

# The neural basis of predicate-argument structure

**James R. Hurford**

*Language Evolution and Computation Research Unit, School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh EH8 9LL, Scotland, United Kingdom.*  
 jim@ling.ed.ac.uk

**Abstract:** Neural correlates exist for a basic component of logical formulae, *PREDICATE(x)*. Vision and audition research in primates and humans shows two independent neural pathways; one locates objects in body-centered space, the other attributes properties, such as colour, to objects. In vision these are the dorsal and ventral pathways. In audition, similarly separable “where” and “what” pathways exist. *PREDICATE(x)* is a schematic representation of the brain’s integration of the two processes of delivery by the senses of the location of an arbitrary referent object, mapped in parietal cortex, and analysis of the properties of the referent by perceptual subsystems.

The brain computes actions using a few “deictic” variables pointing to objects. Parallels exist between such nonlinguistic variables and linguistic deictic devices. Indexicality and reference have linguistic and nonlinguistic (e.g., visual) versions, sharing the concept of attention. The individual variables of logical formulae are interpreted as corresponding to these mental variables. In computing action, the deictic variables are linked with “semantic” information about the objects, corresponding to logical predicates.

Mental scene descriptions are necessary for practical tasks of primates, and preexist language phylogenetically. The type of scene descriptions used by nonhuman primates would be reused for more complex cognitive, ultimately linguistic, purposes. The provision by the brain’s sensory/perceptual systems of about four variables for temporary assignment to objects, and the separate processes of perceptual categorization of the objects so identified, constitute a pre-adaptive platform on which an early system for the linguistic description of scenes developed.

**Keywords:** argument; attention; deictic; dorsal; logic; neural; object; predicate; reference; ventral

## 1. Introduction

This article argues for the following thesis: Neural evidence exists for predicate-argument structure as the core of phylogenetically and ontogenetically primitive (prelinguistic) mental representations. The structures of modern natural languages can be mapped onto these primitive representations.

The idea that language is built onto preexisting representations is common enough, being found in various forms in works such as Bickerton (1998), Kirby (1999; 2000), Hurford (2000b), and Bennett (1976). Conjunctions of elementary propositions of the form *PREDICATE(x)* have been used by Batali as representations of conceptual structure preexisting language in his impressive computer simulations of the emergence of syntactic structure in a population of interacting agents (Batali 2002). Justifying such preexisting representations in terms of neural structure and processes is relatively new.

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 neuroscience. The fact that it is possible to show a correlation between neural processes and logicians’ conclusions about logical form is a step in the unification of science. The discoveries in neuroscience confirm that the logicians have been on the right track, that the two disciplines have some-

thing to say to each other despite their radically different methods, and that further unification may be sought. As the brain has a complexity far in excess of any representation scheme dreamt up by a logician, it is to be expected that the basic *PREDICATE(x)* formalism is to some extent an idealization of what actually happens in the brain. But, conceding that the neural facts are messier than could be captured with absolute fidelity by any formula as simple as *PREDICATE(x)*, I hope to show that the central ideas embodied in the logical formula map satisfyingly neatly onto certain specific neural processes.

The claim that some feature of language structure maps onto a feature of primitive mental representations needs (1) a plausible bridge between such representation and the structure of language, and (2) a characterization of

**JAMES R. HURFORD** is Professor of General Linguistics at the University of Edinburgh. He has a broad interest in reconciling various traditions in linguistics which have tended to conflict. In particular, he has worked on articulating a framework in which formal representation of grammars in individual minds interacts with properties of language as used in communities. The framework emphasizes the interaction of evolution, learning and communication. He is perhaps best known for his computer simulations of various aspects of the evolution of language.

“primitive mental representation” independent of language itself, to avoid circularity. The means of satisfying the first, the “bridge to language” condition, will be discussed in the next subsection. Fulfilling the second condition, the bridge to brain structure and processing, by establishing the language-independent validity of *PREDICATE(x)* as representing fundamental mental processes in both humans and nonhuman primates, will occupy the meat of this article (sects. 2 and 3). The article is original only in bringing together the fruits of others’ labours. Neuroscientists 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 subjects by recent neurological research.

### 1.1. The bridge from logic to language

The relationship between language and thought is, of course, a vast topic, and there is only space here to sketch my premises about this relationship.

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. Neuroscientists 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 neuroscientists’ 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.).

