 12.68; p<.0001) and error rates (F[4, 184] = 12.53; p<.0001), decreased as the numerical distance between the numbers to be compared increased (see Table 3.4).

\textsuperscript{15} Fifty-two children satisfied the selection criteria, however 4 of the older subjects were excluded from the study in order to have an equal number of participants in the two age groups.
All together the results of the preliminary tasks indicate that children who participated in the experimental study were able, not only to recognise and name the single-digit numerals, but to compare them along their magnitude.

<table>
<thead>
<tr>
<th>numerical distance</th>
<th>4-5 y.o.</th>
<th>5-6 y.o.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>one</td>
<td>10.1 (4.8)</td>
<td>18.8</td>
</tr>
<tr>
<td>two</td>
<td>9.9 (5.0)</td>
<td>4.2</td>
</tr>
<tr>
<td>three</td>
<td>9.6 (4.7)</td>
<td>6.2</td>
</tr>
<tr>
<td>four</td>
<td>9.6 (4.6)</td>
<td>4.2</td>
</tr>
<tr>
<td>five</td>
<td>9.2 (4.6)</td>
<td>1.1</td>
</tr>
</tbody>
</table>

*Table 3.4. Preliminary task, number comparison: Mean Reaction times (in seconds), Standard deviations and Error rates as a function numerical distance and age.*

**Stimuli**

The same stimuli of Experiment 1 were used. However, each single pair was presented only once in each condition, for a total of 60 stimuli. The stimuli were presented in pseudorandom order following the same criteria specified in Experiment 1.

**Design**

Overall, the variables manipulated were: task (numerical and physical); congruity conditions (neutral, congruent and incongruent); numerical distance (close and far). Moreover, two between-subject variables were considered: age (4/5 y.o. and 5/6 y.o.) and order of the tasks (numerical first and physical first).

**Procedure**

Subjects were presented with the numerical and physical comparisons in two sessions on different days (an average of seven days apart). The order of presentation of the tasks was counterbalanced between subjects. In the numerical comparison task, they were
instructed to select the larger in magnitude between the two numbers ignoring variation in
the physical size of the stimuli. Children were told that the two numbers were sometimes
written in the same size, sometimes in different sizes. In the physical comparison task
children were instructed to select the larger digit in physical size ignoring the numerical
value of the stimuli. Few examples were given to make sure that children understood the
tasks. Children were told that both speed and accuracy were important.

Each subject was tested individually. Stimuli were printed in single white card. In the
congruent and incongruent conditions the physically larger digit had a font of 48 point,
the physically smaller had a font of 24 point. In the neutral condition they had an
intermediate size (36 point). Latencies were recorded via a stop-watch from the time the
card was shown to the subject until his response was given. The card was removed when
the subject responded. Any verbal comment or overt strategies used by the child was
recorded. Before the experiment began, subjects were presented with few training trials in
order to practice with the tasks and the experimental setting. Each task lasted
approximately 20 minutes.

3.2.3 Results

The overall error rate was 7.8% (individual error rate never exceeded 18%). However,
as clearly indicated by the analyses below, errors were highly frequent in the numerical
comparison and almost negligible in the physical comparison (see Table 3.5). In
particular, the incongruent condition of the numerical comparison yielded an error rate of
48% in the younger children and 25% in the older ones. Despite the “chance level”
performance in this specific condition, we have reason to believe that younger children
did understand the task, thus producing informative results. In fact, the possibility that
younger children were simply selecting the physically larger numeral disregarding the
semantic dimension may be ruled out by the significant distance effect that emerged from
the analyses. All children were more likely to produce an error when numbers were
numerically close than when numbers were numerically far apart. Yet, caution must be
exercised in analysing the data and in interpreting the outcomes. Both latencies and errors
analyses were performed and reported, even though the contribution of the latter had priority in drawing our conclusions.

Because of marked heterogeneity of variance in the data, a reciprocal transformation of reaction times and an arcsine transformation of error proportions were performed prior to the statistical analysis of the results.

For every subject mean correct reciprocal RTs and arcsine-transformed error proportions for each experimental condition were calculated and entered as dependent variables in a series of ANOVA analyses. Five of the younger children had an error rate of 100% (8/8) in incongruent close pairs in the numerical comparison task. The expected values for these cells were calculated by multiple regression and entered in the analyses.

Reaction times (harmonic means) and error rates for each experimental condition are given in Table 3.5.

<table>
<thead>
<tr>
<th></th>
<th>4-5 y.o.</th>
<th>5-6 y.o.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>%Er</td>
</tr>
<tr>
<td><strong>Numerical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Congruent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td>2390</td>
<td>9.4</td>
</tr>
<tr>
<td>far</td>
<td>2100</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Incongruent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td>6760</td>
<td>68.2</td>
</tr>
<tr>
<td>far</td>
<td>4780</td>
<td>27.1</td>
</tr>
<tr>
<td><strong>Neutral</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td>6290</td>
<td>5.2</td>
</tr>
<tr>
<td>far</td>
<td>5520</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>Physical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Congruent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td>1400</td>
<td>0.0</td>
</tr>
<tr>
<td>far</td>
<td>1420</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>Incongruent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close</td>
<td>1480</td>
<td>0.0</td>
</tr>
<tr>
<td>far</td>
<td>1450</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Neutral</strong></td>
<td>1450</td>
<td>2.5</td>
</tr>
</tbody>
</table>

*Table 3.5. Experiment 2: Harmonic means and Error rates of 4 to 5 year-old and 5 to 6 year-olds as a function of task, congruity and numerical distance.*
Reaction time analysis

A first repeated measures ANOVA with task (numerical and physical), congruity condition (congruent, incongruent and neutral) as within-subjects factors and age (4-5 y.o. and 5-6 y.o.) and order of the tasks (numerical first, physical first) as between-subjects factors was carried out.

The main effect of order of the tasks was not significant and did not enter in any significant interaction; data were then collapsed over this factor.

The effect of age was significant (F[1, 46] = 4.43; p<.05), younger and older children had means reaction times of 2074 ms and 1686 ms respectively.

Physical comparison was significantly faster than numerical comparison (F[1, 46] = 223.7; p<.0001). The task effect interacted marginally with age (F[1, 46] = 3.88; p=.05): in fact, the two groups of children differed in numerical comparison only (F[1, 46] = 8.59; p<.01; physical comparison, F[1, 46] = 1.02; p>.3).

The three levels of congruity differed significantly (F[2, 92] = 31.41, p<.0001), with means of 1650 ms, 1970 ms and 2000 ms respectively for congruent, incongruent and neutral trials. Decomposition into contrasts indicated that congruent trials differed significantly from both neutral and incongruent trials (both, p<.0001) while incongruent trials did not differ from neutral ones (F[1, 46]<1). Congruity interacted significantly with age (F[2, 92] = 5.13, p<.01): the congruity effect was magnified in younger children, although in both groups the effect was highly remarkable (both, p<.0001). More importantly, the effect was highly significant in numerical comparison (F[2, 92] = 42.18, p<.0001) and absent in physical comparison (F[2, 92] = 1.85, p>.1), as indicated by the significant congruity x task interaction (F[2, 92] = 33.69; p<.0001).

The triple interaction between congruity, task and age was also significant (F[2, 92] = 4.49, p<.01). Indeed, the two groups performed similarly in the numerical comparison task, while in the physical comparison older children were slowed down in the neutral condition. The neutral trials in the physical task are visually different from both congruent and incongruent trials, i.e., the same number is displayed in two different sizes (e.g., 2
2). This factor may have played a role in slowing down response times, but the weight of this result remains rather limited.

Data from the numerical comparison only were submitted to a repeated measures ANOVA with congruity (congruent, incongruent and neutral) and numerical distance (close and far) as within-subjects factors and age (4-5 y.o. and 5-6 y.o) as between-subjects factor. The main effects of age (F[1, 46] = 9.21, p<.005) and congruity (F[2, 92] = 41.02, p<.0001) were both significant as well as their interaction (F[2, 92] = 4.82, p<.05), replicating the previous analysis. The main effect of numerical distance was significant (F[1, 46] = 8.23, p<.01) as well as its interaction with age (F[1, 46] = 9.1, p<.005). In fact, reaction times were longer for close pairs than for far pairs but this was true for younger children only (4130 ms and 3460 ms for close and far pairs respectively). These results were replicated when data from the two groups were entered in separate ANOVAs. Younger children showed a significant numerical distance effect (F[1, 21] = 10.84, p<.01); on the other hand, older children did not show a distance effect at any level of congruity (F[1, 21]<1).

To directly compare the effect of instructions on the distance effect, a further ANOVA with task (numerical and physical), congruity condition (congruent and incongruent) and numerical distance (close and far) as within-subjects factors and age (4-5 y.o. and 5-6 y.o) as between-subjects factor was carried out. Replicating previous analyses, a numerical distance effect was found in younger children only and it was limited to numerical comparison (1135 ms and 5 ms for numerical and physical comparisons respectively).

Error analysis

Similar ANOVAs were carried out on arcsine-transformed error proportions. The error analysis indicated that no speed-accuracy trade-off affected the outcomes.

A repeated measures ANOVA with task (numerical and physical), congruity condition (congruent, incongruent and neutral) as within-subjects factors and age (4-5 y.o. and 5-6
order of the tasks (numerical first, physical first) as between-subjects factors was carried out.

Replicating the reaction time analysis, no effect of order of the tasks was found; data were then collapsed over this factor.

Younger children made more error than older ones (9.4% and 6.1% for younger and older children respectively; F[1, 46] = 10.85; p<.005). Errors were significantly more frequent in numerical comparison than in physical comparison (F[1, 46] = 139.8; p<.0001). Moreover, the significant interaction task x age (F[1, 46] = 11.2; p<.005) indicated that the error rates for the two groups were similar in the physical comparison task (1% and 1.1% for 4-5 y.o. and 5-6 y.o. respectively) but differed in the numerical comparison task (18.6% and 11.5% for 4-5 y.o. and 5-6 y.o. respectively) (see Figure 3.2).

Figure 3.2. Experiment 2: Mean Error rates for 4 to 5 year-olds and 5 to 6 year-olds as a function of task and congruity.
The three levels of congruity differed significantly ($F[2, 92] = 94.4, p<.0001$) with means of 1.9%, 18.7% and 3.3% for respectively congruent, incongruent and neutral trials. Decomposition into contrasts indicated that incongruent trials were more error prone than both neutral and congruent ones (both, $p<.0001$), while no difference was detected between neutral and congruent trials. Congruity interacted significantly with age ($F[2, 92] = 11.96, p<.0001$): the congruity effect was magnified in younger children, although in both groups the effect was remarkable (both, $p<.0001$). The significant interaction congruity x task ($F[2, 92] = 90.88, p<.0001$), indicated that this pattern of results was limited to numerical comparison; the overall error rate in physical comparison was indeed extremely low and no appreciable differences were observed between different trials.

The triple interaction between congruity, task and age was also significant ($F[2, 92] = 14.69, p<.0001$). Indeed, the two groups performed similarly in the physical comparison task, whereas in numerical comparison younger children showed a larger interference effect (4-5 y.o., $F[1, 23] = 182.7, p<.0001$; 5-6 y.o., $F[1, 23] = 77.8, p<.0001$).

Error rates from the numerical comparison only were submitted to a repeated measures ANOVA with congruity condition (congruent, incongruent and neutral) and numerical distance (close and far) as within-subjects factors and age (4-5 y.o. and 5-6 y.o.) as between-subjects factor. The main effects of age ($F[1, 46] = 12.42, p<.001$) and congruity ($F[1, 46] = 86.6, p<.0001$) were both significant as well as their interaction ($F[2, 92] = 13.82, p<.0001$) replicating the RT analysis. Decomposition into contrasts indicated that incongruent trials were more error prone than both neutral and congruent ones (both, $p<.0001$), while no difference was detected between neutral and congruent trials. The main effect of numerical distance was significant ($F[1, 45] = 102.9, p<.0001$): subjects were more likely to make errors in close pairs (22.8%) than in far pairs (7%). This effect was highly significant in both groups (both, $p<.0001$), but tended to be larger in younger children. Interestingly, congruity interacted significantly with numerical distance ($F[2, 92] = 44.4, p<.0001$): the interference decreased significantly with
numerical distance, even if in both close (F[1, 46] = 262.25, p<.0001) and far pairs (F[1, 46] = 23.0, p<.0001) the effect reached significance (see Figure 3.3).

![Numerical Comparison Task Graph]

**Figure 3.3** Experiment 2: Mean Error rates in the numerical size comparison task as a function of age, congruity and numerical distance.

When the numerical distance effect was directly compared in the two tasks, the main effect of distance was highly significant (F[1, 46] = 97.92, p<.0001), as well as its interactions with task (F[1, 46] = 93.80, p<.0001) and with congruity (F[1, 46] = 46.67, p<.0001). Moreover, the triple interaction task x congruity x distance was significant (F[1, 46] = 41.25, p<.0001). Indeed, the magnitude of the congruity effect was modulated by the numerical distance (F[1, 46] = 85.87, p<.0001 and F[1, 46] = 30.21, p<.0001, for close and far pairs respectively) but this occurred in numerical comparison only.

**Further comments**

During the experimental study, any verbal comments or strategies adopted by the children were recorded. A descriptive analysis of these data indicates a clear-cut
difference between younger and older children. In the numerical comparison, seven of the older children made self-corrections and explicitly commented on their difficulty in ignoring the discrepancy between physical and numerical sizes. For example, Giulia (5.7 y.o.) commented as follows [trial 2 3]: “No.... era il tre mi sbaglio un po’ perché il più piccolo è scritto più in grande” [No, it was the three! I made a mistake because the smallest was written larger]; Emily (5.11 y.o.) commented as follows the third self-correction “Ancora mi hai imbrogliato ..ma perché ne hai fatti alcuni così grandi?” [You still have cheated me...why are some of the numbers written so big?].

Within the younger group, eleven children made occasional comments (2.5 % of the trials) before or after answering, but only one did actually correct himself using a counting strategy. Most of the children commented on the trials justifying their answers with examples from their own everyday experience. For example, Chiara (4.3 y.o.) stated [trial 3 4] “E’ il 4 io l’anno scorso avevo 3 anni adesso 4 dopo 5” [(the largest) is four! Last year I was three years old, this year I am four and the next one I will be five!]; Fabio (4.3 y.o.) [trial 1 6] “Alessia che ha 6 anni va a scuola, a 1 no, si è piccoli, si dorme sempre” [(The largest is 6)...Alessia is 6 year old and she goes to school. When you are one you don’t go to school, you just sleep all time!]; Erica (4.6 y.o.) [trial 9 4] “E’ il 9, per il 9 ci vogliono due mani, il 9 è di più, è quello vicino al 10!” [It is definitely the 9! For nine you need two hands. It is the one close to ten!].

Younger children sometimes used counting strategies for getting the answers (2.1% of the trials). Interestingly, counting strategies were more frequent in incongruent trials (7.6%) than in neutral trials (2.6%), and never occurred in congruent trials. Verbal counting and finger counting were used with the same frequency. Only one of the younger children, Sara (4.10 y.o.) in one occasion, helped herself by counting real objects that were available. She took a bunch of pencils from her table and by counting them formed two groups, one of 4 and one of 9. Then looking at the two sets she stated: “Nine is larger!”.

In the physical task, younger children occasionally made some comments to remark their confidence in doing the task. As for example: “Sono due sei ma questo e’ il piu’
grande!” [These are both ‘six’ but that one is the larger] or “Facile! Non serve contare” [That’s easy! You don’t need to count].

3.2.4 Discussion

The main results of Experiment 2 may be summarised as follows: 1) physical comparison was faster and more accurate than numerical comparison; 2) age differences emerged in the numerical comparison only; 3) congruity had an effect in the numerical comparison only, 4) congruity speeded up answers to congruent trials and yielded an extremely high error rate in incongruent trials; 5) the distance effect emerged in numerical comparison only; 6) the experience in the first task had no effect whatsoever on the performance in the second one.

The inspection of the overall results (see Table 3.5) leads to a first straightforward conclusion: while the physical comparison did not present major difficulties to children, the numerical one was indeed a challenging task. In line with the outcomes of the preliminary tasks, it took children several seconds to select the numerically larger between two Arabic numerals, even when the physical size was kept constant (i.e., neutral trials); this was particularly evident in younger children. We have already emphasised that this result is relatively unsurprising if we take into account that the task consisted of a symbolic comparison, i.e., the Arabic numerals do not provide themselves the information needed for the comparison and a symbolic reconstruction is required. However, in the neutral condition the error rate was relatively low, suggesting that children might have adopted a strict accuracy criterion when physical information was not provided, increasing significantly their response time.

As expected, the variation in the physical dimension of the stimuli had a great effect on the performance. Children benefited from the congruity between numerical and physical dimensions giving their answers significantly faster than in the other conditions; this result indicates that younger children tended to base their responses on the physical size of the stimuli. On the other hand, when the two dimensions mismatched, their performance dropped dramatically as indicated by the extremely high error rate. These effects were
highly significant in both groups, however younger children showed a larger interference effect (measured by the relative error rate in the incongruent condition), in line with the hypothesis that, as the child grows older, the efficiency of selective attention mechanisms increases (e.g., Dempster, 1992) and the degree of interference that physical information may exert over symbolic information gradually diminishes (e.g., Bisanz et al., 1979).

Taken together, the magnitude of the facilitation and interference effects may suggest that children, especially the younger ones, were often ignoring the numerical dimension being driven exclusively by the perceptual features of the stimuli. This possibility would have been even more likely, given the specific direction of the comparison, i.e., chose the larger. In this case, the well known attentional shift towards the physically larger stimulus would have produced clear-cut facilitation and interference effects as the ones observed. However, the significant effect of numerical distance - all children were more likely to make an error in close than in far pairs - as well as its interaction with congruity - close pairs were more prone to interference than far pairs - rule out the possibility that children were treating Arabic numerals as mere meaningless shapes.

The reaction times analysis yielded intriguing results with regard to the numerical distance of the pairs, producing an insignificant effect for older children. The error analysis, however, indicated unquestionably that in both group’s performance in numerical comparison was highly modulated by the numerical distance. In all conditions of congruity, errors were more frequent when the numbers to be compared where numerically close (e.g., 1 2) than when the numbers where far apart (e.g., 1 6). This result, together with the data from the preliminary task, supports the hypothesis that children as young as five, compared numbers by accessing an internal number representation similar to the adult’s one (e.g., Resnick, 1983). However, the informal observations of children’s comments during the task, suggest that this conclusion may not be generalised to the younger group. The nature of these comments indicates the transition from reasoning about quantities to reasoning about numbers (Resnick, 1992). Younger children may well be able to compare two Arabic numerals, but this task is greatly facilitated by transposing the comparison from a formal to a concrete level: in the former, numbers are conceptual entities themselves, in the latter they just described a
property of a physical quantities (e.g., years, pencils). Moreover, some of the verbal comments suggest that occasionally younger children relied on the ordinal meaning rather than on the cardinal meaning to compare numbers. Though both verbal comments and counting strategies were rather infrequent, they suggest that for youngest children it was not always possible to compare numbers at a symbolic level and that their internal number representation was possibly still developing.

The overall results in the physical comparison indicate that this task was a relatively easy one, even for children as young as four years old: the negligible improvement with age corroborated this conclusion. More importantly, the null effect of congruity indicates that children were not disturbed by the mismatch between physical and numerical dimensions. In other words, the irrelevant semantic dimension was not interfering, at any stage, with the execution of the task. Two alternative interpretations may be put forward. First, the numerical dimension was indeed processed but filtered out being irrelevant to the task. Second, the numerical dimension did not interfere because it was not processed in the first place. Given that the first alternative would imply the intervention of efficient inhibitory mechanisms the lack of which is indicated by children's performance in the numerical task, the second interpretation seems the most plausible one. These results seem to suggest that, even though by the age of 5, children are able to compare numbers by their magnitude, this is a relatively unsalient dimension of Arabic numerals.

In summary, these findings suggest that pre-schooler's number representation is, in some respects, refined enough to give rise to a reliable distance effect in a number comparison task, but far from being autonomously accessed whatever the task demands. The question that follows is how long does it take children to establish the strong association between numerals and numerosities that leads adults to be affected by the incongruity between physical and numerical dimensions when this latter is task irrelevant. This issue was addressed in Experiment 3.
3.3 EXPERIMENT 3

To trace the developmental trends of this phenomenon, the numerical Stroop-paradigm was administered to primary-school children of 6, 8 and 10 years of age.

It is well established that children enter school with a substantial amount of numerical competence and informal arithmetical knowledge (e.g., Fuson, 1992; Resnick, 1992). This knowledge mainly derived from their every-day experience with concrete numerical situations such as counting objects. The introduction to formal education requires children to learn the arithmetical symbolism and to gradually progress from reasoning about concrete quantities to reasoning about numbers as abstract entities (Resnick, 1992). As clearly stated by Resnick, in the mathematics of numbers "...numbers have properties rather than being properties of physical material" (p.404). At this stage, children are formally introduced to the written symbols and gradually learn to operate upon them (Hughes, 1986). Over the course of learning, the association between the numerals and the meaning they convey (i.e., their numerical value) is assumed to strengthen and eventually acquire the autonomous property that characterises adults performance. Thus, by administering the number-Stroop paradigm to first-, third and fifth-grade children we should be able to capture the gradual changes in the intentional and autonomous processing of numerical information.

3.3.1 Method

Pre-test

Prerequisite for participating in the experimental study was an accurate performance in the following tasks: 1) reading of Arabic numerals (1-9); 2) verbal counting (forward and backward from 1 to 9); 3) matching Arabic numerals to numerosities (1-9). All recruited children (N=72) satisfied the above criteria.
Subjects

Twenty-four first-grade children (12 female, 12 male), mean age of 6 years, 6 months, twenty-four third-grade children (13 female, 11 male) mean age of 8 years, four months, twenty-four fifth-grade children (12 female, 12 male) mean age of 10 years, 3 months, participated in the study. All children were recruited in primary schools in northern Italy. The testing took place in late spring, thus even the younger participants had almost complete the first year of school and received some arithmetical training.

Stimuli and procedure

The same stimuli and procedure of Experiment 1 were adopted. Children were asked to do the task the best they could do, but they were informed that their performance would not be graded for school in any way.

Apparatus

A Macintosh Powerbook 175 was used to present the stimuli. Psychlab sof.ware (Version. 99, Gum, 1992) was used to display stimuli and record reaction times.

Design

Overall, the variables manipulated were: task (numerical and physical); congruity conditions (neutral, congruent and incongruent); numerical distance (close and far). Moreover, two between-subjects variables were considered: grade (first, third, fifth) and order of the tasks (numerical first, physical first).

3.3.2 Results

Reaction times (means of medians) and error rates for each experimental condition are given in Table 3.6.
### Table 3.6. Experiment 3: Mean Reaction times, Standard deviations and Error rates for the different grades as a function of task, congruity and numerical distance.

<table>
<thead>
<tr>
<th></th>
<th>1st-grade</th>
<th></th>
<th></th>
<th>3rd-grade</th>
<th></th>
<th></th>
<th>5th-grade</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>SD</td>
<td>%Er</td>
<td>RT</td>
<td>SD</td>
<td>%Er</td>
<td>RT</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Numerical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Congruent</strong></td>
<td>close</td>
<td>2079</td>
<td>518</td>
<td>2.3</td>
<td>1410</td>
<td>430</td>
<td>0.5</td>
<td>930</td>
</tr>
<tr>
<td></td>
<td>far</td>
<td>1891</td>
<td>634</td>
<td>0.3</td>
<td>1276</td>
<td>382</td>
<td>0.0</td>
<td>842</td>
</tr>
<tr>
<td><strong>Incongruent</strong></td>
<td>close</td>
<td>2182</td>
<td>667</td>
<td>7.0</td>
<td>1509</td>
<td>359</td>
<td>8.6</td>
<td>1068</td>
</tr>
<tr>
<td></td>
<td>far</td>
<td>2015</td>
<td>463</td>
<td>2.9</td>
<td>1430</td>
<td>363</td>
<td>5.2</td>
<td>1002</td>
</tr>
<tr>
<td><strong>Neutral</strong></td>
<td>close</td>
<td>2130</td>
<td>619</td>
<td>2.3</td>
<td>1503</td>
<td>400</td>
<td>2.3</td>
<td>993</td>
</tr>
<tr>
<td></td>
<td>far</td>
<td>2017</td>
<td>464</td>
<td>0.3</td>
<td>1383</td>
<td>405</td>
<td>0.5</td>
<td>920</td>
</tr>
<tr>
<td><strong>Physical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Congruent</strong></td>
<td>close</td>
<td>1208</td>
<td>226</td>
<td>0.0</td>
<td>958</td>
<td>362</td>
<td>0.3</td>
<td>665</td>
</tr>
<tr>
<td></td>
<td>far</td>
<td>1217</td>
<td>230</td>
<td>1.0</td>
<td>968</td>
<td>323</td>
<td>0.8</td>
<td>683</td>
</tr>
<tr>
<td><strong>Incongruent</strong></td>
<td>close</td>
<td>1192</td>
<td>221</td>
<td>0.0</td>
<td>1003</td>
<td>368</td>
<td>0.5</td>
<td>722</td>
</tr>
<tr>
<td></td>
<td>far</td>
<td>1233</td>
<td>232</td>
<td>1.3</td>
<td>1005</td>
<td>393</td>
<td>1.0</td>
<td>757</td>
</tr>
<tr>
<td><strong>Neutral</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1188</td>
<td>231</td>
<td>0.5</td>
<td>965</td>
<td>359</td>
<td>0.0</td>
<td>677</td>
</tr>
</tbody>
</table>

The overall error rate was 1.7% (the individual error rate never exceeded 0.5%) and no further overall error analysis was performed; however, error distribution fits the RT data. In numerical comparison there were 0.65% errors in congruent trials, 6.4% in incongruent trials and 1.3% in the neutral ones. In the physical comparison there were 0.35% errors in congruent trials, 1.3% in incongruent trials and 0.2% in the neutral ones. Thus, error distributions indicated that no speed-accuracy trade-off affected the outcomes.

For every subject median correct RTs for each condition were calculated and entered as dependent variable in a series of ANOVA analyses.
A first repeated measures ANOVA with task (numerical and physical), congruity condition (congruent, incongruent and neutral) as within-subject factors and grade (first, third and fifth), order of the tasks (numerical first, physical first) as between-subject factors was carried out. The analysis indicated that RTs decreased with grade (F[2, 68] = 40.2; p<.0001), follow-up comparisons revealed that each group differed significantly from the others (Newman-Keuls, p=.01).

The main effect of order of the tasks was not significant.

Physical comparison was significantly faster than numerical comparison (1470 ms versus 956 ms; F[1, 68] = 195.39; p<.0001); this difference was modulated by order of the tasks (F[1, 68] = 5.45; p<.05): although all subjects showed an advantage in the physical comparison task, within each task, the group who performed the task as the first one, tended to be faster than the other group. In other words, whatever the second task, the experience of the first one seemed to have some influence on it. It is plausible that children encountered some difficulties ignoring a dimension than was previously relevant, regardless of its nature.

The effect of task interacted significantly with age (F[2, 68] = 22.23, p<.0001). In fact, the RT-decrease with grade level was much steeper for numerical comparison than for physical comparison, although in both tasks the effect was highly significant (both, p<.0001) (Figure 3.4).

The three levels of congruity differed significantly (F[2, 136] = 55.92, p<.0001) with means of 1172.1 ms, 1253.9 ms and 1213.4 ms for respectively congruent, incongruent and neutral trials. Decomposition into contrasts indicated that each condition differed significantly from all the others (all, p<.0001). These differences were modulated by order of tasks, (F[2, 136] = 3.70, p<.05): for subjects presented with numerical comparison in the first session, all levels of congruity differed between each others (all, p<.0001), while for subjects presented with physical comparison in the first session difference between congruent and incongruent trials as well as between congruent and neutral trials were highly significant (both, p<.001) while the difference between neutral and incongruent trials was significant to a lesser extent (p<.05).
In order to evaluate to what extent the experience in the first task influenced the performance in the second one, the interaction between congruity effect and order of the presentation was tested separately for the two tasks. The size of the congruity effect in numerical comparison was similar in both groups. On the other hand, subjects who performed the physical task after the numerical one showed a larger congruity effect (F[2, 70] = 14.28, p<.0001) than subjects who performed the physical task as the first one (F[2, 70] = 3.79, p<.05).

The effect of congruity was highly modulated by task (F[2, 136] = 22.13, p<.0001). Decomposition into contrasts showed that while in numerical comparison all levels of congruity differed significantly from each others (all, p<.001), in physical comparison incongruent trials differed significantly from both neutral and congruent ones (both, p<.005), but no difference was found between congruent and neutral trials.
Separate analyses were then carried out for the three groups. In all analyses, the main effect of task was highly significant (all, p<.0001); on the other hand, congruity effect and its interaction with task were interestingly modulated by grade level (Figure 3.5).

