

The Neural Basis of Predicate-Argument Structure

James R Hurford,

Language Evolution and Computation Research Unit,
Department of Theoretical and Applied Linguistics,
University of Edinburgh.

Email: jim@ling.ed.ac.uk

URL: <http://www.ling.ed.ac.uk/~jim/>

Keywords: logic, predicate, argument, neural, object, dorsal, ventral, attention, deictic, reference.

1 Introduction

This article presents a step in the establishment of the following hypothesis:

Hypothesis: Basic elements in the structure of modern natural languages are derived (phylogenetically and ontogenetically) from primitive (prelinguistic) mental representations.

This idea is common enough¹, but cannot be taken further without (a) a characterization of ‘primitive mental representation’ independent of language itself, to avoid circularity, and (b) a plausible bridge between such representation and the structure of language. To fulfil this second condition, the strategy here is to assume that a basic element of first order predicate logic notation, *PREDICATE*(*x*), suitably embedded, can be systematically related to natural language structures, in the ways pursued by recent generations of formal semanticists of natural language². Fulfilling the other condition, establishing the language-independent validity of *PREDICATE*(*x*) as representing fundamental mental processes in both humans and non-human primates, will occupy the meat of this article. The article is original only in bringing together the fruits of others’ labours. Neurologists and psychologists will be familiar with much of the empirical research cited here, but I hope they will be interested in my claims for its wider significance. Linguists, philosophers and logicians might be excited to discover a new light cast on their subject by recent neurological research.

1.1 Why start from a logical notation?

Descriptions of the structure of languages are couched in symbolic terms. Although it is certain that a human’s knowledge of his/her language is implemented in neurons, and at an even more basic level of analysis, in atoms, symbolic representations are clearly well suited for the study of language structure. Neurologists don’t need logical formulae to represent the structures and processes that they find. Ordinary language, supplemented by diagrams, mathematical formulae, and neologized technical nouns, verbs and adjectives, is adequate for the expression of neurologists’ amazingly impressive discoveries. Where exotic technical notations are invented, it is for compactness and convenience, and their empirical content can always be translated into more cumbersome ordinary language (with the technical nouns, adjectives, etc.).

¹The idea is found in various forms in works such as Bickerton 1998; Kirby 2000; Kirby 1999; Hurford 2000b; Bennett 1976. Conjunctions of elementary propositions of the form *PREDICATE*(*x*) have been used by Batali as representations of conceptual structure pre-existing language in his impressive computer simulations of the emergence of syntactic structure in a population of interacting agents (Batali 2001). Note that the hypothesis here is not that **all** linguistic structure derives from prelinguistic mental representations. I argue elsewhere (Hurford in press) that in fact very little of the rich structure of modern languages directly mirrors any mental structure pre-existing language.

²For example, Montague 1970; Montague 1973; Parsons 1990; Kamp and Reyle 1993.

Logical notations, on the other hand, were developed by scholars theorizing in the neurological dark about the structure of language and thought. Languages are systems for the expression of thought. The sounds and written characters, and even the syntax and phonology, of languages can also be described in concrete ordinary language, augmented with diagrams and technical vocabulary. Here too, invented exotic notations are for compactness and convenience; which syntax lecturer has not paraphrased $S \rightarrow NP VP$ into ordinary English for the benefit of a first-year class? But the other end of the language problem, the domain of thoughts or meanings, has remained elusive to non-tautological ordinary language description. Of course, it is possible to use ordinary language to express thoughts — we do it all the time. But to say that ‘Snow is white’ describes the thought expressed by ‘Snow is white’ is either simply wrong (because description and expression are not equivalent) or at best uninformative. To arrive at an informative characterization of the relation between thought and language (assuming the relation to be other than identity), you need some characterization of thought which does not merely mirror language. So logicians have developed special notations for describing thought (not that they have always admitted or been aware that that is what they were doing). But, up to the present, the only route that one could trace from the logical notations to any empirically given facts was **back through** the ordinary language expressions which motivated them in the first place. A neurologist can show you (using suitable instruments which you implicitly trust) the synapses, spikes and neural pathways that he investigates. But the logician cannot illuminatingly bring to your attention the logical form of a particular natural sentence, without using the sentence itself, or a paraphrase of it, as an instrument in his demonstration. The mental adjustment that a beginning student of logic is forced to make, in training herself to have the ‘logician’s mindset’, is absolutely different in kind from the mental adjustment that a beginning student of a typical empirical science has to make. One might, prematurely, conclude that Logic and the empirical sciences occupy different universes, and that no wormhole connects them.

Despite its apparently unempirical character, logical formalism is not mere arbitrary stipulation, as some physical scientists may be tempted to believe. One logical notation can be more explanatorily powerful than another, as Frege’s advances show. Frege’s introduction of quantifiers binding individual variables which could be used in argument places was a great leap forward from the straightjacket of subject-predicate structure originally proposed by Aristotle and not revised for over two millennia. Frege’s new notation (but not its strictly graphological form which was awfully cumbersome) allowed one to explain thoughts and inferences involving a far greater range of natural sentences. Logical representations, systematically mapped to the corresponding sentences of natural languages, clarify enormously the system underlying much human reasoning, which, without the translation to logical notation, would appear utterly chaotic and baffling. This paper starts from a very simple component of the Fregean logical scheme, $PREDICATE(x)$, and proposes a neural interpretation for it. This is, to my knowledge, the first proposal of a ‘wormhole’ between the hitherto mutually isolated universes of formal logic and empirical neurology.

1.2 Why start from a formula with a variable as argument?

Here are two formulae of first order predicate logic (FOPL), with their English translations.

$CAME(john)$ (Translation: ‘John came’)

$\exists x[TALL(x) \ \& \ MAN(x) \ \& \ CAME(x)]$ (Translation: ‘A tall man came’)

The canonical fillers of the argument slots in predicate logic formulae are constants denoting

individuals, corresponding roughly to natural language proper names. In the more traditional schemes of semantics, no distinction between extension and intension is made for proper names. On many accounts, proper names have only extensions (namely the actual individuals they name), and do not have intensions (or ‘senses’). “What is probably the most widely accepted philosophical view nowadays is that they [proper names] may have reference, but not sense.” (Lyons 1977) “Dictionaries do not tell us what [proper] names mean — for the simple reason that they do not mean anything” (Ryle 1957) In this sense, the traditional view has been that proper names are semantically simpler than predicates. More recent theorizing has questioned that view.

In a formula such as $CAME(john)$, the individual constant argument term is interpreted as denoting a particular individual, the very same person on all occasions of use of the formula. FOPL stipulates by fiat this absolutely fixed relationship between an individual constant and a particular individual entity. Note that the denotation of the term is a thing in the world, outside the mind of any user of the logical language. It is argued at length by Hurford (2001) that the mental representations of proto-humans could not have included terms with this property. Protothought had no equivalent of proper names. Control of a proper name in the logical sense requires godlike omniscience. Creatures only have their sense organs to rely on when attempting to identify, and to reidentify, particular objects in the world. Where several distinct objects, identical to the senses, exist, a creature cannot reliably tell which is which, and therefore cannot guarantee control of the fixed relation between an object and its proper name that FOPL stipulates. It’s no use applying the same name to each of them, because that violates the requirement that logical languages be unambiguous. See Hurford (2001) and Hurford (1999) for more detailed arguments along these lines.

Interestingly, several modern developments in theorizing about predicates and their arguments complicate the traditional picture of proper names, the canonical argument terms. The dominant analysis in the modern formal semantics of natural languages (e.g. Montague 1970; Montague 1973) does not treat proper names in languages (e.g. *John*) like the individual constants of FOPL. For reasons having to do with the overall generality of the rules governing the compositional interpretation of all sentences, modern logical treatments make the extensions of natural language proper names actually **more complex than**, for example, the extensions of common nouns, which are 1-place predicates. In such accounts, the extension of a proper name is not simply a particular entity, but the set of classes containing that entity, while the extension of a 1-place predicate is a class. Concretely, the extension of *cat* is the class of cats, while the extension of *John* is the set of all classes containing John.

Further, it is obvious that in natural languages, there are many kinds of expressions other than proper names which can fill the NP slots in clauses.

“Semantically then PNs are an incredibly special case of NP; almost nothing that a randomly selected full NP can denote is also a possible proper noun denotation. This is surprising, as philosophers and linguists have often treated PNs as representative of the entire class of NPs. Somewhat more exactly, perhaps, they have treated the class of full NPs as representable . . . by what we may call *individual denoting* NPs.” (Keenan 1987)

This fact evokes one of two responses in logical accounts. The old-fashioned way was to deny that there is any straightforward correspondence between natural language clauses with non-proper-name subjects or objects and their translations in predicate logic (as Russell (1905) did). The modern way is to complicate the logical account of what grammatical subjects (and objects), including proper names, actually denote (as Montague did).

In sum, logical formulae of the type $CAME(john)$, containing individual constants, cannot be plausibly claimed as corresponding to primitive mental representations pre-existing human language. The required fixing of the designations of the individual constants ('baptism' in Kripke (1980)'s terms) could not be practically relied upon. Modern semantic analysis suggests that natural language proper names are in fact more complex than longer noun phrases like *the man*, in the way they fit into the overall compositional systems of modern languages. And while proper names provide the shortest examples of (non-pronominal) noun phrases, and hence are convenient for brief expository examples, they are in fact somewhat peripheral in their semantic and syntactic properties. Such considerations suggest that, far from being primitive, proper names are more likely to be relatively late developments in the evolution of language. In the historical evolution of individual languages, proper names are frequently, and perhaps always, derived from definite descriptions, as is still obvious from many, e.g. *Baker, Wheeler, Newcastle*. Hence, it is unlikely that any primitive mental representation contained any equivalent of a proper name, i.e. an individual constant. We thus eliminate formulae of the type of $CAME(john)$ as candidates for primitive mental representations.

This leaves us with quantified formulae, as in $\exists x[MAN(x) \ \&TALL(x)]$. Surely we can discount the universal quantifier \forall as a term in primitive mental representations. What remains is one quantifier, which we can take to be implicitly present and to bind the variable arguments of predicates. I propose that formulae of the type $PREDICATE(x)$ are evolutionarily primitive mental representations, for which we can find evidence outside language.

2 Neural correlates of $PREDICATE(x)$

It is high time to mention the brain. In terms of neural structures and processes, what justification is there for positing representations of the form $PREDICATE(x)$ inside human heads? We first set out some groundrules for correlating logical formulae, defined denotationally and syntactically, with events in the brain.

Representations of the form $PREDICATE(x)$ are here interpreted psychologically; specifically, they are taken to stand for the mental events involved when a human attends to an object in the world and classifies it perceptually as satisfying the predicate in question. In this psychologicistic view, it seems reasonable to correlate **denotation** with **stimulus**. Denotations belong in the world outside the organism; stimuli come from the world outside a subject's head. A whole object, such as a bird, can be a stimulus. Likewise, the properties of an object, such as its colour or shape, can be stimuli.

The two types of term in the $PREDICATE(x)$ formula differ in their denotations. An individual variable does not have a constant denotation, but is assigned different denotations on different occasions of use; and the denotation assigned to such a variable is some object in the world, such as a particular bird, or a particular stone or a particular tree. A predicate denotes a constant property observable in the world, such as greenness, roundness, or the complex property of being a certain kind of bird. The question to be posed to neurology is whether we can find separate neural processes corresponding to (1) the shifting, ad hoc assignment of a 'mental variable' to different stimulus objects in the world, not necessarily involving all, or even many, of the objects' properties, and (2) the categorization of objects, once they instantiate mental object variables, in terms of their properties, including more immediate perceptual properties, such as colour, texture, and motion, and more complex properties largely derived from combinations of these.