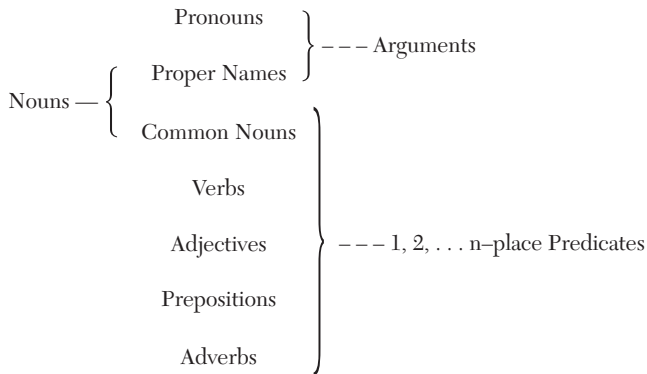
Logical notations, on the other hand, were developed by scholars theorizing in the neurological dark about the structure of language and thought. Languages are systems for the expression of thought. The sounds and written characters, and even the syntax and phonology of languages can also be described 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 of a thought process and expression of a thought 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 neuroscientist 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 straitjacket 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.

It is necessary to note a common divergence of usage, between philosophers and linguists, in the term “subject.” For some philosophers (e.g., Strawson 1959; 1974), a predicate in a simple proposition, as expressed by *John loves Mary*, for example, can have more than one “subject;” in the example given, the predicate corresponds to *loves* and its “subjects” to *John* and *Mary*. In this usage, the term “subject” is equivalent to “argument.” Linguists, on the other hand, distinguish between grammatical subjects and grammatical objects, and further between direct and indirect objects. Thus, in *Russia sold Alaska to America*, the last two nouns are not subjects, but direct and indirect object respectively. The traditional grammatical division of a sentence into Subject + Predicate is especially problematic where the “Predicate” contains several NPs, semantically interpreted as arguments of the predicate expressed by the verb. Which argument of a predicate, if any, is privileged to be expressed as the grammatical subject of a sentence (thus in English typically occurring before the verb, and determining number and person agreement in the verb) is not relevant to the truth-conditional analysis of the sentence. Thus, a variety of sentences such as *Alaska was sold to America by Russia* and *It was America that was sold Alaska by Russia* all describe the same state of affairs as the earlier example. The difference between the sentences is a matter of rhetoric, or appropriate presentation of information in various contextual circumstances, involving what may have been salient in the mind of the hearer or reader before encountering the sentence, or how the speaker or writer wishes to direct the subsequent discourse.

Logical predicates are expressed in natural language by words of various parts of speech, including verbs, adjectives, and common nouns. In particular, there is no special connection between grammatical verbs and logical predicates. The typical correspondences between the main English syntactic categories and basic logical terms are diagrammed below.



Common nouns, used after a copula, as *man* in *He is a man* plainly correspond to predicates. In other positions, although they are embedded in grammatical noun phrases, as in *A man arrived*, they nonetheless correspond to predicates.

The development of formal logical languages, of which first order predicate logic is the foremost example and hardiest survivor, heralds a realization of the essential distance between ordinary language and purely truth-conditional representations of “objective” situations in the world. Indeed, early generations of modern logicians, including Frege, Russell, and Tarski, believed the gap between ordinary language and logical, purely truth-conditional representations to be unbridgeable. Times have changed, and since Montague there have been substantial efforts to describe a systematic mapping between truth conditions and ordinary language. Ordinary language serves several purposes in addition to representation of states of affairs. My argument in this article concerns mental representations of situations in the world, as these representations existed *before* language, and even before communication. Thus, matters involving how information is presented in externalized utterances are not our concern here. The exclusive concern here with pre-communication mental representations absolves us from responsibility to account for further cognitive properties assumed by more or less elaborate signals in communication systems, such as natural languages. For this reason also, the claims to be made here about the neural correlates of *PREDICATE(x)* do not relate at all directly to matters of *linguistic* processing (e.g., sentence parsing), as opposed to the prelinguistic representation of events and situations.

Bertrand Russell was, of course, very far from conceiving of the logical enterprise as relating to how non-linguistic creatures represent the world. But it might be helpful to note that Russell’s kind of flat logical representations, as in  $\exists x [KoF(x) \ \& \ wise(x)]$  for *The king of France is wise*,<sup>1</sup> are essentially like those assumed by Batali (2002) and focussed on in this article. Russell’s famous controversy with Strawson (Russell 1905; 1957; Strawson 1950) centered on the ef-

fect of embedding an expression for a predicate in a noun phrase determined by the definite article. Questions of definiteness only arise in communicative situations, with which Strawson was more concerned. A particular object in the world is inherently neither definite nor indefinite; only when we talk about an object do our referring noun phrases begin to have markers of definiteness, essentially conveying “You are already aware of this thing.”

