Languages treat 1-4 specially:
COMMENTARY on
Stanislas Dehaene's précis of
The Number Sense.

James R Hurford,
Department of Theoretical and Applied Linguistics,
University of Edinburgh.

This commentary is in two sections. The first presents some specifically linguistic evidence regarding Dehaene's statement that, in animals and prelinguistic humans, "only very small numbers (up to about 3) can be represented accurately"[1]. The second section contains some more general reflections on Dehaene's work.

1 Linguistic discontinuity around 3 or 4

Dehaene argues that adequately accurate analog mental representation of very small quantities, in the range 1-4, is ancient. Evidence from language corroborates this. Languages show a tendency to express the first two, three or four numbers differently from slightly higher ones. Some of this difference in behaviour makes the lowest-valued numerals more like adjectives. This suggests that these very low numerosities may be perceived and represented like the meanings of basic adjectives, such as *red, hot* and *round*, as sensible properties of (groups of) objects. Higher numerosities, from about 5 upward, are less directly, or less reliably, observable, and representation of them relies not just on perceptual properties, but probably (at least in part) on their being associated with a particular slot in a conventional recited counting sequence. It seems certain that in the historical development of languages, words for the lower numbers existed before words for higher numbers. The idiosyncrasies which tend to be shown by the very lowest numeral words preserve ancient linguistic patterns. Wittgenstein captures the idea nicely:

> "Our language can be seen as an ancient city: a maze of little streets and squares, of old and new houses, and of houses with additions from various periods; and this surrounded by a multitude of new boroughs with straight regular streets and uniform houses." (*Philosophical Investigations*)

1.1 Grammatical number

Grammarians use the term '(grammatical) number' for the singular/plural distinction marked on nouns in most languages, and supplemented in some languages by categories such as dual, and even trial. Clearly, grammatical number in a language is distinct from its numeral system. Grammatical number tends to be indicated in the morphology of nouns (e.g. by an affix), whereas numerals are distinct words or phrases. Grammatical number in fact has only a somewhat loose relationship with numeral expressions in a language. Nouns marked for grammatical number need not be accompanied by numerals. In Arabic, for instance, *kitaab* means '(a) book' and *kitabeen* means 'two books'. The important point here is that the numerical limit for systems of grammatical number is at around 3 . A few languages have distinct trial forms of nouns, for describing collections of 3; quite a few more languages have dual distinct forms, for describing pairs of things; and of couure, almost all languages distinguish singular from plural.

[1] Much of the linguistic data cited here is taken from Hurford (in press), where further details can be found.

1.2 Idiosyncrasies of words for 1, 2, 3 and 4

In almost all languages, numbers up to the first base number (usually 10) are expressed by single words, and after the base number the system resorts to syntactic combinations. But very often the words for the values 1-10 are not uniform in their grammatical features. It is common for the very lowest numbers, from 1 to about 3 or 4, to be expressed by numeral words with distinctive characteristics and irregularities. The rest of this section documents such idiosyncrasies of the numerals in the range 1 - 4.

1.2.1 Distinct counting numerals

The numeral used attributively to modify a noun is not necessarily the same as the corresponding numeral in the conventional recited counting sequence. There are various degrees of idiosyncratic difference between a counting numeral and a quantifying (attributive) numeral. In some cases the counting numeral is completely dissimilar to the attributive numeral, and in other cases one form iis obviously a morphological variant of the other. Some examples are given below.

LANGUAGE	NUMBER	ATTRIBUTIVE NUMERAL	COUNTING NUMERAL
German	1	*ein, eine* etc.	*eins*
German	2	*zwei*	*zwo*
Maltese	2	*żewg*	*tnejn*
Chinese	2	*liang*	*erh*
Hungarian	2	*két*	*kettö*
Basque	2	*bi*	*biga*

The most common number to be expressed by a distinct counting form is 2. Others are extremely rarely. It might be expected that the occurrence of distinct counting forms would be greatest for the number 1, and decrease thereafter. Although the counting sequence begins at 1, it is only the utterance of a form for 2 after a form for 1 that confirms that the activity of counting is taking place.

1.2.2 Ordinals

Not all languages have a separate series of ordinal numerals. And some languages only have (or in practice their speakers only use) ordinals for a limited low-valued subset of numbers. Maltese only has distinct ordinals up to '4th'. In Spanish, although ordinals for values higher than 20 are available, e.g. *centésimo*, '100th', they are rarely used, and a construction with a cardinal is used instead.