The syntactic structure of the $PREDICATE(x)$ formula combines the two types of term into a unified whole capable of receiving a single interpretation which is a function of the denotations of the

parts; this whole is typically taken to be an event or a state of affairs in the world. The bracketing in the *PREDICATE(x)* formula is not arbitrary: it represents an asymmetric relationship between the two types of information represented by the variable and the predicate terms. Specifically, the predicate term is understood in some sense to operate on, or apply to, the variable, whose value is provided beforehand. The bracketing in the *PREDICATE(x)* formula is the first, lowest-level, step in the construction of complex hierarchical semantic structures, as provided, for example, in more complex formulae of FOPL. The innermost brackets in a FOPL formula are always those separating a predicate from its arguments. If we can find separate neural correlates of individual variables and predicate constants, then the question to be put to neurology about the validity of the whole formula is whether the brain actually at any stage applies the predicate (property) system to the outputs of the object variable system, in a way that can be seen as the bottom level of complex, hierarchically organized brain activity.

2.1 Separate locating and identifying components in vision and hearing

The evidence cited here is mainly from vision. Human vision is the most complex of all sensory systems. About a quarter of human cerebral cortex is devoted to visual analysis and perception. There is more research on vision relevant to our theme, but some work on hearing has followed the recent example of vision research and arrived at similar conclusions.

2.1.1 Dorsal and ventral visual streams

Research on the neurology of vision over the past two decades has reached two important broad conclusions. One important message from the research is that vision is not a single unified system: perceiving an object as having certain properties is a complex process involving clearly distinguishable pathways, and hence processes, in the brain (seminal works are Trevarthen 1968; Ungerleider and Mishkin 1982; Goodale and Milner 1992).

The second important message from this literature, as argued for instance by Milner and Goodale (1995), is that much of the visual processing in any organism is inextricably linked with motor systems. If we are to carve nature at her joints, the separation of vision from motor systems is in many instances untenable. For many cases, it is more reasonable to speak of a number of visuomotor systems. Thus frogs have distinct visuomotor systems for orienting to and snapping at prey, and for avoiding obstacles when jumping (Ingle 1973; Ingle 1980; Ingle 1982). Distinct neural pathways from the frog's retina to different parts of its brain control these reflex actions. A lesion affecting the prey-grabbing system will leave the obstacle-avoidance system intact, and vice-versa.

Distinct visuomotor systems can similarly be identified in mammals:

“In summary, the modular organization of visuomotor behaviour in representative species of at least one mammalian order, the rodents, appears to resemble that of much simpler vertebrates such as the frog and toad. In both groups of animals, visually elicited orienting movements, visually elicited escape, and visually guided locomotion around barriers are mediated by quite separate pathways from the retina right through to motor nuclei in the brainstem and spinal cord. This striking homology in neural architecture suggests that modularity in visuomotor control is an ancient (and presumably efficient) characteristic of vertebrate brains.” (Milner and Goodale 1995)

Coming closer to our species, a clear consensus has emerged in primate (including human) vision research that one must speak of (at least) two separate neural pathways involved in the vision-mediated perception of an object. The literature is centred around discussion of two related

distinctions, the distinction between magno and parvo channels from the retina to the primary visual cortex (V1) (Livingstone and Hubel 1988), and the distinction between dorsal and ventral pathways leading from V1 to further visual cortical areas (Ungerleider and Mishkin 1982; Mishkin et al. 1983). These channels and pathways function largely independently, although there is some crosstalk between them (Merigan et al. 1991; Van Essen et al. 1992), and in matters of detail there is, naturally, complication (e.g. Johnsrude et al. 1999; Hendry and Yoshioka 1994; Marois et al. 2000) and some disagreement (e.g. Franz et al. 2000; Merigan and Maunsell 1993; Zeki 1993). It will suffice for discussion here to use the terms ‘dorsal stream’ and ‘ventral stream’ in a general way to include much of what has been discussed under the headings of ‘magno channel’ and ‘parvo channel’. (See Milner and Goodale (1995) for discussion of the magno/parvo-dorsal/ventral relationship.) It will also suffice here to collapse under the label ‘dorsal stream’ two separate pathways from the retina to posterior parietal cortex; one route passes via the lateral geniculate nucleus and V1, and the other bypasses V1 entirely, passing through the superior colliculus and pulvinar. (See Milner and Goodale (1995) .) The dorsal stream has been characterized as the ‘where’ stream, and the ventral stream as the ‘what’ stream. The popular ‘where’ label can be misleading, suggesting a single system for computing all kinds of spatial location; as we shall see, a distinction must be made between the computing of egocentric (viewer-centred) locational information and allocentric (other-centred) locational information. Bridgeman et al. (1979) use the preferable terms ‘cognitive’ (for ‘what’ information) and ‘motor-oriented’ (for ‘where’ information). Another suitable mnemonic might be the ‘looking’ stream (dorsal) and the ‘seeing’ stream (ventral). Looking is a visuomotor activity, involving a subset of the information from the retina controlling certain motor responses such as eye-movement, head and body orientation and manual grasping or pointing. Seeing is a perceptual process, allowing the subject to deploy other information from the retina to ascribe certain properties, such as colour and motion, to the object to which the dorsal visuomotor looking system has already directed attention.

“... appreciation of an object’s qualities and of its spatial location depends on the processing of different kinds of visual information in the inferior temporal and posterior parietal cortex, respectively.” (Ungerleider and Mishkin 1982)

“... both cortical streams process information about the intrinsic properties of objects and their spatial locations, but the transformations they carry out reflect the different purposes for which the two streams have evolved. The transformations carried out in the ventral stream permit the formation of perceptual and cognitive representations which embody the enduring characteristics of objects and their significance; those carried out in the dorsal stream, which need to capture instead the instantaneous and egocentric features of objects, mediate the control of goal-directed actions.” (Milner and Goodale 1995)

Experimental and pathological data support the distinction between visuo-perceptual and visuo-motor abilities.

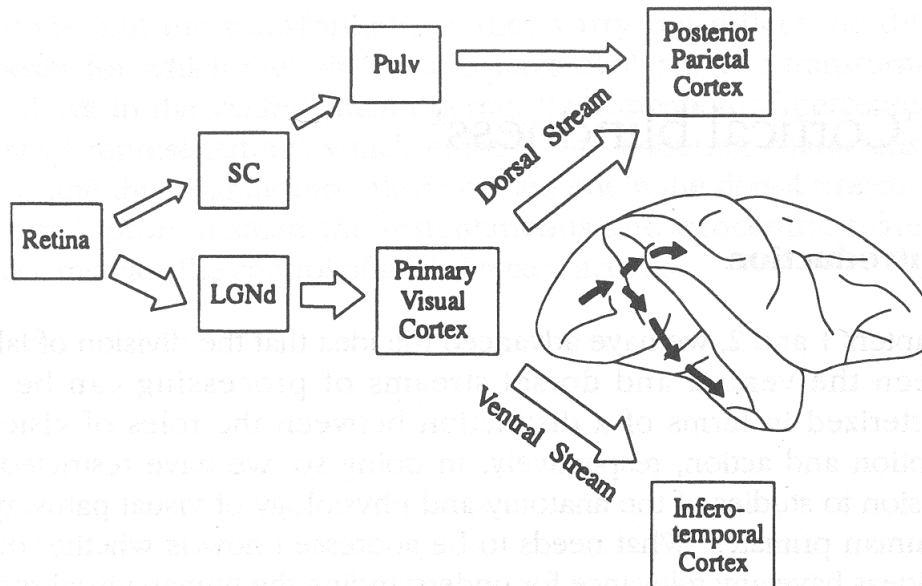


Figure 1. [From Milner and Goodale (1995).] Schematic diagram showing major routes whereby retinal input reaches dorsal and ventral streams. The inset [brain drawing] shows the cortical projections on the right hemisphere of a macaque brain. LGNd, lateral geniculate nucleus, pars dorsalis; Pulv, pulvinar nucleus; SC, superior colliculus.

Patients with cortical blindness, caused by a lesion to the visual cortex in the occipital lobe, sometimes exhibit ‘blindsight’. Sometimes the lesion is unilateral, affecting just one hemifield, sometimes bilateral, affecting both; presentation of stimuli can be controlled experimentally, so that conclusions can be drawn equally for partially and fully blind patients. In fact, paradoxically, patients with the blindsight condition are never strictly ‘fully’ blind, even if both hemifields are fully affected. Such patients verbally disclaim ability to see presented stimuli, and yet they are able to carry out precisely guided actions such as eye-movement, manual grasping and ‘posting’ (into slots). (See Goodale et al. (1994), Marcel (1998), Milner and Goodale (1995), Sanders et al. (1974), Weiskrantz (1986), Weiskrantz (1997).

See also Ramachandran and Blakeslee (1998) for a popular account).

These cited works on blindsight conclude that the spared unconscious abilities in blindsight patients are those identifying relatively low-level features of a ‘blindly seen’ object, such as its size and distance from the observer, while access to relatively higher-level features such as colour and motion is impaired. A complication to this picture arises from work on the recognition of facial expressions by blindsight patients (Morris et al. 1999; de Gelder et al. 2000; de Gelder et al. 1999; Heywood and Kentridge 2000). Facial expressions are complex and are generally thought to require considerable higher-level analysis. Yet detection of facial expressions (e.g. sad, happy, fearful, angry) is possible in some blindsight patients, suggesting that some aspects of this task also are performed via a pathway that, like at least one dorsal pathway, bypasses primary visual cortex. Classic blindsight cases arise with humans, who can report verbally on their inability to see stimuli, but parallel phenomena can be tested and observed in non-humans. Moore et al. (1998) summarizes parallels between residual vision in monkeys and humans with damage to V1.

A converse to the blindsight condition has also been observed, indicating a double dissociation between visually-directed grasping and visual discrimination of objects. Goodale et al.’s patient RV could discriminate one object from another, but was unable to use visual information to grasp odd-shaped objects accurately (Goodale et al. 1994). Experiments with normal subjects also

demonstrate a mismatch between verbally reported visual impressions of the comparative size of objects and visually-guided grasping actions. In these experiments, subjects were presented with a standard size-illusion-generating display, and asserted (incorrectly) that two objects differed in size; yet when asked to grasp the objects, they spontaneously placed their fingers exactly the same distance apart for both objects (Aglioti et al. 1995). Aglioti et al.'s conclusions have recently been called into question by Franz et al. (2000); see the discussion by Westwood et al. (2000) for a brief up-to-date survey of nine other studies on this topic.

Advances in brain-imaging technology have made it possible to confirm in non-pathological subjects the distinct localizations of processing for object recognition and object location (e.g. Aguirre and D'Esposito (1997) and other studies cited in this paragraph). Haxby et al. (1991), while noting the homology between humans and nonhuman primates in the organization of cortical visual systems into "what" and "where" processing streams, also note some displacement, in humans, in the location of these systems due to development of phylogenetically newer cortical areas. They speculate that this may have ramifications for "functions that humans do not share with nonhuman primates, such as language." Similar homology among humans and nonhuman primates, with some displacement of areas specialized for spatial working memory in humans, is noted by Ungerleider et al. (1998), who also speculate that this displacement is related to the emergence of distinctively human cognitive abilities.

The broad separation of visual pathways into ventral and dorsal has been tested against performance on a range of spatial tasks in normal individuals (Chen et al. 2000). Seven spatial tasks were administered, of which three "were constructed so as to rely primarily on known ventral stream functions and four were constructed so as to rely primarily on known dorsal stream functions" (380) For example, a task where subjects had to make a same/different judgement on pairs of random irregular shapes was classified as a task depending largely on the ventral stream; and a task in which "participants had to decide whether two buildings in the top view were in the same locations as two buildings in the side view" (383) was classified as depending largely on the dorsal stream. These classifications, though subtle, seem consistent with the general tenor of the research reviewed here, namely that recognition of the properties of objects is carried out via the ventral stream and the spatial location of objects is carried out via the dorsal stream. After statistical analysis of the performance of forty-eight subjects on all these tasks, Chen et al. conclude

"... the specialization for related functions seen within the ventral stream and within the dorsal stream have direct behavioral manifestations in normal individuals. ... at least two brain-based ability factors, corresponding to the functions of the two processing streams, underlie individual differences in visuospatial information processing." (Chen et al. 2000)

Chen et al. speculate that the individual differences in ventral and dorsal abilities have a genetic basis, mentioning interesting links with Williams syndrome (Bellugi et al. 1988; Frangiskakis et al. 1996).

