COGNITION

Cognition 67 (1998) 365–373

Does language really matter when doing arithmetic? Reply to Campbell (1998)

Marie-Pascale Noel[a,*], Annie Robert[b], Marc Brysbaert[c]

[a]*Cognitive Neuropsychology Department, Université Catholique de Louvain, Louvain, la-Neuve Belgium*
[b]*Institut de Statistique, Université Catholique de Louvain, Louvain, la-Neuve Belgium*
[c]*Department of Experimental Psychology, University of Ghent, Ghent, Belgium*

Abstract

Campbell (1998) has questioned the conclusion of Noël et al. (1997) and has argued that alternative analyses of their data provide strong evidence that arithmetic performance is subject to reading-based interference and provide some support for the language-specificity of number-fact memory. We consider that Campbell reached conclusions different from those we had obtained because (1) he performed his analyses on a different data set (i.e. including also the table-unrelated errors), (2) he has given a double weight to the naming errors and (3) he has multiplied the analyses without correcting the corresponding *P* values. We thus consider that there exist interactions between language and performance in simple multiplication tasks, but that the current data can easily be explained without postulating that such interactions operate at the level of the retrieval stage. In other words, we consider that there are not definitive arguments, as yet, in favour of the hypothesis of modality-specific arithmetical-fact networks.

Keywords: Number-fact memory; Reading-based interference; Modality-specific arithmetical-fact networks

1. Introduction

Noël et al. (1997) replicated and extended Campbell's (1994) experiments to find out whether number format effects in a simple multiplication task are evidence for interactions between language and arithmetical-fact retrieval processes, as claimed by Campbell (1994). They did so by including an extra control task (the number-

* Corresponding author. Departement de Psychologie Cognitive, 10 place du Cardinal Mercier, 1348 Louvain-la-Neuve, Belgique. E-mail: noel@neco.ucl.ac.be

0010-0277/98/$19.00
PII S0010-0277(98)00038-9

matching task of Experiment 1) and by looking at the performance of Dutch-speaking subjects whose language reverses the decades and the units in the number names (Experiment 2). Although Noël et al. replicated Campbell's (1994) interactions between number format and multiplication performance, for two reasons they thought they were not justified interpreting their findings as evidence refuting the hypothesis of McCloskey et al. (1985) that arithmetical-fact retrieval is based on format and language-independent representations. First, very similar interactions were present in the matching task which did not involve retrieval of arithmetical facts; and second, the interactions did not depend on the subjects' language, as would have been predicted by Campbell (1994). It is this latter finding that Campbell (1998) challenges by presenting new analyses which seem to indicate that the predicted language effect is present after all.

Before we discuss Campbell's (1998) interpretation of our data, it may be judicious to stress that we do not question the existence of number format effects on numerical performance. Indeed, interactions between number format and numerical performance can easily be obtained. The question we ask is whether the effects really point to linguistic influences on the arithmetical-fact retrieval system, or whether the effects are primarily due to peripheral input and output processes (see McCloskey et al. (1992) for a similar position). In other words, the challenge is not to find an interaction between number format and arithmetical performance but to establish that the interaction pivotally originates from the arithmetical-fact retrieval stage.

According to Campbell, such evidence is provided by one particular type of error: the operand-intrusion errors (e.g. $4 \times 8 = 28$). These errors are more common when the problem operands are number words (four × eight), and, in addition, the intruding operand tends to keep the same position in the answer as in the problem (e.g. in English, $4 \times \mathit{8} = 2\mathit{8}$ is more frequent than $\mathit{8} \times 4 = 2\mathit{8}$). Campbell (1994) argued that these findings provide convincing evidence for the hypothesis that arithmetic memory depends upon language. Noël et al. tested this view by contrasting the multiplication performance of Dutch- and French-speaking subjects. French-speaking subjects have a number-naming system that largely resembles English, and hence should exhibit a similar pattern of performance, whereas in Dutch, the serial order of decades and units is reversed in the number names (e.g. 24 = 'vingt-quatre' (twenty-four) in French, but 'vier-en-twintig' (four and twenty) in Dutch). According to Campbell's reading-based interference hypothesis, the Dutch number-naming system should result in a different pattern of operand-intrusion errors with many more so-called incongruent intrusion errors (e.g. $6 \times 9 = 36$; six × nine = six and thirty). Noël et al. (1997), however, reported that they failed to obtain such a language effect; and it is this conclusion that Campbell challenges in his reply. Campbell (1998) presents new analyses of the intrusion errors and argues that the position effect of these errors does indeed differ between the Dutch- and the French-speaking subjects. More precisely, the decade intrusions in the word condition would be more often incongruent in the Dutch-speaking group than in the French-speaking group.