In the first-grade, congruity was significant (F[2, 46] = 4.71, p<.05) as well as its interaction with task (F[2, 46] = 4.47, p<.05). Decomposition into contrasts revealed that in numerical comparison, congruent trials were faster than both neutral and incongruent ones (both, p<.005), while in physical comparison no differences were detected at any level of congruity.

In the third-grade, the main effect of congruity was highly significant (F[2, 46]=34.32, p<.0001) as well as its interaction with task (F[2, 46] = 11.26, p<.0001). Decomposition into contrasts revealed that in numerical comparison, congruent trials were faster than neutral ones (p<.0001) and neutral faster than incongruent ones (p<.05). Interestingly, in physical comparison incongruent trials were slower than both neutral and congruent ones (both, p<.05). However, order of the tasks marginally modulated the congruity effect in physical comparison (F[2, 44] = 2.81, p=.07): subjects presented with the physical comparison in the first session showed a reduced congruity effect (p=.1 and p<.01 for first and second sessions respectively).

Fifth-grade children showed a main effect of congruity (F[2, 46] = 83.82, p<.0001) as well as an interaction congruity x task, (F[2, 46] = 22.1, p<.0001). Decomposition into contrasts revealed that in numerical comparison, all levels of congruity differ significantly between them (all, p<.0001); in physical comparison incongruent trials were now significantly slower than both neutral and congruent ones (both, p<.0001). Order of the tasks modulated the congruity effect in physical comparison (F[2, 44] = 4.14, p<.05): subjects presented with the physical comparison in the first session showed a reduced but still significant congruity effect (p<.005 and p<.0001 for first and second sessions respectively).
Figure 3.5. Experiment 3: Mean Reaction times as a function of task, congruity and grade level.

Data from the numerical comparison only were submitted to a repeated measures ANOVA with congruity (congruent, incongruent and neutral) and numerical distance (close and far) as within-subjects factors and grade (first, third and fifth) and order of the tasks as between-subjects factors.

The main effect of order of the tasks was not significant and did not enter in any significant interaction; data were then collapsed over this factor.

The main effects of grade ($F[2, 69] = 44.23$, $p<.0001$) and congruity ($F[2, 138] = 41.25$, $p<.0001$) were both significant replicating the previous analysis. The main effect of numerical distance was significant ($F[1, 69] = 87.85$; $p<.0001$) as well as its interaction
with grade (F[2, 69] = 3.66, p<.05). In fact, reaction times for close pairs (1533.7 ms) were longer than for far pairs (1419.5 ms) and the size of the effect decreased with age. No further significant interactions were observed.

To verify directly whether the congruity effect varies in close and far pairs and to what extent these effects depend on the judged dimension, a further ANOVA with task (numerical and physical), congruity (congruent and incongruent) and numerical distance (close and far) as within-subjects factors and grade (first, third and fifth) as between-subjects factor was carried out.

Replicating previous analyses, a numerical distance effect was found (F[1, 69] = 47.94, p<.0001) and it was modulated by grade (F[2, 69] = 4.05, p<.05). More importantly, numerical distance interacted with task (F[1, 69] = 87.77, p<.0001) as well as with task and grade (F[2, 69] = 4.49, p<.05). Separate analyses for the three groups revealed the following results: first- and third-grade children showed a distance effect in numerical comparison only (first-graders, 2130 ms and 1953 ms for close and far pairs respectively; third-graders, 1460 ms and 1353 ms for close and far pairs respectively). On the other hand, fifth-grade children showed a distance effect in numerical comparison (999 ms versus 922 ms, p<.0001) as well as in physical comparison (693 ms versus 720 ms p<.01); in the latter, however close pairs were answered faster than far ones.

In neither third and fifth-grade children was the congruity effect in the physical task modulated by numerical distance. Only the inspection of the error distributions, suggests that fifth-grade children were more likely to be mislead by the numerical dimension in comparing the physical size of far pairs (3.6% versus 0% for incongruent and congruent trials, respectively).

*Numerical comparison: Error rate analysis*

Arcsine-transformed error proportions from the numerical comparison only were submitted to a repeated measures ANOVA with congruity (congruent, incongruent and neutral) and numerical distance (close and far) as within-subjects factors and grade (first, third and fifth) and order of the tasks as between-subjects factors.
The main effect of order of the tasks was not significant and did not enter in any significant interaction; data were then collapsed over this factor.

The main effect of congruity was significant \( (F[2, 138] = 38.79; \) decomposition into contrasts revealed that errors were significantly more frequent in incongruent trials than in congruent or neutral ones (both, \( p<.0001 \)), but no difference was detected between the latter. The main effect of numerical distance was significant, \( (F[1, 69] = 24.48; p<.0001) \): errors were more frequent in close pairs (4.5%) than in far pairs (1.4%). Moreover the interaction between numerical distance and congruity \( (F[2, 138] = 6.09, p<.01) \) indicated that the disadvantage for incongruent trials was maximised when these were also numerically close.

**Effects size**

In line with other studies focused on age-related changes in Stroop-like tasks (e.g., Spieler et al. 1996), conclusions on the size of the effects may be drawn only once the overall changes in the processing speed are taken into account. To this aim, interference and facilitation ratios were computed for each group, as suggested by Spieler et al (1996; p.467). Interference ratios were calculated by subtracting each participant's median RT in the neutral condition from the median RT in the incongruent condition and dividing it by the median of the neutral condition. Similarly, facilitation ratios were computed by subtracting each participant's median RT in the congruent condition from the median RT in the neutral condition and dividing it by the median of the neutral condition.

Four separate between-subjects ANOVAs for interference and facilitation effects in both numerical and physical tasks, were carried out.

In both tasks, the analysis of the interference ratios only, revealed significant differences between the groups \( (F[2, 69] = 5.14, p<.01 \) and \( F[2, 69] = 3.56, p<.05 \) for numerical and physical comparisons respectively). In both cases, the size of the effect increased with the age of children; however, in numerical comparison, each group differed from the others; in physical comparison fifth-grade only differed from both first- and third-grade (see Table 3.7).
### Table 3.7. Experiment 3: Mean interference and facilitation ratios for numerical and physical comparisons as a function of grade level.

*Interference = (incongruent-neutral)/neutral; Facilitation = (neutral-congruent)/neutral*

<table>
<thead>
<tr>
<th>Task</th>
<th>Grade</th>
<th>Facilitation</th>
<th>Interference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical</td>
<td>first</td>
<td>.020</td>
<td>.040</td>
</tr>
<tr>
<td></td>
<td>third</td>
<td>.033</td>
<td>.070</td>
</tr>
<tr>
<td></td>
<td>fifth</td>
<td>.076</td>
<td>.078</td>
</tr>
<tr>
<td>Physical</td>
<td>first</td>
<td>-.014</td>
<td>.022</td>
</tr>
<tr>
<td></td>
<td>third</td>
<td>.003</td>
<td>.038</td>
</tr>
<tr>
<td></td>
<td>fifth</td>
<td>.008</td>
<td>.076</td>
</tr>
</tbody>
</table>

3.4.3 Discussion

Overall, the results of Experiment 3 show that the number-Stroop paradigm succeeded in capturing the age-related modification in the way in which primary-school children process numerical stimuli. The most interesting result was that developmental changes in both numerical and physical tasks were not simply quantitative, i.e. related to the rate of processing, but also, and more importantly, qualitative.

Physical comparison was always faster than numerical comparison and, at any age, children's performance was almost error-free. On the other hand, the numerical comparison was clearly more difficult. The variation in physical dimension was not the only source of difficulty; indeed, younger children answered neutral trials almost at the same rate as incongruent trials. It is important to note, however, that the error analysis disclosed a significant interference effect for all groups of children. As already emphasised in discussing preschooler's results, the greater response variability in younger children’s RTs, may obscure effects that emerge more neatly in the error analysis.
Yet, the RTs analysis indicated that the source of the congruity effect varied with grade: a facilitation effect was shown by all groups but interference effect emerged from the third-grade only. Moreover, the effect-size analysis indicated that advantage for the congruent trials was steady across ages, while interference increased with age. Thus, despite the immature inhibitory mechanisms (e.g., Dempster, 1992) and the expected primacy of perceptual information, in the present study younger children were less prone to interference than older ones. However, the error analysis seems to suggest that also younger children are disturbed by the mismatch between physical and numerical sizes, although to a lesser extent. These results corroborate the hypothesis that, in the number-Stroop paradigm, the incongruity is determined by the level of "integration" of the conflicting dimensions, i.e. physical size of the Arabic numerals and numerical size represented by them. In line with our predictions, the integration of the two dimensions gradually develops as the children's experience with arithmetical symbols progresses.

As previously reported, all subjects answered faster and more accurately to far pairs than to close pairs and the magnitude of the effect decreased as grade level increased (Duncan & McFarland, 1980; Sekuler & Mierkiewicz, 1977). As for adult subjects (Experiment 1), the RT analysis indicated that the congruity effect was of comparable size in close and far pairs despite the fact that the latter yielded, overall, faster responses. However, additional evidence from the error analysis indicated that children were more likely to be misled by the physical size of the stimuli when they had to compare numerically close numbers. Thus, it may be that when the numerical comparison was more time demanding, there was more opportunity for interference to occur.

The results of the physical task are quite straightforward. The congruity effect was totally absent in first-grader's performance, emerged in the third-graders and was highly significant in fifth-graders. In other words, the mismatch between physical and semantic information was not perceived before the age of 8. Moreover, at this age the sensitivity to irrelevant numerical information was enhanced by previous numerical comparison; while in fifth-grade children the activation of numerical size in physical comparison was a more reliable phenomenon. These findings indicate that access to number meaning becomes autonomous gradually over the course of skill acquisition; the ability to accurately
compare numbers is clearly a prerequisite but does not entail the autonomous activation of their semantics.

Interestingly, when the incongruity comes to slow down the performance, this occurs in the processing of all number pairs: Children were equally disturbed by the irrelevant numerical information, whether the numbers were numerically close or far apart. In the attempt of interpreting this results within Tzelgov et al.’s. approach, we may suggest that the interference effect in children was due to the gradual automatization of the algorithm-based process, responsible for the direct mapping of numbers onto their magnitude representations, rather than to the primacy of a fast and memory-based classification of the numbers in small and large categories. We may tentatively suggest that this latter process is achieved later in the course of development and it reflects a further stage in the refinement of number knowledge. However, additional factors should not be underestimated in the evaluation of the differences between older children’s and adult’s performances. In particular, as age and practice increase, the performance on the task (e.g., physical comparison) speeds up dramatically and the time for interference clearly diminished. Thus, though the interference caused by the numerical dimension increases with age, its impact on the performance is clearly modulated by the general improvement in the task (Schadler & Thissen, 1981). In other words, the age-related differences with regards to the interaction between numerical distance and congruity may partially reflect an overall difference in the speed of processing.

3.4 SUMMARY AND CONCLUSIONS

The study reported in the present Chapter probes the rise of automaticity in accessing number meaning. To this aim we adopted a number-Stroop paradigm consisting of two comparative tasks where the relevant dimension is conceptual (numerical comparison) and physical (physical comparison) size respectively.
The results of Experiment 1 clearly indicate that congruity had an effect in both tasks, suggesting that the irrelevant dimension, either physical or numerical size, was autonomously activated. Critically, adults could not compare two digits along their physical size disregarding their numerical values. In fact, when physical and numerical sizes conveyed conflicting information, the selection of the physical larger digit was significantly slowed down. This result provides support to the hypothesis that numerical information is autonomously activated even if irrelevant to the task (e.g., Henik & Tzelgov, 1982).

To trace the developmental trend of this phenomenon, preschoolers and primary-school children were presented with the same experimental paradigm. This procedure proved to be adequate in capturing both quantitative and qualitative changes that emerge in the course of development. Clearly, these changes are the result of the combination of both general (e.g., speed of processing, attentional development) and specific (e.g., number knowledge) factors. Overall, the pattern of interference and facilitation in a number-Stroop task is primarily determined by the parallel processing of the numerical and the physical dimensions (e.g., Henik & Tzelgov, 1982). We hypothesised that the level of automaticity of the physical and the numerical size changes significantly with age, but, possibly, at different rate; at the same time, attentional abilities and general efficiency in the performance are also improving with practice and they are both likely to exert an influence on the performance (e.g., Rosinsky et al., 1975). Accordingly, the physical size of the stimuli had an impact on the performance in numerical comparison since the very early age, while the numerical size did not influence the physical comparison until the third-grade. Evidence for autonomous processing of number magnitude was thus found only in children above the age of eight.

The present findings strongly suggest that for young children the access to the magnitude representation is far from being automatic. Even if children are able to compare numbers along their magnitude, this is still a potentially negligible dimension of an Arabic numeral; extensive practice with the number symbols is required before their magnitude becomes a salient feature. Thus, we may conclude that the autonomous access
to numerical information arises gradually over the course of learning as children’s numerical skills and arithmetical knowledge progress.
CHAPTER 4

THE AUTONOMOUS PROCESSING OF MAGNITUDE INFORMATION:
INFLUENCE OF STIMULI SELECTION, NOTATION AND TASK DEMANDS.

INTRODUCTION

The results of Experiments 1 and 3 are consistent with the suggestion according to which, once the association between Arabic numerals and the magnitude they represent is fully established, a) magnitude information is autonomously accessed though irrelevant to the task; b) but differs qualitatively from the one activated under intentional conditions. The purpose of the studies reported in the present Chapter was to further explore the extent to which access to magnitude is autonomously activated by adopting an experimental paradigm consisting of two same-different tasks varying in the level of processing required. Experiments 4 (4N and 4P) and 5 (5N and 5P), reported in the first part of the Chapter, investigated the role of the stimuli selection and of the stimuli notation in modulating autonomous semantic access. In the second part of the Chapter, the number matching paradigm was used to further explore the developmental changes in intentional and autonomous processing of numerical information (Experiments 6 (6N and 6P) and 7 (7N and 7P)).

Numbers can be represented by two different written notations, the Arabic notation (e.g., 2) and the verbal notation (e.g., two) that convey the same meaning. Thus, by varying the surface form of the stimuli (Arabic versus verbal), pairs of numerals can be
matched either for physical identity (i.e., at the graphemic level, e.g., 1 ONE constitutes a *different* pair) or for semantic identity\(^{16}\) (e.g., 1 ONE constitutes an *identical* pair). Yet, pairs of non-identical numbers can vary in their semantic similarity, since they can be numerically close (e.g., 1 TWO) or numerically far apart (e.g., 1 NINE); the effect of this variable on the same-different judgement can be considered an index of semantic access. This experimental paradigm was recently adopted by Dehaene and Akhavein (1995) in a study addressing both the issue of autonomy in number processing and the issue of asemantic numerical transcoding. Though a brief summary of their contribution has been previously reported (see Chapter 2), a more thoughtful evaluation is of critical importance and indeed motivated the present research.

Dehaene and Akhavein (1995) used an extension of the Duncan and McFarland (1980) same-different task where pairs of numbers, written in same (pure trials, e.g., 2 2, ONE NINE) or different notations (mixed trials, e.g., 2 TWO, 1 NINE), had to be matched at the semantic level (numerical matching task) or at the physical level (physical matching task). The symbolic distance effect was employed as an index of semantic\(^{17}\) access: both numerically close and numerically far pairs had to be answered as *different*; however, faster responses to the latter would indicate that stimuli were semantically processed. The effect of task (numerical and physical) on the distance effect was assumed to provide an index of the autonomy of semantic access. Moreover, the absence of the distance effect together with a slow and/or inaccurate performance in answering *different* to arabic-verbal pairs representing the same number (e.g., 2 TWO) was considered evidence for asemantic transcoding. All current models postulate distinct input systems for the comprehension of Arabic and verbal numerals (Campbell & Clark, 1988, 1992; Cipolotti & Butterworth, 1995; Dehaene, 1992; McCloskey, 1992). However, two opposite hypotheses have been advanced with regard to the convergence of these two distinct pathways: one assumes that Arabic and verbal numerals are translated into a

---

\(^{16}\) Analogous same-different tasks, requiring to match letter across variations in case or font, were first introduced by Posner & Mitchell (1967). In the original letter-matching paradigm, subjects were required to judge whether two items were physically identical (e.g., A A) or not (e.g., A B) or if they were identical by virtue of having the same name (e.g., A a) or not (e.g., A b).

\(^{17}\) Semantics here refers to magnitude information.
common semantic representation before any further processing may take place (McCloskey et al., 1995; McCloskey, 1992); the second one postulates asemantic transcoding pathways along which Arabic and verbal numerals might be directly mapped without semantic mediation (Campbell & Clark, 1988, 1992; Cipolotti & Butterworth, 1995; Dehaene, 1992; Noel & Seron, 1993; Seron & Deloche, 1983). In the matching paradigm, the existence of an asemantic pathway was probed by testing whether the matching of an Arabic numeral and its corresponding verbal form (e.g., 2 TWO) may be achieved with no evidence of semantic processing. In principle, this mapping is likely to occur at the phonological level, where ‘2’ and ‘two’ have the same ‘name’, but, possibly, also by means of others common codes (Figure 4.1).

\[
\begin{array}{c}
2 \quad \text{two} \\
\downarrow
\end{array}
\]

**Figure 4.1. Schematic representation of the putative processing pathways in a number matching task.**

Dehaene and Akhavein ran two pairs of experiments: in the first one, single digit numbers were used as stimuli; in the second one 1- and 2-digits numbers, chosen according to both lexical and semantic criteria, served as stimuli.

The results of the first pair of experiments (Experiments 1 and 2) may be summarised as follows: in the numerical matching task a significant distance effect was observed in all
trials, suggesting that subjects were always accessing magnitude representations (even if in pure trials they could have answered on the basis of physical or lexical identity). The size of the effect was indeed similar in Arabic, verbal and mixed trials; this evidence was considered in line with other findings of negligible notational effects in number processing (Dehaene et al., 1993; Noel & Seron, 1992; Sokol et al., 1991). To further support the conclusion that magnitude representations were accessed during the task, the authors reported three minor additional effects. First, a significant magnitude effect was reported: within equal and close pairs, trials including small numbers (e.g., 1 1, 2 1) were answered faster than trials including large numbers (e.g., 8 8, 8 9). Second, they claimed that, similarly to the SNARC effect, a preferred left-right ordering of the numbers was found. In particular, far pairs only yielded faster responses when the smaller number was on the left (e.g., 1 9). Third, they reported an effect of parity on the same-different judgements: subjects were slower to answer different to pairs including numbers of the same parity (e.g., 2 8) than to pairs including numbers of different parity (e.g., 1 8); however, this difference reached significance in Arabic pairs only.

In the physical matching task, the distance effect was still significant but in pure trials only; mixed trials were indeed highly dissimilar at a perceptual level and allowed fast different responses. Moreover, in mixed trials, equal pairs were more error prone than unequal pairs even if “...the error trend did not quite reach significance (p=.105)”(p.321). Thus, the authors concluded that subjects were noticing that 2 and TWO represented the same number and that this information occasionally interfered with the decision that these stimuli were physically different. However, the absence of a distance effect in mixed trials (e.g., 1 TWO and 1 NINE yielded similar response times) suggested that this interference arose at an asemantic level.

In the second pair of experiments (Experiments 3 and 4) the authors intended to replicate their results by controlling to what extent the distance effect in the matching tasks may be attributed to lexical rather than to semantic processing. According to this alternative hypothesis, the difficulty discriminating between consecutive numbers would arise from their intralexical associations, established through repetition of the counting sequence, rather than from a similarity in their magnitude. In these terms, the effect
would reflect a lexical distance rather than a semantic distance between the stimuli. The new stimuli were selected so that numerical and lexical distances between the to-be-compared numbers were not confounded; in particular, the lexical distance was kept constant while the numerical distance varied. This criterion was met by virtue of the specific structure of the number lexicon as depicted by current models of number processing (e.g., Deloche & Seron, 1984; McCloskey et al., 1986). Within this framework, numerals are organised in separate stacks for units, teens and tens and, within each stack, the individual entries are specified by their ordinal position. For example, 3 and 9 would occupy the third and ninth positions in the units stack as 13 and 19 the third and ninth positions in the teens stack. In other words, the lexical distance between 3 and 19 and between 9 and 13 is the same, even if the latter pair is numerically closer than the former. Using the new set of stimuli, a distance effect emerged in the numerical matching task only; though this result favours a semantic interpretation of the distance effect, at the same time it calls into question the hypothesis of an autonomous semantic access to magnitude information. The authors invoked several plausible explanations for the reduced effect of semantic information, mainly concerning the numerical and physical properties of the stimuli used. On the one hand, access to magnitude information may be less automatic or slower for larger numbers (e.g., teens and tens) than for smaller ones (i.e., ones), on the other hand, the greater visual dissimilarity between the stimuli may have facilitated the same-different judgements preventing semantic information interfering with the decision.

Although the experimental procedure used by Dehaene and Akhavein proved to be appropriate in testing a series of hypotheses, some caution is required with regard to their findings. Their contribution adds to existing evidence that autonomous access to magnitude representation characterises processing of Arabic numerals (e.g., Duncan & McFarland, 1980) and it suggests that this property extends to processing of verbal numerals. However, the evidence for the latter hypothesis is rather inconclusive. In particular, in the verbal trials, stimuli were not balanced for visual similarity. In Experiments 1 and 2, close pairs were visually more similar (i.e., half of the pairs included
words of the same length, i.e., ONE TWO) than far pairs (e.g. ONE EIGHT). This confounding variable does not allow us to consider the distance effect in pure verbal trials as a conclusive result. It remains to be determined whether the difference in performance in close and far pairs was due to a semantic or simply to a visual factor. Interestingly, in Experiment 4, the mismatch in visual similarity was in favour of close pairs and this advantage was reflected in both reaction time and error distributions (see their Figure 6, p.323). Similarly, the magnitude effect in both verbal and mixed trials may not be considered definitive, given that number names for small (i.e., ONE, TWO) and large numbers (i.e., EIGHT, NINE) were not balanced either for length or for number of syllables. The reliability of other semantic effects, i.e., the relative position of the numbers within a pair and their parity status, is also questionable. Despite the fact that the distance effect in pure trials was of the same size in both numerical and physical matching, suggesting that access to number meaning was not affected by the instructions, these additional effects emerged in the numerical matching task only.

Furthermore, the evidence for an asemantic transcoding process is based on an extremely marginal result. The critical difference between equal (e.g., 2 TWO) and unequal (e.g., 2 NINE) mixed pairs in the physical matching task was not significant, either in the RT analysis or in the error analysis. Error rates were very low and the trend reported was statistically negligible in Experiment 2 (p>1), and marginal in Experiment 4 (p=.062)\(^\text{18}\). In principle, the interference effect from the 'number identity' on the same-different classification of mixed trials may be considered to result from a failure in ignoring irrelevant information of the stimuli. It is conceivable that this interference may be reduced with practice, as observed in several interference-sensitive experimental paradigms (e.g., Roger & Fisk, 1991; Shiffrin, 1988). In this case, it is possible that subjects were indeed disturbed by the 'number identity' between Arabic and verbal numerals (e.g., 3 THREE), but that they may have come to inhibit or to ignore this information over the course of the experiment. This hypothesis may be easily verified by controlling the influence of practice on the performance.

\(^{18}\) Moreover in Experiment 4, unequal pairs were presented with double frequency than equal pairs, a difference that might have affected the relative error rate.
Overall, in Dehaene and Akhavein’s study the most robust result was the significant and comparable distance effect in both numerical matching (Experiment 1) and physical matching (Experiment 2) tasks. This evidence suggests that both Arabic and verbal trials were processed all the way to semantics during the tasks. Even though the physical characteristics of the stimuli might have play a role in the processing of the verbal trials, the distance effect in Arabic trials remains a robust result.

However, Dehaene and Akhavein’s experimental trials included only four digits (i.e., 1, 2, 8, 9) that, within the range of one-digit numbers (i.e., 1-9), correspond to the extreme values. It is well established that within a fixed series (e.g., numbers 1-9) end-terms are processed faster than middle terms (e.g., Banks, 1977; Moyer & Dumais, 1978); this effect is clearly confounded with the distance effect, given that pairs far apart are more likely to include end-terms (i.e., 1 9, 1 8, 2 8, 2 9). It is possible that, also when magnitude is irrelevant to the task, these numbers may be processed faster than the others, as it is the case in standard number comparison tasks (e.g. Banks, 1977). This may be particularly relevant if we consider that numerical information which is accessed in autonomous conditions may qualitatively differ from the one accessed in intentional processing (Tzelgov et al., 1992). It has been suggested that in autonomous conditions a dichotomous small/large classification resulting from the primary encoding stage dominates the process, though a parallel activation to a more elaborated magnitude representation is also initiated. It is plausible that the speed and accuracy of this classification may vary with the relative size of the numbers: for example, 9 yields the stronger association to the category ‘large’ given that, in any combination with the other digits, it is always the larger term; in other words, extreme digits are prototypical in their category (small/large; Tzelgov et al., 1992). It follows that, for larger numerical distance and in particular when far pairs include extreme values, the numerical information may be available earlier. Thus, Dehaene and Akhavein’s selection of the four extreme values as experimental trials might have determined or maximised the occurrence of the distance effect in both numerical and physical matching tasks. Moreover, even if distractor trials with the digits 3 to 7 were included in the experiments, they were few (80 out of 384
experimental trials) and, possibly, not enough to balance the high frequency of the four target numbers. Thus, if in autonomous conditions, access to number meaning is dominated by a categorical small/large classification, the speed of which is in turn modulated by the relative size of the numbers, we would expect semantic effects in same-different judgements to vary with the specific range of numbers presented. In particular, if numerically distant pairs include the smallest (1 or 2) and the largest numbers (8 and 9) within single-digit numbers, the probability of observing a significant distance effect would be maximised; however, if this criterion is no longer met we would expect the interference from semantic information to be greatly reduced. This should apply specifically when the task requirements minimise the effect of facilitation from semantic processing, i.e., in the physical matching task.

Aim of the first study (Experiments 4N and 4P) reported in this Chapter was to replicate Dehaene and Akhavein’s numerical matching experiments, controlling more carefully some variables involved in the tasks. Moreover, the hypothesis that a small/large classification dominates the processing of numbers in autonomous conditions, and that the speed of this classification depends on the absolute characteristics of the numbers, was also tested. To this aim, the comparison between numerical and physical tasks is of critical importance, since the latter provides the stringent condition for measuring the autonomy of semantic access. In the present study, close and far conditions included pairs of equal numerical distance (distance 1 for close pairs and distance 4 for far pairs; in Dehaene and Akhavein’s study, far pairs were of different numerical distances) and laterality was a criterion for selecting the experimental trials: close pairs are all unilateral (both numbers are small - 1, 2, 3 and 4 - or large - 6, 7 and 8 - and far pairs are all bilateral (one number is small and the other is large). Moreover, far pairs never included the number 1 and 9. The set of stimuli was selected so that visual similarity within the verbal trials was balanced across the numerical conditions. With the obvious constrains

19 Indeed, Garner et al (1982) did not observed any distance effect when pair of Arabic numerals, including the numbers 3, 4, 6 and 7, were presented to be matched for physical identity and for parity status.

20 For its status of medium number, 5 was not included.
due to the above criteria, the range of numbers in the experimental trials was as large as possible. Moreover, some additional methodological differences were introduced. In the present experiments, half of the subjects answered *same* with the right-hand key and *different* with the left-hand key; this assignment was reversed in the other half of the subjects. This was done to reduce potential confusion in the answer-key assignment if treated as a within-subjects factor. At the same time, this would allow us to control for practice effects, by simply dividing the experiment in two blocks of trials. Both number of subjects (present experiment 24 versus 10) and number of observations for the single subject, differ from Dehaene and Akhavein's study (present experiment 288 versus 384).

To summarise, the purpose of the following experiments was to evaluate first, the extent to which magnitude information is autonomously activated in same-different judgements varying in the level of processing required by the task; second, whether this property is bound to a specific stimuli code (Arabic, verbal). Finally, by matching numbers written in different codes, the existence of a direct asemantic transcoding route between Arabic and verbal input lexicons was evaluated.

### 4.1 EXPERIMENT 4N
**NUMERICAL MATCHING TASK**

Subjects were presented with pairs of numbers that could be written either in the same notation (*pure trials* e.g., 2 3, four four) or in different notations (*mixed trials*, e.g., 2 three, 4 four). They were asked to judge if the two stimuli were equal or different in terms of numerical value, regardless of notational differences. The symbolic distance effect was employed to measure semantic access: if magnitude information is accessed during the task, the semantic similarity may interfere with the same-different judgements. Different pairs were either numerically close (distance of 1 unit; e.g., 2 3) or far apart (distance of 4 units; e.g., 2 6); if numerals are semantically processed, subjects would be slower to answer *different* to close pairs than to far pairs. In pure trials, the judgement
could be easily carried out on the basis of visual similarity; however, the evidence so far suggests that subjects access magnitude information even if not critical to the task (Duncan & McFarland, 1980; Dehaene & Akhavan, 1995). On the other hand, in mixed trials, stimuli semantically identical are not identical at the physical level, thus, they must be compared at a different level. If one assumes a direct asemantic route between Arabic and verbal lexicons, numerals might be matched at an asemantic level (e.g., phonological) and no distance effect should be observed. Even in this case, however, we may not exclude the possibility that semantic information may be autonomously activated and interfere with the decision.