Milner (1998) gives a brief but comprehensive overview of the evidence, up to 1998, for separate dorsal and ventral streams in vision. For my purposes, Pylyshyn (2000) sums it up best:

"... the most primitive contact that the visual system makes with the world (the contact that precedes the encoding of any sensory properties) is a contact with what have been termed visual objects or proto-objects ... As a result of the deployment of focal attention, it becomes possible to encode the various properties of the visual objects, including their location, color, shape and so on." (Pylyshyn 2000)

2.1.2 Auditory location and recognition

Less research has been done on auditory systems than on vision. There are recent indications that a dissociation exists between the spatial location of the source of sounds and recognition of sounds, and that these different functions are served by separate neural pathways.

Rauschecker (1997), Korte and Rauschecker (1993) and Tian and Rauschecker (1998) investigated the responses of single neurons in cats to various auditory stimuli. Rauschecker concludes

“The proportion of spatially tuned neurons in the AE [= anterior ectosylvian] and their sharpness of tuning depends on the sensory experience of the animal. This and the high incidence of spatially tuned neurons in AE suggests that the anterior areas could be part of a ‘where’ system in audition, which signals the location of sound. By contrast, the posterior areas of cat auditory cortex could be part of a ‘what’ system, which analyses what kind of sound is present.” (Rauschecker 1997)

Rauschecker suggests that there could be a similar functional separation in monkey auditory cortex.

Romanski et al. (1999) have considerably extended these results in a study on macaques using anatomical tracing of pathways combined with microelectrode recording. Their study reveals a complex network of connections in the auditory system (conveniently summarized in a diagram by Kaas and Hackett (1999)). Within this complex network it is possible to discern two broad pathways, with much cross-talk between them but nevertheless somewhat specialized for separate sound-localization and higher auditory processing, respectively. The sound localization pathway involves some of the same areas that are centrally involved in visual localization of stimuli, namely dorsolateral prefrontal cortex and posterior parietal cortex. Kaas and Hackett (1999), in their commentary, emphasize the similarities between visual, auditory and somatosensory systems each dividing along ‘what’ versus ‘where’ lines³. Graziano et al. (1999) have shown that certain neurons in macaques have spatial receptive fields limited to about 30cm around the head of the animal, thus contributing to a specialized sound-location system.

Coming to human audition, Clarke et al. (2000) tested a range of abilities in four patients with known lesions, concluding

“Our observation of a double dissociation between auditory recognition and localisation is compatible with the existence of two anatomically distinct processing pathways for non-verbal auditory information. We propose that one pathway is involved in auditory recognition and comprises lateral auditory areas and the temporal convexity. The other pathway is involved in auditory-spatial analysis and comprises posterior auditory areas, the insula and the parietal convexity.” (Clarke et al. 2000)

We have now considered both visual and auditory perception, and related them to object-location motor responses involving eye-movement, head-movement, body movement, and manual grasping. Given that when the head moves, the eyes move too, and when the body moves, the hands, head and eyes also move, we should perhaps not be surprised to learn that the brain has ways of controlling the interactions of these bodyparts and integrating signals from them into single coherent overall responses to the location of objects. Given a stimulus somewhere far round to one

³Belin and Zatorre (2000) suggest that the dorsal auditory pathway is involved in extracting the verbal message contained in a spoken sentence. This seems highly unlikely, as parsing a sentence appeals to higher-level lexical and grammatical information. The evidence they cite would only be relevant to the early pressure-sequence-to-spectrogram stages of spoken sentence processing.

side, we instinctively turn our whole body toward it; if the stimulus comes from not very far around, we may only turn our head; and if the stimulus comes from quite close to our front, we may only move our eyes. All this happens regardless of whether the stimulus was a heard sound or something glimpsed with the eye. Furthermore, as we turn our head or our eyes, light from the same object falls on a track across the retina, yet we do not perceive this as movement of the object. Research is beginning to close in on the areas of the brain that are responsible for this integrated location ability. Duhamel et al. (1992) found that the receptive fields of neurons in lateral intraparietal cortex are adjusted to compensate for saccades.

“One important form of spatial recoding would be to modulate the retinal information as a function of eye position with respect to the head, thus allowing the computation of location in head-based rather than retina-based coordinates. . . . by the time visual information about spatial location reaches premotor areas in the frontal lobe, it has been considerably recalibrated by information derived from eye position and other non-retinal sources.” (Milner and Goodale 1995)

The evidence that Milner and Goodale (1995) cite is from Galletti and Battaglini 1989; Andersen et al. 1985; Andersen et al. 1990; Gentilucci et al. 1983. Brotchie et al. (1995) present evidence that in monkeys

“ . . . the visual and saccadic activities of parietal neurons are strongly affected by head position. The eye and head position effects are equivalent for individual neurons, indicating that the modulation is a function of gaze direction, regardless of whether the eyes or head are used to direct gaze. These data are consistent with the idea that the posterior parietal cortex contains a distributed representation of space in body-centred coordinates” (Brotchie et al. 1995)

Gaymard et al. (2000) report on a pathological human case which “supports the hypothesis of a common unique gaze motor command in which eye and head movements would be rapidly exchangeable.” (819) Nakamura (1999) gives a brief review of this idea of integrated spatial representations distributed over parietal cortex. Parietal cortex is the endpoint of the dorsal stream, and neurons in this area both respond to visual stimuli and provide motor control of grasping movements (Jeannerod et al. 1995). In a study of vision-guided manual reaching, Carrozzo et al. (1999) have located a gradual transformation from viewer-centered to body-centered and arm-centered coordinates in superior and inferior parietal cortex. Graziano et al. (1997) discovered ‘arm+visual’ neurons in macaques, which are sensitive to both visual and tactile stimuli, and in which the visual receptive field is adjusted according to the position of the arm. Stricanne et al. (1996) investigated how lateral intraparietal (LIP) neurons respond when a monkey makes saccades to the remembered location of sound sources in the absence of visual stimulation; they propose that “area LIP is either at the origin of, or participates in, the transformation of auditory signals for oculomotor purposes.” (2071) Most recently, Kikuchi-Yorioka and Sawaguchi (2000) have found neurons which are active both in the brief remembering of the location of a sound and in the brief remembering of the location of a light stimulus. A further interesting connection between visual and auditory localization comes from Weeks et al. (2000), who find that both sighted and congenitally blind subjects use posterior parietal areas in localizing the source of sounds, but the blind subjects also use right occipital association areas originally intended for dorsal-stream visual processing. Egly et al. (1994) found a difference between left-parietal-lesioned and right-parietal-lesioned patients in an attention-shifting task.

The broad generalization holds that the dorsal stream does not provide all the information about an object that the brain eventually gets, but just about enough to direct attention to its location

and enable some motor responses to it. The ventral stream fills out the picture with further detailed information, enough to enable a judgement by the animal about exactly what kind of object it is dealing with (e.g. flea, hair, piece of grit, small leaf, shadow, nipple, or in another kind of situation brother, sister, father, enemy, leopard, human). A PET scan study (Martin et al. 1996) confirms that the recognition of an object (say, as a gorilla or a pair of scissors) involves activation of a ventral occipitotemporal stream. The particular properties that an animal identifies will depend on its ecological niche and lifestyle. It probably has no need of a taxonomy of pieces of grit, but it does need taxonomies of fruit and prey animals, and will accordingly have somewhat finely detailed mental categories for different types of fruit and prey. I identify such mental categories, along with non-constant properties, such as colour, texture and movement, which the ventral stream also delivers, with **predicates**.

Some information about an object, for example enough about its shape and size to grasp it, can be accessed via the dorsal stream, in a preattentive process. The evidence cited above from optical size illusions in normal subjects shows that information about size as delivered by the dorsal stream can be at odds with information about size as delivered by the ventral stream. Thus we cannot say that the two streams have access to exactly the same property, ‘size’; presumably the same is true for shape. Much processing for shape occurs in the ventral stream, after its divergence from the dorsal stream in V1 (Gross 1992); at the early V1 stage full shapes are not represented, but rather basic information about lines and oriented edges, as Hubel and Wiesel (1968) first showed. Something about the appearance of an object in peripheral vision draws attention to it. Once the object is focally attended to, we can try to report the ‘something’ about it that drew our attention. But the informational encapsulation (in the sense of Fodor (1983)) of the attention-directing reflex means that the more deliberative process of contemplating an object cannot be guaranteed to report accurately. And stimuli impinging on the retinal periphery trigger different processes from stimuli impinging on the fovea. Thus it is not clear whether the dorsal stream can be said to deliver properties, or mental predicates, at all. But the dorsal stream can be said to deliver representations of **objects**, provided we interpret ‘object’ in a minimal sense, to be further discussed in the next section. (At a broad-brush level, the differences between preattentive processes and focal attention have been known for some time, and are concisely and elegantly set out in Ch.5 of Neisser (1967).)

As an interim summary, the formula *PREDICATE*(x) is a simplifying schematic representation of the integration by the brain of two broadly separable processes. One process is the rapid delivery by the senses (visual and/or auditory) of information about the spatial location of a referent object relative to the body, represented in parietal cortex. The eyes, often the head and body, and sometimes also the hands, are oriented to the referent object, which becomes the instantiation of a mental variable. The other process is the somewhat slower analysis of the delivered referent object by the perceptual (visual or auditory) recognition subsystems in terms of its properties. The asymmetric relationship between the predicate and the variable, inherent in the bracketing of the formula, also holds of the two neural processes:

“From the genetical and functional perspectives the two modes of processing are asymmetrically related: while egocentric evaluation of ‘where’ need not take into account the identity of objects, the perception of ‘what’ usually proceeds through an intermediate stage in which objects are dynamically localized.” (Bridgeman et al. 1994)

2.1.3 Digression: An earlier proposal for ‘what’ and ‘where’ systems

Jackendoff and Landau (1992) and Landau and Jackendoff (1993)⁴ noticed the early neurological literature on ventral and dorsal streams and proposed a connection between the ‘where’/‘what’ dichotomy and the linguistic distinction between prepositions and common nouns. They correlate common nouns denoting classes of physical objects with information provided by the ventral stream, and prepositions with information provided by the dorsal stream. L&J emphasize the tentative and suggestive nature of their conclusions, but it will be useful to explain briefly why I believe their proposed correlations are incorrect, and to contrast their suggestions with mine.

Let us start with the proposed noun/ventral correlation. Nouns, as L&J correctly state, encode complex properties, such as being a dog. And categorization of objects, as when one recognizes a particular object as a dog, involves the ventral stream. This much is right. L&J emphasize the striking contrast between the enormous number of nouns in a language and the very restricted number of prepositions. It is this stark quantitative contrast which stands in need of explanation, and for which they invoke the neurological ‘what’/‘where’ distinction. Their reasoning is that the dorsal stream provides a bare minimum of information about the location of an object (no more than is encoded by the small inventory of prepositions in a language), while the ventral stream does all the rest of the work that may be necessary in categorizing it. This characterization of the relative amounts of linguistically expressible information provided by the respective streams certainly goes in the right direction (but is in fact, I will argue, an understatement).

However, a correlation of populous syntactic categories (such as noun) with the ventral stream, and a complementary correlation of sparsely populated categories (such as preposition) with the dorsal stream will not work. Consider adjectives. Adjectives are never as numerous in a language as nouns, many languages have only about a dozen adjectives, and some languages have none at all (Dixon 1982). Taking the numbers of nouns, adjectives and prepositions (or postpositions) across languages as a whole, one would be more likely to group adjectives with prepositions as relatively sparsely populated syntactic categories. But many of the properties typically expressed by adjectives, such as colour, are detected within the ventral stream. L&J might respond with the revised suggestion that the ventral stream processes both noun meanings and adjective meanings, leaving the difference in typical numbers of nouns and adjectives still unexplained, and this is fair enough, but it gets closer to the correlation proposed in the present paper between predicates generally and the ventral stream. Indeed when one considers all syntactic categories, rather than restricting discussion to just nouns and prepositions, it is clear that judgements corresponding to the meanings of many verbs (e.g. *move* and its hyponyms), and many adverbs (e.g. *fast* and similar words) are made in the ventral stream. Verbs are pretty numerous in languages, though not as numerous as nouns, while adverbs are much less numerous, and some languages don’t have adverbs at all. The relative population-size of syntactic categories does not correlate with the ventral/dorsal distinction.