The thesis proposed here is that there were, and still are, pre-communication mental representations which embody the fundamental distinction between predicates and arguments, and in which the foundational primitive relationship is that captured in logic by formulae of the kind *PREDICATE(x)*. The novel contribution here is that the centrality of predicate-argument structure has a neural basis, adapted to a sentient organism’s traffic with the world, rather than having to be postulated as “logically true” or even Platonically given. Neuroscience can, I claim, offer some informative answers to the question of where elements of logical form came from.

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 language, for example, Montague (1970; 1973), Parsons (1990), Kamp and Reyle (1993). The hypothesis here is *not* that *all* linguistic structure derives from prelinguistic mental representations. I argue elsewhere (Hurford 2002) that in fact very little of the rich structure of modern languages directly mirrors any mental structure in pre-existing language.

In generative linguistics, such terms as “deep structure” and “surface structure,” “logical form” and “phonetic form” have specialized theory-internal meanings, but the basic insight inherent in such terminology is that linguistic structure is a *mapping between* two distinct levels of representation. In fact, most of the complexity in language structure belongs to this mapping, rather than to the forms of the anchoring representations themselves. In particular, the syntax of logical form is very simple. All of the complexities of phonological structure belong to the mapping between meaning and form, rather than to either meaning or form *per se*. A very great proportion of morphosyntactic structure clearly also belongs to this mapping – components such as word-ordering, agreement phenomena, anaphoric marking, most syntactic category distinctions (e.g., noun, verb, auxiliary, determiner), which have no counterparts in logic, and focussing and topicalization devices. In this respect, the view taken here differs significantly from Bickerton’s (in Calvin & Bickerton 2000) that modern grammar in all its glory can be derived, with only a few auxiliary assumptions, from the kind of mental representations suitable for cheater detection that our prelinguistic ancestors would have been equipped with; see Hurford (2002) for a fuller argument.

Therefore, to argue, as I will in this article, that a basic component of the representation of meaning preexists language and can be found in apes, monkeys, and possibly other mammals, leaves most of the structure of language (the complex mappings of meanings to phonetic signals) still unexplained in evolutionary terms. To argue that apes have representations of the form *PREDICATE(x)*, does not make them out to be language-capable humans. 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.

The view developed here is similar in its overall direction to that taken by Bickerton (1990). Bickerton argues for a “primary representation system (PRS)” existing in variously developed forms in all higher animals. “In all probability, language served in the first instance merely to label proto-concepts derived from prelinguistic experience” (p. 91). This is entirely consistent with the view proposed here, assuming that what I call “prelinguistic mental predicates” are Bickerton’s “protoconcepts.” Bickerton also believes, as I do, that the representation systems of prelinguistic creatures have predicate-argument structure. Bickerton further suggests that, even before the emergence of language, it is possible to distinguish subclasses of mental predicates along lines that will eventually give rise to linguistic distinctions such as Noun/Verb. He argues that “[concepts corresponding to] verbs are much more abstract than [those corresponding to] nouns” (p. 98). I also believe that a certain basic functional classification of predicates can be argued to give rise to the universal linguistic categories of Noun and Verb. But that subdivision of the class of predicates is not my concern here. Here the focus is on the more fundamental issue of the distinction between predicates and their arguments. So this paper is *not* about the emergence of Noun/Verb structure (which is a story that must wait for another day). (Batali’s [2002] impressive computer simulations of the emergence of some aspects of natural language syntax start from conjunctions of elementary formulae in *PREDICATE(x)* form, but it is notable that they do not arrive at anything corresponding to a Noun/Verb distinction.)

On top of predicate-argument structure, 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 (syntax); (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. It is surely a move forward in explaining the evolution of language to be able to dissect out the separate steps that must be involved, even if these turn out to be more dauntingly numerous than was previously thought. (In parallel fashion, the discovery of the structure of DNA immediately posed problems of previously unimagined complexity to the next generation of biologists.)