The alternative analyses of our data proposed by Campbell are interesting but they differ from our own work in two important aspects: First, Campbell has broken down the data and run multiple smaller analyses, and second, the data set, on which his analyses are based, was different from the one we used. First, we will discuss the impact of these two differences, and then we will return to the more general issue of how to interpret the language effects.

In our analyses, only table-related errors with intrusions (and responses greater than or equal to 17) were considered. Whereas for Campbell, all tabled products (greater than or equal to 17) were taken into account. Thus, his analyses include table-unrelated and naming errors on top of the table-related errors we had considered. In order to test possible interactions between reading processes and the arithmetical-fact retrieval stage, we should focus our attention on intrusions appearing in errors that can nearly unambiguously be attributed to the arithmetical-fact retrieval stage. For instance, when subjects produce an error that does not belong to the multiplication table (e.g. $7 \times 9 = 47$), they may do so because of an error in the production stage or, less probably, in the encoding or arithmetical-retrieval stage. On the other hand, an error belonging to the table of one of the operands (or even both, e.g. $4 \times 8 = 24$) is much more likely to arise from a failure in selecting the correct answer from the arithmetical-fact store. The case of the table-unrelated errors lies in-between (e.g. $3 \times 4 = 14$). In order to test the hypothesis of an interaction between reading processes and arithmetical-fact retrieval processes, one should select the errors which (1) minimise the probability of being due to errors occurring in the encoding or the production stages and (2) maximise the probability of being due to an error occurring in the arithmetical-fact retrieval stage. We consider that the table-related errors are those which best meet these two criteria.

In addition, we do not agree with Campbell that naming errors can be combined with the other intrusion errors, because such practice gives naming errors a double weight: an answer like $4 \times 8 = 48$ is considered both as an intrusion of the first operand in the decade position and of the second operand in the unit position. We consider that this practice is not warranted and, therefore, that naming errors should be analysed separately from the other intrusion errors.

The second difference between Campbell's approach and ours is that instead of running a global analysis of the data, Campbell has preferred to perform multiple smaller analyses. However, by doing so, one automatically increases the probability of falsely rejecting the null hypothesis. This may lead to misleading conclusions. Consequently, a correction, such as Bonferroni's, should be applied.

In order to answer Campbell's new analyses properly, we carried out similar ones but applying the Bonferroni correction and restricting our analyses to the table-related errors with intrusions on the one hand, and to the naming errors on the other.

2. Table-related intrusion errors

In order to avoid any confusion, all the data are presented in Table 1 as mean number of errors in each category for each format and each group of subjects. A first way of splitting the data of table-related errors was to separate the congruent (e.g. $8 \times \mathit{4} = 2\mathit{4}$) and incongruent (e.g. $\mathit{4} \times 8 = 2\mathit{4}$) intrusions. If one does that (with group and format as between- and within-subject factors), all the *P* values have to

Table 1
Mean number of errors, standard deviation and total frequencies of table-related errors with intrusions as a function of format, intruder position and lexical class