Now consider L&J’s proposed dorsal/preposition correlation. Prepositions express predicates, many of which give spatial information, both egocentric and allocentric. L&J’s article naturally depended on the literature available at the time it was written, especially the classic Ungerleider and Mishkin (1982), which gave the impression of a distinction between ‘object vision’ and a single system of ‘spatial vision’. In a later very detailed critique of this work, Milner and Goodale (1995) devote several chapters to accumulating evidence that an egocentric system of “visual guidance of gaze, hand, arm or whole body movement”(118) is located in the posterior parietal region, while many other kinds of visual judgement, including computation of allocentric spatial information,

⁴Landau and Jackendoff (1993) is a more detailed version of Jackendoff and Landau (1992); I will refer here to the later paper, L&J.

are made using occipito-temporal and infero-temporal regions of cortex. “Perhaps the most basic distinction that needs to be made in thinking about spatial vision is between the locational coordinates of some object within the visual field and the relationship between the loci of more than one object.” (Milner and Goodale 1995) . Prepositions do not respect this distinction, being used indiscriminately for both egocentric (e.g. *behind me*) and allocentric (e.g. *behind the house*) information. Only information of the egocentric kind is computed in the dorsal stream.

Of course, as Bryant (1993) points out, there must be interaction between the systems for egocentric location and the building of allocentric spatial maps. Galati et al. (2000) is a recent fMRI study which begins to relate egocentric and allocentric functions to specific regions of cortex.

Both nouns and prepositions are predicates. I have argued that the categorical judgements of properties and relations involved in the application of all predicates to attended-to objects are mediated by the ventral stream. The key logico-linguistic distinction is between predicates and individual variables, not between different syntactic subdivisions of the class of predicates. Thus the logico-linguistic distinction correlated here with the neurological dorsal/ventral distinction is considerably more fundamental, and hence likely to be evolutionarily more primitive, than the distinction on which L&J focus. This idea is close to what I believe Bridgeman (1993), in his commentary on L&J, states: “... cognitive and [motor-oriented] spatial systems can be distinguished on a lower level than that of Landau & Jackendoff, a level that differentiates linguistic from nonlinguistic coding.” (240) Predicates are coded linguistically; the vast majority of words in a language correspond to predicates. In languages generally, only a tiny inventory of words, the deictic pronouns, such as *this* and *that* could be said to correlate directly with the individual variables x , y , z of simple formulae such as $\exists x[LION(x)]$, loosely translatable as *That is a lion*. In more complex examples, a case can be made that the logical variables correspond to anaphoric pronouns, as in *There was a lion and it yawned*. The deictic nature of the variables whose instantiations are delivered to posterior parietal cortex by the sensory ‘where’ systems will be the subject of section 4.

3 Attention to locations, features or objects?

Thus far, I have correlated logical predicates with perceived features, such as colour or shape, or more complex combinations of features, such as make up a particular face; and I have correlated the instantiations of individual variable arguments of predicates with whole objects attended to, such as a particular bird, stone or tree. But, one might ask, isn’t the representation of an object nothing more than the representation of a bundle of features?⁵ The notion of an object, as opposed to its features, is important for the central claim of this article, that modern neurology has revealed close correlates of the elements of the logical $PREDICATE(x)$ formula. In FOPL, individual variables are instantiated by whole objects, not by properties. Substantial evidence now exists that the primary targets of attentive processes are indeed whole objects, and not properties or features.

Beside the object/feature distinction, the object/location distinction must also be mentioned. Preattentive processes, operating largely through the dorsal stream, direct saccades to a location represented in a mental spatial map defined in terms of parts of the body. So, in a sense, attention is directed to a place, rather than to an object. But, except in cases of illusion or stimuli which vanish as soon as they are noticed, what the mind finds at the location to which attention is directed is an object. So what is held in attention, the object, or the location? Evidence has accumulated

⁵Bertrand Russell at times espoused the view that particulars are in reality nothing but bundles of properties (Russell 1940; Russell 1948; Russell 1959). See also Armstrong (1978). There is also a phenomenalist view that “the view that so-called material things, physical objects, are nothing but congeries of sensations” (Copi 1958).

in recent years that what is held in attention are objects, and not locations.

A paper by Duncan (1984), while by no means the first on this topic, is a good place to start a survey of recent research. Duncan distinguishes between object-based, discrimination-based and space-based theories of visual attention. “*Object-based* theories propose a limit on the number of separate objects that can be perceived simultaneously. *Discrimination-based* theories propose a limit on the number of separate discriminations that can be made. *Space-based* theories propose a limit on the spatial area from which information can be taken up.” (501) Space-based theories have been called ‘mental spotlight’ theories, as they emphasize the ‘illumination’ of a small circle in space. Duncan experimented with brief exposures to narrow displays, subtending less than one degree at the eye, consisting of two overlapping objects, an upright box (small or large) with a line (dotted or dashed) passing down through it. The box always had a small gap in one side, to left or right, and the line always slanted slightly to the right or the left. Subjects had to report judgements on two dimensions at a time, from the four possible dimensions *box(size)*, *box(gap)*, *line(tilt)* and *line(texture)*. “It was found that two judgments that concern the same object can be made simultaneously without loss of accuracy, whereas two judgments that concern different objects cannot. Neither the similarity nor the difficulty of required discriminations, nor the spatial distribution of information, could account for the results. The experiments support a view in which parallel, preattentive processes serve to segment the field into separate objects, followed by a process of focal attention that deals with only one object at a time.” (501) “The present data confirm that focal attention acts on packages of information defined preattentively and that these packages seem to correspond, at least to a first approximation, to our intuitions concerning discrete objects.” (514)

Duncan notes that object-based, discrimination-based and space-based theories are not mutually exclusive. This idea is repeated by some later writers (e.g. (Vecera and Farah 1994; Egly et al. 1994)), who discuss the possibilities of distinct systems of attention operating at different stages or levels (e.g. early versus late) or in response to different tasks (e.g. expectancy tasks versus selection tasks). The experimental evidence for space-based attention provided by these authors involves a different task from the task that Duncan set his subjects (although the experimental materials were very similar). Duncan asked his subjects for judgements about the objects attended to. The experiments suggesting space-based attention involved subjects being given a ‘precue’ (mostly valid, sometimes not) leading them to expect a stimulus to appear in a certain area, or on a certain object, and their task was simply to press a button when the stimulus appeared. Reaction times were measured and compared. Vecera and Farah (1994) suggest “Instead of attention being a single limitation or a single system, there may be different types of limitations or different types of attention that depend on the representations used in different tasks.” (153) This way of expressing it seems to me to depart from the useful distinction between **preattentive** processes and **focal attention**. Duncan’s subjects gave judgements about what was in their focal attention. In the precued experiments, the reaction times measured the subjects’ preattentive processes. As Egly et al. (1994) note, “. . . previous findings revealed evidence for both space-based and object-based components to visual attention. However, we note that these two components have been identified in very different paradigms.” (173) I will continue on the assumption that the cued reaction-time paradigm in fact tests preattentive processes. My question here is whether focal attention operates on objects, locations or features⁶.

⁶Egly et al. (1994) state “. . . we found evidence for both space-based and object-based components to covert visual orienting in normal observers. Invalid cues produced a cost when attention had to be shifted from the cue to another location within the same object, demonstrating a space-based component to attention. However, the costs of invalid cues were significantly larger when attention had to be shifted an equivalent distance and direction to part of another object, demonstrating an object-based component as well.” (173) This again conflates attention-shifting, a

A series of papers (Baylis and Driver 1993; Gibson 1994; Baylis 1994) takes up Duncan's theme of whether focal attention is applied to objects or locations. As with Duncan's experiments, subjects were required to make judgements about what they saw, but in this case reaction times were measured. In most of the experiments, the displays shown to subjects could be interpreted as either a convex white object against a black ground, or two partly concave black objects with a white space between them. Subjects had to judge which of two apices in the display was the lower. The apices could be seen as belonging to the same (middle) object, or to two different (flanking) objects.

"Position judgments about parts of one object were more rapid than equivalent judgments about two objects even though the positions to be compared were the same for one- and two-object displays. This two-object cost was found in each of five experiments. Moreover, this effect was even found when the one- and two-object displays were physically identical in every respect but parsed as one or two objects according to the subjects' perceptual set. . . . We propose that spatial information is routinely represented in two different ways in the visual system. First, a scene-based description of space represents the location of objects within a scene. Second, an object-based description is produced to describe the relative positions of parts of each object. Such a hierarchical representation of space may parallel the division of the primate visual system into a scene-based dorsal stream and an object-based ventral stream." (466-467)

Gibson (1994) suggested that these results could have been caused by a confound between the number of objects perceived and the concavity or convexity of the objects. Baylis (1994) replied to this objection with further experiments controlling against this possible confound, reinforcing the original conclusion that making a judgement about two objects is more costly than making a judgement about a single object, even when the displays are in fact physically identical.

Luck and Vogel (1997) presented subjects with visual arrays, with a slight delay between them, and asked them to report differences between the arrays. They summarize their conclusion as follows:

" . . . it is possible to retain information about only four colours or orientations in visual working memory at one time. However, it is also possible to retain both the colour and the orientation of four objects, indicating that visual working memory stores integrated objects rather than individual features. Indeed, objects defined by a conjunction of four features can be retained in working memory just as well as single-feature objects, allowing sixteen individual features to be retained when distributed across four objects. Thus, the capacity of visual working memory must be understood in terms of integrated objects rather than individual features." (279)

Valdes-Sosa et al. (1998)

" . . . studied transparent motion defined by two sets of differently colored dots that were interspersed in the same region of space, and matched in spatial and spatial frequency properties. Each set moved in a distinct and randomly chosen direction. We found that simultaneous judgments of speed and direction were more accurate when they

preattentive (and postattentive) process, with attention itself. These experiments relate only to attention-shifting, as the title of Egly et al. (1994)'s article implies. (Further, it would be interesting to know whether the distribution of RTs for the invalidly cued 'within-object' attention shifts was in fact bimodal. If so, this could suggest that subjects were sometimes interpreting the end of a rectangle as a **different object** from the rectangle itself, and sometimes not. In this case, the responses taken to indicate a space-based process could in fact have been object-based.)

concerned only one set than when they concerned different sets. Furthermore, appraisal of the directions taken by two sets of dots is more difficult than judging direction for only one set, a difficulty that increases for briefer motion. We conclude that perceptual grouping by common fate exerted a more powerful constraint than spatial proximity, a result consistent with object-based attention.” (B13)

The most recent and most ingenious experiment comparing object-based, feature-based and location-based theories of attention is Blaser et al. (2000). In this experiment, subjects were presented with a display consisting of two patterned patches (‘Gabors’), completely spatially superimposed. The trick of getting two objects to seem to occupy the same space at the same time was accomplished by presenting the patches in alternate video frames. The patches changed gradually, and with a certain inertia, along the three dimensions of colour, thickness of stripes and orientation of stripes. Subjects had to indicate judgements about the movements of these patches through ‘feature space’. In one experiment it was shown that observers are “capable of tracking a single object in spite of a spatially superimposed distractor”. In a second experiment, “observers had both an instruction and a task that encouraged them to attend and track two objects simultaneously. It is clear that observers did much worse in these conditions than in the within-object conditions, where they only had to attend and track a single object.”

The story so far, then, is that the brain interprets relatively abrupt discontinuities, such as change of orientation of a line, change of colour, change of brightness, together as constructing wholistic visual objects which are expected to share a ‘common fate’. It is these whole objects that are held in attention. A shift of attention from one object to another is costly, whereas a shift of attention from one feature of an object to another feature of the same object is less costly. This is consistent with the view underlying FOPL that the entities to which predicates apply are objects, and not properties (nor locations). In accepting this correlation between logic and neuropsychology we have, paradoxically, to abandon an ‘objective’ view of objects. No perceptible physical object is ever the same from one moment of its existence to the next. Every thing changes. Objects are merely slow events. What we perceive as objects is entirely dependent on the speed our brains work at. An object is anything that naturally attracts and holds our attention. But objects are what classical logicians have had in mind as the basic entities populating their postulated universes. The tradition goes back at least to Aristotle, with his ‘primary substances’ (= individual physical objects).