## 1.2. Prelinguistic predicates

In the view adopted here, a predicate corresponds, to a first approximation, to a judgement that a creature can make about an object. Some predicates are relatively simple. For a simple predicate, the senses provide the brain with input allowing a decision with relatively little computation. On a scale of complexity, basic colour predicates are near the

simple end, while predicates paraphrasable as *sycamore* or *weasel* are much more complex. Mentally computing the applicability of complex predicates often involves simpler predicates, hence relatively more computation.

Some ordinary languages predicates, such as *big*, depend for their interpretation on the prior application of other predicates. Generically speaking, a big flea is not big; this is no contradiction, once it is admitted that the sentence implicitly establishes two separate contexts for the application of the adjective *big*. There is “big, generically speaking,” that is, in the context of consideration of all kinds of objects and of no one kind of object in particular; and there is “big for a flea.” This is semantic modulation. Such modulation is not a solely linguistic phenomenon. Many of our higher-level perceptual judgements are modulated in a similar way. An object or substance characterized by its whitish colour (like chalk) reflects bright light in direct sunlight, but a light of lower intensity in the shade at dusk. Nevertheless, the brain, in both circumstances, is able to categorize this colour as whitish, even though the lower intensity of light is reflected by a greyish object or substance (like slate) in direct sunlight. In recognizing a substance as whitish or greyish, the brain adjusts to the ambient lighting environment. Viewing chalk in poor light, the visual system returns the judgement “Whitish, for poor light”; in response to light of the same intensity, as when viewing slate in direct sunlight, the visual system returns the judgement “Greyish, for broad daylight.” A similar example can be given from speech perception. In a language such as Yoruba, with three level lexical tones, high, mid, and low, a single word spoken by an unknown speaker cannot reliably be recognized as on a high tone spoken by a man or a low or mid tone spoken by a woman or child. But as soon as a few words are spoken, the hearer recognizes the appropriate tones in the context of the overall pitch range of the speaker’s voice. Thus, the ranges of external stimuli which trigger a mental predicate may vary, systematically, as a function of other stimuli present.

This article will be mainly concerned with 1-place predicates, arguing that they correspond to perceived properties. There is no space here to present a fully elaborated extension of the theory to predicates of degree greater than one, but a few suggestive remarks may convince a reader that in principle the theory may be extendable to *n*-place predicates ( $n > 1$ ).

Prototypical events or situations involving 2-place predicates are described by *John kicked Fido* (an event) or *The cat is on the mat* (a situation). Here I will take it as given that observers perceive events or situations as unified wholes; there is some psychological reality to the concept of an atomic event or situation. In a 2-place predication (barring predicates used reflexively), the two participant entities involved in the event or situation also have properties. In formal logic, it is possible to write a formula such as  $\exists x \exists y [kick(x, y)]$ , paraphrasable as *Something kicks something*. But I claim that it is never possible for an observer to perceive an event of this sort without also being able to make some different 1-place judgements about the participants. Perhaps the most plausible potential counterexample to this claim would be reported as *I feel something*. Now this could be intended to express a 1-place state, as in *I am hungry*; but if it is genuinely intended as a report of an experience involving an entity other than the experiencer, I claim that there will always be some (1-place) property of

this entity present to the mind of the reporter. That is, the “something” which is felt will always be felt as having some property, such as sharpness, coldness or furriness. Expressed in terms of a psychologically realistic logical language enhanced by meaning postulates, this amounts to the claim that every 2-place predicate occurs in the *implicans* of some meaning postulate whose *implicatum* includes 1-place predicates applicable to its arguments. The selectional restrictions expressed in some generative grammars provide good examples; the subject of *drink* must be animate, the object of *drink* must be a liquid.

In the case of asymmetric predicates, the asymmetry can always be expressed in terms of one participant in the event or situation having some property which the other lacks. And, I suggest, this treatment is psychologically plausible. In cases of asymmetric actions, as described by such verbs as *hit* and *eat*, the actor has the metaproperty of being the actor, cashed out in more basic properties such as movement, animacy, and appearance of volition. Likewise, the other, passive, participant is typically characterized by properties such as lack of movement, change of state, inanimacy, and so forth (see Cruse 1973 and Dowty 1991 for relevant discussion). Cases of asymmetric situations, such as are involved in spatial relations as described by prepositions such as *on*, *in*, and *under*, are perhaps less obviously treatable in this way. Here, I suggest that properties involving some kind of perceptual salience in the given situation are involved. In English, while both sentences are grammatical, *The pen is on the table* is commonplace, but *The table is under the pen* is studiously odd. I would suggest that an object described by the grammatical subject of *on* has a property of being taken in as a whole object comfortably by the eye, whereas the other object involved lacks this property and is perceived (on the occasion concerned) rather as a *surface* than as a whole object.