### 4.2.1 Method

**Subjects**

24 postgraduate and undergraduate right-handed students (11 male and 13 female) from University College London participated in the experiment and received payment. Participants were native English speakers; mean age 21 (range from 18 to 30). All reported normal or corrected to normal vision.

**Stimuli**

Two variables were systematically varied to defined the experimental trials: number notation (AA, VV, AV, VA) and numerical disparity (0, 1, 4) (see Table 4.1). The levels of numerical disparity were used with equal frequency; pairs of numbers were chosen so that visual similarity (length of the verbal numerals) was matched in close and far pairs.

The actual pairs used were: 1 1, 2 2, 3 3, 4 4, 6 6, 7 7, 8 8, for equal pairs; 1 2, 2 1, 3 4, 4 3, 7 8, 8 7 for close pairs; 2 6, 6 2, 3 7, 7 3, 4 8, 8 4, for far pairs. Each single number appeared with the same frequency with the exception of 1 and 6; but overall, the frequency of the single number was balanced in close and far conditions (1 was included in a close pair; 6 was in a far pair). Each pair was presented four times in each condition.

---

21 This terminology is adopted after Dehaene and Akhavan.
with the exception of the pairs ‘1 1’ and ‘6 6’ which were presented twice in order to equalise the frequency of the single numbers in equal and unequal pairs.

Numbers were presented either in Arabic or verbal notation, defining four different categories of trials: Arabic trials (e.g., 2 6), verbal trials (e.g., two six) Arabic-verbal trials (e.g., 2 six) and verbal-Arabic trials (e.g., two 6). Thus overall, numerical disparity and notation defined twelve experimental trials, each presented 24 times, for a total of 288 trials.

Overall, same and different responses constituted respectively 1/3 and 2/3 of the total answers; however, this difference in the frequency of presentation was not assumed to affect the critical comparison between close and far pairs.

<table>
<thead>
<tr>
<th>NUMERICAL DISPARITY</th>
<th>NOTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA</td>
</tr>
<tr>
<td>equal</td>
<td>6 6</td>
</tr>
<tr>
<td>unequal close</td>
<td>2 3</td>
</tr>
<tr>
<td>unequal far</td>
<td>2 6</td>
</tr>
</tbody>
</table>

Table 4.1. Examples of the stimuli used in Experiments 4-7.

Apparatus

A Macintosh LCII running PsychLab 1.0 software (version .99; Gum, 1996) was used to display the stimuli and record RTs. Arabic numerals and lowercase verbal numerals were displayed in white Geneva bold 24-point font onto a black background.

Procedure

Each trial began with a fixation point displayed at the centre of the screen for 500 ms and was followed, 500 ms later, by a pair of stimuli to be compared. Stimuli stayed on the screen for 200 ms. The interstimulus interval was of 2500 ms; the entire experiment lasted approximately 25 minutes. Half of the participants were instructed to press the left-
hand key for same numerosity, the other half the right-hand key. Instructions emphasised both speed and accuracy.

**Design**

Experimental trials were assigned to two blocks of 144 trials each. Two pseudorandom orders for each block were obtained with the following criteria: 1) no same numbers in the same notation in consecutive trials; 2) no same trials consecutively; 3) no more than three same (same or different) correct answers in a row; 4) no more than three equal numerical disparities in a row; 4) no more than two trials in the same notation consecutively. The order of the blocks was counterbalanced between subjects. Each block started with 20 random training trials.

**4.1.2 Results**

Reaction times (means of medians) and error rates for each experimental condition are given in Table 4.2 and plotted in Figure 4.2.\(^{22}\)

The overall error rate was 4.9%; individual error rate did not exceed 12.8%.

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th></th>
<th></th>
<th>VV</th>
<th></th>
<th></th>
<th>AV-VA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>SD</td>
<td>%ER</td>
<td>RT</td>
<td>SD</td>
<td>%ER</td>
<td>RT</td>
<td>SD</td>
</tr>
<tr>
<td><strong>equal</strong></td>
<td>573.4</td>
<td>61.0</td>
<td>3.3</td>
<td>611.7</td>
<td>67.1</td>
<td>3.6</td>
<td>684.0</td>
<td>87.6</td>
</tr>
<tr>
<td><strong>close</strong></td>
<td>615.5</td>
<td>72.7</td>
<td>1.2</td>
<td>682.7</td>
<td>80.7</td>
<td>2.9</td>
<td>704.8</td>
<td>97.3</td>
</tr>
<tr>
<td><strong>far</strong></td>
<td>594.3</td>
<td>66.5</td>
<td>1.9</td>
<td>663.5</td>
<td>91.1</td>
<td>2.4</td>
<td>671.2</td>
<td>93.3</td>
</tr>
</tbody>
</table>

*Table 4.2. Experiment 4N: Mean Reaction times, Standard deviations and Error rates as a function of notation and numerical disparity.*

\(^{22}\) In both tables and figures data for AV and VA pairs are collapsed given their substantial similarity. Any difference in the processing of these two conditions is reported in the text.
Reaction time analysis

For every subject median correct RTs for each condition were calculated and entered in a repeated measures ANOVA with block (first and second), notation (AA, VV, AV, VA) and disparity (equal, close and far) as within-subjects factors.

The overall performance differed marginally in the two blocks ($F[1, 23] = 4.11$, $p=.05$) being 18 ms faster in the second block.

The main effect of notation was significant ($F[3, 69] = 108.47$, $p<.0001$).

Decomposition into contrasts indicated that pure trials (AA and VV) were answered significantly faster than mixed trials (63.2 ms; $F[1, 23] = 227.84$, $p<.0001$). Not surprisingly, Arabic trials were answered significantly faster than verbal trials (58.2 ms;
F[1, 23] = 96.82, p<.0001; in mixed trials the order of the stimuli (Arabic numeral in first or second position) did not have any effect on reaction times (F[1, 23] <1).

The main effect of disparity was significant (F[2, 46] = 13.91, p<.0001), as well as its interactions with block (F[2, 46] = 6.71, p<.005) and with notation (F[6, 138] = 8.97, p<.0001). This effect was decomposed contrasting first, equal versus unequal pairs and then, close versus far pairs.

*Same* responses were 25.2 ms faster than *different* responses (F[1, 23] = 14.97, p<.0005); however this difference was significant in the first block only (block 1, 44.3 ms; block 2, 6.1 ms). Moreover, this contrast interacted with notation (F[3, 69] = 13.50, p<.005) indicating that *same* responses were 46 ms faster than *different* ones in pure trials (p<.0001) but only 4 ms faster in mixed trials (p>.1).

Within *different* responses, far pairs were answered significantly faster than close pairs (26.9 ms; F[1, 23] = 12.84; p<.001) and this effect reached significance in all notations (all, p<.05).

*Error analysis*

A similar ANOVA was carried out on arcsine-transformed error proportions. The error analysis replicated the RT analysis and indicated that no speed-accuracy trade-off affected the outcomes. The main effect of notation was significant (F[3, 69] = 19.82, p<.0001); decomposition into contrasts indicated that subjects made more errors in mixed than in pure trials (7.4% versus 2.6%; F[1, 23] = 58.47, p<.0001). The effect of disparity was also significant (F[2, 46] = 23.98, p<.0001) as well as its interaction with notation (F[6, 138] = 9.10, p<.0001). Decomposition into contrasts indicated that equal pairs were more error prone than unequal pairs (9% versus 2.95%; F[1, 23] = 47.51, p<.0001) but this effect was significant in mixed trials only (F[1, 23] = 134.05, p<.0001). Finally, error rates for close and far pairs did not differ (F[1, 23] <1).

*Additional analyses.* Following Dehaene and Akhavein, the existence of minor effects reflecting semantic processing was tested. As previously pointed out, the magnitude effect may be safely evaluated in pure Arabic trials only. The comparison of the two
subset of trials, easily classifiable as small (i.e., 1 1, 2 2, 1 2, 2 1) and large (i.e., 7 7, 8 8, 7 8, 8 7), did not reveal any significant difference (F[1, 23] <1). Similarly, the relative position of the numbers within a pair did not have any effect on the RTs (F[1, 23] <1). The parity effect was not testable because numbers within far pairs always matched for parity.

4.1.3 Discussion

Overall, the outcomes of Experiment 4N replicated Dehaene and Akhavein's results. A significant distance effect was found in all conditions of notation, suggesting that numerical stimuli were always semantically processed. The size of the effect (26 ms) was smaller than the one reported by Dehaene and Akhavein (Experiment 1, 42 ms); this result was expected given that the numerical distance between close and far pairs was smaller in the present experiment (3 versus >5). While close pairs in both studies included consecutive numbers (e.g., 1 2), far pairs differed by 4 units in the present experiment (e.g., 2 6), and by 6 units or over in the Dehaene and Akhavein's study (e.g., 1 9). This factor might have greatly facilitated a different answer, also considering that in their experiment far pairs always included extreme-values. Moreover, in the Dehaene and Akhavein's study, it is possible that the visual dissimilarity in favour of the numerically far pairs in verbal notation might have played a role in increasing the magnitude of the distance effect.

We also found an error trend for responding different to mixed-trials suggesting that visual similarity had an effect on the same-different judgements. The distance effect in pure trials indicates that subjects were accessing the magnitude of the numbers even if they could carry out the task relying exclusively on the physical features of the stimuli. This was equally true for Arabic and verbal trials. Similarly, in mixed trials the distance effect ruled out the possibility that subjects matched the stimuli bypassing semantics.

In the following experiment subjects were explicitly instructed to ignore numerical information and to attend exclusively to the physical characteristics of the stimuli. This task provides a more stringent test for the autonomy of semantic access.
4.2 EXPERIMENT 4P  
PHYSICAL MATCHING TASK

The same material and procedure of Experiment 4N were used in the following experiment, however the subjects were now required to judge the two stimuli exclusively on the basis of their physical similarity. They were asked to answer *same* to physically identical stimuli (e.g., 2 2; two  two) and different otherwise. Critically, subjects were now required to answer *different* to pairs numerically identical but written in different notations (e.g., 2 two).

4.2.1 Method

*Subjects*

24 right-handed postgraduate and undergraduate students (11 male and 13 female) from University College London participated in the experiment and received payment. Participants were native English speakers; mean age 21 (range from 18 to 48). All reported normal or corrected to normal vision.

*Stimuli and apparatus*

The stimuli and apparatus were identical to those used in Experiment 4N. Since instructions required to answer *same* to physically identical stimuli only, *same* answers now constituted only one-sixth of the total answers. This difference in frequency was very likely to affect the relative speed of the *same* and *different* answers, rather than the critical comparisons within the *different* answers (equal versus unequal mixed trials, close versus far pairs in all notations).

*Procedure and design*

The procedure was identical to Experiment 4N. The experimental design was also similar with the only exception that one of the randomization criteria did no longer apply (no more than three same-*same* or different-*correct* answers).
4.2.2 Results

Reaction times (means of medians) and error rates for each experimental condition are given in Table 4.3 and plotted in Figure 4.3.

The overall error rate was 2.6%, individual error rate did not exceed 9.7%.

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th></th>
<th>VV</th>
<th></th>
<th>AV-VA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>SD</td>
<td>%ER</td>
<td>RT</td>
<td>SD</td>
<td>%ER</td>
</tr>
<tr>
<td>equal</td>
<td>571.8</td>
<td>82.9</td>
<td>7.5</td>
<td>614.8</td>
<td>111.4</td>
<td>10.4</td>
</tr>
<tr>
<td>close</td>
<td>515.3</td>
<td>76.8</td>
<td>1.0</td>
<td>577.8</td>
<td>91.2</td>
<td>2.4</td>
</tr>
<tr>
<td>far</td>
<td>520.6</td>
<td>74.9</td>
<td>1.0</td>
<td>565.5</td>
<td>90.8</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 4.3 Experiment 4P: Mean Reaction Times, Standard Deviations and Error rates as a function of notation and numerical disparity.

Figure 4.3. Experiment 4P: Mean Reaction times and Error rates as a function notation and numerical disparity [* white data points correspond to “same” answers; black data points to “different” answers].
Reaction time analysis

For every subject median correct RTs for each condition were calculated and entered in a repeated measures ANOVA with block (first and second), notation (AA, VV, AV, VA) and disparity (equal, close and far) as within-subjects factors.

Subjects were 45 ms faster performing the second block compared to the first one, (F[1, 23] = 36.37; p<.0001).

The main effect of notation was significant (F[3, 69] = 46.87, p<.0001) as well as its interaction with block (F[3, 69] = 7.38, p<.001). In fact, all trials were answered faster in the second block, but this effect was more pronounced for mixed trials (547.5 ms versus 489.3 ms) than for pure trials (577.3 ms versus 544.7 ms). Decomposition into contrasts indicated that mixed trials (AV and VA) were now answered significantly faster than pure trials (518.4 ms versus 560.9 ms; F[1, 23] = 82.67; p<.0001). This difference reached significance in both blocks, however it was greater in the second one (55.4 ms and 29.8 ms for block 1 and block 2 respectively). Arabic trials were answered significantly faster than verbal trials (50.1 ms; F[1, 23] = 57.26, p<.0001); while within mixed trials, AV and VA pairs did not differ significantly (F[1, 23]<1).

The main effect of disparity was significant (F[2, 46] = 27.00, p<.0001), as well as its interaction with notation (F[6, 138] = 4.87, p<.001). This effect was decomposed contrasting, first, equal versus unequal pairs in both pure and mixed trials and, then, close versus far pairs. Within pure trials, equal pairs were answered 48.5 ms slower than unequal pairs (F[1, 23] = 85.17 p<.0001), not surprisingly due to the low frequency of the same answers compared to the different ones. More interestingly, this difference was also significant in mixed trials (F[1, 23] = 7.71, p<.01): subjects were 14.6 ms slower in answering different to pairs of stimuli written in different notations but representing the same quantity (e.g., 4 four) compared to stimuli different on both dimensions (e.g., 2 six). Moreover, the triple interaction between block, notation and disparity (F[6, 138] = 3.96; p<.005) indicated that in mixed trials this effect was highly significant in block 1 (32.8 ms; F[1, 23] = 29.81, p<.0001) but disappeared in block 2 (3.7 ms in the opposite direction, F[1, 23]<1). Contrasts also indicated that close and far pairs did not differ significantly, and this was true for all notations (all, p>.1).
Error analysis

A similar ANOVA was carried out on arsine-transformed error proportions. The results indicated that no speed-accuracy trade-off affected the outcomes. The main effect of notation was significant ($F[3, 69] = 11.63, p<.0001$); decomposition into contrasts indicated that subjects made more errors in pure than in mixed trials (3.9% versus 1.3%, $F[1, 23] = 30.29, p<.0001$), and that verbal trials were more error prone than Arabic trials ($F[1, 23] = 4.57, p<.05$), while there was no difference between AV and VA trials ($F[1, 23] <1$).

The pure-mixed effect interacted with block ($F[3, 69] = 3.95, p<.05$): in the second block the error rate increased in pure trials ($F[1, 23] = 8.82, p<.01$) and tended to decrease in mixed trials ($F[1, 23] = 3.03; p=.08$).

The effect of disparity was also significant ($F[2, 46] = 28.14, p<.0001$) as well as its interaction with notation ($F[6, 138] = 7.00, p<.0001$). Decomposition into contrasts indicated that equal pairs were more error prone than unequal pairs (5.6% versus 1.1%) ($F[1, 23] = 55.91, p<.0001$). This effect was highly significant in pure trials ($F[1, 23] = 116.96, p<.0001$) and significant to a lower extent in mixed trials ($F[1, 23] = 4.37, p<.05$). Finally, error rates for close and far pairs did not differ ($F[1, 23] <1$) and this was true for all notations (all, $p>.1$).

4.2.3 Discussion

The two main results in Experiment 4P are the dramatic decline of the distance effect together with the interference of the numerical identity in mixed trials.

The first result is in sharp contrast with Dehaene and Akhavein’s results; in their study, subjects, even when requested to match the stimuli for physical identity, answered faster to numerically distant pairs than to numerically close pairs, as long as they were written in the same notation (e.g. 1 2; ONE TWO). These contrasting results cannot be attributed to a different accuracy, in particular with regard to a possible response bias to different answers (due to the low frequency of the same response, only one sixth of the
overall responses) that might have contributed to speed up the response times and make the distance effect disappeared, since error rates were similar in the two studies (2.6% and 2.7%).

The absence of the distance effect in pure trials seems to suggest that subjects were carrying out the task ignoring successfully semantic information. Thus, according to our hypothesis, with a larger range of numbers, and in particular, when far pairs did not included end-terms only, access to semantic information was possibly initiated but not completed in time for interfering with the response. Given that far pairs were formed by a small (2, 3 or 4) and a large (6, 7 or 8) digit, the activation of a dichotomous classification would have still determined a distance effect. However, if we assume that the speed of this classification is indeed modulated by the relative size of the numbers (i.e., 1 yields the strongest association with ‘small’, 9 the strongest association with ‘large’), the observed results would be expected.

Mixed trials were physically highly dissimilar and thus easily to discriminate; it is plausible that subjects were able to answer different to those pairs before any semantic access could take place. However, our results indicate that subjects were slower and tended to make more error in responding different to mixed pairs representing the same quantity (e.g., 1 one) compared to pairs representing different quantities (e.g., 2 six). This suggests that subjects were actually recognising that the two stimuli represented the same number and this information interfered with the decision that these were physically different stimuli. Further, the lack of distance effect suggests that this mapping did not occur at the semantic level. Hence, this result seems to support the hypothesis of an additional asemantic transcoding route between Arabic and verbal lexicons.

Critically, the interference effect was modulated by practice, in fact it disappeared in the second block of trials; thus, the mismatch between number-identity and physical-difference had only a confined effect. It is worth noticing that in the present study, subjects required very little training in order to overcome interference. The effect of practice in reducing interference effects in selective tasks is a well known phenomenon (e.g., Reisberg, Baron & Kemler, 1980; Melara & Mounts, 1993). The source of this effect is not always clear and multiple and nonexclusive explanations may be proposed.
Subjects may have learned to inhibit or ignore the irrelevant dimension of the stimuli, e.g., the number-identity, and/or selectively attend the relevant one, i.e., their physical features; this latter may partially result from the adoption of peripheral strategies. It is beyond the purpose of this study to favour one or the other of these explanations; however, age-related effects on these attentional mechanisms will be later on explored and this phenomenon clarified (See Section 4.5).

4.3 EXPERIMENTS 5N AND 5P

EXACT REPLICATION OF DEHAENE AND AKHAVEIN’S STUDY

Our results suggest that autonomous access to magnitude information may be modulated by the specific selection of the stimuli and by the task demands. The distance effect observed in the numerical matching task points to the primacy of semantic information in number processing; however, we found that instructions had a major influence in determining the level of processing of the stimuli. In the physical matching task, subjects were explicitly requested to ignore semantic information and their performance was, not surprisingly, much faster than in the numerical matching task. In the latter, even if semantic processing was unintentionally initiated, the meaning of the numbers was not always available before the subject responded. In summary, when numbers are highly distinctive and their assignment to a small or large code unambiguous, this semantic information seems to interfere with the same-different judgements. However, when this coding is more time-demanding, as when a larger range of numbers is presented and medium terms included, the semantic information does not affect the performance.

To further corroborate our considerations, and safely attribute the source of the differences between Dehaene and Akhavein’s study and the present one to the critical selection of the stimuli, we decided to replicate their experiments with the same material and experimental design adopted by them.
4.3.1 Method

Subjects

20 right-handed postgraduate and undergraduate students (12 male and 8 female) from University College London participated in the experiment and received payment. Participants were native English speakers; mean age 26 (range from 19 to 45). All reported normal or corrected to normal vision. Equal number of participants was randomly assigned to the numerical matching and to the physical matching tasks.

Stimuli

The same material used by Dehaene and Akhavein was adopted. Two variables were systematically varied to defined the experimental trials: number notation (AA, VV, AV, VA) and numerical disparity (0, 1, >5). The actual pairs used were: 1 1, 2 2, 8 8, 9 9, for equal pairs; 1 2, 2 1, 8 9, 9 8 for close pairs; 1 8, 8 1, 9 1, 2 8, 8 2, 2 9, 9 2 for far pairs. To equalise the overall frequency of each disparity condition the four equal pairs and the four close pairs were presented four times each in each condition, while the eight far pairs were presented only twice each. Numbers were presented either in Arabic or verbal notation, defining four different categories of trials: Arabic trials (e.g., 2 1), Verbal trials (e.g., TWO ONE) Arabic-verbal trials (e.g., 2 ONE) and verbal-Arabic trials (e.g., TWO 6). Thus, overall, numerical disparity and notation defined twelve experimental trials, each presented 16 times, for a total of 192 experimental trials. 40 distracting pairs using the numbers 3-7 were also included.

Apparatus

The same apparatus of Experiment 4N was used.

Procedure

Each trial began with a fixation point displayed at the centre of the screen for 500 ms and was followed 500 ms later by a pair of stimuli to be compared. Stimuli stayed on the screen for 200 ms. The interstimulus interval was of 2500 ms; the entire experiment lasted approximately 30 minutes. Two blocks of stimuli were presented to each subject.
In one block, *same* responses were assigned to the left-hand key and *different* responses to the right-hand key. This assignment was reversed in the second block; and order of the block was counterbalanced between participants. Instructions emphasised both speed and accuracy.

**Design**

Experimental trials were assigned to two blocks of 242 trials each. Two pseudorandom orders per each block were obtained with the following criteria: 1) no same numbers in the same notation in consecutive trials; 2) no same trials consecutively; 3) no more than three equal numerical distance in a row; 4) no more than two pairs in the same notation in a row; 5) no more than three same correct -same or different- answers in consecutive trials (this criterion applied to the numerical matching task only). The order of presentation of the randomised blocks was counterbalanced between subjects. Each block started with 20 random training trials. Both training and distracting trials were disregard in the analyses.

**4.3.2 Results**

*Experiment 5N: Numerical matching task.*

Reaction times (means of medians) and error rates for each experimental condition are given in Table 4.4 and plotted in Figure 4.4

The overall error rate was 3.3 %; individual error rate did not exceed 6%.

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th></th>
<th></th>
<th>VV</th>
<th></th>
<th></th>
<th>AV-VA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>SD</td>
<td>%ER</td>
<td>RT</td>
<td>SD</td>
<td>%ER</td>
<td>RT</td>
<td>SD</td>
</tr>
<tr>
<td><em>equal</em></td>
<td>629.8</td>
<td>87.5</td>
<td>3.4</td>
<td>661.3</td>
<td>94.1</td>
<td>2.8</td>
<td>754.4</td>
<td>102.6</td>
</tr>
<tr>
<td><em>close</em></td>
<td>710.4</td>
<td>84.1</td>
<td>1.9</td>
<td>799.1</td>
<td>118.8</td>
<td>3.4</td>
<td>813.4</td>
<td>121.3</td>
</tr>
<tr>
<td><em>far</em></td>
<td>685.5</td>
<td>81.9</td>
<td>1.6</td>
<td>751.9</td>
<td>91.4</td>
<td>0.3</td>
<td>761.2</td>
<td>92.5</td>
</tr>
</tbody>
</table>

*Table 4.4. Experiment 5N: Mean Reaction times, Standard deviations and error rates as a function of notation and numerical disparity.*
Figure 4.4. Experiment SN: Mean Reaction times and Error rates as a function of notation and numerical disparity [* white data points correspond to “same” answers; black data points to “different” answers].

Reaction Times analysis

For every subject median correct RTs for each condition were calculated and entered in a first repeated measures ANOVA with block (first and second), notation (AA, VV, AV, VA) and disparity (equal, close and far) as within-subjects factors and order of the blocks as between-subjects factor.

Order of the blocks was not a significant effect and did not enter in any significant interaction. Data were then collapsed over this factor.

The main effect of block was not significant and did not enter in any significant interaction. Given that the assignment for the response-keys was reversed in the second block, we did not expect subjects to show any practice effect, as we observed in Experiment 4N.
The main effect of notation was significant \( (F[3, 27] = 26.22, \ p<.0001) \). Decomposition into contrasts indicated that pure trials were answered significantly faster than mixed trials \( (70.2 \text{ ms}; F[1, 9] = 57.48; p<.0001) \). Arabic pairs were answered significantly faster than Verbal pairs \( (62.3 \text{ ms}; F[1, 9] = 21.16, p<.0001) \); while no significant difference was found within mixed trials \( (AV-VA; F[1, 9]<1) \).

The main effect of disparity was significant \( (F[2, 18] = 13.96, p<.0001) \), as well as its interaction with notation \( (F[6, 54] = 2.87, p<.05) \). This effect was decomposed contrasting, first, equal versus unequal pairs in both pure and mixed trial and then, close versus far pairs.

*Same* responses were 62 ms faster than *different* responses \( (F[1, 9] = 20.50, p<.001) \). Moreover, this contrast interacted with notation \( (F[3, 27] = 6.13, p<.005) \) indicating that *same* responses were 91 ms faster than *different* ones in pure trials \( (p<.0001) \), but only 33 ms faster \( (p>.01) \) in mixed trials.

Within *different* responses, far pairs were answered significantly faster than close pairs \( (44.4 \text{ ms}; F[1, 9] = 7.41; p<.05) \) and this effect reached significance in all notations \( (all, p<.05) \).

**Error analysis**

A similar ANOVA was carried out on arcsine-transformed error proportions. The main effect of notation was significant \( (F[3, 27] = 3.66, p<.05) \); decomposition into contrasts indicated that subjects made more errors in mixed than in pure trials \( (4.3\% \text{ versus } 2.3\%; F[1, 9] = 10.76, p<.005) \). The effect of disparity was also significant \( (F[2, 18] = 9.38, p<.005) \) as well as its interaction with notation \( (F[6, 54] = 2.55, p<.05) \). Decomposition into contrasts indicated that equal pairs were more error prone than unequal pairs \( (5.6\% \text{ versus } 2.1\%; F[1, 9] = 15.76, p<.001) \), but this effect was significant in mixed trials only \( (F[1, 9] = 31.23, p<.0001 \text{ and } F[1, 9] = 1.64, p>.2, \text{ for mixed and pure trials respectively}) \). Finally, error rates for close pairs tended to be higher than for far pairs \( (F[1, 9] = 3.01, p=.09) \).
Further analysis. The existence of minor effects reflecting semantic processing was tested. Within Arabic trials (only equal and close pairs were analysed) pairs with smaller numbers (i.e., 1 1, 2 2, 1 2, 2 1) yielded faster RTs than pairs with larger numbers (i.e., 8 8, 9 9, 8 9, 9 8) ($F[1, 9] = 32.6, p<.0001$). However, the relative position of the numbers within a pair did not have any effect on the RTs ($F[1, 9]<1$). Similarly, there was non reliable effect of the parity of the numbers on the speed of same-different judgements.

Experiment 5P: Physical Matching task

Reaction times (means of medians) and error rates for each experimental condition are given in Table 4.5 and plotted in Figure 4.5.

The overall error rate was 2.9%; individual error rate did not exceed 6.7%.

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>VV</th>
<th>AV-VA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>equal</td>
<td>534.3</td>
<td>547.4</td>
<td>484.6</td>
</tr>
<tr>
<td>close</td>
<td>554.7</td>
<td>597.3</td>
<td>480.6</td>
</tr>
<tr>
<td>far</td>
<td>526.6</td>
<td>545.5</td>
<td>482.5</td>
</tr>
</tbody>
</table>

Table 4.5. Experiment 5P: Mean Reaction times, Standard Deviations and Error rates as a function of notation and numerical disparity.

Reaction times analysis

For every subject median correct RTs for each condition were calculated and entered in a first repeated measures ANOVA with block (first and second), notation (AA, VV, AV, VA) and disparity (equal, close and far) as within-subjects factors and order of the blocks as between-subjects factor.
Figure 4.5. Experiment 5P: Mean Reaction times and Error rates as a function of notation and numerical disparity [* white data points correspond to “same” answers; black data points to “different” answers].

The effect of order of the blocks was not significant and did not enter in any significant interaction. Data were then collapsed over this factor.

Subjects were 32 ms faster performing the second block compared to the first one, (F[1, 9] = 12.90; p<.01). Despite the reversed assignment, subjects benefited from practice in the physical matching task.

The main effect of notation was significant (F[3, 27] = 60.69, p<.0001). Decomposition into contrasts indicated that mixed trials (AV-VA) were now answered significantly faster than pure trials (65 ms; F[1, 9] = 168.06; p<.0001). Arabic pairs were
answered significantly faster than Verbal pairs (24.8 ms; F[1, 9] = 12.12, p<.005); while in mixed trials AV and VA pairs did not differ significantly (F[1, 9] <1).