3.1 Word learning and ‘dumb’ attentional mechanisms

Finally in the this section, I mention research on children’s word-learning which invokes ‘dumb’ attentional mechanisms. In a functioning high-level organism, the information provided by the dorsal and ventral streams can be expected to be well coordinated, except in the unusual circumstances which generate illusions. Thus it would be surprising if a human being did not have mental predicates corresponding at least roughly to information of the type used by the dorsal stream. But humans have an enormous wealth of other predicates as well, bearing only indirect relationships to salient attention-drawing traits of objects. Humans classify and name objects (and substances) on the basis of properties at all levels of concreteness and salience. Landau et al. (1988), Smith et al. (1996), Landau et al. (1998a) and Landau et al. (1998b) report a number of experiments on adults’ and children’s dispositions to name familiar and unfamiliar objects. There are clear differences between children and adults, and between children’s responses to objects that they in some sense understand and to those that are strange to them. Those subjects with least conceptual knowledge of the objects presented, that is the youngest children, presented with strange objects,

tended to name objects on the basis of their shape. Smith et al. (1996) relate this disposition to the attention-drawing traits of objects:

“Given that an adult is attending to a concrete object and producing a novel name, children may interpret the novel name as referring to ‘whatever it is about the object that most demands attention.’ An attentional device that produces this result may work well enough to *start* a child’s learning of a specific object name.” (Smith et al. 1996)

This is not unexpected. Higher-level features and categories are learned, and once learned, can be applied in extending names to things. The youngest humans, having learned few or no higher-level categories, have only the most basic features to appeal to, those corresponding to information gleaned by the dorsal stream. See Bloom (2000) for a recent commentary on this literature, emphasizing a different theme, but not inconsistent with the hypothesis that children’s **earliest** naming tendencies capitalize strongly on attention-drawing traits of objects.

4 Computing deictic variables in vision, action and language

The previous section concerned the holding in attention of single whole objects. We can deal with several different objects in a single task, and take in scenes containing more than one object. How do we do this, and what are the limits on the number of different objects we can manage to ‘keep in mind’ at any one time?

The idea of objects of attention as the temporary instantiations of mental computational variables has been developed by Kahneman and Treisman (1992), Ballard et al. (1995), Ballard et al. (1997) and Pylyshyn (2000), drawing on earlier work including Kahneman and Treisman (1984), Ullman (1984), Agre and Chapman (1987) and Pylyshyn (1989). The idea behind this work is that the mind, as a computational device for managing an organism’s interactions with the world, has available for use at any time a small number of ‘deictic’ or ‘indexical’ variables. Pylyshyn (1989) calls such variables ‘FINSTs’, a mnemonic for ‘INSTantiation FINger’.

“A FINST is, in fact, a reference (or index) to a particular feature or feature cluster on the retina. However, a FINST has the following additional important property: because of the way clusters are primitively computed, a FINST keeps pointing to the ‘same’ feature cluster as the cluster moves across the retina. . . . The FINST itself does not encode any properties of the feature in question, it merely makes it possible to locate the feature in order to examine it further if needed.” (Pylyshyn 1989) :69-70 .

“This is precisely what the FINST hypothesis claims: it says that there is a primitive referencing mechanism for *pointing to* certain kinds of features, thereby maintaining their distinctive identity without either *recognizing* them (in the sense of categorizing them), or explicitly encoding their locations.” (Pylyshyn 1989) , [italics in original]

All practical tasks involve analysis of the scene of the task in terms of the principal objects concerned. The simple scene-descriptions of predicate logic, such as $\exists x, y[MAN(x) \& DOG(y) \& BEHIND(y, x)]$ (translated as *A dog is behind a man*) have direct counterparts in examples used by vision researchers of what happens in the brain when analyzing a visual scene. An early example from Ullman is:

“ Suppose, for example, that a scene contains several objects, such as a man at one location, and a dog at another, and that following the visual analysis of the man figure we shift our gaze and processing focus to the dog. The visual analysis of the man figure has been summarized in the incremental representation, and this information is still available at least in part as the gaze is shifted to the dog. In addition to this information we keep a spatial map, a set of spatial pointers, which tell us that the dog is at one direction, and the man at another. Although we no longer see the man clearly, we have a clear notion of what exists where. The ‘what’ is supplied by the incremental representations, and the ‘where’ by the marking map.” (Ullman 1984)

Since this passage was written in the early 1980’s, vision research has substantially developed the idea of separate ‘where’ and ‘what’ neural pathways, dorsal and ventral respectively, as surveyed above.

The everyday tasks of primates are plausibly envisaged in such terms. Activities such as fishing for termites with a stick and eating them, or building a sleeping nest in a tree, or collaborating with others in a hunt, all involve attention to different objects while performing the task. During the task, immediate attention is shifted from one thing to another, but the small number of principal things involved in the task are not put out of mind. Crucial information about them is stored as the contents of variables, or computational pointers. The termite-fishing chimpanzee at one moment attends to the termites caught on its stick, and guides them to its mouth. Meanwhile, it still holds, as part of the ongoing larger task, information about the hole in the termite mound, though it is not visually attending to it while putting the termites in its mouth. After eating the termites, visual attention is switched back to the hole in the termite mound, and the stick is manually guided into the hole. The chimpanzee need not rediscover the properties of the hole (e.g. its size and orientation), because these properties have been stored as the contents of a computational variable.

(Managing scenes with several objects necessitates control of sameness and difference. The ape doing some practical task with several objects does not need to be able to distinguish these objects in principle from all other objects in the world, but certainly does need to distinguish among the objects themselves. This is the simple seed from which the more advanced concept of a unique-in-the-world individual may grow.)

An idea very similar to Pylyshyn’s FINSTs, but slightly different in detail, is proposed by Kahneman and Treisman (1984) and Kahneman and Treisman (1992). These authors hypothesize that the mind sets up temporary ‘object files’ in which information about objects in a scene is stored. The object files can be updated, as the viewer tracks changes in an object’s features or location. It is emphasized that the information stored in temporary object files is not the same as that which may be stored in long term memory. But the information in object files can be matched with properties associated with objects in long term memory, for such purposes as object recognition. When (or shortly after) objects disappear from the current scene, their object files are discarded. A file full of information is not a variable. In discussing the relationship between object files and Pylyshyn’s FINSTs, Kahneman and Treisman (1992) suggest that “a FINST might be the initial phase of a simple object file before any features have been attached to it”. (217) This correspondence works well, apart from a reservation, which Kahneman and Treisman (1992) note, involving the possibility of there being objects with parts that are also objects. This is a detail that I will not go into here. An ‘empty’ object file, available for information to be put into it, is computationally an uninstantiated variable, provided that it can be identified and distinguished from other such files that are also available and that may get different information put into them. The fact that object files can be updated, are temporary, and can be discarded for re-use with completely new values, underlines their status as computational variables used by the mind for the

short-term grasping of scenes.

Kahneman and Treisman (1992) “assume that there is some limit to the number of object files that can be maintained at once”. (178) Ballard et al. stress that computational efficiency is optimized if the number of such variables is small. Luck and Vogel (1997) demonstrate a limit of four objects in visual working memory (and propose an interesting explanation in terms of the “oscillatory or temporally correlated firing patterns among the neurons that code the features of an object” (280)). Pylyshyn assumes “a pool of four or five available indexes” (Pylyshyn 2000) . It is perhaps at first helpful to concretize these ideas by identifying the available variables in the same way as logicians do, by the letters w , x , y and z . Neither logicians nor vision researchers wish to be tied to the claim that the mind can only handle a maximum of four variables, but hardly any examples given by them ever involve more than four separate variables. So it would seem for many practical purposes that about four variables are enough. In performing an everyday task, then, a creature such as a primate mentally juggles a parsimonious inventory of variables, w , x , y , z ,

This small inventory of variables can explain other known size-limitations in humans and non-human primates. The upper limit of subitizing in humans is around 4; given a quick glance at a group of objects, a human can guess accurately how many there are, without explicit counting, up to a limit of about 4 or 5 (see Gelman and Gallistel (1978), Antell and Keating (1983), Starkey and Cooper (1980), Russac (1983), Schaeffer et al. (1974), Mandler and Shebo (1982) for some relevant studies). Both Ullman (1984) and Pylyshyn (2000) make the connection between subitizing (which Ullman calls ‘visual counting’) and the marking or indexing of locations in a scene. Trick and Pylyshyn (1994, Trick and Pylyshyn (1993) explain for the natural limit of subitizing in terms of the number of objects that can be involved in ‘pre-attentive’ processing in vision. Dehaene (1997), in work on the numerical competences of many species, finds a natural difference between low numerosities up to about 3 or 4, and higher ones. For details of how this natural discontinuity at around 4 in the number sequence is reflected in the numerals, adjectives and nouns of many human languages, see Hurford (2000a), Hurford (1987).

The simple clauses of human languages are constrained to a maximum of about 4 or 5 core arguments; indeed most clauses have fewer than this. Presumably this reflects the structure of the underlying mental propositions. Conceivably, one could analyze the content of a complex sentence, such as *The cat chased the mouse that stole the cheese that lay in the house that Jack built* as having a single predicate *CHASE-STEAL-LIE-BUILD* and five arguments (the cat, the mouse, the cheese, the house and Jack). But it is more reasonable to suppose that the grammatical structure of such embedded natural language clauses reflects a mental structure involving a nesting of separate propositions, each with its own simple predicate expressing a relation between just two arguments (which may be shared with other predicates)⁷.

Ballard et al. give grounds why the number of variables juggled in computing practical tasks must be small (typically no more than three). Of course most sentences in human languages are not direct representations of any practical task on the part of the speaker, like ‘Put the stick in the hole’. Humans exchange declarative information about the world for use at later times, e.g. ‘Your mother’s coming on Tuesday’. But mental scene-descriptions are necessary for carrying out practical tasks of the kind that primates are capable of, and therefore pre-exist language phylogenetically. It is plausible that the type of scene-descriptions used by non-human primates would be reused for more complex cognitive, and ultimately linguistic, purposes. I suggest that the limitation of elementary propositions to no more than about three arguments, and the typical use of even fewer

⁷There is presumably a complex ecological balance between the information carried by a mental predicate and its frequency of use in the mental life of the creature concerned. Complex relations, if occurring frequently enough, might be somehow compressed into unitary mental predicates. An analogous case in language would be the common compressing of *CAUSE(a, PRED(b))* into a form with a single causative verb.

arguments, derives from the considerations of computational efficiency advanced by Ballard et al.⁸

The marking, or indexing, of spatial locations in a visually analyzed scene, as described by Ullman and Pylyshyn, has a direct analog in human signed languages. Where spoken languages establish the existence of discourse referents with noun phrases, and subsequently use definite pronouns and descriptions to re-identify these referents, signed languages can use a directly visuo-spatial method of keeping track of discourse referents. A user of British Sign Language, for instance, on telling a story involving three participants, will, on introducing them into the discourse, assign them a position in the signing space around him. On referring back to these individuals, he will point to the appropriate spatial position (equivalent to saying ‘this one’ or ‘that one’).

“[In many sign languages] Anaphoric pronouns can only occur following the localization of the referent noun in the location assigned to the pronoun. Nouns articulated in the space in front of the body are, for example, moved to third person space; nouns located on a body part would be followed by an indexing of third person space. This assignment of location to a referent . . . then continues through the discourse until it is changed. To indicate anaphoric reference, the signer indexes the location previously assigned to that referent. . . .

The operation of anaphora . . . can be seen in the following BSL example ‘The woman keeps hitting the man’. In this, the sign MAN is articulated with the left hand, followed by the ‘person’ classifier, located to fourth person space. The left hand remains in the ‘person’ classifier handshape and fourth person location, while the remainder of the sentence is signed. The sign WOMAN is articulated with the right hand, followed by the ‘person’ classifier, located to third person space. The verb HIT, an agreement verb, is then articulated, moving on a track from the subject (third person) to object (fourth person).” (Woll and Kyle 1994)

See also Liddell (1990), McDonald (1994), Padden (1990). For the sign language recipient, the experience of decoding a signed scene-describing utterance closely parallels the visual act of analyzing the scene itself; in both cases, the objects referred to are assigned to different locations in space, which the recipient/observer marks.