In the case of symmetric predicates, as described by *fight each other* or *as tall as*, the arguments are not necessarily distinguished by any properties perceived by an observer.

I assume a version of event theory (Davidson 1980; Parsons 1990), in which the basic ontological elements are whole events or situations, annotated as *e*, and the participants of these events, typically no more than about three, annotated as *x*, *y*, and *z*. For example, the event described by *A man bites a dog* could be represented as  $\exists e, x, y, \text{bite}(e), \text{man}(x), \text{dog}(y), \text{agent}(x), \text{patient}(y)$ . In clumsy English, this corresponds to “There is a biting event involving a man and a dog, in which the man is the active volitional participant, and the dog is the passive participant.” The less newsworthy event would be represented as  $\exists e, x, y, \text{bite}(e), \text{man}(x), \text{dog}(y), \text{agent}(y), \text{patient}(x)$ . The situation described by *The pen is on the table* could be represented as  $\exists e, x, y, \text{on}(e), \text{pen}(x), \text{table}(y), \text{small\_object}(x), \text{surface}(y)$ .

In this enterprise it is important to realize the great ambiguity of many ordinary language words. The relations expressed by English *on* in *An elephant sat on a tack* and in *A book lay on a table* are perceptually quite different (though they also have something in common). Thus, there are at least several mental predicates corresponding to ordinary language words. When, in the histories of natural languages, words change their meanings, the overt linguistic forms become associated with *different* mental predicates. The predicates which I am concerned with here are prelinguistic mental predicates, and are not to be simply identified with words.

Summarizing these notes, it is suggested that it may be possible to sustain the claim that *n*-place predicates ( $n > 1$ ) are, at least in perceptual terms, constructible from 1-place predicates. The core of my argument in this article concerns formulae of the form *PREDICATE*(*x*), that is, 1-place predications. My core argument in this article does not stand or fall depending on the correctness of these suggestions about  $n > 1$ -place predicates. If the suggestions about  $n > 1$ -place predicates are wrong, then the core claim is limited to 1-place predications, and some further argument will need to be made concerning the neural basis of  $n > 1$ -place predications. A unified theory relating all logical predicates to the brain is methodologically preferable, so there is some incentive to pursue the topic of  $n > 1$ -place predicates.

### 1.3. Individual variables as prelinguistic arguments

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

*CAME*(*john*) (Translation: “John came”)  
 $\exists x[\text{TALL}(x) \ \& \ \text{MAN}(x) \ \& \ \text{CAME}(x)]$  (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, p. 219). “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 protohumans 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 re-identify, 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. More detailed arguments along these lines are given in Hurford (1999; 2001), but it is worth repeating here the counterargument to the most common objection to this idea. It is commonly asserted that animals can recognize other animals in their groups.

The following quotation demonstrates the *prima facie* at-

traction of the impression that animals distinguish such individuals, but simultaneously gives the game away.

The speed with which recognition of individual parents can be acquired is illustrated by the "His Master's Voice" experiments performed by Stevenson et al. (1970) on young terns: these responded immediately to tape-recordings of their own parents (by cheeping a greeting, and walking towards the loudspeaker) but ignored other tern calls, even those recorded from other adult members of their own colony. (Walker 1983, p. 215)

Obviously, the tern chicks in the experiment were *not* recognizing their individual parents – they were being fooled into treating a loudspeaker as a parent tern. For the tern chick, anything that behaved sufficiently like its parent was "recognized" as its parent, even if it wasn't. The tern chicks were responding to very finely-grained *properties* of the auditory signal, and apparently neglecting even the most obvious of visual properties discernible in the situation. In tern life, there usually aren't human experimenters playing tricks with loudspeakers, and so terns have evolved to discriminate between auditory cues just to the extent that they can identify their own parents with a high degree of reliability. Even terns presumably sometimes get it wrong.

Animals respond in mechanical robot-like fashion to key stimuli. They can usually be "tricked" into responding to crude dummies that resemble the true, natural stimulus situation only partially, or in superficial respects. (Krebs & Dawkins 1984, p. 384; quoted in Hurford 2001)

The logical notion of an individual constant permits no degree of tolerance over the assignment of these logical constants to individuals; this is why they are called "constants." It is an *a priori* fiat of the design of the logical language that individual constants pick out particular individuals with absolute consistency. In this sense, the logical language is practically unrealistic, requiring, as previously mentioned, Godlike omniscience on the part of its users, the kind of omniscience reflected in the biblical line "But even the very hairs of your head are all numbered" (Matthew, Ch.10).