		Digit problems			Word problems		
Position	Lex. class	Decade	Unit	Total	Decade	Unit	Total
		Dutch sample					
First	Mean	1.83	2.04		2.13	3.83	
	SD	2.24	3.01		3.11	4.56	
	Sum	44	49	93	51	92	143
Second	Mean	1.17	2.96		3.08	5.17	
	SD	1.95	2.99		3.55	4.82	
	Sum	28	71	99	74	124	198
Total		72	120	192	125	216	341
		French sample					
First	Mean	1.04	1.38		0.63	2.79	
	SD	1.55	1.58		0.88	2.41	
	Sum	25	33	58	15	67	82
Second	Mean	0.17	2.04		0.63	3.79	
	SD	0.48	2.14		0.92	3.40	
	Sum	4	49	53	15	91	106
Total		29	82	111	30	158	188

be multiplied by two. Consequently, for the congruent intrusions, a major format effect is observed ($F(1,46) = 14.55$, MSe = 6.06, $P = 0.0008$) but no evident group effect ($F(1,46) = 4.27$, MSe = 29.52, $P = 0.088$) nor any interaction between group and format ($F(1,46) = 1.35$, MSe = 6.06, $P = 0.50$). For the incongruent intrusions, both the group ($F(1,46) = 629$, MSe = 25.45, $P = 0.03$) and the format effects ($F(1,46) = 30.00$, MSe = 6.23, $P = 0.0001$) are obtained, but once more, the critical interaction between group and format fails to reach statistical significance ($F(1,46) = 6.23$, MSe = 6.23, $P = 0.15$). Thus, applying the Bonferroni correction and restricting the analysis to the table-related errors leads to a failure to observe any interaction between the group and the format.

Nevertheless, Campbell pursues his analyses by decomposing this time the data into a finer level, i.e. intrusions as being congruent or incongruent and appearing in decades or in units. Only one of these four analyses gives rise to a significant interaction between the format and the group, i.e. the one carried on the incongruent intrusions in the decade position. Fig. 1 is a scatter plot of the difference between congruent and incongruent intrusions in decades only, with our own data. According to Campbell, Dutch subjects produce fewer congruent than incongruent intrusions in words (i.e. a negative difference in words) and more congruent than incongruent intrusions in digits (i.e. a positive difference in digits); Dutch subjects should therefore be in the upper left square of the plot. In contrast, French subjects are described as producing more congruent than incongruent intrusions in words but with the same difference in words as in digits; French subjects should therefore be on the diagonal and within the upper right square of the plot. Yet, such a conclusion is not supported