There is a further parallel between linguistic deictic terms and the deictic variables invoked by vision researchers. As we have seen, Pylyshyn postulates “a pool of four or five available indexes”, and Ballard et al. emphasize that most ordinary visually guided tasks can be accomplished with no more than three deictic variables. The deictic terms of natural languages are organized into internally contrastive subsystems: English examples are *here/there*, *now/then*, *yesterday/today/tomorrow*, *Past-tense/non-Past-tense*, *this/that*, *these/those*. Some languages are slightly richer in their deictic systems than English. Japanese, for instance, distinguishes between three demonstratives, *kono* (close to the speaker), *sono* (close to the listener, or previously referred to), and *ano* (reasonably distant from both speaker and listener); this three-way distinction in demonstrative adjectives is paralleled by three-way distinctions in *kore/sore/are* (demonstrative pronouns) and *koko/soko/asoko* and *kochira/sochira/achira* (adverbs of place and direction respectively). Spanish likewise makes a three-way distinction in demonstratives, *este/ese/aquel*, with slightly different meanings from the Japanese. No language, however, provides more than five contrasting deictic terms in any subsystem. This corresponds nicely to the ‘available pool’ of visual deictic variables. In an utterance entirely concerning objects in the vicinity of the speech-situation, none of which are identified by any predicate/property, there is a limit to how many separate things a speaker or hearer can keep

⁸The claim in the text is not about memory limitations involved in parsing linguistic strings; it is about how many arguments the elementary propositions in the mind of a pre-linguistic creature could have.

track of, with expressions equivalent to ‘this one near me’, ‘that one near you’, ‘that one yonder’, and so on. Pylyshyn (1989) explicitly relates his FINST devices to the indexical pronouns *here* and *there*, and suggests that FINSTs provide a semantics for such expressions.

The provision by the brain’s sensory/perceptual systems of a pool of about four or five variables for ad hoc deictic assignment to objects in the accessible environment, and the separate processes of perceptual categorization of the objects so identified, constitutes an early system for the representation of scenes. This system was based on multiple instances of (or conjunctions of) propositions of the form *PREDICATE(x)*, involving up to about four different variables. An example of such a scene-description might be

APE(x) & STICK(y) & MOUND(z) & HOLE(w) & IN(w, z) & PUT(x, y, w)

translating to *An ape puts a stick into a hole in a mound*. This translation is given here just for the convenience. So far, we have made no move to suggest how such non-linguistic mental representations came to be externalized in the shared communication system of a community. If we are talking about language at all, it is, so far, only private language. Nevertheless, given the genetic homogeneity of communities of primates, it is highly likely that what happens in the brain of one animal on seeing a scene is represented very similarly in the brains of its fellow troop members. The simply structured internal representations provide a preadaptive platform on which a simple public language could develop⁹

5 Common ground of neurology, linguistics and philosophy

I have made the connection between neural processing of visual scenes and mental representations of propositions as expressed by simple natural language clauses. This same connection is everywhere heavily implicit, though not explicitly defended, in the writing of the vision researchers cited here. In particular, the four terms, ‘deictic’, ‘indexical’, ‘refer’ and ‘semantic’, borrowed from linguistics and the philosophy of language, have slipped with remarkable ease and naturalness into the discussion of visual processing. ‘Deictic’ as a grammatical term has a history going back to the Greek grammarians (who used *δεικτικος* ‘deiktikos’; see Lyons (1977):636 for a sketch of this history), indicating a ‘pointing’ relationship between words and things. ‘Deictic’ and ‘indexical’ are equivalent terms. Agre and Chapman (1987) apply the term ‘indexical’ to computational entities invoked by a program designed for fast, efficient, planning-free interaction with a model world. These entities “...are not logical categories because they are indexical: their extension depends on the circumstances. In this way, indexical-functional entities are intermediate between logical individuals and categories” (Agre and Chapman 1987).¹⁰ The parallels between efficient computing for fast local action and the efficient fast analysis of visual scenes, using deictic or indexical entities, are later taken up by a small but growing number of writers (e.g. Ballard et al. 1995; Ballard et al. 1997; Pylyshyn 2000) arguing the advantages of reorientating perceptual and cognitive research along ‘situated’ or ‘embodied’ lines.

Similarly, the term ‘refer’ is typically used in ordinary language, and consistently in the more technical discourse of linguists and philosophers, with a linguistic entity, such as a word, as one of its arguments, and a thing in the world as another argument, as in ‘*Fido* refers to my dog’. Strawson’s classic article “On Referring” (Strawson 1950) is all about statements and sentences of ordinary

⁹See Batali (2001) for a computer simulation of the emergence of public language from representations of exactly this form.

¹⁰Agre and Chapman (1987) do not, as stated by Ballard et al. (1995), use the term ‘deictic’.

languages; for Searle (Searle 1979) and other speech act theorists, referring is a speech act. Linguists prefer to include a third argument, the speaker, as in ‘He referred to me as *Jimmy*’. Manually pointing to an object, without speaking, might be considered by some linguists and philosophers to be at best a marginal case of referring, especially where the intention is to draw attention of another to the object. But notice how easily this and other originally linguistic terms (‘demonstrative’, ‘indexical’) are interpreted when applied to a visual, entirely non-linguistic process:

“... the visual system ... needs a special kind of direct reference mechanism to refer to objects without having to encode their properties. ... This kind of direct reference is provided by what is referred to as a demonstrative, or more generally, an indexical¹¹.”
(Pylyshyn 2000)

The central idea involved in linguistic and vision-oriented and activity-oriented uses of the terms ‘deictic’, ‘indexical’ and ‘refer’ is **attention**. In all cases, be it a monkey swivelling its eyes toward a target, an ape grasping for an object, or a human referring to an object with a demonstrative pronoun, the organism is attending to an object. This is the archetypal sense of ‘refer-’; the linguist’s preferred usage of ‘refer-’, involving a speaker, is closer to the archetypal sense than the 20th century logician’s, for whom reference is a relation between words and things, without mediation by any agent’s mind. But the linguist’s and the philosopher’s restriction of ‘referring’ to a necessarily linguistic act misses what I claim is the phylogenetic, prelinguistic origin of referring.

Classically, semantics is said to involve a relation between a representation and the world, without involvement of any user of that representation (e.g. a speaker) (Carnap 1942; Morris 1938; Morris 1946). Thus the relation of denotation between a proper name and its referent, or between a predicate and a set of objects, is traditionally the concern of semantics. Vision researchers use the term ‘semantic’ with no sense of a relation involving linguistic entities. Jeannerod et al. (1995) identify events in the dorsal stream with pragmatics and events in the ventral stream with semantics:

“In humans, neuropsychological studies of patients with lesions to the parietal lobule confirm that primitive shape characteristics of an object for grasping are analyzed in the parietal lobe, and also demonstrate that this ‘pragmatic’ analysis of objects is separated from the ‘semantic’ analysis performed in the temporal lobe.” (Jeannerod et al. 1995)

Likewise Milner and Goodale (1995) write of the “content or semantics” of non-verbal interactions with the world, such as putting an object in a particular place. Further, “... even after objects have been individuated and identified, additional semantic content can be gleaned from knowing something about the relative location of the objects in the visual world.” (Milner and Goodale 1995) The central idea linking linguists’, philosophers’ and vision researchers’ use of ‘semantic’ is the idea of **information** or **content**. For us modern humans, especially the literate variety, language so dominates our lives that we tend to believe that language has a monopoly of information and content. But of course there is, potentially, information in everything. And since the beginning of the electronic age, we now understand how information can be transmitted, transformed and stored with wires, waves and neurons. Information about the relative location of the objects in a visual scene, or about the properties of those objects, represented in a perceiver’s brain, has the same essential quality of ‘aboutness’, a relation with an external world, that linguists and philosophers identify with the semantics of sentences. Those philosophers and linguists who have insisted that semantics is a relation between a language and the world, without mediation by a

¹¹Indeed, this quoted sentence contains the stem ‘refer-’ four times, three times alluding to a visual process and once to a linguistic convention; probably few readers remark on the coincidence as in any way disturbing.

representing mind have eliminated the essential middleman between language and the world. The vision researchers have got it more right, in speaking of the ‘semantics’ of neural representations, regardless of whether any linguistic utterance is involved. It is on the platform of such neural representations that language can be built.

An evolutionary history of reference can be envisaged, in which reference as a relation between the mind and the world is the original.

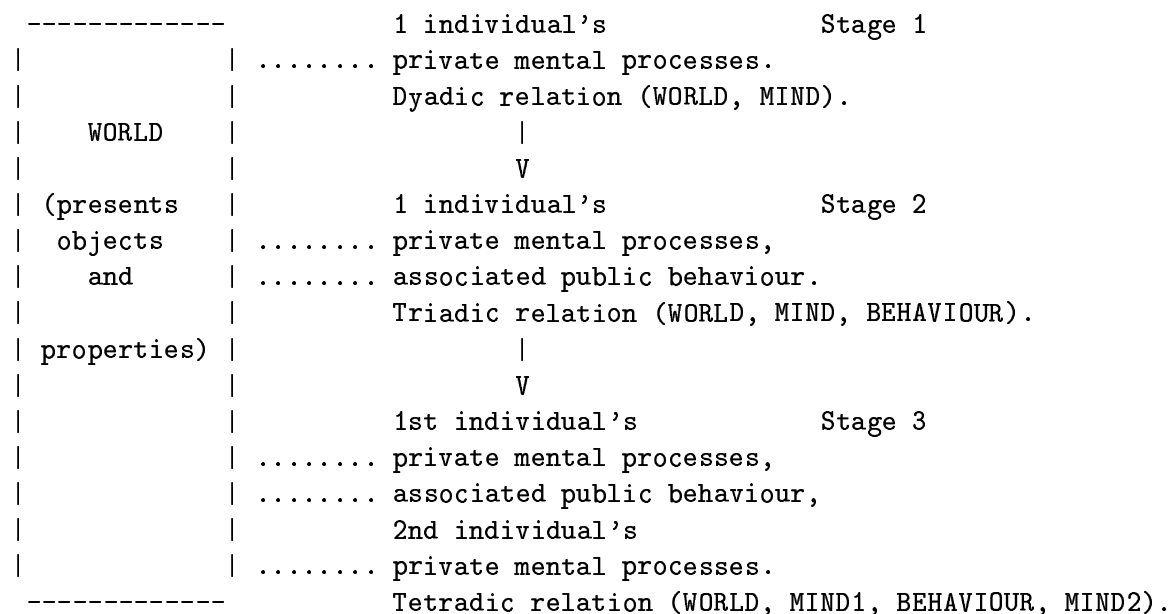


Figure 2: The evolution of reference. The relationship between mental processes and the world is the original and enduring factor. The last stage is successful reference as understood by linguists, and as manifested by people speaking natural languages. The stages may overlap, in that further evolution of one stage may continue to complexify after evolution of a later stage has commenced.

At present, the dual use of such terms as ‘deictic’ and ‘refer’ for both linguistic and visual processes is possibly no more than a metaphor. The mere intuitive plausibility of the parallels between the visual and the linguistic processes is not as good as empirical evidence that the brain in some way treats linguistic deictic variables and visual deictic variables in related ways. Possibly the right kind of evidence could be forthcoming from imaging studies, but the picture is sure to be quite complicated.

6 Wrapping up

The neural correlates of *PREDICATE(x)* can be found not only in humans but also in primates and probably many other higher mammals. Thus, as far as human evolution is concerned, this form of mental representation is quite ‘primitive’, an early development not unique to our species. It can be seen as building on an earlier stage (evident, for example, in frogs) in which the only response to an attention-drawing stimulus was some immediate action. A fundamental development in higher mammals was to augment, and eventually to supplant, the immediate motor responses of

a sensorimotor system with internalized, judgmental responses which could be a basis for complex inferential processes working on material stored in long term memory. Rather than ‘If it moves, grab it’, we begin to have ‘If it catches your attention, inspect it carefully and figure out what do to with it’, and later still ‘If you notice it, remember what is important about it for later use.’