The main effect of disparity was significant (F[2, 18] = 7.26, p<.005), as well as its interaction with notation (F[6, 54] = 3.99, p<.005). This effect was decomposed contrasting, first, equal versus unequal pairs in both pure and mixed trials and then, close versus far pairs. In pure trials, equal pairs were answered 10 ms faster than unequal pairs, but this difference did not reach significance (F[1, 9] = 1.84 p=. 18). In mixed trials, equal (e.g., 4 four) and unequal (e.g., 2 six) pairs yielded similar response latencies. However, planned comparison revealed that in the first block, subjects tended to be slower in answering different to pairs of stimuli written in different notations but representing the same quantity (19 ms, F[1, 9] = 3.47, p=.06) compared to stimuli different on both dimensions; but this trend disappeared in the second block (F[1, 9]<.1).

Close and far pairs differ significantly (F[1, 9] = 14.44, p<.0001) and this contrast interacted with notation (F[1, 10] = 17.75, p<.001). The distance effect was highly significant in both Arabic (28 ms, F[1, 9] = 8.58, p<.005) and verbal pairs (52 ms, F[1, 9] = 24.49, p<.0001) but did not emerged in mixed trials (F[1, 9] <1).

**Error analysis**

A similar ANOVA was carried out on arcsine-transformed error proportions. The error analysis replicated the RT analysis and indicated that no speed-accuracy trade-off affected the outcomes. The main effect of notation was significant, (F[3, 27] = 6.84, p<.005); decomposition into contrasts indicated that subjects made more errors in pure than in mixed trials (4.8% versus 1.0%; F[1, 9] = 18.61, p<.001), but no differences were observed within pure or within mixed trials.

The effect of disparity was also significant (F[2,18] = 10.13, p<.001) as well as its interaction with notation (F[6, 54] = 3.07, p<.01). Decomposition into contrasts indicated that equal pairs were more error prone than unequal pairs (5% versus 1.6%; F[1, 9] = 16.71, p<.001); however this effect was significant in pure trials only (F[1, 9] = 36.05, p<.0001; mixed trials, F[1, 9] <1). Overall, error rates for close and far pairs did not differ (F[1, 9] = 2.08, p>.16). However, in both Arabic (F[1, 9] = 3.17, p=08) and verbal
pairs \( F[1, 9] = 3.16, p=0.08 \) error tended to be more frequent in close than in far pairs; no difference emerged for close and far pairs within mixed pairs \( F[1, 9] <1 \).

4.3.3 Discussion

Experiments 5N and 5P successfully replicated Dehaene and Akhavein's results. Though the overall level of accuracy was similar across the studies, the RT profiles were remarkably longer in the present experiments. We cannot provide any plausible explanation for this and we merely attribute this difference to random variability in the recruited populations. The relative difference between same and different responses also varied across the studies, possibly due to interindividual difference in performing the tasks: in the present study in both numerical and physical matching tasks (despite the low frequency of the same responses in the latter) same responses yielded extremely faster RTs compared to different responses; yet, the error rates were comparable.

With regard to the more critical results, in the numerical matching task a significant distance effect was observed in all notations. The effect was indeed of the same size as the one reported by Dehaene and Akhavein \( 44 \text{ ms versus } 42 \); but note that the overall performance differs in the two studies, \( 741 \text{ ms versus } 619 \text{ ms} \). This evidence seems to suggest that subjects were processing numerical stimuli all the way to semantics and that the numerical information did interfere with the same-different judgements. However, we failed to observe two of the three additional semantic effects, i.e., the effect of the relative position of the digits within a pair and the effect of the parity of the numbers within a pair. These results point to the relative reliability of these minor effects.

Critically, in the physical matching task, close pairs still yielded longer different responses than far pairs. However, this effect was limited to pure trials. The size of the effect in the verbal pairs was of 52 ms, and we strongly believe that the mismatch in visual similarity was a critical factor in determining such a remarkable difference between
close and far pairs. Yet, the source of the distance effect in Arabic pairs may be located at the semantic level.

However, a careful inspection of the data suggests caution in drawing firm conclusions from these results. In fact, an item analysis carried out on Arabic pairs only, revealed that one (i.e., 8 9) of the two numerically close pairs (i.e., 1 2, 8 9) was responsible for the significant distance effect. In other words, while ‘8 9’ yielded significantly longer RTs than any other far trials, (614 ms versus 530 ms for ‘8 9’ and far pairs (i.e., 1 8, 1 9, 2 8, 2 9) respectively; this effect was highly consistent across subjects being shown by all of them), subjects tended to be even faster in answering to trial ‘1 2’ than to far pairs (505.1 ms versus 530 for ‘1 2’ and far pairs respectively). We may not exclude that the high visual similarity of the digits 8 and 9 (both are curvilinear; their upper halves are symmetrical) may be responsible for this effect. It is plausible, and rather unsurprising, to assume that visual similarity plays a role in any same-different task requiring to attend to the physical characteristics of the stimuli. In this respect, the assumption that “Because the digits 1 and 2 are not more visually similar than are 1 and 9, it is clear that the distance effect must reflect access to the sequential structure of the numbers.” (Dehaene and Akhavein, p.321) seems rather simplistic. Clearly, only targeted studies where the visual similarities between the stimuli is independently controlled may clarify the impact of visual factors in the matching tasks (e.g., Bagnara, Boles. Simion & Umilta’, 1983).

Overall, the results from Experiments 5N and 5P corroborate the hypothesis that the numerals employed, i.e., 1, 2, 8, 9, may constitute a rather special subset of stimuli within the range of single-digit numbers. We do not exclude that their distinctiveness was further enhanced by the rather limited range of distracting trials included in the task.

The interference effect yielded by mixed trials representing the same number (e.g., 3 THREE) was now significantly reduced compared to Experiment 4P. Still, in the first half of the experiment, this effect fell short of significance (p=.06). This reduction in the size of the effect compare to Experiment 4P, may be simply explained by the larger number of trials included in the first block in the present experiment (242 versus 144). When an
analysis was performed on the first 144 trials only, this effect reached significance ($F[1, 9] = 6.77, p<.01$). There is no reason to assume that the interference might be moderated by the specific numerals presented (e.g., 9 NINE and 3 THREE should yield similar response time). However, it is possible that the relative frequency of the individual items modulates the speed in which subjects learn to inhibit the interference. The critical pairs were now only four (distracting trials of this type were only 5 in total), and subjects might have come to recognise and control the irrelevant information in a shorter time.

**INTRODUCTION TO THE DEVELOPMENTAL STUDY**

The study reported in Chapter 3 strongly suggests that access to number meaning acquires autonomous property over the course of learning: even though it has been suggested that by the age of 6, children have a magnitude representation similar to the adult’s one (e.g., Duncan & McFarland, 1980; Sekuler & Mierkeiwicz, 1977), it is only with extensive practice with written numerals that these representations may be autonomously activated. The paradigm used in Experiments 4 and 5 may well provide further test for this hypothesis. Moreover, an analogous same-different judgement task was originally adopted by Duncan and McFarland (1980) in testing children at different ages; their results favour the hypothesis that magnitude information dominates number processing since a very early age. However, to our knowledge, no other studies have followed up this issue or replicated their results.

In Duncan and McFarland’s study, children were required to decide whether or not two Arabic numerals were identical, and same and different responses were equally frequent. In the present study, both Arabic and verbal numerals were presented and the proportion of same and different responses varied with the task (numerical and physical matching). These methodological differences are likely to modulate the performance and are properly considered in the discussion of our results.
In the present paradigm, the factorial combination of notation (Arabic, verbal or mixed) and numerical disparity (equal, close and far) allowed for multiple comparisons. By asking children to match the stimuli for numerical or physical identity we could evaluate the extent to which the access to magnitude information is autonomously activated. In the numerical matching task, children were requested to attend to the numerical value of the stimuli, i.e. they had to respond *same* to pairs such as “2 TWO”; yet, in pure pairs, a simple visual matching was sufficient to take a decision. Results from Experiments 4N and 5N indicate that for adults this was not the case, and that semantic information was accessed and used to facilitate their decision, whether the numerals were in Arabic or verbal notation. Even though these findings do not constitute by themselves very conclusive evidence for autonomous semantic access, they certainly point to the dominance of magnitude information in number processing. However, it is likely that this phenomenon emerges gradually over the course of learning, and that for young children the access to numerical information may only result from intentional processing. However, by the time magnitude information is automatically accessed we do not expect notational effects to modulate this process, similarly to adults.

In the physical matching task, children were required to specifically attend to the physical characteristics of the stimuli and disregard semantic information; this constitutes a more stringent test for the autonomy of semantic access. Indeed, we found that evidence for semantic processing was limited to Experiment 5P, where a set of stimuli corresponding to the extreme-values within the range of 1 to 9, was presented. The occurrence of the distance effect was not only dependent upon the specific range of stimuli used, but also upon their notation: mixed trials were easily classified as *different* on the basis of their physical features. In this condition, however, it seemed that subjects were not completely able to disregard the fact that equal pairs, such as 1 and ONE, did represent the same number. In the absence of a distance effect it has been suggested that the mapping between Arabic and verbal numerals occurred at an asemantic level. Furthermore, this interference effect was modulated by practice: subjects were disturbed by the mismatch between number-identity and physical-difference but they learned to disregard this information over the course of the experiment. If this is the result of
efficient mechanisms of selective attention, we may expect children to be less successful in benefiting from practice with the task, and to show a more pronounced and stable interference effect (e.g., Tipper et al., 1989).

A major procedural modification was adopted in order to present primary-school children with the matching tasks: the stimuli presentation was terminated by the subject's response instead of lasting 200 ms as in the previous experiments. This modification may crucially affect the size of both interference and distance effects, preventing a direct comparison with adult's performance; thus, we first run the numerical and physical matching tasks with adult participants (Experiments 6N and 6P), and then we presented both experiments to third- and fifth-graders. Since the testing of primary-school children took place in Italy, a new set of stimuli was designed, following the same criteria adopted in Experiments 4N and 4P (see Stimuli section).

4.4 EXPERIMENTS 6N AND 6P

The following experiments were designed in order to a) test the feasibility of the modified experimental procedure, b) replicate the results of Experiments 4N and 4P with a new set of stimuli. Stimuli were chosen with the following criteria: the frequency of the single numerals was balanced across conditions, close and far conditions corresponded respectively to distance 1 and distance 4, the visual similarity between verbal numeral pairs was balanced in close and far pairs. The pairs of stimuli that satisfied these criteria were only two for each numerical distance and included the numbers 2, 3, 6 and 7. As in Experiments 4N and 4P, end-terms were not included, thus we expected a semantic processing to emerge only in the numerical matching task. Distractor trials were included to match the frequency of the experimental conditions with Experiments 4N and 4P and to allow us to directly compare practice effect in the two experiments.
4.4.1 Method

Subjects

20 undergraduate and postgraduate right-handed students (6 male and 14 female) participated in the experiments and received payment. Participants were all native Italian speakers; mean age 26 (range from 20 to 34). All reported normal or corrected to normal vision. Equal number of participants was randomly assigned to the numerical matching and to the physical matching tasks.

Stimuli

The actual pairs used were: 2 2, 3 3, 6 6, 7 7, for equal pairs; 2 3, 3 2, 6 7, 7 6 for close pairs; 2 6, 6 2, 3 7, 7 3 for far pairs. Each single number appears with the same frequency in the different conditions and each single pair was presented four times in each condition.

Numbers were presented either in Arabic or verbal notation, defining four different categories of pairs: Arabic trials (e.g., 2 6), Verbal trials (e.g., DUE SEI [two six] Arabic-verbal trials (e.g., 2 SEI) and verbal-Arabic trials (e.g., DUE 6). Thus overall, numerical disparity and notation defined twelve experimental trials, each presented 16 times, for a total of 192 trials.

To prevent participant’s attention to be focused exclusively on numbers 2, 3, 6, 7 and to equalise the presentation of the different experimental conditions to Experiments 4N and 4P, distractor trials including the remaining numbers (1, 4, 5, 8 and 9) were included (N=96). Thus, subjects were still presented with the same number of trials per each experimental condition: this is clearly a critical factor for comparing, across the experiments, the extent to which practice modulates the performance.
Apparatus

A Macintosh LCII running PsychLab 1.0 software (version.99; Gum, 1996) was used to displayed the stimuli and record RTs. Arabic numerals and uppercase verbal numerals were displayed in white Geneva bold 36-point font onto a black background.

Procedure

Each trial began with a fixation point displayed at the centre of the screen for 500 ms and was followed 500 ms later by a pair of stimuli to be compared. The stimuli display was terminated by the subject’s response. The interstimulus interval was of 2500 ms; the entire experiment lasted approximately 25 minutes. Half of the participants were instructed to press the left-hand key for same numerosity, the other half the right-hand key. Instructions emphasised both speed and accuracy.

Design

Experimental trials were assigned to four sets of 72 trials each. Two pseudorandom orders per each set were obtained with the following criteria: 1) no same numbers in the same notation in consecutive trials; 2) no same trial consecutively; 3) no more than three equal numerical distances in a row; 4) no more than two pairs in the same notation in consecutive trials, 4) no more than three same correct responses - same or different- consecutively (this criterion applied to the numerical matching task only). The order of the sets was counterbalanced between subjects. Before starting the experiment, a training block of 24 random trials was presented.

---

23 Verbal numerals were presented in uppercase since children are first introduced to this case and thus, more familiar with them.

24 The presentation of 4 shorter blocks was considered more adequate to the attentional resource of young children.
4.4.2 Results

*Experiment 6N: Numerical Matching task*

Reaction times (means of medians) and error rates for each experimental condition are given in Table 4.6 and plotted in Figure 4.6.

The overall error rate was 3%; the individual error rate did not exceed 8%.

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>VV</th>
<th>AV-VA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>SD</td>
<td>%ER</td>
</tr>
<tr>
<td>equal</td>
<td>535.9</td>
<td>77.5</td>
<td>2.0</td>
</tr>
<tr>
<td>close</td>
<td>560.0</td>
<td>47.4</td>
<td>2.0</td>
</tr>
<tr>
<td>far</td>
<td>542.7</td>
<td>41.7</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Table 4.6. Experiment 6N: Mean Reaction times, Standard deviations and Error rates as a function of notation and numerical disparity.

*Reaction time analysis*

For every subject median correct RTs for each condition were calculated and entered in a first repeated measures ANOVA with block (first -set 1 and 2- second -set 3 and 4), notation (AA, VV, AV, VA) and disparity (equal, close and far) as within-subjects factors.

The main effect of block was significant (22 ms, F[1, 9] = 5.64, p<.05): subjects tended to answer faster in the second block.

The main effect of notation was significant (F[3, 27] = 41.55, p<.0001). Decomposition into contrasts indicated that pure trials were answered significantly faster than mixed trials (60 ms; F[1, 9] = 61.13; p<.0001). Arabic pairs were answered significantly faster than verbal pairs (86.5 ms; F[1, 9] = 63.23, p<.0001; while in mixed trials the order of the stimuli (AV-VA) did not have any effect on RTs (F[1, 9] <1).
Figure 4.6. Experiment 6N: Mean Reaction times and Error rates as a function of notation and numerical disparity [* white data points correspond to “same” answers; black data points to “different” answers].

The main effect of disparity fell short of significance, (F[2, 18] = 2.70, p=.09); however, disparity interacted significantly with block (F[2, 18] = 5.7, p<.05) and with notation (F[6, 54] = 3.37, p<.01). Same responses were faster than different responses in the first block only (F[1, 9] = 31.2, p<.0001 and F[1, 9] < 1, for block 1 and block 2 respectively). On the other hand, the contrast between close and far pairs did not interact with block (F[1, 9] < 1). In pure trials, same responses were 43 ms faster than different responses (F[1, 9] = 21.69, p<.0001); in mixed trials, no difference between same and different responses emerged (F[1, 9] <.1).Within different responses, far trials were
answered faster than close trials (F[1, 9]=9.61, p<.05) and this effect reached significance in all notations (all, p<.05).

**Error analysis**

A similar ANOVA was carried out on arcsine-transformed error proportions. The main effect of notation was significant, (F[3, 27] = 3.92, p<.05); decomposition into contrasts indicated that subjects made more errors in mixed than in pure trials (4.1% versus 1.4%) (F[1, 9] = 11.63, p<.01). In mixed trials only, *same* responses were more error prone than *different* responses (F[1,9]=3.81, p=.05). No other effect resulted significant.

**Further analysis.** The magnitude effect was evaluated for Arabic pairs only. When equal and close pairs were analysed, no difference was detected between pairs with smaller numbers (i.e., 3 3, 2 2, 3 2, 2 3) and pairs with larger numbers (i.e., 6 6, 7 7, 6 7, 7 6), F[1, 9] = 1.16, p>.3). Relative order of the numbers within a pair did not have any effect on the performance.

**Experiment 6P: Physical Matching task**

Reaction times (means of medians) and error rates for each experimental condition are given in Table 4.7 and plotted in Figure 4.7.

The overall error rate was 2.1%; individual error rate did not exceed 4%.

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>VV</th>
<th>AV-VA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>SD</td>
<td>%ER</td>
</tr>
<tr>
<td>equal</td>
<td>538.6</td>
<td>54.6</td>
<td>5.0</td>
</tr>
<tr>
<td>close</td>
<td>514.9</td>
<td>71.2</td>
<td>3.1</td>
</tr>
<tr>
<td>far</td>
<td>506.3</td>
<td>71.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

*Table 4.7. Experiment 6P: Mean Reaction Times, Standard deviations and Error rates as a function of notation and numerical disparity.*
Figure 4.7. Experiment 6P: Mean Reaction times and Error rates as a function of notation and numerical disparity (* white data points correspond to “same” answers; black data points to “different” answers).

Reaction times analysis

For every subject's median correct RTs for each condition were calculated and entered in a repeated measures ANOVA with block (first - set 1 and 2- and second - set 3 and 4), notation (AA, VV, AV, VA) and disparity (equal, close and far) as within-subjects factors.

Subjects were 32.3 ms faster performing the second block compared to the first one, (F[1, 9] = 4.00; p=.07).

The main effect of notation was significant (F[3, 27] = 15.39, p<.0001). Decomposition into contrasts indicated that mixed trials were answered significantly
faster than pure trials (56 ms; F[1, 9] = 21.64; p<.0001). Arabic pairs were answered significantly faster than verbal pairs (88 ms; F[1, 9] = 27.36, p<.0001) while within mixed trials, AV and VA pairs did not differ significantly (F[1, 9] <1).

The main effect of disparity was significant (F[2, 18] = 5.36, p<.05). Contrasts indicated that equal pairs were answered 23 ms slower than unequal pairs (F[1, 9] = 10.02, p<.01), while no difference was found between close and far pairs (F[1, 9] <1).

The interaction between notation and block was also significant (F[3, 27] = 4.93, p<.01): all pairs were answered faster in the second block, but this effect was more pronounced in mixed pairs (block 1, 532.4 versus block 2, 484.9) than in pure pairs (block 1, 573.4 versus block 2, 554.6).

Equal and unequal pairs in both pure and mixed trials were then contrasted. In pure trials, equal pairs were answered 36.3 ms slower than unequal pairs (F[1, 9] = 10.64, p<.001); this effect was significant in both blocks but maximised in the second one (16.7 ms versus 37 ms). In mixed trials this difference was also significant (F[1, 9] = 4.45; p<.05): subjects were 17.5 ms slower in answering different to pairs of stimuli written in different notations but representing the same quantity (e.g., 6 SEI) compared to stimuli different on both dimensions (e.g., 2 SEI). However, this effect was significant in block 1 only (block 1, 26 ms; F[1, 9] = 7.32, p<.01; block 2, 9 ms, F[1, 9] <1).

The difference between close and far pairs was not significant and this contrast did not interact with notation (F[3, 27]< 1).

**Error analysis**

A similar ANOVA was carried out on arcsine-transformed error proportions. The error analysis replicated the RT analysis and indicated that no speed-accuracy trade-off affected the outcomes. The main effect of disparity was significant, (F[2,18] = 25.04; p<.0001); decomposition into contrasts indicated that subjects made more errors in unequal trials than in equal trials (5% versus 0.8%; F[1, 9] = 48.79, p<.0001). This effect reached significance in both pure (F[1, 9] = 15.45, p<.001) and mixed pairs, (F[1, 9] = 9.08, p<.01). However, in mixed pairs the effect was again limited to the first block.
(block 1, F[1, 9] = 12.31, p<.001; block 2, F[1, 9] = 2.31, p>.1). There was no significant
difference between error rates in close and far pairs (p>.27).

4.4.3 Discussion

The findings replicated the most critical results of Experiments 4N and 4P. In the
numerical matching task, a distance effect was observed in all conditions of notation
suggesting that number magnitude was accessed during the task. The size of the effect
was comparable to Experiment 4N (indeed, the same numerical distance - 1 and 4-
defined close and far pairs in both Experiments); thus, the unlimited presentation of the
stimuli had no influence on this effect. Similarly to Experiment 4N, we did not observe
any additional semantic effects. Consistently with both previous numerical matching tasks
(Experiments 4N and 5N), pure pairs were answered faster than mixed pairs and these
latter yielded relatively frequent errors in the same condition; these findings suggest that
subjects were sensitive to the physical similarity of the stimuli, though they clearly did not
rely merely on this dimension to carry out the task.

More critically, in the physical matching task the distance effect did not emerge in any
condition of notation. Thus, subjects were not affected by the semantic dimension of the
stimuli, not only when pairs where physically highly dissimilar (mixed trials) but also
when pairs were written in the same notation. Again, there was no sign of response bias
or accuracy trade/off that might have influence the level of processing of the stimuli and
thus contribute to the absence of the distance effect. The main factor in determining the
difference between the present findings and the ones reported by Dehaene and Akhavein,
may, thus, safely be attributed to the specific range of numerals presented.

Similarly to Experiment 4P, an interference effect, induced by the number-identity of
physically different numerals (e.g., 1 ONE), emerged in both RTs and error rates, but it
was again systematically observed only in the first block of trials. Thus, the unlimited
presentation of the stimuli did not have any critical effect neither on the size not on the
gradual decrease of the interference. No distance effect was observed in mixed pairs suggesting that the interference emerged at an asemantic level.

4.5 EXPERIMENTS 7N AND 7P

4.5.1 Method

Preliminary tasks

Third- and fifth- grade children participated in the study. The presentation of written verbal numerals to be quickly and accurately recognised was the criteria for not including younger children. Indeed, developmental research on word recognition suggests that by the age of 7/8 children show automatic semantic processing of printed words (e.g., Rosinsky et al., 1975; Shadler & Thissen, 1981) and their reading and comprehension skills may be considered adequate for the purpose of the present study.

All recruited children underwent a preliminary assessment where some basic numerical skills were evaluated. These tasks included: reading aloud the numbers from 1 to 9 both in Arabic and in verbal notations; counting forward and backward from 1 to 20; matching Arabic numerals to numerosities (N=6); selecting the larger between two Arabic numerals (hand-times RTs; N=12). Selection criteria were a flawless performance in reading, counting and matching tasks and an error rate lower than 0.25 in numerical comparison. All recruited children (N=106) satisfied the above criteria.

The chronometric data from the number comparison task were analysed in a 2x2 ANOVA with grade (third and fifth) as between-subjects factor and distance (close and far) as within-subjects factor. Both main effects resulted significant: younger children were significantly slower than older ones (F[1, 104] = 11.00; p<.005); more importantly, both groups showed a normal distance effect (F[1, 104] =11.59; p<.001).

The overall error rate was low (1.3%) and error distribution fits the RT data (Table 4.8).
Table 4.8. Preliminary numerical comparison task: Reaction times, Standard deviations and error rates as a function of grade and of numerical distance.

<table>
<thead>
<tr>
<th></th>
<th>3rd Grade</th>
<th></th>
<th></th>
<th>5th Grade</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>SD</td>
<td>%Er</td>
<td>RT</td>
<td>SD</td>
<td>%Er</td>
</tr>
<tr>
<td>close</td>
<td>1101</td>
<td>228</td>
<td>4.1</td>
<td>972</td>
<td>138</td>
<td>0.6</td>
</tr>
<tr>
<td>far</td>
<td>1052</td>
<td>217</td>
<td>0.0</td>
<td>953</td>
<td>131</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Subjects

Twenty-four third-grade children (12 female and 12 male) with an average age of 8 years and 8 months and twenty-four fifth-grade children (11 female and 13 male) with an average age of 10 years and 8 months, participated in the numerical matching task (Experiment 7N). Twenty-eight third-grade children (17 female and 11 male) with an average age of 8 years and 9 months, and thirty fifth-grade children (14 female and 16 male) with an average age of 10 years and 8 months participated in the physical matching task (Experiment 7P). All subjects were right-handed and reported normal or corrected to normal vision.

Stimuli and procedure

The same stimuli, procedure and experimental design of Experiments 6N and 6P were used.

Apparatus

A Macintosh PowerBook 175 running PsychLab 1.0 software (version.99; Gum, 1996) was used to displayed the stimuli and record RTs. Arabic numerals and uppercase verbal numerals were displayed in white Geneva bold 36-point font onto a black background.
4.5.1 Results

Experiment 7N: Numerical matching task

Reaction times (means of medians) and error rates for each experimental condition in the two groups are showed in Table 4.9 and plotted in Figure 4.8.

The overall error rate was 2.3% for the third-graders and 1.75% for the fifth-graders (individual error rate did not exceed 7.9%)\(^25\).

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>VV</th>
<th>AV-VA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>SD</td>
<td>%ER</td>
</tr>
<tr>
<td>Grade 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>equal</td>
<td>1207.5</td>
<td>252</td>
<td>1.0</td>
</tr>
<tr>
<td>close</td>
<td>1329.1</td>
<td>296</td>
<td>0.5</td>
</tr>
<tr>
<td>far</td>
<td>1356.9</td>
<td>334</td>
<td>1.3</td>
</tr>
<tr>
<td>Grade 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>equal</td>
<td>941.5</td>
<td>135</td>
<td>1.8</td>
</tr>
<tr>
<td>close</td>
<td>1038.6</td>
<td>169</td>
<td>0.3</td>
</tr>
<tr>
<td>far</td>
<td>1005.1</td>
<td>163</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 4.9. Experiment 7N: Mean Reaction times, Standard deviations and Error rates as a function of grade, notation and numerical disparity.

\(^{25}\) This surprisingly very low error rate has to be explained by the general rule explicitly taught to children in school which emphasises the importance of accuracy over speed of execution. Thus, even if instructions to these experiments specifically address the relative importance of errors, children were clearly guided by their standard approach to any tasks.
Figure 4.8. Experiment 7N: Mean Reaction times and Error rates as a function of grade, notation and numerical disparity [* white data points correspond to “same” answers; black data points to “different” answers].

Reaction times analysis

For every subject median correct RTs for each condition were calculated and entered in a repeated measures ANOVA with grade (third and fifth) as between-subjects factor and block (first - set 1 and 2- and second -set 3 and 4), notation (AA, VV, AV, VA) and disparity (equal, close and far) as within-subjects factors.

The overall performance differed significantly in the two grades (F[1, 46] = 21.52; p<.0001) with younger children being 309 ms slower than older ones (1457 ms and 1147 ms, for third- and fifth- grade respectively).

The main effect of block was not significant, nor its interaction with grade. The main effect of notation was significant (F[3, 138] = 128.37, p<.0001). Decomposition into contrasts indicated that pure trials (AA and VV) were answered significantly faster than mixed trials (172.3 ms; F[1, 46] = 288.45; p<.0001). Arabic trials were answered
significantly faster than verbal trials (136.7 ms; F[1, 48] = 96.63 p<.0001); in mixed trials
the order of the stimuli (Arabic number in first or second position) did not have any effect
on reaction times (F[1, 46] <1).

The main effect of disparity was significant (F[2, 92] = 33.65, p<.0001), as well as its
interactions with block (F[2, 92] = 3.23, p<.05) and with notation (F[6, 276] = 9.76,
p<.0001). This effect was decomposed contrasting first, equal versus unequal pairs and
then, close versus far pairs.

Same responses were 93.2 ms faster than different responses (F[1, 46] = 59.58,
p=.0001); this difference was more pronounced in the first block than in the second one
(block 1, 113 ms; block 2, 73.5 ms), and it was more pronounced in pure trials (141.3
ms, p<.0001) than in mixed trials (45.2 ms; p<.001).

Within different responses, far pairs were answered significantly faster than close
pairs (38.8 ms; F[1, 46] = 7.73; p<.01) but this effect reached significance in mixed trials
only (83.4 ms; F[1, 46] = 23.68, p<.0001). For older children, however, this difference
was also significant in pure trials (F[1, 23] = 7.42, p<.01).