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; 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, p. 464).

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 nonproper 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's [1980] 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 (nonpronominal) 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, such as *Baker*, *Wheeler*, *Newcastle*. It is very rare for languages to lack proper names, but such languages do exist. Machiguenga (or Matsigenka), an Arawakan language, is one, as several primary sources (Johnson 2003; Snell 1964) testify.

A most unusual feature of Matsigenka culture is the near absence of personal names (Snell 1964, pp. 17–25). Since personal names are widely regarded by anthropologists as a human universal (e.g., Murdock 1960, p. 132), this startling assertion is likely to be received with skepticism. When I first read Snell's discussion of the phenomenon, before I had gone into the field myself, I suspected that he had missed something (perhaps the existence of secret ceremonial names) despite his compelling presentation of evidence and his conclusion:

I have said that the names of individual Machiguenga, when forthcoming, are either of Spanish origin and given to them by the white man, or nicknames. We have known Machiguenga Indians who reached adulthood and died without ever having received a name or any other designation outside of the kinship system. . . . Living in small isolated groups there is no imperative need for them to designate each other in any other way than by kinship terminology. Although there may be only a few tribes who do not employ names. I conclude that the Machiguenga is one of those few. (Snell 1964, p. 25)

Experience has taught me that Snell was right. Although the Matsigenka of Shimaa did learn the Spanish names given them, and used them in instances where it was necessary to refer to someone not of their family group, they rarely used them otherwise and frequently forgot or changed them. (Johnson 2003)

Joseph Henrich, another researcher on Machiguenga tells me "This is a well established fact among Machiguenga researchers" (personal communication).



In this society there is very little cooperation, exchange or sharing beyond the family unit. This insularity is reflected in the fact that until recently they didn't even have personal names, referring to each other simply as "father," "patrilineal same-sex cousin," or whatever. (Douglas 2001, p. 41)

The social arrangements of our prelinguistic ancestors probably involved no cooperation, exchange, or sharing beyond the family unit, and the mental representations which they associated with individuals could well have been kinship predicates or other descriptive predicates.

In Australian languages, people are usually referred to by descriptive predicates.

Each member of a tribe will also have a number of personal names, of different types. They may be generally known by a nickname, describing some incident in which they were involved or some personal habit or characteristic, e.g., "[she who] knocked the hut over," "[he who] sticks out his elbows when walking," "[she who] runs away when a boomerang is thrown," "[he who] has a damaged foot." But each individual will also have a sacred name, generally given soon after birth. (Dixon 1980, p. 27)

The extensive anthropological literature on names testifies to the very special status, in a wide range of cultures, of such sacred or "baptismal" proper names, both for people and places. It is common for proper names to be used with great reluctance, for fear of giving offense or somehow intruding on a person's mystical selfhood. A person's proper name is sometimes even a secret.

The personal names by which a man is known are something more than names. Native statements suggest that names are thought to partake of the personality which they designate. The name seems to bear much the same relation to the personality as the shadow or image does to the sentient body. (Stanner 1937, quoted in Dixon 1980, p. 28)

It is hard to see how such mystical beliefs can have become established in the minds of creatures without language. More probably, it was only early forms of language itself that made possible such elaborate responses to proper names.

Hence, it is unlikely that any primitive mental representation contained any equivalent of a proper name, that is, 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? I first set out some ground rules 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 psychologistic 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 neuroscience 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 neuroscience 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 Goodale & Milner 1992; Trevarthen 1968; Ungerleider & Mishkin 1982).

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; 1980; 1982). Distinct neural pathways from the frog's retina to different parts of its brain control these reflex actions.

Distinct visuomotor systems can similarly be identified in mammals, as Milner and Goodale (1995) report:

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 & Goodale 1995, pp. 18–19)

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 & Hubel 1988), and the distinction between dorsal and ventral pathways leading from V1 to further visual cortical areas (Mishkin et al. 1983; Ungerleider & Mishkin 1982). 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., Hendry & Yoshioka 1994; Johnsrude et al. 1999; Marois et al. 2000) and some disagreement (e.g., Franz et al. 2000; Merigan & Maunsell 1993; Zeki 1993). See Milner and Goodale (1995, pp. 33–39, 134–36) for discussion of the magno/parvo-dorsal/ventral relationship. (One has to be careful what one understands by “modular” when quoting Milner & Goodale [1995]. In real brains, modules are neural entities that modulate, compete, and cooperate, rather than being encapsulated processors for one “faculty” [Arbib 1987b].) It will 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 & Goodale 1995, p. 68).