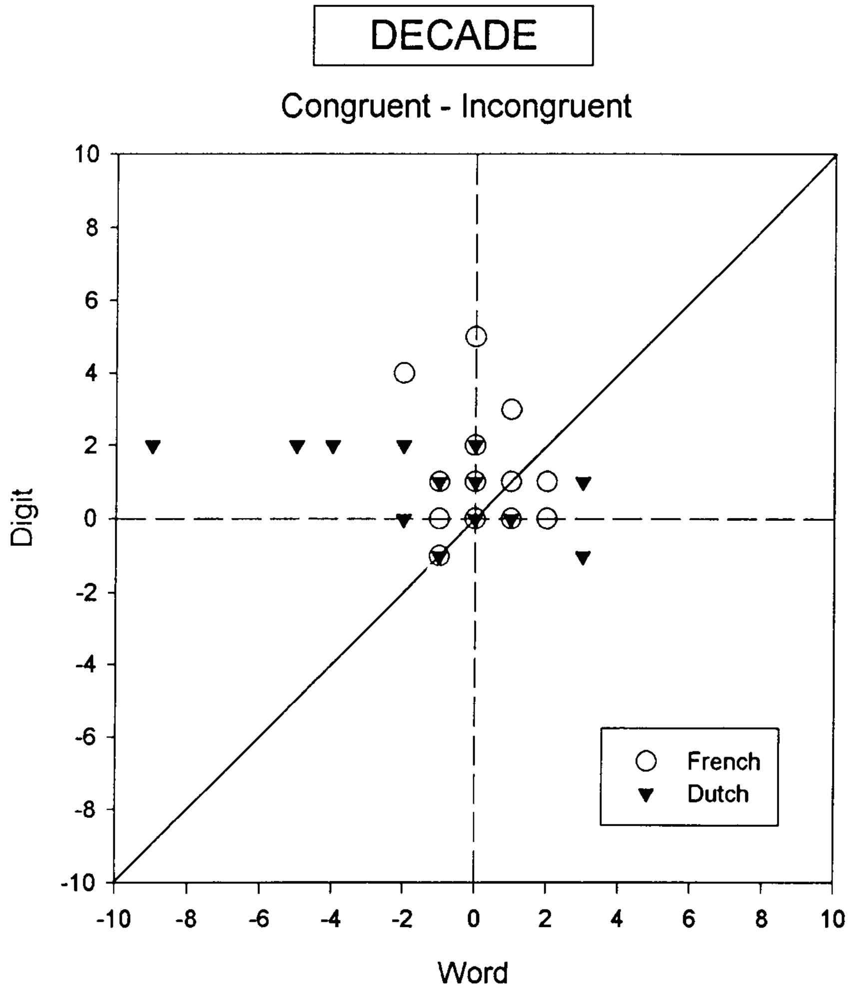

Fig. 1. Scatter plot of the differences between congruent and incongruent intrusions as a function of the presentation format of the problem and the group of the subjects.

by data points of Fig. 1. Furthermore, this figure does not indicate the presence of two different clouds due to a group effect.

This observation is supported by the statistical analysis. If we consider the only cell in which an interaction between the format and the group was found by Campbell, i.e. the incongruent intrusions in the decade position, a new analysis gives the following results (with corrected P values): a main effect of the group ($F(1,46) = 10.99$, MSe = 6.53, $P = 0.007$), a main effect of the format ($F(1,46) = 15.28$, MSe = 2.21, $P = 0.0012$) but only a marginally-significant interaction between group and format ($F(1,46) = 5.76$, MSe = 2.21, $P = 0.08$). Looking at the data, we could argue, as Campbell did, that the difference of incongruent intrusions between the word and digit formats is bigger, although not significantly, in the

Dutch sample (difference of 46 (=74 − 28)) than in the French sample (difference of 11 (=15 − 4)). The mean difference between word and digit format is thus 4.2 times (46/11) higher in Dutch than in French but the deviation of the difference is also 3.4 times higher (see higher dispersions of errors within each subdivision of Table 1)! Furthermore, doing so is quite misleading because it does not take into account the fact that, generally, Dutch subjects produced many more errors than French subjects. If we now take this point into consideration and calculate the percentage of incongruent intrusions in the word format relative to both the word and the digit format, we may come up with the opposite conclusion: the percentage of incongruent intrusions in the word format is slightly more important in the French sample (78.9%) than in the Dutch sample (72.5%). However, an analysis of these percentages fails to reach significance ($t(29) = -0.82$, $P = 0.42$; Note that for this analysis, only subjects having produced at least one of these intrusions are included, i.e. 20 Dutch and 11 French subjects.

3. Naming errors

If we now consider the naming errors, there are two possibilities: either the subject gives a unit-response (e.g. $6 \times 9 = 6$; $9 \times 5 = 5$) or a two-digit-response (e.g. $4 \times 8 = 48$). Since we are interested in the reversal effect between decade and unit in French- and Dutch-speaking samples, only the second type of naming error should be considered. These errors can be distributed into four categories, according to the presentation format of the problem and the matching of the position of the operands in the problem and in the response (i.e. congruent ($4 \times 8 = 48$) or incongruent ($4 \times 8 = 84$)). These results are presented in Table 2. The data seem to suggest, as claimed by Campbell, that in the digit condition, both groups produce more congruent intrusions than incongruent intrusions, whereas in the word format, French subjects show the same rate of congruent and incongruent intrusions but the Dutch show the reverse pattern, i.e. more incongruent intrusions than congruent

Table 2
Frequency of naming errors as a function of the presentation format and congruency status, as well as Wilcoxon-test and *P*-value associated