Simple early communicative utterances could be reports of a *PREDICATE(x)* experience. For example, the vervet chatter could signify that the animal is having a *SNAKE(x)* experience, i.e. has had its attention drawn to an object which it recognizes as a snake. Primitive internal representations, I have claimed, contain two elements, a deictic variable and a categorizing predicate. Nowhere in natural non-human communication do we find any two-term signals in which one term conveys the deictic element and the other conveys the mental predicate. But some simple sentences in some human languages have just these elements and no other. Russian and Arabic provide clear examples.

eto čelovek
 DEICTIC MAN
 “This is a man.” (Russian)

di sahl
 DEICTIC EASY
 “That is easy” (Egyptian Arabic)

Even if the internal representations of animals are structured in the *PREDICATE(x)* form, there would be no evolutionary pressure to structure the corresponding signals into two parts until the number of possible mental combinations of predicates and variables exceeded the total number of predicates and variables, counted separately (Nowak et al. 2000). If the category of things that are pointed to in a given direction is always the same, there is no pressure for the signal to differentiate the direction from the category.

Possession of the *PREDICATE(x)* form of representation is evidently not sufficient to propel a species into full-blown syntactic language. There is much more to human language than predicate-argument structure, but predicate-argument structure is the semantic foundation on which all the rest is built. A number of other factors need to come together for language to evolve. Only the sketchiest mention will be given of such factors here, but they include (a) the transition from private mental representations to public signals; (b) the transition from involuntary to voluntary control; (c) the transition from epigenetically determined to learned and culturally transmitted systems; (d) the convergence on a common code by a community; (e) the evolution of control of complex hierarchically organized signalling behaviour; (f) the development of deictic here-and-now talk into definite reference and proper naming capable of evoking events and things distant in time and space.

I have argued that *PREDICATE(x)* is a reasonable schematic way of representing what happens in an act of perception. It is another step, not taken here, to show that a similar kind of logical form is also appropriate for representing stored episodic memories. A form in which only individual variables can be the arguments of predicates might be too restrictive. Here, let me, finally, mention the ‘Aristotle problem’. Aristotle and his followers for the next two millennia took the basic semantic representation to be Subject+Predicate, where the same kind of term could fill both the Subject slot and the Predicate slot. Thus, for example, a term such as *man* could be the subject of *The man died* and the predicate of *Plato is a man*. Kant’s characterization of analytic judgements relies on subject terms being of the same type as predicate terms. “Analytical judgments express nothing in the predicate but what has been already actually thought in the concept of the subject, though

not so distinctly or with the same (full) consciousness". (Kant 1905)¹². FOPL is more distanced from the surface forms of natural languages, and the same terms cannot be both arguments (e.g. subjects) and predicates. It remains to provide an explanation for the typical structure of modern languages, organized around the Noun/Verb dichotomy. I suspect that an explanation can be provided in terms of a distinction between predicates which denote invariant properties of objects, such as being a dog, and more ephemeral properties, such as barking. But that is another story.

References

- Aglioti, S., M. A. Goodale, and J. F. X. DeSouza (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology* 5, 679–685.
- Agre, P. E. and D. Chapman (1987). Pengi: an implementation of a theory of activity. *Proc., AAAI 87*, 268–272.
- Aguirre, G. K. and M. D’Esposito (1997). Environmental knowledge is subserved by separable dorsal/ventral neural areas. *The Journal of Neuroscience* 17(7), 2512–2518.
- Andersen, R. A., R. M. Bracewell, S. Barash, J. W. Gnadt, and L. Fogassi (1990). Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7A of macaque. *Journal of Neuroscience* 10, 1176–1196.
- Andersen, R. A., G. K. Essick, and R. M. Siegel (1985). Encoding of spatial location by posterior parietal neurons. *Science* 230, 456–458.
- Antell, S. E. and J. P. Keating (1983). Perception of numerical invariance in neonates. *Child Development* 54, 695–701.
- Armstrong, D. M. (1978). *Nominalism and Realism: Universals and Scientific Realism, Volume I*. Cambridge: Cambridge University Press.
- Ballard, D. H., M. M. Hayhoe, and J. B. Pelz (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience* 7:1, 66–80.
- Ballard, D. H., M. M. Hayhoe, P. K. Pook, and R. P. N. Rao (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences* 20(4), 723–xxx.
- Batali, J. (2001). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In T. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.
- Baylis, G. C. (1994). Visual attention and objects: Two-object costs with equal convexity. *Journal of Experimental Psychology: Human Perception and Performance* 20(1), 208–212.
- Baylis, G. C. and J. Driver (1993). Visual attention and objects: evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance* 19(3), 451–470.
- Belin, P. and R. J. Zatorre (2000). ‘What’, ‘where’ and ‘how’ in auditory cortex. *Nature Neuroscience* 3(10), 965–966.
- Bellugi, U., H. Sabo, and J. Vaid (1988). Spatial deficits in children with Williams syndrome. In J. Stiles-Davis, M. Kritchevsky, and U. Bellugi (Eds.), *Spatial Cognition: Brain Bases and Development*, pp. 273–298. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

¹²Notice in this quotation from Kant a faint forerunner of the idea developed in this article, that predicates are associated with processes more accessible to consciousness than arguments.

- Bennett, J. (1976). *Linguistic Behaviour*. Cambridge: Cambridge University Press.
- Bickerton, D. (1998). Catastrophic evolution: the case for a single step from protolanguage to full human language. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (Eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge: Cambridge University Press.
- Blaser, E., Z. Pylyshyn, and A. O. Holcombe (2000). Tracking and object through feature space. *Nature* 408, 196–199.
- Bloom, P. (2000). *How Children Learn the Meanings of Words*. Learning, Development and Conceptual Change. MIT Press.
- Bridgeman, B. (1993). Spatial and cognitive vision differentiate at low levels, but not in language. *Behavioral and Brain Sciences* 16(2), 240.
- Bridgeman, B., S. Lewis, G. Heit, and M. Nagle (1979). Relationship between cognitive and motor-oriented systems of visual position perception. *Journal of Experimental Psychology: Human Perception and Performance* 5, 692–700.
- Bridgeman, B., A. H. C. van der Heijden, and B. M. Velichkovsky (1994). A theory of visual stability across saccadic eye movements. *Behavioural and Brain Sciences* 7(2), 247–292.
- Brotchie, P. R., R. A. Anderson, L. H. Snyder, and S. J. Goodman (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* 375, 232–235.
- Bryant, D. J. (1993). Frames of reference in the spatial representation system. *Behavioral and Brain Sciences* 16(2), 241–242.
- Carnap, R. (1942). *Introduction to Semantics*. Cambridge, MA: MIT Press.
- Carrozzo, M., J. McIntyre, M. Zago, and F. Lacquaniti (1999). Viewer-centered and body-centered frames of reference in direct visuomotor transformations. *Experimental Brain Research* 129(2), 201–210.
- Chen, J., J. Myerson, S. Hala, and A. Simon (2000). Behavioral evidence for brain-based ability factors in visuospatial information processing. *Neuropsychologia* 38, 380–387.
- Clarke, S., A. Bellmann, R. A. Meuli, G. Assal, and A. J. Steck (2000). Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. *Neuropsychologia* 38, 797–807.
- Copi, I. (1958). Artificial languages. In P. Henle (Ed.), *Language, Thought and Culture*. Ann Arbor, Michigan: University of Michigan Press.
- de Gelder, B., J. Vroomen, G. Pourtois, and L. Weiskrantz (1999). Non-conscious recognition of affect in the absence of striate cortex. *NeuroReport* 10(18), 3759–3763.
- de Gelder, B., J. Vroomen, G. Pourtois, and L. Weiskrantz (2000). Affective blindsight: are we blindly led by emotions? *Trends in Cognitive Sciences* 4(4), 126–127.
- Dehaene, S. (1997). *The Number Sense*. New York: Oxford University Press.
- Dixon, R. M. W. (1982). *Where have all the Adjectives gone? : and other Essays in Semantics and Syntax*. Berlin: Mouton.
- Duhamel, J.-R., C. Colby, and M. Goldberg (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* 255, 90–92.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General* 113(4), 501–517.

- Egely, R., J. Driver, and R. D. Rafal (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General* 123(2), 161–177.
- Fodor, J. A. (1983). *The Modularity of Mind : an Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Frangiskakis, J. M., A. K. Ewart, C. A. Morris, C. B. Mervis, J. Bertrand, B. F. Robinson, B. P. Klein, G. J. Ensing, L. A. Everett, E. D. Green, C. Proschel, N. J. Gutowski, M. Noble, D. L. Atkinson, S. J. Odelberg, and M. T. Keating (1996). Lim-kinase 1 hemizyosity implicated in impaired visuospatial constructive cognition. *Cell* 86, 59–69.
- Franz, V. H., K. R. Gegenfurtner, H. H. Bühlhoff, and M. Fahle (2000). Grasping visual illusions: no evidence for a dissociation between perception and action. *Psychological Science* 11(1), 20–25.
- Galati, G., E. Lobel, G. Vallar, A. Berthoz, L. Pizzamiglio, and D. L. Bihan (2000). The neural basis of egocentric and allocentric coding of space in humans: a functional magnetic resonance study. *Experimental Brain Research* 133, 156–164.
- Galletti, C. and P. P. Battaglini (1989). Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *Journal of Neuroscience* 9, 1112–1125.
- Gaymard, B., I. Siegler, S. Rivaud-Pechoux, I. Israel, C. Pierrot-Deseilligny, and A. Berthoz (2000). A common mechanism for the control of eye and head movements in humans. *Annals of Neurology* 47(6), 819–822.
- Gelman, R. and C. R. Gallistel (1978). *The Child’s Understanding of Number*. New York: Academic Press.
- Gentilucci, M., C. Scandolara, I. N. Pigarev, and G. Rizzolatti (1983). Visual responses in the postarcuate cortex (area 6) of the monkey that are independent of eye position. *Experimental Brain Research* 50, 464–468.
- Gibson, B. S. (1994). Visual attention and objects: One versus two or convex versus concave? *Journal of Experimental Psychology: Human Perception and Performance* 20(1), 203–207.
- Goodale, M. A., J. P. Meenan, H. H. Bühlhoff, D. A. Nicolle, K. J. Murphy, and C. I. Racicot (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology* 4(7), 604–610.
- Goodale, M. A. and D. Milner (1992). Separate visual pathways for perception and action. *Trends in Neurosciences* 15(1), 20–25.
- Graziano, M. S. A., X. T. A. Hu, and C. G. Gross (1997). Visuospatial properties of ventral premotor cortex. *Journal of Neurophysiology* 77(5), 2268–2292.
- Graziano, M. S. A., L. A. J. Reiss, and C. G. Gross (1999). A neuronal representation of the location of nearby sounds. *Nature* 397(6718), 428–430.
- Gross, C. G. (1992). Representation of visual stimuli in inferior temporal cortex. *Philosophical Transactions of the Royal Society of London, series B – Biological Sciences* 335(1273), 3–10.
- Haxby, J. V., C. L. Grady, B. Horwitz, L. G. Ungerleider, M. Mishkin, R. E. Carson, P. Herscovitch, M. B. Schapiro, and S. I. Rapoport (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences* 88, 1621–1625.
- Hendry, S. H. C. and T. Yoshioka (1994). A neurochemically distinct third channel in the macaque dorsal lateral geniculate nucleus. *Science* 264, 575–577.