While it is not obvious that both divergences pertain to the same functional role, the proposals made here are not so detailed or subtle as to suggest any relevant discrimination between these two branches of the route from retina

to parietal cortex. 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 & Mishkin 1982, p. 578)

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 & Goodale 1995, pp. 65–66).

Figure 1 shows the separation of dorsal and ventral pathways in schematic form.

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

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 & Goodale 1995; Sanders et al. 1974; Weiskrantz 1986; 1997. See also Ramachandran & 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 some aspects of motion is impaired.<sup>2</sup> 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 nonhumans. Moore et al. (1998)



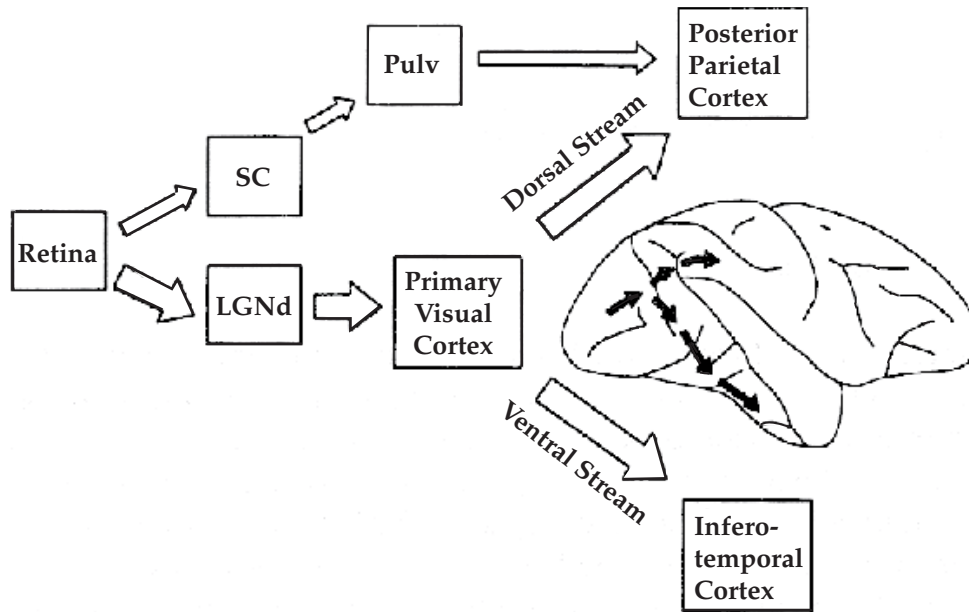


Figure 1. Schematic diagram showing major routes whereby retinal input reaches dorsal and ventral streams. The inset [brain drawing] shows the cortical projections on the right hemisphere of a macaque brain. LGNd, lateral geniculate nucleus, pars dorsalis; Pulv, pulvinar nucleus; SC, superior colliculus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

summarize 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 R.V. 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 nonpathological subjects the distinct localizations of processing for object recognition and object location (e.g., Aguirre & 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 dis-

placement 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" (p. 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" (p. 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, p. 386)

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, p. 206)

**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, p. 35)

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 & 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” lines.<sup>3</sup> Graziano et al. (1999) have shown that certain neurons in macaques have spatial receptive fields limited to about 30 cm 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, p. 805)

Evidence from audition is less central to my argument than evidence from vision. My main claim is that in predicate-argument structure, the predicate represents some judgement about the argument, which is canonically an attended-to object. There is a key difference between vision and hearing. What is seen is an *object*, typically enduring;