Group	Word		*Z* (Wilcoxon)	*P* value[a]
	Congruent	Incongruent		
Dutch	5	11	1.61	0.11
French	2	2	0	1
	Digit			
	Congruent	Incongruent	*Z* (Wilcoxon)	*P* value[a]
Dutch	6	0	1.604	0.10
French	4	1	−1.134	0.26

[a]*P* values are not corrected for multiple analyses.

intrusions, thus confirming his theory. However, statistical analyses do not confirm this impression. Wilcoxon tests comparing the number of congruent and incongruent intrusions in each format for each population do not reach statistical significance (see Table 2).

In summary, we consider that Campbell reached conclusions different from those we had obtained because (1) he has aggregated the table-related and table-unrelated errors, (2) he has given a double weight to the naming errors, and (3) he has multiplied the analyses without correcting the corresponding *P* values. Consequently, we maintain our conclusion that French- and Dutch-speaking subjects show the same profile of position effects in the intrusions, although we agree that in the case of naming errors we may be dealing with a problem because the sample of errors produced is too small. In the next section we will outline why we do not consider this a major problem, notwithstanding.

4. More general comments on the involvement of language when doing arithmetic

If we return to the broader issue discussed in Noël et al.'s paper, the whole question was whether or not the linguistic factors affected the arithmetical-fact retrieval stage. Noël et al. argued this was not the case, because the Dutch-speaking subjects did not show the pattern of errors predicted by Campbell (i.e. a reversed position effect in the intrusion errors). As we have seen above, Campbell's new, alternative analyses do not reject this interpretation if performed properly (Campbell, 1998).

This same issue was at the centre of another work by Brysbaert et al. (1998), in which Dutch- and French-speaking subjects were required to realise simple additions involving a two-digit number and a one-digit number (e.g. 20 + 4 = ?). In half of the trials, the two-digit number was the first operand (e.g. 24 + 3 =); in the other half, it was the second operand (e.g. 3 + 24 =); the presentation format was also manipulated (digits or words). We were interested to see whether the presentation order of the two operands would have an effect on the response time, and whether the effect would vary as a function of the subject's language.

We will not recapitulate the findings of Brysbaert et al. at great length here, because they are readily available, but one important aspect for the present discussion should be mentioned. In their first experiment, Brysbaert et al. obtained a series of results which seemed to indicate that, whereas French-speaking subjects performed better when the two-digit number preceded the single-digit number (i.e. 20 + 4 was faster than 4 + 20; and 24 + 3 was faster than 3 + 24), the same was not true for the Dutch-speaking subjects. These subjects showed a much less pronounced order effect for the problems of the type 20 + 4, and even a reversed effect for the problems of the type 24 + 3 (i.e. Dutch-speaking subjects were faster to provide the solution to the problem 3 + 24 = ? than to the problem 24 + 3 = ?). Needless to say, such findings are more in line with models that postulate language effects on the arithmetical-fact retrieval system, than with models that do not.

However, Brysbaert et al. noticed that by asking the subjects to give the answer aloud, they also varied the format of the answer between the two groups. In particular, the French-speaking started their answer with the value of the decade, whereas the Dutch-speaking started their answer with the value of the unit. To control for this factor, Brysbaert et al. added an experiment, in which French- and Dutch-speaking subjects had to type in the answer to the problem and, thus, both had to start with the value of the decade. In this experiment, virtually all language differences disappeared, showing that production characteristics may have a profound impact on arithmetical performance.