Error analysis

A similar ANOVA was carried out on arcsine-transformed error proportions. The
error analysis replicated the RT analysis and indicated that no speed/accuracy trade-off
affected the outcomes. The main effect of notation was significant, (F[3, 138] = 13.55;
p<.0001); decomposition into contrasts indicated that subjects made more errors in mixed
than in pure trials (2.5% versus 1%; F[1, 46] = 38.36, p<.0001). The effect of disparity
was also significant (F[2, 92] = 14.79, p<.0001) as well as its interaction with notation
(F[6, 276] = 3.43, p<.005). Decomposition into contrasts indicated that equal pairs were
more error prone than unequal pairs (3.79% versus 0.9%; F[1, 46] = 29.20, p<.0001),
but this effect was significant in mixed trials only (F[1, 46] = 54.58, p<.0001). Overall,
error rates for close and far pairs did not differ (F[1, 46] <1); yet, this effect was
marginally significant in mixed trials (F[1, 46] = 2.97, p=.08).
**Experiment 7P: Physical matching task**

Reaction times (means of medians) and error rates for each experimental condition are showed in Table 4.10 and plotted in Figure 4.9.

The error rates were 0.8% for the third-graders and 0.7% for the fifth-graders (individual error rate did not exceed 3.1%). Error rates were extremely low and were not further analysed.

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>VV</th>
<th>AV-VA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>SD</td>
<td>%ER</td>
</tr>
<tr>
<td>Grade 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>equal</td>
<td>1163.8</td>
<td>258</td>
<td>1.6</td>
</tr>
<tr>
<td>close</td>
<td>1212.2</td>
<td>385</td>
<td>0.9</td>
</tr>
<tr>
<td>far</td>
<td>1202.3</td>
<td>323</td>
<td>0.4</td>
</tr>
<tr>
<td>Grade 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>equal</td>
<td>993.4</td>
<td>224</td>
<td>1.6</td>
</tr>
<tr>
<td>close</td>
<td>992.2</td>
<td>209</td>
<td>0.6</td>
</tr>
<tr>
<td>far</td>
<td>974.5</td>
<td>247</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 4.10. Experiment 7P: Mean Reaction times, Standard deviations and Error rates as a function of grade, notation and numerical disparity.

**Reaction times analysis**

For every subject median correct RTs for each condition were calculated and entered in a repeated measures ANOVA with block (first and second), notation (AA, VV, AV, VA) and disparity (equal, close and far) as within-subjects factors, and grade (third, fifth) as a between-subjects factor.

The two grades differ significantly, older subjects were 228 ms faster than younger ones, \((F[1, 56] = 11.06; p<.001)\).
Overall, reaction times were slower in the first block (1158.7 ms) than in the second block (1107.9 ms), \(F[1, 56] = 5.35; p<.05\).

The main effect of notation was significant \(F[3, 168] = 14.35, p<.0001\): decomposition into contrasts indicated that Arabic pairs were answered significantly faster than Verbal pairs (89 ms; \(F[1, 56] = 42.74; p<.0001\)) while within mixed trials, AV and VA pairs did not differ significantly \(F[1,56] < 1\). Overall pure (AA and VV) and mixed trials (AV-VA) did not differ \(F<1\). Notation also interacted with block \(F[3, 168] = 7.00, p<.001\): all pairs were answered faster in the second block, but this effect was maximised for mixed trials (mixed trials, 79.5 ms; pure trials, 21.9 ms). Moreover, within pure trials, Verbal pairs only benefited from practice (41.5 ms versus 2 ms). A marginally significant triple interaction between block, notation and grade \(F[3, 168] = 2.53; p=.059\) suggested that this pattern modified with age. Decomposition into contrasts indicated that
for both grades, all notations were answered faster in the second block than in the first one with the exception of Arabic trials: for no clear reason, younger children answered slower to Arabic pairs in the second block (p=.069).

The main effect of disparity was significant (F[2, 112] = 4.84, p<.01), as well as its interaction with notation (F[6, 336] = 4.45, p<.0005). This effect was decomposed contrasting, first, equal versus unequal pairs in both pure and mixed trials and then, close versus far pairs. In pure trials, equal and unequal pairs did not differ despite the unbalanced frequency of same and different answers (F[1, 56]<1). However, in mixed trials this difference was highly significant (F[1,56] = 33.13; p<.0001): subjects were 68.3 ms slower in answering different to pairs of stimuli written in different notations but representing the same quantity (e.g., 2 DUE) compared to stimuli different on both dimensions (e.g., 2 SEI). This effect was highly significant in both grades (both, p<.01). Contrasts did not reveal any difference between close and far pairs (F[1, 56] <1) in any condition of notation.

4.5.3 Discussion

Overall, age-related differences were both quantitative and, to a lesser extent, qualitative. Older children's performance was faster than younger one, possibly due a general improvement in encoding and answer-selection skills; accuracy on the other hand, did not differ in the two groups.

In the numerical matching task, the two groups performed differently, not only quantitatively but also qualitatively. Third-Grade children did not show any distance effect in the processing of pure trials: when requested to decide whether two numerals were identical or not they seemed to answer relying on the physical features of the stimuli. In older children, however, this effect was significant, indicating a developmental trend toward an autonomous access to magnitude information. This result is in contrast with the evidence reported by Duncan and McFarland (1980). In their study, children were presented with pairs of Arabic numerals to be classified as identical or not; a
distance effect was observed in children as young as six years old. However, their task differed from the present one in several methodological aspects. Further to a different set of stimuli pairs and a different proportion of same/different responses, their experiment included Arabic pairs only and, quite critically, half of the trials were presented in degraded condition, i.e., a diagonal line grid was superimposed over the digit pairs. Possibly this manipulation, introduced to disentangle the functional level at which the symbolic distance effect arises, increased the difficulty of recognising the stimuli, thus preventing the subjects from performing a simple visual matching. Though a same-different decision on the non-degraded pairs might have been easily taken on the basis of visual features, this would have required children to switch, from trial to trial, from one level of processing to a different one.

In mixed trials, where a simple perceptual matching was not enough to answer, the distance effect emerged, indicating that numerals were semantically processed; this effect was qualitatively the same in both Grades.

Thus, the absence of a distance effect in the processing of pure trials in younger children’s performance provide further support for the hypothesis that the semantic dimension of the numerals (i.e., the magnitude represented by Arabic and verbal numerals) becomes salient over the course of learning; we assume this to be a prerequisite for autonomous access to magnitude information. These results are thus consistent with the evidence reported in Chapter 3, where a similar developmental trend towards autonomous access to magnitude information was observed using a different experimental paradigm.

In the physical matching task no distance effect was found in any condition of notation; for older children, as for adults, the instructions modulated the depth of processing of the stimuli. To directly verify the extent to which the distance effect was modulated by the instructions, data from Experiments 7N and 7P were entered in an ANOVA with notation (AA, VV, AV-VA) and distance (close and far) as within-subjects factors and task (numerical and physical) as between-subjects factor. The analysis was carried out for older children only. For fifth-graders, there was a significant effect of
numerical distance, \((F[1, 52] = 16.64; p<.001)\): reaction times were faster for far pairs (1089 ms) than for close pairs (1060 ms). The significant interaction between numerical distance and task, \((F[1, 52] = 4.88; p<.05)\) indicated that the size of the effect differed in the two tasks: in the numerical matching task, far pairs were answered significantly faster (1133.3 ms) than close pairs (1180.7 ms), while in the physical matching task, this difference was no longer significant (far pairs, 1003.4 versus close pairs 1017.8 ms; \(p>.1\)).

Moreover, in the physical matching task a significant interference effect was observed all over the experiment. Children took longer to answer *different* to pairs of stimuli representing the same number (e.g., 2 TWO) compared to stimuli representing different numbers (e.g., 2 SIX). Though mixed pairs were physically highly dissimilar, children were disturbed by the mismatch between number-identity and physical difference. The long lasting interference effect suggests that children, contrary to adults, did not benefit from practice: the immature selective attention mechanisms may be responsible for this failure (e.g., Tipper et al., 1989). Moreover, the absence of age differences in the size of this phenomenon suggests that, in the course of developmental, the improvement of inhibitory mechanisms co-occurs with the increasing impact of symbolic information in numerical processing.

### 4.6 Summary and Conclusions

The aims of the Experiments presented in this chapter were 1) to evaluate the extent to which access to magnitude information is autonomously activated, 2) to examine whether this process occurs irrespective of the stimuli notation and 3) to gather evidence for the existence of a direct route between Arabic and verbal lexicons. Finally, the developmental trends of these phenomena were investigated by testing third-grade and fifth-grade children.

The significant distance effect observed in Experiments 4N, 5N and 6N strongly suggests that access to magnitude information is critical in number processing (e.g.,
Dehaene et al., 1993; Fias et al., 1996; Tzelgov et al., 1992). Even if the same-different judgements could have been carried out on the basis of the visual characteristics of the stimuli, nonetheless stimuli were semantically processed. This effect was similar across different notations, supporting the hypothesis that both Arabic and verbal numerals have equal access to an internal magnitude representation (Dehaene, 1992; McCloskey, 1992). These data however, may not contribute to clarify whether Arabic and verbal numerals have direct access to the semantic system (e.g., McCloskey 1992) or if alternatively, they are converted into another format before semantic access takes place (e.g., Dehaene, 1992; Noel & Seron, 1993).

In the numerical matching task, however, for half of the experimental trials, i.e., all mixed pairs, a visual matching would have not be sufficient to answer correctly; this fact possibly favour a semantic processing of all stimuli. In this respect, the physical matching task provided a more stringent test for autonomous processing of magnitude information; the attention of the participants was explicitly drawn to the physical features of the stimuli (e.g., equal pairs such as ‘2 two’ have now to be answered different).

In fact, we found that in this condition, evidence for semantic processing was modulated by the nature of the stimuli (see Table 4.11). In pure trials, a significant distance effect was found only when pairs included extreme values and the distance between far pairs was maximised (Experiment 5P). However, when these conditions were no longer met (Experiments 4P and 6P) the distance effect was not significant. These results suggest that for larger numerical distances and, in particular, for numbers which are unambiguously associated with small and large codes, semantic information might be available earlier, thus having the time to interfere with the response. In all other conditions, access to magnitude information was possibly initiated but not completed before the fast physical discrimination took place (see also Experiment 4 in Dehaene and Akhavein, 1995).
<table>
<thead>
<tr>
<th>NUMERICAL</th>
<th>AA</th>
<th>VV</th>
<th>AV-VA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>close</td>
<td>far</td>
<td>close</td>
</tr>
<tr>
<td>Exp. 4N</td>
<td>616</td>
<td>22</td>
<td>594</td>
</tr>
<tr>
<td>Exp. 5N</td>
<td>710</td>
<td>24</td>
<td>686</td>
</tr>
<tr>
<td>Exp. 6N</td>
<td>560</td>
<td>17</td>
<td>543</td>
</tr>
<tr>
<td>Dehaene and Akhavein Exp. 1</td>
<td>578</td>
<td>28</td>
<td>550</td>
</tr>
<tr>
<td>PHYSICAL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp. 4P</td>
<td>515</td>
<td>n.s.</td>
<td>521</td>
</tr>
<tr>
<td>Exp. 5P</td>
<td>555</td>
<td>28</td>
<td>527</td>
</tr>
<tr>
<td>Exp. 6P</td>
<td>515</td>
<td>n.s.</td>
<td>506</td>
</tr>
<tr>
<td>Dehaene and Akhavein Exp. 2</td>
<td>466</td>
<td>26</td>
<td>440</td>
</tr>
</tbody>
</table>

Table 4.11. Distance effect in numerical and physical matching tasks as a function of notation and Experiment (RTs in ms, n.s. = not significant; c-f = RT close - RT far)

However, evidence from other experimental paradigms (e.g., Fias et al., 1996) seems to suggest that all single-digit numbers may indeed access semantics in an autonomous fashion; moreover, it is also well established that, when intentionally activated, the absolute magnitude of the number greatly modulate the speed of this process (e.g., Dehaene & Mehler, 1993). Nonetheless, we believe that task demands and the presentation mode (presentation of single digit versus pair of digits) may play an important role in modulating the occurrence of semantic effects. Experimental paradigms, as the matching tasks adopted in the present study, where semantic effects are measured by an index of semantic similarity between two numbers (i.e., distance effect), may be more sensitive to the range of numbers used and, by definition, to their semantic proximity. In agreement with this interpretation, Garner et al. (1982) did not report any effect of numerical distance on the speed of different responses when pairs of Arabic numerals including the numbers 3, 4, 6 and 7, had to be matched for physical identity and
for parity status. Evidently, the selection of the stimuli is more important than is generally assumed.

The existence of an asemantic transcoding route between Arabic and verbal lexicons, as postulated by some of current models of number processing (e.g., Cipolotti & Butterworth, 1995; Dehaene, 1992, Noel & Seron, 1993) was probed by evaluating the extent to which subjects might have recognised that 2 and TWO represented the same number, without showing any index of semantic processing. This effect emerged in the physical matching tasks and in both reaction times and error rates. Interestingly, this interference was reduced with practice, similarly to what has been reported in several interference-sensitive tasks (e.g., Stroop, 1935; Roger & Fisk, 1991). However, the evidence for asemantic transcoding may not be considered conclusive and an alternative explanation may be offered. Even within semantics, same-different relations might be more relevant than close-far relations, thus yielding stronger interference (McCloskey, quoted by Deheane & Akhavein, 1995). According to this interpretation, the significant same-different effect and the insignificant distance effect may have both arisen from the semantic level. In this regard, children's performance may be of particular relevance. Overall, children presented a reduced effect of semantic information in their performance, yet they consistently showed a great interference effect from mixed pairs representing the same number. As expected by their immature inhibitory mechanisms, children were not able to benefit from practice; the absence of a developmental trend in this phenomenon seems to suggest that, over the course of learning, symbolic information may gradually become more relevant and increase its power as potential source of interference (Bisanz et al., 1979).
CHAPTER 5

SEMANTIC EFFECTS IN VERBAL NUMBER PRODUCTION:
A CASE STUDY

INTRODUCTION

One of the issue addressed in the studies reported in Chapter 4 concerns the existence of an asemantic route mediating the transcoding between Arabic and verbal numerals. This hypothesis was experimentally tested by evaluating if subjects could map an Arabic numeral (e.g., 2) onto its corresponding verbal form (e.g., two) without semantic mediation. Indeed, the absence of a distance effect together with a slow and inaccurate performance in answering different to arabic-verbal pairs representing the same quantity (e.g., 2 TWO), seemed to favour the above hypothesis. Yet, as previously discussed, this evidence may be alternatively explained by calling into question the differential strength of same-different relations and close-far relations within number semantics: the former, being more salient, may give rise to stronger interference effect. Thus, the evidence for asemantic transcoding provided by Experiments 4, 5 & 6 may not be considered conclusive.

As already discussed in the literature review, the semantic/aseastic debate has received major input from the neuropsychological investigations, where clear-cut pattern of preserved and impaired performances in tasks apparently mediated by the same underlying process, pointed to the necessity of postulating both semantic and asemantic

---

26 The study reported in this Chapter has been published in Neurocase (Delazer & Girelli, 1997).
routes for numerical transcoding. In the present Chapter we report a neuropsychological study that contributes to this debate providing further evidence that semantics may facilitate number processing (e.g., production of spoken verbal numerals), in particular showing that this facilitation may be modulated by the specific numerical meaning invoked by the task. Thus, while the experimental studies reported in Chapter 3 and Chapter 4 focused on the impact of magnitude information in numerical processing, in the present study we explored the role and effect of different numerical meanings.

The case-study reported in the present Chapter investigates the production of spoken verbal numerals in an aphasic and dyslexic patient, ZA. Several models of transcoding numerals from one format to another (e.g. from Arabic numerals to spoken verbal numerals) have been advanced in recent years. Besides single route models (Deloche & Seron, 1982a, 1982b; McCloskey et al., 1985; McCloskey et al., 1986), multiple-route models have been proposed, which parallel models of alphabetical reading (Cipolotti, 1993; Cipolotti, 1995; Cipolotti & Butterworth, 1995; Cohen et al., 1994; Dehaene & Cohen, 1995). These models assume semantic transcoding routes accessing semantic information and direct transcoding routes, which bypass the semantic system. However, as Seron and Noel (1995) state, multi-route models are still imprecise with regard to the type of underlying representation and to the mechanisms involved in numerical processing. In the present study we examine the effect of different notations (Arabic and alphabetic), tasks (reading, counting, calculation etc.), and task instructions on ZA's production of spoken verbal numerals.

Several studies were concerned with the transcoding from one number format to another (e.g. reading aloud Arabic numerals or writing Arabic numerals from dictation). Deloche and Seron (1982a, b; Seron & Deloche, 1983; 1984) studied the number transcoding in aphasic patients and found typical errors described as lexical (e.g., the written verbal numeral fifty-two is transformed to the Arabic numeral 54) or syntactic (e.g., the written verbal numeral fifty-two is transformed to the Arabic numeral 502). These errors were thought to indicate a selective breakdown in specific stages (lexical or
syntactic) of a transcoding algorithm, which was assumed to be asemantic in its nature. Deloche and Seron’s model of asemantic transcoding has been challenged by McCloskey et al. (1985). In the McCloskey model all transcoding processes pass through a common semantic representation (Figure 5.1). At the comprehension stage, all numerical inputs are transformed into abstract semantic representations via notation specific modules. At the production stage, abstract semantic representations activate notation specific modules which produce spoken verbal numerals, written verbal numerals or Arabic numerals. The comprehension and production mechanisms include syntactic and lexical stages which can be selectively disrupted, as shown in several neuropsychological case studies (McCloskey et al., 1985, 1986; McCloskey, 1992).

Recent studies proposed multiple routes, semantic and asemantic ones, for number transcoding, drawing a parallel to the semantic and asemantic routes identified in word reading. Cohen et al. (1994) reported a dissociation between preserved reading of meaningful numbers and impaired reading of numbers without specific meaning in a deep dyslexic patient. They proposed a model where familiar (meaningful) and unfamiliar numerals are processed via different pathways, the first along a lexical, the latter along a non-lexical route (Figure 5.2).
For example, the number 1789 (date of the French revolution) would be processed along the lexical route by French subjects, but not the number 2781. Familiar numerals would be recognised in the numerical input lexicon and would point to the semantic knowledge associated with the number; on the other hand, unfamiliar numerals would be processed along the non-lexical surface route. Cohen et al. (1994) assumed that their patient had a specific deficit in the surface route, but could still processed familiar numerals along the lexical pathway. This pattern of results received an alternative explanation from Seron and Noel (1995). They suggested that the familiarity effect intervenes at the verbal output mechanisms: Familiar numbers may be preassembled in the verbal output lexicon (as other verbal formula) and may thus be easier to produce than
other, unfamiliar numerals. A further distinction between different reading routes for Arabic numerals has been suggested by Cohen and Dehaene (1995) in study of Arabic number reading in two alexic patients. These patients were more accurate in reading numbers in a number comparison task than in reading the same numbers in an addition task. Cohen and Dehaene (1995) argued that two different pathways are employed in the two tasks, one direct asemantic pathway, active in reading numerals in the addition tasks, and one pathway accessing analogue magnitude representations, active in the comparison task (the assumption of asemantic transcoding in the addition task relies on the number processing model of Dehaene and Cohen (1995, Cohen et al., 1994) which claims that simple calculations are processed verbally without access to semantic elaboration).

Thus, Cohen and Dehaene (1995; Cohen et al., 1994) proposed more than one semantic route for reading numerals: one lexical semantic route for familiar Arabic numerals accessing a store of semantic number knowledge (Cohen et al., 1994) and one non-lexical semantic route allowing access to the magnitude representation associated to any well-formed numerals. Finally, an asemantic route would allow direct conversion from Arabic input to spoken output.

A multiple route model for number processing was also proposed by Cipolotti (1995; Cipolotti & Butterworth, 1995). Patient SF (Cipolotti, 1995) was able to answer questions concerning numerical knowledge and cognitive estimations correctly in verbal form, but frequently failed to read Arabic numerals (a comprehension deficit was excluded by additional tasks). Cipolotti proposed two independent pathways for the production of verbal numerals - one asemantic pathway, activated in reading aloud and one semantic pathway mediating tasks assumed to rely on number meaning (Figure 5.3). Furthermore, the activation of the damaged asemantic route (in reading aloud) would inhibit the activation of the semantic route (which cannot therefore compensate for the deficit).
Seron and Noel (1995) mention similar reservations as in the case of Cohen et al. (1994). Again, specific tasks could be facilitated by the activation of preassembled number words. Another patient, SAM (Cipolotti & Butterworth, 1995) was impaired in transcoding tasks requiring verbal and written output (e.g., writing from dictation). In contrast, he was able to produce matched numerals, even multi-digit numerals, in written calculation (it should be noted, however, that the production of multi-digit Arabic numerals in calculation does require only minimal syntactic elaboration; indeed, it is sufficient to compute the answer step by step without processing the numeral as a whole) and 2-digit numerals in verbal calculation. Similarly, he could also correctly produce spoken verbal numerals in answering questions related to numerical knowledge. Cipolotti and Butterworth (1995) suggested four different asemantic routes...
in addition to the semantic route proposed by McCloskey et al. (1985). Most importantly, direct routes for reading Arabic numerals aloud and for writing Arabic numerals on dictation were assumed. The two other asemantic routes, which allow the repetition of number words and the reading of written number words, may rely on general language processing mechanisms which are not specific for numbers. In fact, repetition of verbal stimuli does not require access to their meaning (McCarthy & Warrington, 1984); similarly, the transcoding of written to spoken verbal numerals can be processed without meaning (Shallice et al., 1983; see Cipolotti & Butterworth, 1995 for discussion).

In summary, the case studies published in the last few years argue for multiple-routes in number processing and challenge the single route assumption proposed in the McCloskey model. However, the proposed distinction between asemantic and semantic pathways meets with both methodological and theoretical problems, and has not yet been conclusively established (for an extensive discussion, see Seron & Noel, 1995). First, no double dissociation between asemantic and semantic processing has been observed till now. All reported cases performed better in tasks relying on semantic processing and no one showed the reverse dissociation as, for example, impaired number comparison (i.e., a task that, if characterised by a normal distance effect, implies preserved number comprehension)\(^{27}\) and preserved numbers reading. It is thus possible that semantic tasks are simply "easier" than asemantic tasks. In fact, semantic tasks mostly require the production of familiar numerals, which are possibly preassembled at the output level and thus do not require any syntactic elaboration. Similarly, a better production of multi-digit numerals in written calculation than in a writing task could simply reflect the minimal syntactic competence required by the former. Second, the distinction between semantic and asemantic tasks is theoretically questionable. All numerals represent a numerosity and are thus by definition meaningful. Each so-called

\(^{27}\) Recently, Delazer and Butterworth (1997) reported a patient whose pattern of preserved (e.g., counting) and impaired numerical abilities (e.g., simple arithmetic) indicated a striking dissociation between the sequence meaning of number terms and the cardinal meaning of them. The patient was still able to compare two numbers but only by counting from one number to the other; this strategy yielded an inverse distance effect.
as a semantic task, such as reading Arabic numerals, can be transformed into a semantic task by accessing its corresponding numerosity. Thus it is hardly possible to prove that a task is performed without any access to semantics. Third, the repeatedly claimed parallels between the semantic and asemantic reading routes employed in alphabetical material and the reading routes employed in reading Arabic notation are questionable. As Seron and Noel (1995) state, every Arabic numeral can be transcoded through a non-lexical route, since no irregular numerals exist. Thus, even an asemantic surface route should transcode all possible numerals correctly. On the other hand, if the asemantic surface route is damaged, the semantic route should still be able to interpret all numerals, since they all represent a numerosity. Every possible combination of single digits represents a legal multi-digit numeral, denoting a numerosity, while the same is not true for alphabetical material, where randomly generated combinations of letters can result in nonwords.

The study reported in this Chapter contributes to this debate providing further evidence that semantics may facilitate the production of spoken verbal numerals, and in particular, showing that this facilitation may be modulated by the specific numerical meaning invoked by the task.

The study consists of two parts, the second part was carried out three years after the first one. In the first part we investigated whether reading difficulties for words and nonwords were associated with an impairment in reading numbers. We also assessed whether the type of notation (Arabic or alphabetical) had a critical effect in reading numbers. The analysis of error types in reading Arabic numerals and written verbal numerals allowed us to draw conclusions on the functional localisation of the reading impairment at a lexical stage. As repeatedly described (Deloche & Seron, 1982a; McCloskey et al., 1985 and following work), distinct and functionally independent mechanisms are devoted to the processing of the syntactic structure and lexical elements of multi-digit numerals at the production level.

The second part of the study consists of the reading tests used in the first testing session and of several tasks assessing the production of spoken verbal numerals in
semantic and asemantic tasks. The repetition of the reading tests after three years allowed us to assess quantitative and qualitative changes in ZA’s performance. Of special interest is the differential improvement in different tasks over time. We expected the reading of written verbal numerals to improve along with the reading of words and nonwords, since these three types of stimuli share the same type of notation (i.e., alphabetical). On the other hand, the reading of Arabic numerals would have possibly shown a different pattern of recovery.

5.1 CASE DESCRIPTION

ZA was a 46 year old right-handed man with 13 years of formal education. Before his illness he worked as a director of a professional education centre. In February 1991 he suffered a left cerebrovascular accident, followed by right hemiparesis and non-fluent aphasia. A CT scan showed a large hypodense area in the territory of the left middle cerebral artery.

The patient was first-seen as an out-patient at the Clinica Neurologica in Padova 20 months after the onset of his illness, when the first part of the tests was performed. The second part of the tests was performed 5 years after the CVA.

5.2 FIRST PART OF THE INVESTIGATION: TESTS IN 1992

5.2.1 Background tests

ZA performed within the average in the Raven test (Raven Standard Progressive Matrices 46/48 correct; Spinnler & Tognoni, 1987). His digit span was 4 (raw score) and his visuo-spatial span, assessed by the Corsi block tapping test was 5. ZA’s spontaneous speech was non-fluent with frequent anomias and severe articulatory problems. In the Aachener Aphasie Test (Versione Italiana; Luzzatti, Willmes and
DeBleser, 1991) ZA scored very poorly in the Token subtest (41/50), but showed only mild difficulties in the comprehension subtest (97/120). In repetition he had problems with long words and with sentences (medium deficit, 112/150). In naming a medium deficit was found (63/120).

The assessment of ZA's calculation and number processing abilities was carried out through the administration of the Miceli and Capasso battery (1991; see Table 5.1 and 5.2).

<table>
<thead>
<tr>
<th>TASK</th>
<th>Items correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison of dot patterns</td>
<td>10/10</td>
</tr>
<tr>
<td>Dot pattern seriation (seriation of dot patterns according to their numerosity)</td>
<td>5/5</td>
</tr>
<tr>
<td>Number comparison (1-to 6-digit numerals)</td>
<td></td>
</tr>
<tr>
<td>Arabic numerals</td>
<td>33/34</td>
</tr>
<tr>
<td>Written verbal numerals</td>
<td>33/34</td>
</tr>
<tr>
<td>Spoken verbal numerals</td>
<td>32/34</td>
</tr>
<tr>
<td>Recognition of arithmetic signs</td>
<td></td>
</tr>
<tr>
<td>Visual</td>
<td>10/10</td>
</tr>
<tr>
<td>Verbal</td>
<td>8/10</td>
</tr>
<tr>
<td>Written calculation (multi-digit calculation)</td>
<td></td>
</tr>
<tr>
<td>Addition</td>
<td>20/20</td>
</tr>
<tr>
<td>Subtraction</td>
<td>20/20</td>
</tr>
<tr>
<td>Multiplication</td>
<td>3/20</td>
</tr>
<tr>
<td>Mental calculation (presentation verbally, answer in Arabic digits)</td>
<td></td>
</tr>
<tr>
<td>Addition</td>
<td>20/20</td>
</tr>
<tr>
<td>Subtraction</td>
<td>20/20</td>
</tr>
<tr>
<td>Multiplication</td>
<td>5/15</td>
</tr>
</tbody>
</table>

*Table 5.1. Non-numerical tasks, number comparison, calculation and recognition of arithmetic signs in 1992 (tasks from Miceli and Capasso, 1991).*
ZA's performance was nearly perfect in number comparison (Arabic numerals, written verbal numerals, spoken verbal numerals) and perfect in the comparison and seriation (i.e., ordering by magnitude) of dot patterns. Repetition, writing from dictation and transcoding from written verbal numerals to Arabic numerals were good for one- and 2-digit numbers. With 3-digit numbers ZA scored at zero level in all tasks, except in repetition (5/10; errors were due to articulatory problems) and in transcoding from written verbal numerals to Arabic numerals (7/10). ZA was unable to answer 4- to 6-digit numbers in any task, so we did not present the whole range of items of the battery. In reading both Arabic and verbal numerals a specific error pattern was found. Errors frequently consisted of other numbers of the same number class (Ones were substituted by Ones, Teens by Teens, Tens by Tens; e.g., 9 was read as *quattro* [four], 12 as *quindici* [fifteen], and so forth). As in repetition, errors in reading long numerals (in both verbal and Arabic notations) were not clearly classifiable due to articulation problems.