Irregularity in the formation of ordinals occurs in various degrees. Some idiosyncratic modification of the cardinal stem may accompany an otherwise regular morphological process, as in English *five/fifth*. Or the ordinal may bear some unpredictable resemblance to the cardinal, as with English *three/third*. The extreme case is suppletion, which typically involves the very low numbers, 1, 2, and perhaps 3. Examples with 1/1st are: Greek *ena/protos*, Welsh *un/cyntaf*, Italian *uno/primo*, Finnish *yksi/ensimmäinen*, and English *one/first*. Examples with 2/2nd are Greek *djo/devteros*, Welsh *dau/ail*, Italian *due/secondo* and Finnish *kaksi/toinen*.

The following table summarizes the formation of ordinals in a sample of 17 mainly European languages, from many different language families. It shows clearly how various kinds of irregularity are clustered in the lower range of numbers, with most irregularity in the range 1-4.

Number	Suppletive	Irregular	Regular	No distinct ordinal form
‘1’	14		3	
‘2’	6	3	8	
‘3’		6	11	
‘4’		3	14	
‘5’		2	14	1
‘6’		1	15	1
‘7’		1	15	1
‘8’		1	15	1
‘9’		1	15	1
‘10’			16	1

Table 1: Suppletive and irregular ordinals.

1.2.3 Distinct numerals for different objects

Some languages have several sequences of numeral words, and choice of the appropriate numeral depends on the noun being modified. In some cases, such distinct numerals only exist for a small range of numbers, typically in the low range. For the numbers 2 - 6, Bulgarian has a series of numerals, applicable only to male humans. The similar forms for 7 - 10 are marginal or archaic.

1.2.4 Word order

The very lowest-valued numerals in a language sometimes have a different order from the rest. Basque, Arabic, Hebrew and Maltese place the numeral for 1, exceptionally, after the noun, like an adjective, and all other numerals precede the noun (in Vizcayan Basque the numeral for 2 also, exceptionally, follows the noun).

1.2.5 Marking for (in)definiteness

In Albanian, there are separate definite and indefinite numerals for the numbers 1-4. The paradigms for ‘1’ are given below.

Albanian: Definite versus indefinite numerals, 1-4.

	INDEFINITE		DEFINITE	
	Masc.	Fem.	Masc.	Fem.
NOM	një	një	njëri	njëra
GEN	njëri	njëre	njërit	njërës
DAT	njëri	njëre	njërit	njërës
ACC	një	një	njërin	njërën
ABL	njëri	njëre	njërit	njërës

Lambertz’ (1959) grammar, from which this information comes, also mentions distinct definite and indefinite forms for the numerals 2-4, but many, though not all, of his examples do not involve morphological variants of the numerals themselves, but rather a definite particle preposed to the numeral. My own informant, Eranda Kabashi, a speaker of the Geg dialect of Kosovo, recognizes fewer morphological distinctions among the numerals than Lambertz, but nevertheless has some distinct forms for definites and indefinites, for example:

dy vajza
'two girls'

të dyja vajzat
'the two girls'

German has distinct indefinite genitive forms just for the numerals denoting 2 and 3, namely *zweier* and *dreier*.

1.2.6 Distinct case forms

The table below shows the maximum number of cases distinguished idiosyncratically on numerals, that is, instances where case-marking on numerals is not by a regular productive process which also applies to other parts of speech, such as adjectives or nouns. The numbers show, for each language, how many distinct rows there are in the case paradigm for each numeral.

Numeral	Albanian	German	Greek	Icelandic	Romany	Russian	Zürich German
'1'	3	4	3	4	2	6	2
'2'	3	2	1	4	2	4	1
'3'	1	2	2	4	2	4	1
'4'	3	1	2	4	2	4	1
'5'	1	1	1	1	2	3	1
'6'	1	1	1	1	2	3	1
'7'	1	1	1	1	1	3	1
'8'	1	1	1	1	1	3	1
'9'	1	1	1	1	1	3	1
'10'	1	1	1	1	2	3	1

Table 2: Maximum number of grammatical cases distinguished idiosyncratically on numerals. Thus '1' in a cell indicates **no** distinctions of case, '2' indicates a single distinction, and so on. This table does not count languages in which there is fully regular productive affixing of case morphemes which can also apply to non-numerals, e.g. in Hungarian. The Albanian is that of Lambertz (1959).