It is not inconceivable that the same production characteristic affected the data of Noël et al. (1997). Having to start the response with the value of the unit may lead to some very specific errors. Thus, Noël et al. (1997) already reported (see Table 7, p. 363) striking differences between French- and Dutch-speaking subjects in the miscellaneous errors: whereas the French-speaking participants made a considerable number of errors in which they reported a decade-name only, the Dutch-speaking participants almost never did. The reverse was true for the wrong responses in which only a unit-name was reported. In all those cases, subjects initiated an answer (by a decade name in French or a unit-name in Dutch), then realised that their answer was incorrect and thus stopped emitting the response. A similar phenomenon may account for differences in naming errors in which only the value of one digit is given: 'seven × five = five (and don't know any more)'. The same may be true for the naming errors presented in Table 2 (i.e. $4 \times 8 = 84$). For these specific instances, we have acknowledged that the difference between Dutch- and French-speaking subjects might reach significance if more subjects were included in the experiment. But what do these errors tell us? That when confronted with a word problem such as 'vier × acht' (four × eight) subjects sometimes say 'vier and tachtig' (four and eighty) because they cannot inhibit a reading process. This does not seem very strange given that written words are known to activate the corresponding phonological representation (either by grapheme–phoneme conversion rules or by direct links between the graphemic input lexicon and the phonological output lexicon). Reading a word-problem could thus activate two processes in parallel: a reading process and an arithmetical-fact retrieval process, which are much more likely to compete with one another at the output stage than at the arithmetical-fact retrieval stage. In the case of these particular naming errors, one could very well imagine that the product of the reading process reaches the production stage before the response of the arithmetical-fact retrieval process and that it seriously affects the final, spoken, answer.

In summary, we started our work on format and language effects some years ago with the strong expectation that we would find many format and language effects, which would argue against McCloskey et al.'s position that arithmetical-fact retrieval is based on a single, language-independent and format-independent, representation system. We indeed found the interactions, but became more and more sceptical about their power to reject the hypothesis of a single representation. When one looks carefully at the interactions reported in the literature, ours included, they all display the following characteristics: (1) very similar interactions are found for other tasks than arithmetic (e.g. the number matching task of Noël et al. (1997)), (2) the interactions are virtually restricted to those cases where the input perfectly matches the operation to be performed (e.g. VII = 5 + 2, but not VII = 6 + 1, Noël and Seron (1997); 'een + twintig = een-en-twintig', but not 'een + drie-en-twintig = vier-en-twintig', Brysbaert et al. (1998)) and (3) the interactions are mostly restricted to verbal responses (e.g. Brysbaert et al. (1998)). These three observations made us believe that such interactions need not be a sign of language influences at the arithmetical stage, but could very well be due to interactions at the phonological output stage, if one assumes that the (verbal) input spreads automatically to the phonological output stage in a cascaded fashion.

So, coming back to the question 'does language really matter when doing arithmetic', we would answer, 'yes', there exist interactions between language and performance in simple multiplication tasks. However, the current data can easily be explained without postulating that such interactions operate at the level of the retrieval stage. In other words, we consider that there are not definitive arguments, as yet, in favour of the hypothesis of modality-specific arithmetical-fact networks.

Acknowledgements

We would like to thanks Xavier Seron for his helpful comments on this paper. This work has benefited from a grant from the Belgian Federal government: P.A.I.U. on temporal control of dynamic task situations and the nature of human knowledge representation.

References

Brysbaert, M., Fias, W., Noel, M.-P., 1998. The Whorfian hypothesis and numerical cognition: is 'twenty-four' processed in the same way as 'four-and-twenty'? Cognition 66, 51–77.

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

Campbell, J.I.D., 1998. Notational and linguistic influences in cognitive arithmetic: comment on Noël, Fias and Brysbaert (1997). Cognition 67, in press.

Noël, M.-P., Fias, W., Brysbaert, M., 1997. About the influence of the presentation format on arithmetical-fact retrieval processes. Cognition 63, 335–374.

Noël, M.-P., Seron, X., 1997. On the existence of intermediate representations. Journal of Experimental Psychology: Learning, Memory and Cognition 23, 697–720.

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

McCloskey, M., Macaruso, P., Whetstone, T., 1992. The functional architecture of numerical processing mechanisms: defending the modular model. In: Campbell, J.I.D. (Ed.), The Nature and Origin of Mathematical Skills, Elsevier, Amsterdam, pp. 493–537.