- Heywood, C. A. and R. W. Kentridge (2000). Affective blindsight? *Trends in Cognitive Sciences* 4(4), 125–126.
- Hubel, D. H. and T. N. Wiesel (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology* 195, 215–243.
- Hurford, J. R. (1987). *Language and number: the emergence of a cognitive system*. Oxford: Basil Blackwell.
- Hurford, J. R. (1999). Individuals are abstractions. *Behavioral and Brain Sciences* 22,4, 620–621.
- Hurford, J. R. (2000a). Languages treat 1-4 specially. *Mind and Language* xxx, xxx–xxx.
- Hurford, J. R. (2000b). Social transmission favours linguistic generalization. In C. Knight, M. Studdert-Kennedy, and J. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pp. 324–352. Cambridge: Cambridge University Press.
- Hurford, J. R. (2001). Protothought had no logical names. In J. Trabant and S. Ward (Eds.), *New Essays on the Origin of Language*. Berlin: de Gruyter.
- Hurford, J. R. (in press). The roles of expression and representation in language evolution. In J.-L. Dessalles, C. Knight, and A. Wray (Eds.), *The Transition to Language*.
- Ingle, D. J. (1973). Two visual systems in the frog. *Science* 181, 1053–1055.
- Ingle, D. J. (1980). Some effects of pretectum lesions on the frog’s detection of stationary objects. *Behavioural Brain Research* 1, 139–163.
- Ingle, D. J. (1982). Organization of visuomotor behaviors in vertebrates. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior*. Cambridge, MA: MIT Press.
- Jackendoff, R. and B. Landau (1992). Spatial language and spatial cognition. In R. Jackendoff (Ed.), *Languages of the Mind: Essays on Mental Representation*, pp. 99–124. Cambridge, MA: MIT Press.
- Jeannerod, M., M. A. Arbib, G. Rizzolatti, and H. Sakata (1995). Grasping objects - the cortical mechanisms of visuomotor transformation. *Trends in Neurosciences* 18:(7), 314–320.
- Johnsrude, I. S., A. M. Owen, J. Crane, B. Milner, and A. C. Evans (1999). A cognitive activation study of memory for spatial relationships. *Neuropsychologia* 37, 829–841.
- Kaas, J. H. and T. A. Hackett (1999). ‘What’ and ‘where’ processing in auditory cortex. *Nature Neuroscience* 2(12), 1045–1047.
- Kahneman, D. and A. Treisman (1984). Changing views of attention and automaticity. In R. Parasuraman and D. A. Davies (Eds.), *Varieties of Attention*. New York: Academic Press.
- Kahneman, D. and A. Treisman (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology* 24, 175–219.
- Kamp, H. and U. Reyle (1993). *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht, Holland: Kluwer Academic.
- Kant, I. (1905). *Prolegomena to any Future Metaphysics*. LaSalle, Illinois: Open Court. Translated by Paul Carus.
- Keenan, E. (1987). Facing the truth. In E. Keenan (Ed.), *Universal Grammar: 15 Essays*. London: Croom Helm.

- Kikuchi-Yorioka, Y. and T. Sawaguchi (2000). Parallel visuospatial and audiospatial working memory processes in the monkey dorsolateral prefrontal cortex. *Nature Neuroscience* 3(11), 1075–1076.
- Kirby, S. (1999). Syntax out of learning: the cultural evolution of structured communication in a population of induction algorithms. In D. Floreano, J. D. Nicoud, and F. Mondada (Eds.), *Advances in Artificial Life*, Number 1674 in Lecture notes in computer science. Springer.
- Kirby, S. (2000). Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In C. Knight, M. Studdert-Kennedy, and J. R. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pp. 303–323. Cambridge: Cambridge University Press.
- Korte, M. and J. P. Rauschecker (1993). Auditory spatial tuning of cortical neurons is sharpened in cats with early blindness. *Journal of Neurophysiology* 70(4), 1717–1721.
- Kripke, S. (1980). *Naming and Necessity*. Oxford: Blackwell.
- Landau, B. and R. Jackendoff (1993). ‘What’ and ‘where’ in spatial language and spatial cognition. *Behavioral and Brain Sciences* 16(2), 217–238.
- Landau, B., L. B. Smith, and S. S. Jones (1988). The importance of shape in early lexical learning. *Cognitive Development* 3, 299–321.
- Landau, B., L. B. Smith, and S. S. Jones (1998a). Object perception and object naming in early development. *Trends in Cognitive Sciences* 2(1), 19–24.
- Landau, B., L. B. Smith, and S. S. Jones (1998b). Object shape, object function and object name. *Journal of Memory and Language* 38, 1–27.
- Liddell, S. K. (1990). Four functions of a locus: re-examining the structure of a space in ASL. In C. Lucas (Ed.), *Sign Language Research: Theoretical Issues*. Washington, D.C.: Gallaudet University Press.
- Livingstone, M. and D. Hubel (1988). Segregation of form, color, movement, and depth: anatomy, physiology and perception. *Science* 240, 740–769.
- Luck, S. J. and E. K. Vogel (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281.
- Lyons, J. (1977). *Semantics (2 vols.)*. Cambridge: Cambridge University Press.
- Mandler, G. and B. J. Shebo (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General* 111, 1–22.
- Marcel, A. J. (1998). Blindsight and shape perception: deficit of visual consciousness or of visual function? *Brain* 121, 1565–1588.
- Marois, R., H. chung Leung, and J. Gore (2000). A stimulus-driven approach to object identity and location processing in the human brain. *Neuron* 25, 717–728.
- Martin, A., C. L. Wiggs, L. G. Ungerleider, and J. V. Haxby (1996). Neural correlates of category-specific knowledge. *Nature* 379, 649–652.
- McDonald, B. H. (1994). Sign language morphology. In R. E. Asher and J. M. Y. Simpson (Eds.), *The Encyclopedia of Language and Linguistics*, pp. 3917–3919. Oxford: Pergamon Press.
- Merigan, W. H., C. E. Byrne, and J. H. R. Maunsell (1991). Does primate motion perception depend on the magnocellular pathway? *The Journal of Neuroscience* 11(11), 3422–3429.

- Merigan, W. H. and J. H. R. Maunsell (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience* 16, 369–402.
- Milner, A. D. (1998). Streams and consciousness: visual awareness and the brain. *Trends in Cognitive Sciences* 2(1), 25–30.
- Milner, A. D. and M. A. Goodale (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.
- Mishkin, M., L. G. Ungerleider, and K. A. Macko (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neuroscience* 6, 414–417.
- Montague, R. (1970). English as a formal language. In B. Visentini (Ed.), *Linguaggi nella Società e nella Tecnica*, pp. 189–223. Milan: Edizioni di Comunità.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In J. Hintikka, J. Moravcsik, and P. Suppes (Eds.), *Approaches to Natural Language*. Dordrecht, Holland: Reidel.
- Moore, T., H. R. Rodman, and C. G. Gross (1998). Man, monkey, and blindsight. *Neuroscientist* 4(4), 227–230.
- Morris, C. W. (1938). *Foundations of the Theory of Signs*. Chicago: Chicago University Press.
- Morris, C. W. (1946). *Signs, Language, and Behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Morris, J. S., A. Ohman, and R. J. Dolan (1999). A subcortical pathway to the the right amygdala mediating ‘unseen’ fear. *Proceedings of the National Academy of Sciences of the United States of America* 96(4), 1680–1685.
- Nakamura, K. (1999). Auditory spatial discriminatory and mnemonic neurons in rat posterior parietal cortex. *Journal of Neurophysiology* 82(5), 2503–2517.
- Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.
- Nowak, M. A., J. B. Plotkin, and V. A. A. Jansen (2000). The evolution of syntactic communication. *Nature* 404, 495–498.
- Padden, C. (1990). The relation between space and grammar in asl verb morphology. In C. Lucas (Ed.), *Sign Language Research: Theoretical Issues*. Washington, D.C.: Gallaudet University Press.
- Parsons, T. (1990). *Events in the Semantics of English : a Study in Subatomic Semantics*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition* 32, 65–97.
- Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Sciences* 4(5), 197–207.
- Ramachandran, V. S. and S. Blakeslee (1998). *Phantoms in the Brain: Human Nature and the Architecture of the Mind*. London: Fourth Estate.
- Rauschecker, J. P. (1997). Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngologica Supplement* 532, 34–38.
- Romanski, L. M., B. Tian, J. Fritz, M. Mishkin, P. S. Goldman-Rakic, and J. P. Rauschecker (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience* 2(12), 1131–1136.
- Russac, R. J. (1983). Early discrimination among small object collections. *Journal of Experimental Child Psychology* 36, 124–138.

- Russell, B. (1905). On denoting. *Mind* 14, 479–493.
- Russell, B. (1940). *An Inquiry into Meaning and Truth*. London: Allen and Unwin.
- Russell, B. (1948). *Human Knowledge, its Scope and Limits*. London: Allen and Unwin.
- Russell, B. (1959). *My Philosophical Development*. London: Allen and Unwin.
- Ryle, G. (1957). The theory of meaning. In C. A. Mace (Ed.), *British Philosophy in the Mid-Century*. London: Allen and Unwin.
- Sanders, M. D., E. K. Warrington, J. Marshall, and L. Weiskrantz (1974). ‘Blindsight’: Vision in a field defect. *Lancet* 20, 707–708.
- Schaeffer, B., V. H. Eggleston, and J. L. Scott (1974). Number development in young children. *Cognitive Psychology* 6, 357–379.
- Searle, J. R. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.
- Smith, L. B., S. S. Jones, and B. Landau (1996). Naming in young children: a dumb attentional mechanism? *Cognition* 60, 143–171.
- Starkey, P. and R. G. Cooper (1980). Perception of numbers by human infants. *Science* 210, 1033–1035.
- Strawson, P. F. (1950). On referring. *Mind* 59, 320–344.
- Stricanne, B., R. A. Andersen, and P. Mazzoni (1996). Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *Journal of Neurophysiology* 76(3), 2071–2076.
- Tian, B. and J. P. Rauschecker (1998). Processing of frequency-modulated sounds in the cat’s posterior auditory field. *Journal of Neurophysiology* 79(5), 2629–2642.
- Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung* 31, 299–337.
- Trick, L. M. and Z. W. Pylyshyn (1993). What enumeration studies tell us about spatial attention: evidence for limited capacity pre-attentive processing. *Journal of Experimental Psychology – Human Perception and Performance* 19(2), 331–351.
- Trick, L. M. and Z. W. Pylyshyn (1994). Why are small and large numbers enumerated differently? a limited capacity pre-attentive stage in vision. *Psychological Review* 10, 1–23.
- Ullman, S. (1984). Visual routines. *Cognition* 18, 97–157.
- Ungerleider, L. G., S. M. Courtney, and J. V. Haxby (1998). A neural system for human visual working memory. *Proceedings of the National Academy of Sciences* 95, 883–890.
- Ungerleider, L. G. and M. Mishkin (1982). Two cortical visual systems. In D.J.Ingle, M.A.Goodale, and R.J.W.Mansfield (Eds.), *Analysis of Visual Behavior*, pp. 549–586. Cambridge, MA: MIT Press.
- Valdes-Sosa, M., A. Cobo, and T. Pinilla (1998). Transparent motion and object-based attention. *Cognition* 66, B13–B23.
- Van Essen, D. C., C. H. Anderson, and D. J. Felleman (1992). Information processing in the primate visual system: an integrated systems perspective. *Science* 255, 419–423.
- Vecera, S. P. and M. J. Farah (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General* 123(2), 146–160.

- Weeks, R., B. Horwitz, A. Aziz-Sultan, B. Tian, C. M. Wessinger, L. G. Cohen, M. Hallett, and J. P. Rauschecker (2000). A positron emission tomographic study of auditory localization in the congenitally blind. *Journal of Neuroscience* 20(7), 2664–2672.
- Weiskrantz, L. (1986). *Blindsight: a Case Study and Implications*. Oxford: Oxford University Press.
- Weiskrantz, L. (1997). *Consciousness Lost and Found : a Neuropsychological Exploration*. Oxford: Oxford University Press.
- Westwood, D. A., C. D. Chapman, and E. A. Roy (2000). Pantomimed actions may be controlled by the ventral visual stream. *Experimental Brain Research* 130, 545–548.
- Woll, B. and J. Kyle (1994). Sign language. In R. E. Asher and J. M. Y. Simpson (Eds.), *The Encyclopedia of Language and Linguistics*, pp. 3890–3912. Oxford: Pergamon Press.
- Zeki, S. (1993). *A Vision of the Brain*. Oxford: Blackwell Scientific Publications.