ZA showed no difficulties in recognising arithmetical signs. He made no errors in simple addition and subtraction (presented verbally, results given in Arabic numerals), while multiplication was clearly compromised (only 33.3% correct). In written multi-digit calculation tasks, again, addition and subtraction were perfect, while many errors occurred in multiplication (15% correct; errors were due to incorrect fact retrieval). ZA had no problems composing visually presented Arabic numerals with poker chips of different values (e.g., 52 should be composed by a chip of 50 and two chips of 1; 24/24

---

**Table 5.2. Transcoding tasks in 1992 (tasks from Miceli and Capasso, 1991).**

<table>
<thead>
<tr>
<th>Task</th>
<th>Ones</th>
<th>Teens</th>
<th>2-digit numbers</th>
<th>3-digit numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetition of verbal numerals</td>
<td>10/10</td>
<td>5/5</td>
<td>10/10</td>
<td>5/10</td>
</tr>
<tr>
<td>Reading Arabic numerals aloud</td>
<td>10/10</td>
<td>3/5</td>
<td>5/10</td>
<td>0/10</td>
</tr>
<tr>
<td>Reading written verbal numerals aloud</td>
<td>10/10</td>
<td>5/5</td>
<td>2/10</td>
<td>0/10</td>
</tr>
<tr>
<td>Writing Arabic numerals on dictation</td>
<td>10/10</td>
<td>5/5</td>
<td>8/10</td>
<td>0/10</td>
</tr>
<tr>
<td>Transcoding written verbal numerals</td>
<td>10/10</td>
<td>5/5</td>
<td>9/10</td>
<td>7/10</td>
</tr>
<tr>
<td>to Arabic numerals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
correct; correct answers consisted of 2 to 6 chips). He answered correctly 15/15 cardinal number facts giving the answers in Arabic numerals (e.g. How many months there are in a year?).

ZA's nearly perfect performance in number comparison (in different notations and modalities), in addition and subtraction problems (with written answer) and in composing numbers with poker chips, indicated his intact ability to encode and understand Arabic numerals as well as written and spoken verbal numerals. In reading Arabic numerals and written verbal numerals, ZA showed particular difficulties, errors frequently resulting in lexical substitutions. In 1992 an investigation of his reading of both numerical stimuli (Arabic digits and written verbal numerals), and other alphabetical material was performed.

5.2.2 Reading tasks

Type of stimuli and number of items included in the different tasks are specified in the Tables.

Written words (different grammatical and semantic classes), nonwords (orthographically legal four to seven letter strings), verbal numerals and Arabic numerals were presented in the reading tasks. One part of the stimuli was presented in mixed lists (words, Arabic digits, verbal numerals), the other part was presented in pure lists. Since the performance did not differ between pure and mixed lists, data from both lists were collapsed. The patient was instructed to read the stimuli aloud.

Reading words

Reading was good for nouns, adjectives, proper names (common Italian forenames) and compound nouns (see Table 5.3; between 98.5.% and 89.9% correct). More errors occurred in function words and verbs (72.3% and 74.5% correct).
Reading errors were classified as follows: Visual errors (sharing more than 50% of the letters with the target and resulting in a legal Italian word; n=24), phonological paraphasias (substitutions of single phonemes; n=2), semantic paraphasias (Italian word with a semantic relation to the target and sharing less than 50% of the target’s letters; n=1), verbal paraphasias (Italian word without semantic relation to the target and sharing less than 50% of the target’s letter; n=5), inflectional errors (incorrect verb form; n=5) and omissions (n=2).

Reading nonwords

Sixty nonwords (orthographically legal four to seven letter strings) were used as stimuli. Half of them were constructed by changing one letter of an Italian word.

ZA read only 3 out of 60 nonwords correctly. The high error rate can not be attributed to articulation problems since ZA repeated the same nonwords without error in a verbal repetition task. He was also able to distinguish words from nonwords and he performed perfectly on a lexical decision task (40/40 correct; items taken from the nonword reading task).

Errors in reading nonwords were classified as lexicalisations (substitutions resulting in existing Italian words; n=29), phonological paraphasias (substitutions of single phonemes; n=4) and omissions (n=24).
Reading of Arabic numerals

ZA failed to read 85 out of 321 Arabic numerals (26.5%). The accuracy clearly dropped for 3-digit numerals (see Table 5.4).

Errors were classified as lexical or syntactic, as firstly proposed by Deloche and Seron (1982 a, b). Lexical errors concern one element of the numeral and result in substitutions within one number class, e.g. the number 62 may be misread as sixty-four or fifty-two. Syntactic errors concern the syntactic structure of the numeral and result in numerals which are generally of a different magnitude from the target number, e.g. 62 may be misread as six hundred and two. Additional classes consisted in mixed errors (substitution of elements in syntactically wrong numbers) and omissions.

The answers in reading Arabic numerals included 60 lexical errors, 10 syntactic errors, 9 mixed errors and 6 omissions.

Reading of verbal numerals

ZA failed to read 51 out of 266 verbal numerals (19.2%). ZA performed better with Ones and Teens than with more complex numbers (see Table 5.4). Errors were classified as for reading Arabic numerals; 39 lexical, 2 syntactic and 10 mixed errors were observed.

<table>
<thead>
<tr>
<th>READING ARABIC NUMERALS</th>
<th>% correct</th>
<th>READING VERBAL NUMERALS</th>
<th>% correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONES</td>
<td>92.1%</td>
<td>ONES</td>
<td>94%</td>
</tr>
<tr>
<td>TEENS</td>
<td>77.9%</td>
<td>TEENS</td>
<td>95.4%</td>
</tr>
<tr>
<td>TENS</td>
<td>70.4%</td>
<td>TENS</td>
<td>66.7%</td>
</tr>
<tr>
<td>2-digit</td>
<td>60.8%</td>
<td>2-digit</td>
<td>65.2%</td>
</tr>
<tr>
<td>3-digit*</td>
<td>23.0%</td>
<td>3-digit**</td>
<td>50%</td>
</tr>
<tr>
<td>Ones+ mult</td>
<td>80%</td>
<td>Ones+ mult.</td>
<td>92.8%</td>
</tr>
<tr>
<td>Total</td>
<td>73.5 %</td>
<td>Total</td>
<td>80.8%</td>
</tr>
</tbody>
</table>

Table 5.4. Reading verbal numerals and Arabic numerals in 1992. Percentage of correct answers. (*two with an internal 0; one with 0 at the end** one with an internal 0, one with 0 at the end)
5.2.2a Comments on the reading tasks

ZA's reading accuracy was good for nouns, adjectives, proper names and compounds (89.9% or more correct). Errors for these word classes were visually related, but semantically unrelated paraphasias. More errors occurred in reading function words and verbs. ZA was unable to read nonwords, instead producing similar real words. Problems in reading nonwords cannot be due to verbal output processes, since ZA repeated nonwords without difficulty. ZA had no problems recognising words, and lexical decision was performed with no errors. His inability to read nonwords indicates an impairment in sublexical grapheme-phoneme conversion processes. ZA thus showed the typical reading problems described in the literature as phonological dyslexia (Beauvois & Derousne, 1979; Funnell, 1983; Patterson, 1982; Sartori, Barry & Job, 1984; Shallice & Warrington, 1980).

ZA showed clear difficulties in reading complex numbers, both in Arabic and alphabetic notation. A highly significant correlation between error rate and number of words to be produced was found (Spearman rank coefficient $r=0.8627$; $p=0.006$; e.g., the numeral 123 requires the production of three number-words cento venti tre [hundred twenty three]). Lexical substitution was the dominant error type in both types of script. Moreover, errors could not be attributed to comprehension problems, since ZA performed tasks requiring number comprehension with no difficulty (number comparison, composing numbers with poker chips, answering calculations in written form). All these tasks required the correct encoding and comprehension of Arabic numerals, written and spoken verbal numerals. ZA's overall performance, thus, indicates a problem in the output mechanisms for spoken verbal numerals.

Although ZA showed reading difficulties for both Arabic and verbal notation, his reading of written verbal numerals was significantly better than his reading of Arabic numerals ($\chi^2 (1) = 4.36; p<0.05$). Thus, despite his reading problems, which included a total inability to use grapheme-phoneme conversion routines, ZA was better in processing alphabetic than Arabic notation.
5.3 FOLLOW-UP INVESTIGATION: TESTS IN 1996

5.3.1 Hypotheses for the follow-up investigation

In the follow-up study we investigated whether an improvement in reading words and nonwords in alphabetical script was accompanied by an improvement in reading verbal numerals. We expected such an improvement for the three types of stimuli sharing the same notation, although reading of Arabic numerals may possibly show a different pattern of recovery. Thus, we hypothesised that an improvement in reading alphabetical notation, and in particular in using sublexical grapheme-phoneme conversion rules, should have a positive effect on reading verbal numerals, but no effect on reading numerals in Arabic notation. Similarly, we also investigated whether error patterns between the two types of notation dissociate or not. In fact, improved reading of alphabetical stimuli could possibly change not only the error rate, but also the error type in reading verbal numerals.

Recently, it has been claimed that the production of spoken verbal numerals can dissociate in "semantic" and "asemantic" tasks (Cipolotti, 1995; Cipolotti & Butterworth, 1995; Cohen et al., 1994); this motivated the inclusion of these different types of tasks in the follow-up study. If these tasks rely specifically on asemantic and semantic routes, then differences in the performance in the two types of tasks are possible. For example, if asemantic tasks are processed exclusively along asemantic pathways and these pathways are disrupted, difficulties in asemantic tasks should appear (see Cipolotti & Butterworth, 1995). If, however, all tasks are processed along a single route (as proposed by McCloskey et al., 1985), then no processing dissociations between asemantic and semantic tasks should be found. In McCloskey's model all tasks have access to a common semantic component and, if they require a verbal answer, they are processed along the same output route to spoken verbal numerals. Only the repetition of verbal numerals and the reading of written verbal numerals would access asemantic pathways (which are, however, not specific for numerical stimuli, but are part
of the general language processing system). These asemantic pathways should operate on verbal numerals just as on any other verbal stimulus.

The proposed distinction between "semantic" and "asemantic" tasks is, however, difficult to test and theoretically questionable. First, there is no single "meaning" of a number, but different meanings depending on the context and the specific reference of the numeral (Fuson, 1988; Seron & Noel, 1995). A numeral can refer to a quantity, to a measure, to a date, to a bus number etc. It can also represent a cardinal value in calculation or part of a sequence in counting. Indeed, recent neuropsychological evidence suggests that different number meanings may be selectively lost after cerebral-lesion (Delazer & Butterworth, 1997). "Semantic tasks" therefore can assess completely different meanings for the same numeral. The number 24, for example, can represent Christmas eve, midnight, the bus from central London to Hampstead, the price of a book, the result of 3x8, the half of 48, the number after 23 or the collection of 24 units. It is possible that processing dissociations between tasks assessing different number meanings (i.e. within "semantic tasks") may be found. We presented three types of tasks tapping different aspects of number meaning (dot transcoding, calculation, encyclopaedic knowledge) and we compared the outcome of these "semantic tasks" with the performance in reading Arabic numerals.

First, production of spoken verbal numerals was tested in a 'dot transcoding task', which required the transcoding from dot patterns to spoken verbal numerals. Second, production of spoken verbal numerals was tested in calculation tasks. There is considerably discrepancy between authors as to whether calculation problems are processed along semantic or asemantic pathways. The McCloskey model (1985) postulates that all calculation tasks, both complex and easy ones, are performed on abstract semantic representations. This assumption is also implied in the Cipolotti and Butterworth's model (1995). On the other hand, Dehaene (1992) claimed that arithmetic facts (e.g., 3x4) are answered by means of verbal associations and therefore do not require access to semantic transcoding pathways; Yet, the computation of complex calculations would require semantic elaboration (Dehaene, 1992; Dehaene & Cohen,
In the present study we presented arithmetic fact problems as well as 2- and 3-digit addition tasks which necessarily require access to number semantics. Third, the production of spoken verbal numerals was assessed in tasks dealing with encyclopaedic and autobiographical knowledge. The different tasks allowed us to assess (a) the influence of encyclopaedic information on verbal number production, (b) the effect of meaning attributed by the reader (but not presented in the stimulus material) thus differentiating between the effects of semantic facilitation, visual familiarity and facilitation of verbal output and (c) the effect of priming encyclopaedic associations.

5.3.2 Background tasks

ZA attended speech therapy from 1992 to 1996. His verbal fluency in spontaneous speech improved during this period, as well as his reading abilities. He still showed a moderate deficit in naming (41 out of 60 items of the Boston naming test; Kaplan et al., 1983). However, he was able to produce long words and named correctly 20 out of 24 items in a compound naming task (confrontation naming of pictures representing compounds). His performance in an Arabic numerals comparison task was error free (1-to 6-digit numbers; 32/32 correct) and he could repeat 98/100 spoken verbal numerals correctly (1- to 6-digit number words). He answered 10/10 addition, 10/10 subtraction, 8/10 multiplication and 8/10 division problems in written form (presentation and answer in Arabic digits; all factors were single-digit numbers). These results suggest that ZA could correctly encode Arabic numerals and that his arithmetical facts knowledge was intact. Moreover, his performance in repetition of spoken numerals indicated improved speech output mechanisms.

5.3.3 Reading tasks

Type of stimuli and number of items included in the different tasks are specified in the Tables.
Reading words

340 written words (different grammatical classes, high and low frequency words; including all items presented in 1992) were presented one by one in a reading task.

Reading was perfect for nouns, abstract nouns, compounds and proper names. A few errors occurred in reading adjectives, verbs, function words (see Table 5.5). Reading accuracy improved significantly from 1992 to 1996 ($\chi^2(1) = 11.09; p<0.01$).

Reading errors were classified as in 1992. Errors consisted in 5 visual errors (legal Italian word sharing more than 50% of the letters with the target word) and 4 inflectional errors.

<table>
<thead>
<tr>
<th></th>
<th>High frequency</th>
<th>Low frequency</th>
<th>% correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>30/30</td>
<td>30/30</td>
<td>100%</td>
</tr>
<tr>
<td>Abstract nouns</td>
<td>30/30</td>
<td>30/30</td>
<td>100%</td>
</tr>
<tr>
<td>Compounds</td>
<td></td>
<td>40/40</td>
<td>100%</td>
</tr>
<tr>
<td>Adjectives</td>
<td>29/30</td>
<td>30/30</td>
<td>98.3%</td>
</tr>
<tr>
<td>Verbs</td>
<td>28/30</td>
<td>26/30</td>
<td>90%</td>
</tr>
<tr>
<td>Function words</td>
<td>15/15</td>
<td>13/15</td>
<td>93.3%</td>
</tr>
<tr>
<td>Proper names</td>
<td>30/30</td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 5.5 Percentage of correct answers in reading high and low frequency words in different grammatical classes in 1996.

Reading nonwords

Sixty nonwords (the same items as in 1992) were presented to be read aloud. ZA read 22 out of 60 nonwords correctly. Performance improved significantly from 1992 to 1996 (McNemar $\chi^2 = 14.42$, p<0.01). Errors were classified as in 1992: ZA produced 22 lexicalisations and 16 phonological paraphasias.

His performance was perfect in a lexical decision task (80/80).
Reading Arabic numerals

ZA failed to read 85 out of 337 Arabic numerals. The error rate did not differ between 1992 and 1996 ($\chi^2(1)=0.4$, n.s.). Analysing each number class separately, a significant improvement was found only for Ones ($\chi^2(1)=4.14; p<0.05$), but not for any of the others classes (Teens, Tens, 2-digit numbers, 3-digit numbers, Ones + multipliers).

Answers were classified as lexical, syntactic, mixed and omission (Table 5.6). Lexical errors appeared frequently for 2-digit numbers, while syntactic errors were found almost exclusively for 3- and 4-digit numbers. These syntactic errors often consisted in the segmentation of the stimulus (e.g. 736 was read as sette, tre, sei [seven, three, six]). Accuracy decreased with the complexity of the stimulus.

<table>
<thead>
<tr>
<th>Errors/n</th>
<th>% correct</th>
<th>Lex.</th>
<th>Synt.</th>
<th>Mixed</th>
<th>Omis.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONES</td>
<td>0/30</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEENS</td>
<td>5/30</td>
<td>83.3%</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TENS</td>
<td>2/21</td>
<td>90.4%</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-digit</td>
<td>44/154</td>
<td>71.4%</td>
<td>42</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3-digit*</td>
<td>24/64</td>
<td>62.5%</td>
<td>12</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>4digit**</td>
<td>8/20</td>
<td>60%</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Ones+mult</td>
<td>2/18</td>
<td>88.8%</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>85/337</td>
<td>74.7%</td>
<td>63</td>
<td>16</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5.6 Number of errors in reading Arabic numerals and distribution of errors across classes (lexical errors, syntactic errors, mixed errors and omissions). The third column gives the percentage correct answers. (*21 had an internal 0, 20 a 0 at the end; **6 had an internal 0 and 9 a 0 at the end).

Reading of verbal numerals

ZA failed to read 30 written verbal numerals out of 287 stimuli. No errors appeared for Ones and Teens and only one error occurred in Tens and in Ones+multiplier numbers. More errors were found in reading verbal numerals, corresponding to 2-digit and 3-digit numerals (see Table 5.7).
Overall, the performance improved significantly from 1992 ($\chi^2(1)= 8.53; p<0.01$). Lexical errors were the most frequent ones (n=27), only two syntactic and one mixed errors were observed (see Table 5.7).

<table>
<thead>
<tr>
<th>Errors/n</th>
<th>% correct</th>
<th>Lex.</th>
<th>Synt.</th>
<th>Mixed</th>
<th>Omis.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONES</td>
<td>0/20</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEENS</td>
<td>0/20</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TENS</td>
<td>1/13</td>
<td>92.3%</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-digit</td>
<td>11/82</td>
<td>86.6%</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-digit*</td>
<td>13/105</td>
<td>87.6%</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-digit**</td>
<td>4/30</td>
<td>86.7%</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Ones+mult</td>
<td>1/17</td>
<td>94.1%</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30/287</td>
<td>89.5%</td>
<td>27</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.7 Number of errors in reading written verbal numerals and distribution of answers across different classes (lexical errors, syntactic errors, mixed errors and omissions). The third column gives the percentage of correct answers. (*10 had an internal 0, 11 a 0 at the end; **12 had an internal 0 and 8 a 0 at the end)

5.3.3a Comments on the reading tasks

ZA’s ability to read alphabetical stimuli improved significantly from 1992 to 1996 (reading words; $\chi^2(1) = 22.6; p<0.01$), though he still showed the error pattern typical of phonological dyslexia. His reading of nonwords improved significantly, and this result indicates a better use of sublexical reading routines ($\chi^2(1)=14.41; p<0.01$). Similarly, ZA's performance in reading verbal numerals was significantly better in 1996 than in 1992 ($\chi^2(1)=8.53; p<0.01$), On the other hand, reading of Arabic numerals did not improve. The error rate in reading Arabic numerals and verbal numerals increased with the length of the stimuli, i.e. with the number of words to be produced (Spearman rank coefficient r=0.517; p=0.06). In both testing sessions, 1992 and 1996, reading of verbal
numerals was better than reading of Arabic numerals (1992: $\chi^2(1)=4.36; p<0.05$; 1996: $\chi^2(1)=19.36; p<0.01$).

Thus, ZA showed a dissociation between the reading of alphabetical notation and Arabic notation which emerged clearly in the second session of testing. While the reading of verbal numerals improved from 1992 to 1996, reading of Arabic numerals did not. We assume that reading of written verbal numerals improved along with the general reading performance. Possibly, the better functioning of the sublexical reading route was likely to give ZA additional information about the verbal numeral to be read.

The main error type observed in reading Arabic numerals and verbal numerals was the substitution of lexical elements. These errors indicate a problem in lexical output mechanisms, responsible for selecting the appropriate item within the different number classes.

Frequently, in the attempt to produce a spoken verbal numeral ZA used a step-wise approach leading to a correct or an incorrect result (e.g. 32 was read as venti, trenta, trentuno [twenty, thirty, thirty-one]). These strategies were more frequent for Arabic numerals than for written verbal numerals ($\chi^2(1) = 90.76; p<0.01$). Searching strategies in verbal numerals were not only less frequent than in reading Arabic numerals, they also differed in complexity, consisting of only two steps (e.g., trenta [thirty] was read as venti, trenta [twenty, thirty]).

5.3.5 Production of spoken numerals: calculation tasks

Verbal number production in response to calculations was tested in three tasks, two multiplication and one addition task. The first multiplication task consisted of 74 verbally presented multiplications (64 multiplications with operands between 2 and 9 plus 10 multiplications including 0 or 1 as an operand). The second multiplication task consisted of the same problems presented in visual modality. These two tasks allowed us to control for the effect of input modality on the production of spoken verbal
numerals in response to arithmetical fact. Each problem (in both tasks) was first answered verbally, and then in Arabic format. This procedure allowed us to differentiate between calculation errors (which would appear also in written form) and pure verbal production errors (which would appear in the verbal response, but not in the written answer) (This distinction is based on the fact that ZA had no difficulty in writing 1- to 3-digit numerals at the time of the 1996 testing; he was thus able to write the intended calculation result in Arabic numerals).

**Multiplications presented verbally, answered verbally and in Arabic numerals**

66 out of 74 problems were answered correctly in written form (error rate 10.8%). The 8 errors were also incorrect when answered verbally. Out of the 66 items correct in written form, 17 were answered incorrectly in verbal form.

11 lexical errors and 6 omissions were found in pure verbal errors (see Table 5.8).

**Multiplications presented visually, answered verbally and in Arabic numerals.**

Out of 74 problems 69 were answered correctly in written form (error rate 6.8%). The 5 errors were also answered incorrectly in verbal form. Out of the 69 items correct in written form, 22 were answered incorrectly in verbal form.

Pure verbal errors consisted of 17 lexical errors and 5 omissions (see Table 5.8).

<table>
<thead>
<tr>
<th>Verbal errors in verbally presented multiplications</th>
<th>Verbal errors in visually presented multiplications</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Lex</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>ONES</td>
<td>16</td>
</tr>
<tr>
<td>TEENS</td>
<td>17</td>
</tr>
<tr>
<td>TENS</td>
<td>6</td>
</tr>
<tr>
<td>2-digit</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
</tr>
</tbody>
</table>

*Table 5.8 Verbal errors to simple multiplications presented (a) verbally and (b) visually. Only items answered correctly in written form are taken into account. Lexical substitutions and omissions were found.*
In order to assess modality specific effects of input we compared the accuracy in the two multiplication tasks. Accuracy in verbally and visually presented multiplications did not differ significantly, neither for written answers (McNemar $\chi^2=0.72$), nor for verbal answers (taking into account only the items answered correctly in written form; McNemar $\chi^2=2.96$). These results indicate that modality of input did not affect significantly the performance.

A qualitative analysis of the errors was also performed. Only 7 operand errors (correct answers to problems sharing one operand with the target problem; e.g., 3x4=16) appeared and 2 table errors (correct answers to other multiplication problems; e.g., 4x4=25). Non-table errors (answer corresponding to numbers not included in the tables; e.g. 4x5=31) were more frequent (n=11), as well as close-miss errors (n=8; answers close in magnitude to the correct result; e.g., 6x6=37) and omissions (n=11). 26 errors were not classifiable, since ZA gave more than one answer (never the correct one).

The production of verbal numerals in both multiplication tasks was compared to the reading of the same numbers written in Arabic notation. The performance in reading Arabic numerals did not differ from the performance in answering multiplications (answering visually presented multiplications 68% correct, reading task 75% correct; answering verbally presented multiplications 74% correct, reading task 74% correct). There was neither a significant difference between reading and answering verbally presented multiplications (McNemar $\chi^2=0.05$), nor between reading and answering visually presented multiplications (McNemar $\chi^2=0.26$).

*Additions presented visually, answered verbally and in Arabic numerals.*

43 additions were presented in Arabic numerals and were answered first in verbal form and then in Arabic numerals. Answers consisted in Teens (n=9), Tens (n=5), 2-digit (n=21) and 3-digit (n=8) numbers.
Although all additions were answered correctly in written form (43/43), 12 errors were found in the verbal answers. Errors consisted in 11 lexical errors (1 Teen, 2 Tens, 6 2-digit numbers, 2 3-digit numbers) and one mixed error (3-digit number).

Numbers corresponding to the correct answers to the addition task were presented in Arabic format to be read aloud. No significant difference was found between verbal production in answering additions (72% correct) and reading Arabic numerals (79% correct; McNemar $\chi^2 = 1.14$).

5.3.5a Comments on the calculation tasks

ZA showed no facilitation of verbal output processes in calculation tasks, either in automatically retrieved multiplication or in complex addition which required elaboration via semantic processes.

We also assessed whether the frequency of searching strategies differed between verbal production in calculation and when reading Arabic numerals: searching strategies were more frequent in reading Arabic numerals ($\chi^2(1) = 17.75; p<0.05$).

Moreover, it should be noted that a short term memory deficit cannot be responsible for ZA verbal errors, since he was able to write the correct answers after having produced a verbal error.

5.3.6 Production of spoken numerals: ‘semantic’ tasks

Dot transcoding

An original task was developed for testing the production of verbal numerals in a transcoding task where the input format was neither Arabic nor verbal. ZA was asked to produce 52 spoken numerals in a dot transcoding task. Answers consisted of 1- to 4-word verbal numerals corresponding to 1- to 4-digit numerals (9 Ones, 10 Teens, 4 Tens, 16 2-digit numbers, 10 3-digit numbers and 3 Ones+multipliers). Dots of different
sizes and colours were used in order to represent the values 1, 5, 10, 50, 100 (e.g., a big blue dot would represent 100, a small green dot would represent 5). The instructions were, first, to name the sum of the dots and, second, to write the answer in Arabic format (the written result allowed us to control whether an erroneous verbal response was determined by a verbal production error or by an error in the computation itself). ZA wrote all results correctly, though he produced one lexical error in answering in verbal modality (for a 2-digit number).

The same set of target numbers used in the dot transcoding task were presented as stimuli in an Arabic numeral reading task; ZA failed to read 13 out of 52 stimuli. The performance differed significantly between the reading and the dot transcoding task (McNemar $\chi^2 = 11.09; p<0.01$).

**Influence of attributed meaning in reading and in semantic questions**

A set of 15 numbers was used in three different tasks. In the first task, ZA answered, both verbally and in written form, to general questions requiring a numerical answers; in a second task he read Arabic numerals aloud (with no particular instructions), in a third task he was asked to read the same Arabic numerals attributing a specific meaning to them. The set of numbers included 3 Teens, 3 2-digit numbers, 4 3-digit numbers and 5 4-digit numbers.

**Questions of numerical knowledge**

15 questions concerning encyclopaedic and autobiographical information (How many weeks in a year? When did the World War I end? When was your son born? etc.) were presented to ZA. He first answered verbally, then in Arabic numerals. In the verbal answer one lexical error occurred (the beginning of the World War I was answered with *mille nove cento undici* [one thousand nine hundred eleven] (In Italian dates are always given in the Thousands-Hundreds-Tens-Ones structure (*mille nove cento novanta sei*) and never in the Teens-Hundreds-Tens-Ones structure as in English (*nineteen hundred ninety nine*), French or German). All questions were answered correctly in written form.
Reading Arabic numerals (‘asemantic’ reading)

The same 15 numbers (corresponding to the correct answers to the general questions) were presented in a reading task (Arabic numerals) mixed with 15 distractors of matched length and syntactic structure.

ZA made 10 errors in reading the 15 “meaningful” numbers (6 lexical, 3 mixed, 1 syntactic). The error rate (66.6%) corresponded exactly to the error rate ZA had in reading the distractor items; i.e. ZA treated "meaningful" and "non meaningful" numbers in the same way.

Reading with meaning attributed by the patient.

The second reading task was performed 2 months after the Numerical knowledge task. 15 Arabic numerals (corresponding to the correct answers to the Numerical knowledge task) were used as stimuli. ZA was told that each number could be linked to a specific meaning and that he should try to associate a meaning to each stimulus before reading the numeral aloud. He was given the following example: “If you see the number 17 you could think of bad luck” (an association commonly known in Italy). He was not instructed to formulate his associations verbally, but sometimes he commented on the numbers aloud.

ZA attributed meanings to 14 out of 15 numerals and read those 14 numerals correctly. For one numeral (100) he claimed to have no semantic association and failed to read it (100 was read as mille.[thousand]).