Clearly, distinctive case marking on numerals, where it occurs, is largely concentrated on the range 1-4.

1.2.7 Grammatical effect on noun

In Russian, the numerals for 2, 3, and 4 assign genitive singular to a sister noun, whereas the numerals for 5 - 10 assign genitive plural.

1.2.8 Distinct gender forms

Gender marking on attributive numerals is always in agreement with the inherent gender of the sister noun. Examples from Zurich German (Weber, 1964:132,133) are:

zwee Mane
two men

zwoo Fraue
two women

zwäi Chind
two children

drei Mane
three men

drüü Chind
three children

Idiosyncratic gender-marking is typically restricted to just the first few numerals, as the following table shows.

N	Alb_L	Alb_N	Blg	Fr	Grm	Grk	Ice	Mlt	Rus	ScGl	Wel	ZD
'1'	2	1	2	2	3	3	3	2	3	1	1	3
'2'	2	1	2	1	1	1	3	1	2	1	2	3
'3'	2	2	1	1	1	2	3	1	1	1	2	2
'4'	2	1	1	1	1	2	3	1	1	1	2	1
'5'	1	1	1	1	1	1	1	1	1	1	1	1
'6'	1	1	1	1	1	1	1	1	1	1	1	1
'7'	1	1	1	1	1	1	1	1	1	1	1	1
'8'	1	1	1	1	1	1	1	1	1	1	1	1
'9'	1	1	1	1	1	1	1	1	1	1	1	1
'10'	1	1	1	1	1	1	1	1	1	1	1	1

Table 3: Maximum number of grammatical genders distinguished idiosyncratically on numerals. This table does not count languages in which there is fully regular productive affixing of gender. Alb_L refers to the Albanian of Lambertz (1959); Alb_N refers to the Albanian of Newmark et al. (1982).

2 General Comments

2.1 Number Platonism, Psychology and Language

Dehaene's critique of Platonism is entirely harmonious with the view expounded in Hurford (1987, especially chapters 3 and 4). This work is missing the crucial neurological underpinning that

Dehaene has begun to establish. But it outlines how the complete modern educated human grasp of whole numbers has evolved as a product of the interaction among simple linguistic signs, the disposition to recite ritual sequences of words, the utility of precise reference to collections of objects of various cardinalities, and recruitment of certain pre-existing syntactic processes (e.g. conjunction and pluralization). The pre-existing apparatus attributed to the child in respect of the very low numbers is very similar for both Dehaene and Hurford. In Hurford's account, concepts of numbers as abstract entities emerge from a process of abstraction from the more concrete meanings of linguistic expressions denoting collections of observable objects. It would be interesting to see to what extent Dehaene could agree with the details of this outline.

2.2 Neurology is Complex

I am not qualified to judge Dehaene's neurological claims. But I would not be surprised if, ten years from now, we will have found that Dehaene's neat correspondences between brain areas and the three different formats (analogical, verbal and visual/Arabic) are prone to substantial individual variation. This has been the story with the once-popular "language areas" of the brain, Broca's and Wernicke's. Although there is a strong statistical correlation between lesions in these areas and specific patterns of aphasia, we now know that there can be lesions to these areas without the aphasias and patients can be diagnosed with the specific aphasias without having the localized lesions.

References

Hurford J R 1987 *Language and Number: the emergence of a cognitive system.* Basil Blackwell, Oxford.

Hurford J R in press "The interaction between numerals and nouns" in F.Plank (ed.) *Noun Phrase Structure in the Languages of Europe*, Walter de Gruyter, Inc.

Lambertz, Max 1959 *Lehrgang des Albanischen: Teil III, Grammatik der Albanischen Sprache.* Halle (Saale): Max Niemeyer Verlag.

Newmark, Leonard, Philip Hubbard, and Peter Prifti 1982 *Standard Albanian: A reference grammar for students.* Stanford, Calif.: Stanford University Press.

Weber, Albert 1964 *Zürichdeutsche Grammatik: Ein Wegweiser zur Guten Mundart.* Zürich: Schweizer Spiegel Verlag.