Verbal production when answering numerical knowledge questions differed significantly from reading Arabic numerals in the asemantic reading task (14/15 versus 5/15, McNemar $\chi^2 = 5.33; p<0.05$), but not from reading Arabic numerals in the “semantic” reading task (14/15 versus 14/15, McNemar $\chi^2 = 0.5; n.s.$). On the other hand, the performance in the semantic reading task differed significantly from the performance in the asemantic reading task (McNemar $\chi^2 = 7.27; p<0.01$).
Semantic priming

ZA’s performance in answering questions on numerical knowledge and in the “semantic” reading task, suggested that the activation of number meaning can improve the retrieval of the verbal number form. The following task was designed to investigate whether the presence of a semantic prime would help to activate the correct verbal number form, without explicit instruction to focus on the meaning of numbers. In the priming task, words and Arabic numerals were presented on single stimulus cards in alternating order. The patient’s task was to read the presented words or Arabic numerals. The stimuli consisted of two identical sets of numerals (presented in Arabic format) and of two different sets of words. One set of Arabic numerals was preceded by words providing semantic cues, the other set was preceded by distractors without semantic relation. E.g. the target 7 would be preceded by the word nani [dwarfs] in the “semantic” condition and by the word luce [light] in the neutral condition. Semantic and neutral pairs were mixed in a pseudo-random order. Originally, each set consisted of 64 items (all together 128 numerals and 128 words). After the administration of a control task, in which ZA had to answer questions regarding the used prime-target relation in Arabic numerals (e.g., How many dwarfs appear in Snow-white story?), eight items were excluded from the analysis due to the mismatch between ZA’s answer and the number used in the task. Answers to 56 items (16 Ones, 11 Teens, 18 2-digit numerals and 11 3-and 4-digit numerals), presented once with a semantic prime and once with a neutral prime, contributed to the final analysis (112 numerals, 112 words).

ZA produced 4 errors in the semantic condition and 10 errors in the neutral condition. No difference was found for Ones, Teens and 2-digit numbers, where ZA’s performance was highly accurate in both semantic and neutral conditions. A significant difference, however, was found for longer stimuli (3-digit and 4-digit Arabic numerals) where semantic primes showed a facilitation effect (McNemar $\chi^2 = 6.85; p<0.01$). For example, the numeral 164 was read correctly when preceded by the semantic cue "Alfa Romeo", but not when preceded by the neutral cue "Barilla" (Italian brand of pasta).

Errors in the semantic condition were mixed (2), lexical (1) and syntactic (1). Errors in the neutral condition were mixed (5), lexical (1) and syntactic (4).
5.3.6a Comments on the semantic tasks

ZA performed significantly better in transcoding pattern of dots to spoken verbal numerals than in reading aloud Arabic numerals. The stimuli used in the dot transcoding task did not provide particular clues that could have facilitated verbal number production: Dots represented different numerical values according to their colour and size, and the number of dots included in a specific pattern always differed from the represented sum. Thus, the task did not consist in naming numerosities, but also included a calculation component (e.g. a pattern including one dot representing 50 and a second dot of value 20 would require the answer ‘seventy’).

Moreover, ZA showed a facilitation effect in the production of spoken verbal numerals when answering encyclopaedic and autobiographical knowledge questions.

ZA performed better in the “semantic” reading task (attributing meaning to numerals) than in the asemantic reading task. This result indicates that semantic facilitation does not necessarily depend on the presented context, but can also be achieved by the meaning a subject attributes to a numeral. The asemantic control task allowed us to better define the functional level of the facilitation effect. The facilitation appeared in the reading task where the patient had to imagine a number meaning, but not in the asemantic reading task, although in both tasks the same numerals were presented. The facilitation effect thus neither resulted from a visual familiarity effect (in the number input lexicon), nor from a verbal familiarity effect (in the verbal output lexicon), since visual input and verbal output were the same in the asemantic and in the semantic reading tasks. The difference in results found between the reading tasks strongly suggest that facilitation is linked to semantic information and not to input or output mechanisms.

In the semantic priming task no explicit associations between numerals and numerical meanings were asked for. In fact, ZA read all the items at the same rate, without any difference between semantic and neutral items. Nevertheless, since ZA showed a facilitation effect we assume that the effect does not depend on the active recollection of numerical knowledge (as in the questions task or in the “semantic” reading task), but can be achieved by presenting words semantically related to the numerical stimulus. Interestingly, ZA also reached a good accuracy level for 2-digit
Numerals when they were preceded by neutral primes. Possibly the task led him to use a semantic approach also when reading items preceded by neutral primes.

In all tasks searching strategies (e.g. *venti, trenta, trentuno* [twenty, thirty, thirty-one] for thirty-one) were observed, though these strategies were not always successful. However, in order to exclude the possibility that the semantic facilitation effect found in dot transcoding, answering questions, reading with attributed meaning and semantic priming stems just from a more frequent use of searching strategies as a result of an higher level of attention, we compared the number of searching strategies observed in these tasks with the number of searching strategies observed in the Arabic reading task. The frequency of searching strategies did not differ between the tasks. \( \chi^2(1)=0.09; \text{n.s.} \).

### 5.4 Discussion

The results of this study can be summarised as follows: In 1992, ZA's reading performance showed the typical pattern of phonological dyslexia. His reading difficulties extended to numerical stimuli in both alphabetical and Arabic notations. These problems were not attributable to faulty input mechanisms or comprehension problems (in fact, number comparison tasks were error free in all modalities). Difficulties in reading numerical stimuli were influenced by the type of notation, reading alphabetical notation being better preserved than reading Arabic notation. In both notations, lexical substitutions were the most frequent errors. In 1996, ZA was significantly better at reading words, nonwords and verbal numerals, while reading Arabic numerals did not improve significantly. Thus, we found a differential pattern of improvement for alphabetical and Arabic notation over the three years.

In 1996 ZA's production of spoken verbal numerals was tested in several tasks which were thought to rely on the activation of number meaning. The performance in these semantic tasks was compared to the reading of Arabic notation which, by some
authors, is assumed to be processed along asemantic pathways (Cipolotti, 1995; Cipolotti & Butterworth, 1995; Cohen & Dehaene, 1995; but see McCloskey et al., 1985). ZA performed significantly better in a task where he had to name the quantity denoted by a set of dots (representing different values) than in a reading task (Arabic numerals). On the other hand, no facilitation of verbal number production was found in the three calculation tasks. No difference was found between answers to verbally presented multiplications and answers to visually presented multiplications, i.e. the modality of input did not influence his performance. Importantly, the verbal number production in both multiplication tasks did not differ from the verbal number production in an Arabic numeral reading task. Similar results were found in an addition task, where the verbal production of 2- and 3-digit numerals was not better than in an Arabic numeral reading task. However, verbal number production was consistently better in tasks involving encyclopaedic knowledge than in asemantic reading tasks. ZA was able to answer questions concerning famous dates and autobiographical events. He was also quite accurate when reading Arabic numerals after being instructed to associate each numeral with a specific meaning before reading aloud. He performed significantly worse with the same numerals when no specific instruction was given (numerals were mixed with distractors). Finally, in a semantic priming task, we found a significantly better reading performance for Arabic numerals preceded by related words than for Arabic numerals preceded by unrelated words. This result may indicate that semantic facilitation is not necessarily bound to the active retrieval of number knowledge.

The following points were raised by the results of the study, each of which will be considered in turn: First, the dissociations between the processing of Arabic and alphabetic notation. Second, how single-route and multiple-route models of number processing can account for the results of the asemantic and semantic tasks used and in particular for the dissociations within semantic tasks. Finally why access to different number meanings may have a differential facilitation effect on verbal number production.
In 1992 and 1996 ZA showed a better performance when reading verbal numerals than when reading Arabic numerals and in both notations lexical errors were the most frequent. More importantly, he showed a significant improvement in reading verbal numerals from 1992 to 1996, but not in reading Arabic numerals. Though ZA's error rates differ between the two notations, we nevertheless assume that his difficulties in reading verbal numerals and Arabic numerals emerge from the same source. For both notations the dominant error type was the substitution of lexical elements in the verbal output. These substitutions cannot be explained by encoding or comprehension problems. We thus suggest that the problem lies in the activation of lexical output mechanisms which select the appropriate item within the different number classes (Ones, Teens, Tens etc.). All current models assume such lexical mechanisms (Cipolotti & Butterworth, 1995; Deloche & Seron, 1982a, b; McCloskey et al., 1985).

The differential improvement of alphabetic and Arabic notation seems to indicate a dissociation between the processing of the two notations, as previously claimed by several authors (Cipolotti, 1995; Anderson et al., 1990; Noel & Seron, 1993). However, one has to take into account that different factors may influence the reading of Arabic numerals and verbal numerals: The nature of the notation itself (ideographic or alphabetic; Cipolotti, 1993), the actual processing stages required for reading the two notations and the compensation mechanisms that the notations offer.

Verbal numerals can be read by matching graphemic input (lexical units or also single graphemes) to spoken output (lexical units or single phonemes), just as other words can. Each complex verbal numeral can be separated into lexical units which can be read sequentially (e.g. hundred-twenty-six can be read as "hundred", "twenty", "six"). Reading Arabic numerals, on the other hand, generally requires the activation of syntactic rules both in comprehension and production which allow one to generate the spoken output (126 has to be transformed into "hundred", "twenty", "six" taking into account the power of ten represented by the single digits). A direct mapping from graphemic input to spoken output (via lexical units or via grapheme-phoneme conversion rules) is not possible. Dissociations between Arabic and alphabetic notation may therefore arise from deficits to processing components which are essential for the
reading of one notation, but not for the other (e.g., a syntactic deficit will disrupt reading of Arabic notation, but not reading of alphabetical notation). Dissociations may also arise from the compensation possibilities one notation, but not the other, offers (e.g., using a Grapheme-Phoneme route in reading alphabetical notation). ZA's improvement in reading verbal numerals is thus explicable in terms of a general improvement of alphabetical reading (as shown in the reading of words) and by an improved functioning of sublexical reading routines which may support the reading of verbal numerals, but not the reading of Arabic numerals (each single digit maps to a complete lexical unit).

We found heterogeneous results in tasks assumed to involve semantic elaboration of numerals. While assessing a quantity in the dot transcoding task facilitated verbal output as compared to the verbal production in reading Arabic numerals, calculation tasks did not. All tasks involving encyclopaedic knowledge, on the other hand, showed a clear facilitation effect. We will consider in turn how these results can be accounted for by current models of verbal number production.

McCloskey et al. (1985) assume a central semantic system which is accessed in all transcoding tasks and in all tasks dealing with numerical input. Since ZA had no difficulties in comprehending numerical input nor in retrieving correct answers in various semantic tasks we can assume that he was able to retrieve the correct abstract semantic representation of the target numbers and that the difficulties emerged in the transcoding from abstract semantic representations to spoken verbal numerals. The McCloskey model predicts that, once the correct semantic representation has been accessed, verbal number production should not vary between tasks since all tasks share the same pathway from semantic representation to spoken verbal numerals. However, the dissociations found in ZA's verbal number production are not consistent with this prediction.

Cipolotti's (1995; Cipolotti & Butterworth, 1995) multiple route model assumes that different asemantic and semantic pathways underlie number transcoding and that these different routes are accessed depending on task demands. While ZA's reading performance would be compatible with the assumption of a damaged asemantic route, it is, however, difficult to interpret ZA's performance in the semantic tasks within the
Cipolotti and Butterworth (1995) assume one central semantic system that is accessed in calculation, in dot counting and in tasks tapping encyclopaedic knowledge. The output of the semantic system (results of the calculation tasks, answers to question etc.) is processed along one code specific route into spoken verbal numerals. The dissociation between better preserved verbal number production in tasks concerning encyclopaedic knowledge and impaired verbal number production in calculation tasks thus argues against a uniform processing along a common semantic route.

Cohen et al. (1994; Cohen & Dehaene, 1995; Dehaene & Cohen, 1995) proposed a multi-routes model that differs from the Cipolotti and Butterworth model in some respects, most importantly by assuming different semantic routes. One lexical semantic route would access encyclopaedic knowledge from the visual input lexicon and, in turn, would activate the phonological output lexicon (Cohen et al., 1994). Non-lexical semantic routes would access magnitude representations of numbers from the visual Arabic number form and would again activate the phonological output lexicon. In addition to these semantic routes, a direct asemantic route would allow one to transform Arabic numerals into spoken verbal numerals.

ZA showed a consistent facilitation of verbal number production in all tasks involving autobiographical or encyclopaedic knowledge of numbers (semantic reading, numerical questions, semantic priming). Cohen et al., (1994) reported a similar facilitation effect for a patient who performed significantly better with familiar numerals (e.g., famous dates) than with unfamiliar ones and proposed a specific pathway accessing encyclopaedic knowledge from the visual input lexicon. The performance of ZA, however, indicates that the facilitation effect emerges from the semantic level and not from the input lexicon. In the semantic reading task ZA was able to attribute meaning to Arabic numerals and in this way considerably improved his verbal production as compared to reading the same numbers without specific instruction. The familiarity of the visual input form and the verbal output form did not influence the performance critically, since they were identical in the semantic and asemantic reading. The "semantic" instruction, however, did change the performance significantly.
ZA showed also good verbal number production in the dot transcoding task. One may hypothesise that ZA's good verbal production in assessing the quantity of dot patterns was due to access to the magnitude representation as specified in the Dehaene and Cohen model (1995). However, we have two reservations concerning this hypothesis. Firstly, ZA had to give the exact value represented by the dots, while the magnitude representation in their model works approximately. Secondly, the task also included a calculation component (e.g., one dot representing 50 and another dot representing 10 which had to be summed) which cannot be answered by reference to magnitude representations only.

Finally, ZA's calculation performance is difficult to interpret within the Cohen and Dehaene's model. They assume that answering simple calculations (arithmetic facts) does not require any access to semantics, but would rely on access to stored verbal associations (Dehaene, 1992; Dehaene & Cohen, 1995). More complex problems, however, would require semantic elaboration (Dehaene & Cohen, 1995). ZA's multiplication performance was neither facilitated by a verbal input form, nor were the verbal answers more accurate than the written ones. He usually succeeded to write the correct answer to a multiplication problem, but failed to produce the correct verbal answer. This finding is not compatible with an exclusively phonological storage of arithmetic facts, as proposed by Dehaene (1992). However, as Dehaene and Cohen (1995) note, the model can account for better written than verbal performance when the verbal storage of arithmetic facts is located at a pre-phonological lemma level and verbal errors occur due to failures of lexical access. ZA's production of verbal numerals was equally impaired in response to additions as in response to multiplications and in asemantic reading of Arabic numerals. Thus, two calculation tasks, one thought to be asemantic, one thought to be semantic (Dehaene & Cohen, 1995) lead to the same difficulties in verbal number production.

Overall, ZA's results confirm the assumption of recent models that numbers are processed differently according to the task demands (Cipolotti, 1995; Cipolotti & Butterworth, 1995; Cohen & Dehaene, 1995; Dehaene & Cohen, 1995). His verbal
number production was better in most semantic tasks than in asemantic reading tasks. However, ZA's performance indicates that the semantic/asemic distinction proposed in current models of number processing is not sufficiently specified. ZA's verbal number production was facilitated in tasks involving encyclopaedic knowledge and in a task where quantities had to be assessed, but not in calculation tasks. All three types of task require access to number meaning and can thus be considered as "semantic". However, the meanings involved in the tasks are clearly different in nature and thus possibly have a differential semantic facilitation effect - the dot transcoding task requires to assess a quantity, the calculation task requires the manipulation of cardinal values and the encyclopaedic knowledge task requires the retrieval of dates, measures or numerical labels. A possible explanation for the better performance in the encyclopaedic knowledge task might be that encyclopaedic number knowledge is always linked to general semantic knowledge - numbers here refer to non-numerical information, as, for example, to a specific year or a type of car. Answers to calculation, on the other side, are more abstract and are not linked to general semantic knowledge. Dot transcoding requires the assessment of a quantity and may be facilitated by the visually presented dot pattern.

In conclusion, the dissociations between the three groups of semantic tasks indicate that different types of semantic information have a differential facilitation effect on the verbal output mechanisms which select the appropriate verbal numerals.

None of the proposed model completely explains ZA performance. However, our results argue against an exclusively asemantic transcoding model (e.g., Deloche & Seron, 1982a,b) and are compatible with dual-route models (Cipolotti & Butterworth, 1995; Dehaene & Cohen, 1995). Our results clearly show that semantics influences verbal number production.
5.5 SUMMARY AND CONCLUSIONS

The case-study reported in the present Chapter investigated the effect of different notations (Arabic/alphabetical) and of different tasks on the production of spoken verbal numerals in an aphasie and phonological dyslexic patient.

The qualitative analysis of his reading performance indicated that his functional deficit lied in the activation of the lexical output mechanism. Over a period of three years, ZA showed a significant improvement in the reading of alphabetical stimuli (verbal numerals, words, nonwords), but not of Arabic numerals, in line with other evidence of a processing dissociation between Arabic and verbal notations. Yet, we suggested that the differential processing stages and compensation mechanisms offered by the two scripts may be critical factors in determining the observed dissociation.

Following the recent proposal that number processing may dissociate between semantic and asemantic tasks (Cipolotti & Butterworth, 1995; Cohen et al., 1994; Cohen & Dehaene, 1995), ZA's verbal number production was assessed in a variety of tasks assumed to rely on different aspects of number semantics. ZA's verbal number production was significantly better in semantic tasks that involved encyclopaedic knowledge of numbers and the transcoding of quantities than in asemantic reading tasks. On the other hand, the verbal number production in response to calculations (both arithmetic facts and complex calculations) did not differ from the performance in reading tasks. Thus, the results clearly point to the importance of distinguishing between different types of number meanings.

Overall, these results provide evidence in favour of the existence of both asemantic and semantic pathways mediating number transcoding. Moreover, the study contributes to elucidate the mechanisms underlying semantic and asemantic transcoding by evaluating to what extent semantic effects may be determined by visual familiarity and/or verbal output facilitation rather than to the semantic information itself, and by exploring the role the differential facilitation effect determined by intentional and unintentional semantic activation.
**CHAPTER 6**

**GENERAL DISCUSSION**

The underlying issue of the studies reported in this thesis is the role of semantic information in number processing.

The term semantic information is conventionally used to refer to the magnitude represented by a numerical symbol, although this is by no means an exhaustive definition of number semantics. For example, we do not simply know that 9 is larger than 8, but also that it is the power of 3, an odd number and the last number within the ones. These multiple aspects of number semantics are likely to exert some influence at different stages of number processing, even though none of the current models explicitly address this issue. Numbers may also convey non-numerical meaning, like when they are used in labelling context, such as to refer to a specific model of car (e.g., Peugeot 506), to a bus line or to a date (e.g., 1945). This corpus of information includes those numbers that have become 'meaningful' for being associated to specific semantic information (e.g., the number 1945 is the end of the World War II and is thus associated to this event). The issue of the multiple meanings a number may convey depending on the context in which it is presented has been only recently addressed by neuropsychologists (Cohen et al., 1994; Delazer & Butterworth, 1998), even though it is certainly not new to the developmental literature (Fuson, 1988, 1992).

The issue of access to number meaning is related to the debate of whether number transcoding may be only mediated by semantic processing or also by asemantic pathways. According to the first hypothesis, transcoding from one numerical format to another would always require the intermediate activation of a semantic representation (McCloskey, 1992); alternatively, it has been proposed that transcoding processes may be
also accomplished by means of additional asemantic pathways (Cipolotti & Butterworth, 1995; Dehaene, 1992; Seron & Deloche, 1982a,b).

Most experimental studies concerned with number semantics, specifically address the issue of the magnitude represented by the numerical symbols (Arabic digits and verbal numerals); similarly, the experiments reported in Chapter 3 and 4 focused on the impact of magnitude information, i.e., to the specific numerical value represented by the written symbol, in number processing. On the other hand, the role and effect that different numerical meanings have on number processing were investigated in the case-study reported in Chapter 5. Furthermore, this neuropsychological investigation also contributed to the semantic/asemantic debate, by evaluating the effect of different tasks on numerical transcoding.

In the following sections, the issues addressed in the present thesis are reviewed to provide a critical evaluation of the results and of their theoretical implications. Finally, the limitations of the thesis are discussed and possible implications for future research are drawn.

6.1 ON THE AUTONOMY OF NUMBER SEMANTIC PROCESSING

Though the magnitude is only one of the semantic aspects of a number, it is certainly the most salient and distinctive attribute of it. This claim is further supported by the evidence indicating that number recognition may hardly occur without accessing magnitude information. In Chapter 2, the experimental evidence in favour of the hypothesis of an autonomous access to magnitude information was reviewed and attention was paid to identify advantages and disadvantages of different experimental paradigm used to explore this issue.

In the present work, two different paradigms were used to investigate the extent to which number magnitude is autonomously processed: a number-Stroop paradigm and a same-different judgement paradigm. The requirements of the tasks are clearly different and this factor has to be taken into account when discussing the overall results. The
number-Stroop paradigm includes two tasks, a numerical size and a physical size comparison. In the latter, which is the critical task for the evaluation of the autonomous access to magnitude information, subjects are required to select the physically larger of two displayed digits. The numerical dimension is irrelevant to the task, i.e., it is not needed for an accurate performance, nevertheless it is semantically related to the relevant dimension (small/large relations are equally relevant to numerical and physical comparisons); moreover, it conveys a meaning that conflicts with the meaning of the attended dimension (similarly to the standard word-colour Stroop task). On the other hand, in the same-different judgement tasks, pairs of numbers have to be matched for numerical or physical identity. In the former, the numerical dimension is still relevant to the decision process, yet is not in conflict with the required judgement (close-far and same-different are not contrasting relations); in the physical identity task, the role of magnitude information is critically reduced and the task provides the most stringent test for autonomous semantic access.

Experiment 1 successfully replicated the standard results reported in the number-Stroop literature. As previously pointed out, only three other studies adopted both numerical and physical sizes as comparative dimensions (e.g., Henik & Tzelgov, 1985; Peereman & Holender, 1984; Tzelgov et al., 1992) and the present experiments provide further support to their conclusions. In both tasks, the irrelevant dimension had an effect on the performance, indicating that both physical and numerical sizes were autonomously activated during the tasks. However, when numbers had to be compared along their numerical size, variation in the physical dimension produced both interference and facilitation effects, i.e., the performance was slowed down in the incongruent condition and speeded up in the congruent one. On the other hand, when the comparison was made along the physical dimension, the irrelevant numerical size only yielded interference effect. The lack of facilitation effect in the physical comparison is rather unsurprising if we consider that the control condition yielded already very fast responses, thus impeding any further improvement. Yet, the significant interference effect clearly indicated that Arabic numerals were semantically processed: their numerical value was accessed and
interfered with the physical comparison. Interestingly, this effect was significant whether or not subjects were previously requested to attend to the numerical dimension. This result provides support to the hypothesis according to which numerical information is autonomously activated even if irrelevant to the task. Moreover, the size of the interference was modulated by numerical distance: the effect was maximised when a small and a large number were contrasted (e.g., 3 7) compared to pairs including numbers close in magnitude, i.e., both belonging to the small (1 to 4) or large (6 to 9) categories (e.g., 3 4, 7 8). This result is in line with Tzelgov et al.'s (1992) suggestion that, when autonomously activated, numerical information is less refined than in conditions of intentional activation. In the former, a dichotomous small/large classification, based on the overlearned association between the 1 to 9 digits and their relative sizes derived by a primary encoding stage, would dominate the process. The activation of a more precise representation of the number magnitude would be also initiated, but it requires longer to be completed. Within this framework, the incongruity between numerical and physical dimensions would be more neatly perceived when the numbers have a different initial encoding, i.e., when a small and a large number are contrasted. However, alternative explanations to the larger congruity effect in far pairs are equally plausible. It is conceivable that in autonomous conditions, the semantic activation may be less strong or proceed slower than in conditions of intentional processing. In this case, the discrimination between numerically close numbers would be further penalised, preventing the numerical information from being available in time to interfere with the response. Moreover, the differentiation between consecutive numbers is always more time-demanding (i.e., distance effect) than that of numerically distant numbers, and it is possible that in close pairs the comparison of the physical size is performed before the numerical discrimination is completed.

The hypothesis of an autonomous activation of magnitude information was further investigated in Experiments 4, 5 and 6, where subjects were requested to perform same-different judgements for pairs of numerals written in either Arabic or verbal notations (e.g., 1 2, ONE TWO, ONE 2). The sameness rule was the magnitude of the numbers
in the numerical matching task and their physical identity in the physical matching task. In both cases, numerically close and far pairs had to be answered as different and a significant distance effect was considered index of semantic processing. This procedure was first adopted by Dehaene and Akhavein (1995), and their results suggested the autonomous activation of magnitude information, regardless of the task demands and of the stimuli format. In the numerical matching task, a significant distance effect was found in all conditions of notation. More critically, Arabic and verbal pairs yielded this effect also in the physical matching task, when the participants' attention was clearly addressed to the physical feature of the stimuli and semantic processing had no facilitation effect on the performance (e.g., 2 TWO had to be answered as different). The only condition where no distance effect was observed was when the stimuli pairs were visually highly dissimilar, i.e. in mixed pairs (e.g., 3 TWO), and the physical match quickly performed.

In the present study, the strength of Dehaene and Akhavein’s conclusions were called into question by evaluating more carefully the selection of the adopted stimuli set. In particular, the inclusion of only extreme-values, corresponding to the smallest (i.e., 1, 2) and largest (i.e., 8, 9) within the range of single-digit numbers, was assumed to have maximised the probability of observing semantic effects. In line with the suggestion that end-terms are processed faster than middle-terms (e.g., Banks, 1977; Moyer & Dumais, 1978), and that these numbers yield the strongest associations to the small and large categories, we hypothesised that the access to semantic information would be faster for this subset of stimuli compared to any other subset. This hypothesis was corroborated by the findings reported in Experiments 4N and 4P, where the above experimental procedure was adopted with a set of stimuli that included a larger range of numbers, minimised the impact of end-terms in defining close and far pairs and controlled for visual similarity within pairs across the numerical conditions. Indeed, in Dehaene and Akhavein’s experiments (Experiments 1 and 2) verbal numerals were not balanced for visual similarity, i.e., close pairs were visually more similar than far pairs, preventing to attribute the distance effect to the numerical rather than to the visual similarity between the stimuli. When these factors were controlled, a different pattern of results emerged. In the numerical matching task (Experiment 4N), a significant distance effect was still observed,
and the size of the effect was smaller, as expected by the reduced numerical distance between close and far pairs. The effect did not differ whether numerical stimuli were two Arabic, two verbal or an Arabic and a verbal numeral. The absence of format-effects in accessing magnitude information is in line with other evidence pointing to the minimal impact of the input format in number processing (Dehaene et al., 1993; LeFevre et al., 1988; Noel & Seron, 1992; Sokol et al., 1991). Yet, when the instructions emphasised the importance of the physical similarity between the stimuli, no distance effect emerged either in Arabic or verbal pairs. Subjects were now successfully disregarding the numerical meaning of the stimuli and performed the matching task on the basis of a mere visual analysis. Thus, when the numerical distance between close and far pairs was reduced, and, more critically, no end-terms were included in the far pairs, the semantic access was possibly initiated but not completed in time for interfering with the response. The critical role of the stimuli range in determining these different patterns of results was further tested by replicating both experiments with the same material adopted by Dehaene and Akhavein (Experiments 5N and 5P), as well as with a new set of stimuli selected according to the same criteria as in Experiments 4 (Experiments 6N and 6P). The major findings were replicated: when the experimental pairs included the numbers 1, 2, 8 and 9, a significant distance effect was observed in all conditions of notation in the numerical matching task, and in both Arabic and verbal pairs in the physical matching task. However, when the numbers 2, 3, 6 and 7 were used as stimuli, the distance effect was limited to the numerical matching task (Experiment 6N).

Overall, these results undermine the legitimacy of Dehaene and Akhavein's conclusions. In the physical matching task, which constitutes the most stringent test for the autonomy of numerical semantic processing, the occurrence of semantic effects was strongly dependent on the specific set of stimuli used. The only set of numbers that yielded a significant distance effect included the four extreme-values (e.g., 1, 2, 8, 9) within the single digit numbers. We accounted for this result by postulating that these numbers constitute a rather special subset: they are more distinctive than middle terms and unambiguously associated to small or large codes. Thus, even in autonomous
conditions, they are more likely to quickly activate numerical information, possibly in the form of an unrefined small/large categorisation (Tzelgov et al., 1992). Based on the results from this limited stimuli set, Dehaene and Akhavein have drawn the general conclusion that access to number meaning is autonomously activated and not dependent upon task demands. However, our findings indicate that these conclusions may not be generalised. Whether the results from the numerical matching tasks indicate that magnitude information is a core feature in number processing, the outcomes of the physical matching tasks suggest that numerical semantic access may be modulated by factors such as instructions and/or attentional demands and the specific range of numbers presented.

On the basis of these results we favour the hypothesis that semantic access in number processing is "partially" rather than completely autonomous (Zbrodoff & Logan, 1986). Autonomous processes are defined as processes that a) begin without intention, and b) can run on to completion without intention, despite effort to inhibit it (Zbrodoff & Logan, 1986). However, it has been suggested that a process may be more or less autonomous, satisfying only one of the two criteria or both, to a different extent. In the context of numerical processing, semantic access has been so far defined as automatic (e.g., Dehaene, 1992, Sudevan & Taylor, 1987) and autonomous (e.g., Dehaene & Akhavein, 1995; Tzelgov et al., 1992) and the absolute nature of these properties has not been questioned. Overall, the evidence suggest that a magnitude-related representation plays a central role in a variety of numerical tasks and that is autonomously accessed even when irrelevant to the task. Yet, the present findings favour the hypothesis that numerical semantic access must be considered only partially autonomous, being modulated by factors such as stimuli selection and task demands.

6.2 On the rise of automaticity in accessing number semantics

Even though the reported findings do not allow us to consider the access to number meaning as a completely autonomous process, the evidence suggests that magnitude is a
salient and distinctive attribute of a number, whether or not relevant to the specific context and to the task demands. The results from the physical number-Stroop task (Experiment 1) and from the numerical matching tasks (Experiments 5, 7, 9) are consistent with the hypothesis that numerical semantic access may be initiated and completed in an autonomous fashion. One of the aims of the present work was to trace the developmental trend of this phenomenon. It might be suggested that, by the time children attribute to Arabic numerals a symbolic value and they successfully operate upon these symbols, they may exhibit the ubiquitous primacy of the magnitude information, similarly to adults. This issue has so far received very little attention, though the few existing contributions seem to favour the hypothesis that autonomous access to number meaning is quite an early achievement (e.g., Duncan & McFarland, 1980; LeFevre et al., 1991; Lemaire et al., 1994). In the present work, this issue has been explored by presenting children of different ages with the same experimental paradigms in which adult’s performance proved the autonomy of numerical semantic access.

In Chapter 3, a number-Stroop paradigm has been used to test both pre-school and primary-school children. This paradigm, being sensitive to the effect of the irrelevant numerical information, was considered a useful and novel measure for capturing developmental changes in the processing of Arabic numerals. Children were presented with pairs of Arabic numerals varying along both the numerical and the physical dimension. In the numerical comparison, they had to select the numerically larger digit, ignoring variation in the physical size; in the physical comparison they had to select the physically larger digit, ignoring variation in the numerical size. Thus, developmental changes in the performance were assumed to result from general factors, such as increasing efficiency in selective attention mechanisms and in the processing speed, and more importantly, from age-related differences in number processing. We hypothesised that two factors could critically modulate the pattern of facilitation and interference in the tasks: first, the speed and efficiency in processing the numerical and physical dimensions, and second, the degree of integrality of these two dimensions. In particular, if the impact of the irrelevant dimension on the performance is function of its degree of automaticity (e.g., MacLeod & Dunbar, 1988), we expected the pattern of facilitation and interference
to differ across tasks and ages. The physical size, being a perceptual dimension, is likely to exert a great influence on the number comparison task since a very early age; yet, this effect may be further modulated by the fact that physical and numerical sizes are still separable dimensions for younger children given their rather limited familiarity with Arabic digits in numerical context. On the other hand, the effect of the numerical size on the physical comparison will possibly not emerge until the association between the Arabic numeral and the meaning it conveys (i.e., the numerosity it represents) is fully established. This is considered a pre-requisite condition for the autonomous activation of the symbolic dimension and it is assumed to develop gradually over the course of skill acquisition.

The study carried out with pre-school children (Experiment 2) raises several interesting points of discussion. First, only a minority (31%) of the 4 and 5 year-olds recruited for the study were able, not simply to recite the number sequence, but also to name the digits, to associate the Arabic symbols to the corresponding numerosity and to compare them. Though the screening procedure only aimed to select those children whose numerical competence was adequate for participating in the experimental study and does not allow us to draw any general conclusions, this informal observation is in line with the hypothesis that knowledge of the verbal number sequence precedes the recognition of Arabic numerals (Byalistock, 1992) as well as cardinality (Gelman & Gallistel, 1978, Fuson, 1988), and that the association between the Arabic numerals and their symbolic values is acquired gradually as learning progresses (e.g., Hughes, 1986).

Overall, the outcomes of Experiment 2 are quite straightforward. As expected, the numerical comparison was a quite difficult task, especially for younger children, and the irrelevant variation in the physical size of the stimuli had a great impact on the performance. Children benefited from the congruity between numerical and physical dimensions giving their answers significantly faster than in other conditions; on the other hand, the mismatch of the two dimensions increased their error rates dramatically. In younger children, the size of these effects was so remarkable so as to suggest that they were disregarding the numerical values performing the task on the basis of the perceptual features of the stimuli. This hypothesis was ruled out by the significant distance effect
showed by all children: errors were more likely to occur in numerically close pairs (e.g., 2–3) than in numerically far pairs (e.g., 2–7). This effect clearly indicates that children did not ignore the numerical dimension, but the physical size of the stimuli had a substantial influence on their decision. The pattern was similar in 4 and 5 year-olds, though the size of the interference tended to decrease in the older group. These results are consistent with the hypothesis according to which, as the child grows older, the efficiency of selective attention mechanisms increases (e.g., Dempster, 1992) and, at the same time, the degree of interference that physical information exerts over symbolic information gradually diminishes (e.g., Bisanz et al., 1979). Moreover, errors were more frequent in the comparison of numerically close pairs than in numerically far pairs. This result supports the hypothesis that children compared Arabic numerals by accessing an internal number representation similar to the adults’ one (e.g., Resnik, 1983). However, the informal observation of children’s verbal comments and the occasional use of counting strategies indicated that for the youngest children (4 year-old) this representation was still developing.

Children’s performance in the physical comparison was relatively fast and almost error free. More importantly, the congruity between physical and numerical sizes had no effect on their responses; children attended to the physical dimension of the stimuli undisturbed by the meaning of the numerals. Thus, even though by the age of 5, children are able to compare numbers along their magnitude, this is still a discernible and potentially negligible dimension of an Arabic numeral.

The administration of the number-Stroop paradigm to primary-school children of 6, 8, and 10 years of age, aimed to trace the developmental changes in the processing of Arabic numerals that would gradually lead to the autonomous activation of their symbolic value. Interestingly, our results disclosed that developmental changes in these tasks were quantitative as well as qualitative. In the numerical comparison task, both facilitation and interference effects were reported by all groups; however, once the overall changes in the processing speed were taken into account, an analysis of the effect sizes indicated that the advantage for the congruent pairs was steady across ages, while the interference was more pronounced in older children. This result is in line with the hypothesis that the
amount of interference in Stroop-like paradigms is mostly determined by the 'integrality' of the conflicting dimensions, i.e., the extent to which the two dimensions may not be selectively attended (Gamer et al., 1982; Algom et al., 1996). If the association between an Arabic numeral and its symbolic value develops over the course of learning, we may assume that its physical and numerical sizes are more separable dimensions for younger than for older children. Following this prediction, the larger interference effect exhibited by older children is a rather unsurprising result. However, this interpretation may not account for the pattern of results exhibited by pre-school children. Preschoolers' performance in numerical comparison reflected the absolute primacy of the physical information over symbolic information, possibly enhanced by their limited proficiency in operating upon numerical representations. Yet, after only one year of schooling, children showed a rather different response profile in the processing of Arabic numerals. Though the physical size exerted an influence on the comparison of numerical size, the relation between these two dimensions was now modulated by children's increasing practice with Arabic numerals as purely numerical symbols and by the gradual acquisition of numerical skills. Thus, the developmental changes observed in the numerical comparison task were clearly the result of the combination of several factors; the rather different profile observed in pre-school and primary-school children seems to suggest that these modifications are not merely function of their chronological age, but of the children's growing experience with the Arabic numerals: in this respect, the introduction to formal education constitutes certainly a critical step.

Children's performance in the physical comparison task is in line with the above interpretations. The size congruity effect, determined by the mismatch between physical and semantic information, was totally absent in first-graders, emerged in the third-graders and was highly significant in fifth-graders. Moreover, while in third-grade children the sensitivity to irrelevant numerical information was enhanced by having previously performed the numerical comparison, in fifth-grade children the activation of numerical size in physical comparison was a more robust phenomenon, not being modulated by the order of the tasks. As observed in adult's performance, the irrelevant numerical size yielded interference but no facilitation effect.
These findings indicate that access to number meaning acquires autonomous properties over the course of skills acquisition. Children's ability to manipulate and operate upon Arabic numerals gradually develops over the school years, but it is only after extensive experience with the arithmetical symbols that the magnitude represented by a digit becomes such a distinctive and salient dimension so as to exert an influence on whatever the task requirement.

The results reported in Experiment 7, further corroborate these conclusions. Third and fifth-grade children were requested to judge whether or not two numerals written in the same (pure trials, e.g., 2 3, TWO THREE) or different (mixed trials, e.g., 2 THREE) notations represented the same numerosity. Though in pure trials a simple visual matching may be sufficient to take a decision, adults showed an inability to disregard semantic information (Experiments 4N, 5N and 6N). As previously discussed, evidence for semantic processing comes from the longer reaction times for answering different to numerically close pairs (e.g., 2 3) than to numerically far pairs (e.g., 2 6). However, children's performance revealed a rather different pattern of results. While both groups showed a distance effect in mixed pairs, where a decision may not be possibly taken on the basis of the visual characteristics of the stimuli, Arabic and verbal pairs were differentially processed by third- and fifth-graders. Younger children did not show any distance effect, clearly they were simply relying on physical cues in order to respond. On the other hand, fifth-grade children tended to answer slower to close compared to far pairs; this result indicated the increasing importance that magnitude information acquires over the course of learning. Whether the primacy of the perceptual information in third-grader's performance does not contrast with the evidence emerged from the number-Stroop task might be questioned. In the physical size comparison task, third-grade children were indeed affected by the irrelevant numerical information when performing a physical comparison. These results may be accounted for by considering the different extent to which numerical information is irrelevant to the tasks: as already discussed, in the Stroop-paradigm the numerical dimension conveys meaning that conflicts with the meaning of the attended dimension and, for this reason, it might constitute a more powerful source of interference.
The present findings differ from the ones reported in Duncan and McFarland's study (1980). In an analogous same-different judgement of Arabic numerals, they observed a distance effect in the performance of children as young as 6 years old. However, several methodological differences suggest caution in directly comparing the effects across the studies. Among others, one factor seems to have played a critical role in leading Duncan and McFarland's subjects to favour a semantic processing of the stimuli. Half of the trials were presented in degraded condition, i.e., a diagonal line grid was superimposed over the digit pairs; this manipulation interfering with the stimuli encoding, possibly prevented the children relying on visual cues to match the digits.

To summarise, our results suggest that access to number magnitude becomes autonomous over the course of skill acquisition. Though preschoolers may well recognise Arabic numerals and compare their magnitude, the association between the written symbols and their numerical value is still developing and they can easily ignore it when irrelevant to the task. Thus, the ability to accurately compare numbers magnitude is clearly a pre-requisite but does not entail the obligatory activation of their meaning. It is over the school years that children gradually learn to manipulate and operate upon Arabic numerals, as they progress towards reasoning about numbers as abstract entities (Resnick, 1992). Yet, it is only after several years of practice with formal arithmetic that the magnitude becomes the core dimension of a numeral so to determine its autonomous activation.

These conclusions are consistent with the theory according to which automatization is achieved gradually as the acquisition of a domain-specific knowledge progresses (Logan, 1988). Within the numerical domain, the assumption that practice plays an important role in producing automaticity has received support from studies on children’ mental arithmetic. In particular, it has been shown that, as the associations between a number pair and its sum or product become stronger over the course of learning, their activation may be initiated autonomously producing interference in a variety of tasks (e.g., LeFevre et al., 1991; Lemaire et al., 1994). The evidence reported in the present study suggests that even an apparently simple achievement as the mapping between an
Arab numeral and the numerosity it represents, requires several years of practice to acquire autonomous properties.

### 6.3 On the Levels of Representation in Number Processing

A further issue addressed in the experimental investigation reported in Chapter 4, was whether the mapping between an Arabic numeral and a verbal numeral can occur with no reference to their meaning. This hypothesis was tested by evaluating if subjects could map an Arabic numeral (e.g., 2) onto its corresponding verbal form (e.g., two) without semantic mediation. A significant distance effect in answering different to arabic-verbal pairs (e.g., ‘2 one’ would yield longer reaction times than ‘1 nine’) was considered evidence for semantic access. At the theoretical level, the question is whether Arabic and verbal lexicons may communicate not only by means of the semantic system, but also via a direct asemantic route (e.g., Cipolotti & Butterworth, 1995; Dehaene, 1992; Noel & Seron, 1993). The experimental testing of this hypothesis is rather difficult taking into account that semantic activation may proceed in autonomous fashion, thus masking any potential asemantic processing (Deheane and Akhavein, 1995).

Yet, the physical matching task, where number pairs had to be matched by their visual features and semantic processing was explicitly discouraged, provided some evidence in favour of asemantic transcoding. Here, though all Arabic-verbal pairs (e.g., 2 SIX) had to be answered as different, subjects tended to be slower and less accurate when the Arabic and the verbal numerals represented the same quantity (e.g., 2 TWO) than when pairs represented different quantities (e.g., 2 ONE or SIX 2). Thus, subjects could hardly disregard the fact that the two stimuli represented the same number. Yet, there was no evidence that they were accessing semantic information to perform the task since no distance effect was observed. Hence, the interference seemed not to arise from semantic processing, but more likely from an asemantic mapping where 1 and ONE were recognised as identical. Whether this mapping occurs by means of phonological
representations (Dehaene, 1992), intermediate semantic representation (Noel & Seron, 1997) or through unspecified inter-codes associations (Campbell, 1992) remains to be determined.

In principle, the interference from the 'number-identity' on the same-different judgements is the result of an inability to ignore irrelevant information and/or selectively attend to the stimuli dimension relevant to the task. We reasoned that, as observed in several interference-sensitive paradigms (e.g., Roger & Fisk, 1991; Shiffrin, 1991), practice with the task could modulate the size of the interference. In other words, though subjects might have suffered from the mismatch between name-identity and physical-difference, they would have learned to inhibit or ignore this information as the task proceeds. By comparing the performance in the first and second blocks of trials, our hypothesis was confirmed: In the physical matching task, subjects were slowed down and less accurate when answering different to numerically equal pairs (e.g., 2 two), but this effect was confined to the first block of trials. Thus, practice modulated significantly the interference effect and this result was replicated in all experiments (Experiments 4P, 5P and 6P). This result indicates that, though the direct mapping between Arabic and verbal numerals may be unintentionally activated, its effect may be inhibited or modulated by control mechanisms. The effect of practice in reducing interference is a well known phenomenon (e.g., Reisberg, Baron & Elmer, 1980; Melara & Mounts, 1993); it is also established that this effect is modulated by the efficiency of attentional mechanisms and thus sensitive to developmental changes. This hypothesis was supported by the results of Experiment 7P, where primary-school children showed a significant and long-lasting interference effect when performing the physical identity task. Though they had to match the stimuli on the basis of their physical features, children could not ignore their numerical identity (e.g., 1 ONE).

At first sight, this result seems in contrast with the evidence from the numerical identity task where their performance was characterised by the lack of distance effect in pure trials. This finding was attributed to the primacy of perceptual over symbolic information in children' numerical processing. However, while in the numerical identity

220
task the interference was expected to occur from the appreciation of close-far relations (e.g., 2 and 3 are less different than 2 and 6), in the physical identity the interference was caused by same-different relations (e.g., 2 and TWO represented the same number). It is reasonable to assume that same-different relations are much more distinctive to children than close-far relations.

Indeed, following a similar line of reasoning, the pattern of results observed in the physical matching tasks can be alternative explained. If one assumes that, within the semantic system, identity relations are much stronger than similarity relations, and, in turn, more likely to yield interference, the occurrence of a significant same-different effect and a non significant distance effect could have arisen at the semantic level (McCloskey quoted by Dehaene and Akhavein, 1995). According to this approach, the same-different interference would be a further index of autonomous semantic activation and the practice effect we observed would gain further significance. In particular, this finding would suggest that semantic access to number meaning may be prevented or modulated by attentional factors or strategies development, providing further support to its categorisation as a “partially” rather than “completely” autonomous process (Zbrodoff & Logan, 1986).

Overall, the evidence for asemantic transcoding is not conclusive. Neuropsychological studies may provide more explicit evidence and greatly contribute to the single versus multiple-routes debate.

6.4 ON THE SEMANTIC EFFECTS IN VERBAL NUMBER PRODUCTION

In Chapter 5, the issue of semantic effects in numerical processing was explored by means of a different approach. We reported a single-case study of an aphasic and dyslexic patient who showed specific difficulties in the production of spoken verbal numerals. We investigated the effect of different notations (Arabic versus verbal numerals), tasks (reading, calculation etc.) and task instructions on the patient’s ability to produce spoken
verbal numerals and we specifically addressed the issue of distinct semantic aspects within number knowledge.

ZA was a phonological dyslexic patient with severe difficulties in reading both Arabic and verbal numerals. His reading performance was better for verbal than for Arabic numerals and this notational effect remained stable over a period of three years: In fact, his reading of verbal numerals improved along with his alphabetical reading, while reading of Arabic numerals did not. His errors consisted mainly of lexical substitutions and in absence of encoding and comprehension problems, the functional deficit was located at the level of the activation of lexical output mechanisms (e.g., McCloskey et al., 1986). Dissociations between reading of alphabetic and Arabic notations have been previously reported (Anderson et al., 1990; Cipolotti, 1995, Noel & Seron, 1993); however, we suggested that the differential processing stages and compensation mechanisms that distinguish the processing of the two scripts are critical factors in determining the observed dissociation. Furthermore, following the recent proposal that number processing may dissociate between semantic and asemantic tasks (Cipolotti & Butterworth, 1995; Cohen et al., 1994; Cohen & Dehaene, 1995), ZA's verbal number production was assessed in a variety of tasks assumed to rely on different aspects of number semantics.

The distinction between numerical meanings has been only recently acknowledged by neuropsychologists due to the observation that number processing abilities in cerebral-lesioned patients may vary accordingly to the context a number is referring to. For example, Cohen et al, (1994; see also Beauvois & Derousne', 1979) described a deep dyslexic patient who could no longer read numerical stimuli unless they were 'meaningful', i.e., associated with encyclopaedic or semantic knowledge (e.g., 1789 as the date of the French revolution). Further dissociations have been described between the ability to process numerals in tasks that specifically required access to number magnitude (e.g., number comparison, complex calculation) and the ability to process the same numerals in a simple reading aloud task (e.g., Cohen & Dehaene, 1995; Cipolotti & Butterworth, 1995).
On the basis of these observations, dual-routes models for number transcoding have been recently proposed, by analogy to models of alphabetic reading (e.g., Cohen et al., 1994; Cipolotti & Butterworth, 1995). At the general level, these models assume semantic transcoding routes accessing semantic information and direct asemantic routes, which bypass the semantic system. For example, 'meaningful' numbers would be processed along a lexical route that accesses a store of semantic knowledge (Cohen et al., 1994); while a semantic route addressing magnitude information would mediate number processing in magnitude-related tasks (e.g., Cipolotti & Butterworth, 1995; Cohen & Dehaene, 1995). However, Seron and Noel (1995) point out that the asemantic/semantic distinction within number processing, meets with both theoretical and methodological problems that have been extensively discussed in Chapter 5.

The present study contributes to this debate by evaluating a) to what extent semantic effects may be attributed to visual familiarity and/or verbal output facilitation rather than to the semantic information itself; b) the role of intentional versus unintentional activation in semantic processing and c) the differential semantic facilitation effect exerted by different number meanings.

The same number stimuli were presented in two reading tasks: in the first one, no specific instructions were given to ZA; in the second one he was requested to associate the stimulus to a specific meaning before reading it. In this latter condition, his reading performance improved significantly and this result allowed us to locate the facilitation effect at the semantic level rather than at the input or output levels. Within the Cohen et al. model (1994), the facilitation was assumed to arise from the visual input lexicon, where familiar numerals were recognised and then processed through the lexical pathway. The present finding, however, undermines the strength of this assumption and indicates that the facilitation originates at a more central level.

Furthermore, in a priming task we evaluated if the 'unintentional' activation of an associated meaning could also facilitate the production of spoken verbal numerals. This hypothesis was indeed supported by the results: the presentation of a semantic prime (e.g., dwarfs) facilitated the activation of the correct verbal number form (e.g., seven),
and, to a lesser extent, increased the overall performance, possibly by inducing the patient to adopt a ‘semantic approach’ throughout the task.

Finally, ZA’s verbal number production was compared across tasks assumed to rely on different number meanings. In particular, his performance was significantly better in tasks involving encyclopaedic and autobiographical knowledge as well as tasks based on quantity information (dot-transcoding task) compared to calculation tasks. Clearly, all these tasks require access to number ‘meaning’, though each of them addresses a distinct aspect of number semantics. Numbers refereed to dates, measures or labels in the first type of task, they represented quantities in the second type of task and they were used as cardinals in calculation. Thus, different types of semantic information proved to differentially facilitate verbal number production, possibly by virtue of their different nature, i.e., abstractness, imageability, general semantic knowledge.

Overall, the case of ZA provides evidence in favour of the existence of both asemantic and semantic pathways mediating number transcoding. Type of tasks and instructions proved to be critical in activating one or the other of these alternative routes. Moreover, given that the same task (reading aloud Arabic numerals) and stimuli set was better performed when the patient adopted a semantic strategy, we may wonder to what extent this activation may not be intentionally modulated. In other words, ZA’s reading abilities seemed to benefit from attributing a specific meaning to the numerical stimuli. In particular, encyclopaedic and autobiographical associations proved to be the most efficient to facilitate his verbal number production. It is very likely that, both the type of semantic information and the amount of facilitation, vary across different individuals; we believe, however, that the present findings may provide useful hints for the development of rehabilitative methods. Thus, beyond their theoretical relevance, the observed semantic effects in numerical processing skills may also have practical implications. Though this is at present only a tentative suggestion, future research may prove its value.

Finally, this case-study clearly indicates the importance of distinguishing between different types of number meaning, an aspect that, at present, is unsatisfactorily handled by all models. Within current theories, number semantics mainly refers to magnitude.
information; though other aspects of numerical knowledge are not totally neglected (e.g., Dehaene et al., 1993; Delazer & Butterworth, 1997) they are not explicitly incorporated in any models. Further studies are clearly needed for clarifying the way in which these distinct numerical meanings are represented and exert an influence on numerical processing.

6.5 LIMITATIONS OF THE REPORTED STUDIES AND DIRECTION FOR FUTURE RESEARCH

Though the experimental research presented in this thesis yielded clear-cut results, a number of methodological and theoretical questions remain to be clarified.

On the autonomy in accessing number magnitude

The studies reported in Chapter 3 and 4 provided evidence for autonomous access to magnitude information by probing the existence of semantic-effects in tasks where magnitude information was irrelevant to the performance. In the same-different judgement tasks adopted in the present studies, the distance effect was used as an index of semantic access and a repeated-set of number pairs defined the experimental trials. Indeed, we observed that the selection of the stimuli was a critical factor in modulating semantic processing. This result bears interesting theoretical implications that were previously discussed, but at the same time, points to a rather intrinsic limitation of any paradigm that uses a circumscribed range of numbers as stimuli. Unfortunately, in these experimental paradigms the stimuli selection is constrained by several different criteria (e.g., frequency of the individual numbers, numerical distance, frequency of the numerical distance) and the limited range of experimental trials is unavoidable. However, caution should be exercised in drawing general conclusions from single experiment, in particular when the stimuli set is rather small.
An additional methodological point must be addressed. Both the number-Stroop paradigm and the matching paradigm consisted in two tasks that differ only along the dimension the subjects had to attend to. In both cases, we could evaluate the influence of the instructions in modulating the level of processing of the stimuli. However, in the Stroop-paradigm a within-subjects design was applied, i.e., the same subjects participated in both numerical and physical size comparisons, while in the matching paradigm different groups of subjects took part in the numerical-matching and physical-matching tasks. These different designs were adopted following previous studies employing the same paradigms (e.g., Dehaene & Akhavein, 1995; Henik & Tzelgov, 1985; Tzelgov et al., 1992) to facilitate a direct comparison with them. Given the nature of the tasks, both designs have specific advantages and disadvantages. The use of a within-subjects design allows us a more straightforward comparison of the effects across the tasks and reduces the possibility that variability within the groups may play a role in determining critical differences. However, given that the two tasks only differ for the instructions, subjects may be greatly influenced that the experience of the first task in performing the second one. Of course, as we did in Experiment 1, this factor may be controlled for by taking the order of tasks as independent variable. Indeed, evaluating the influence of previous instructions on the performance yields, in itself, interesting results. Overall, we believe that the within-subjects design adopted in Chapter 3 should be preferred when the effect of the task instructions is critically evaluated.

From a theoretical point of view, our results favour the hypothesis that semantic access in number processing is “partially” rather than “completely” autonomous (Zbrodoff & Logan, 1986). Thus, though magnitude information plays a central role in number processing, its influence on the performance may be modulated by several factors, among which task instructions are of critical importance. The extent to which access to magnitude information is autonomously activated may be further investigated by asking subjects to compare numerical stimuli along non-numerical dimensions, e.g., colour, texture, font, each presumed to demand a different level of processing of the stimuli. These tasks would provide stringent tests for the autonomy of semantic
processing, given that these perceptual dimensions are assumed to be processed at relatively early stages of visual processing, reducing the chance for numerical information to interfere with the performance.

Furthermore, though some of the reported findings are in line with the suggestion that autonomous and intentional processing of numerical information are qualitatively different (Tzelgov et al., 1992), the experimental paradigms adopted in the present studies do not provide conclusive evidence to this regard, and the issue certainly deserves more targeted investigations. In particular, the differential contribution of a dichotomous small/large classification process and of a precise mapping of numbers into their magnitude representations could be only disentangle by controlling the different impact of numerical distance and laterality on the performance.

**Developmental trends in autonomous access to number magnitude**

Findings from the developmental studies strongly suggest that access to number magnitude acquires autonomous properties over the course of skill acquisition. In the present study, we consider the factor of age as an indirect index for numerical skills; yet, it would be extremely interesting to compare the rise of automaticity of numerical semantic access with the gradual refinement of other numerical abilities, such as arithmetic and calculation. In fact, the autonomous activation of number magnitude may be considered an index of a thorough understanding of the numerical symbols. In other words, once the association between Arabic numerals and the numerosities they represent is fully established, Arabic numerals may be said to be truly meaningful and, in principle, this should facilitate operating upon them. In other words, children who come to show a mature comprehension of the arithmetic symbols should deal with them more efficiently, as for example, in mastery arithmetic knowledge or in the use of adequate calculation strategies. Moreover, it seems plausible to assume that access to number magnitude precedes, and possibly constitutes a developmental prerequisite to autonomous activation of mental arithmetic. However, in our attempt to compare our results with the evidence from arithmetical studies (Lefevre et al., 1991; Lemaire et al., 1994), we noticed that the rise of these two processes seems to occur in parallel. Clearly, this conclusion is rather
tentative and only the intraindividual evaluation of these two phenomena, i.e., testing the same children in both tasks, may clarify the developmental trend of these competencies.

Finally, it has been recently suggested that a slow and inefficient access to numerical information may be a critical factor underlying arithmetic learning disabilities (Hitch & McAuley, 1991; Koontz & Berch, 1996). In this regard, the evaluation of autonomous access to magnitude information may provide a diagnostic index for immature or delayed number processing abilities and disclose difficulties that could not show up clearly in standardised arithmetic tests.

Though the issue has been so far overlooked, we believe that further developmental investigations on the autonomy of number processing may be extremely fruitful and greatly contribute to a better understanding of the multiple changes that characterise the acquisition of numerical skills.

On the neuropsychological evidence for multiple routes and multiple meanings in numerical processing

The neuropsychological investigation reported in Chapter 5 raises two major points of concern. First, the case of ZA indicates the importance of distinguishing between different types of number meanings, and in particular, provides additional evidence for the special status of encyclopaedic number knowledge (e.g., Cohen et al., 1994). This term is used to refer to those numbers, mostly used in non-numerical context, that have become 'meaningful' for being associated to specific semantic information (e.g., dates, labels). Yet, we know very little about the way in which this subset of number knowledge is normally processed and to what extent the semantic effects observed with patients equally affect normal number processing. This lack of experimental evidence is rather unfortunate; one way for future research to compensate for it, would be to present normal subjects with priming tasks analogous to the one described in Chapter 5.

Second, our findings overall favour the hypothesis that both asemantic and semantic pathways mediate number transcoding. However, the case of ZA, as well as all the other ones described so far in the literature (e.g., Cohen et al., 1994; Cipolotti, 1995; Cipolotti & Butterworth, 1995), performs better in tasks relying on semantic elaboration than in
'asemantic' tasks and no evidence for the reverse dissociation is yet reported. However, it is well established that neuropsychological evidence is much more definitive when double dissociations are obtained. As extensively discussed elsewhere (section 2.1.7) the disadvantage in asemantic tasks may be always alternatively interpreted on the basis of differential impairment within the semantic route (Seron and Noel, 1995). Thus, though our findings are compatible with dual-route models for number processing (e.g., Cipolotti & Butterworth, 1985; Dehaene & Cohen, 1995), we acknowledge that only the observation of a case that presents the reverse dissociation (e.g., intact transcoding of Arabic numerals and impaired comprehension of them) can unequivocally disentangle the semantic/asemantic debate.
REFERENCES


Boucart, M., Humphreys, G.W. & Lorenceau, J. (1995). Automatic access to object identity: Attention to global information, not to particular physical dimensions, is


240