what is heard is an *event*, typically fleeting. If language is any guide (which it surely is, at least approximately) mental sound predicates can be broadly subdivided into those which simply classify the sound itself (rendered in English with such words as *bang*, *rumble*, *rush*), and those which also classify the event or agent which caused the sound (expressed in English by such words as *scrape*, *grind*, *whisper*, *moan*, *knock*, *tap*). (Perhaps this broad dichotomy is more of a continuum.) When one hears a sound of the first type, such as a bang, there is no object, in the ordinary sense of “object,” which “is the bang.” A bang is an ephemeral event. One cannot attend to an isolated bang in the way in which one directs one’s visual attention to an enduring object. The only way one can simulate attention to an isolated bang is by trying to hold it in memory for as long as possible. This is quite different from maintained visual attention, which gives time for the ventral stream to do heavy work categorizing the visual stimuli in terms of complex properties. Not all sounds are instantaneous, like bangs. One can notice a continuous rushing sound. But again, a rushing sound is not an object. Logically, it seems appropriate to treat bangs and rushing sounds either with zero-place predicates, that is, as predicates without arguments, or as predicates taking event variables as arguments. (The exploration of event-based logics is a relatively recent development.) English descriptions such as *There was a bang* or *There was a rushing* tend to confirm this.

Sounds of the second type, classified in part by what (probably) caused them, allow the hearer to postulate the existence of an object to which some predicate applies. If, for example, you hear a miaow, you mentally classify this sound as a miaow. This, as with the bang or the rushing sound, is the evocation of a zero-place predicate (or alternatively a predicate taking an event variable as argument). Certainly, hearing a miaow justifies you in inferring that there is an object nearby satisfying certain predicates, in particular *CAT(x)*. But it is vital to note that the English word *miaow* is two-ways ambiguous. Compare *That sound was a miaow* with *A cat miaowed*, and note that you can’t say *\*That sound miaowed* or *\*That cat was a miaow*. Where the subject of *miaow* describes some animate agent, the verb actually means “cause a miaow sound.”

It is certainly interesting that the auditory system also separates “where” and “what” streams. But the facts of audition do not fit so closely with the intuitions, canonically involving categorizable enduring objects, which I believe gave rise to the invention by logicians of predicate-argument notation. The idea of zero-place predicates has generally been sidelined in logic (despite their obvious applicability to weather phenomena); and the extension of predicate-argument notation to include event variables is relatively recent. (A few visual predicates, like that expressed by English *flash*, are more like sounds, but these are highly atypical of visual predicates.)

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 body parts and integrating signals from them into single coherent overall responses to the location of objects. Given a stimulus somewhere far round to one

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 & Goodale 1995, p. 90)

The evidence that Milner and Goodale (1995) cite is from Galletti and Battaglini (1989), Andersen et al. (1985; 1990), and 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, p. 232)

Gaymard et al. (2000, p. 819) 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.” 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, p. 2071) 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.” 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 orig-

inally 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 provides very little of 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*.

## 2.2. “Dumb” attentional mechanisms and the object/property distinction

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 argued, or possibly about certain 3D aspects of shape (Lehky & Sejnowski 1988). 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 on this “something.” 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 any properties, or mental predicates, at all. It may not be appropriate to speak of the dorsal stream delivering *representations* (accessible to report) of the nature of objects. Nevertheless, in a clear sense, the dorsal stream does *deliver objects*, in a minimal sense of “object” to be discussed below. What the dorsal stream delivers, very fast, is information about the egocentric location of an object, which triggers motor responses resulting in the orientation of focal attention to the object. (At a broad-brush level, the differences between preattentive processes and focal atten-



tion have been known for some time, and are concisely and elegantly set out in Ch. 5 of Neisser 1967.)

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, although predicates/properties are delivered by the ventral stream, it would not be surprising if a few of the mental predicates available to a human being did not also correspond at least roughly to information of the type used by the dorsal stream. But humans have an enormous wealth of other predicates as well, undoubtedly accessed exclusively via the ventral stream, and 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; 1998a; 1998b) and Smith et al. (1996) 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, p. 169)

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 consistent with the hypothesis that children's *earliest* naming tendencies capitalize strongly on attention-drawing traits of objects.

But doesn't talk of "attention-drawing traits of objects" undermine my central argument, by locating some "traits" (alias properties) within the class of information delivered by the dorsal stream? A position diametrically opposed to mine would be that ultimately there is no distinction at all to be made between objects and properties. A philosophical argument for such a position might appeal to English terms such as "objecthood," meaning the property of being an object. Advanced logical systems can play havoc with basic ontological categories, such as object and property, by various devices such as type-raising. Such devices may be appropriate in the analysis of elaborated human languages and the systems of thought that they make available. Yes, humans can treat properties as objects, by reification, and objects as properties (by "Pegasizing Pegasus," as Quine put it). But I would claim that an ape's mental traffic with the world is in terms of two broadly noninterconvertible ontological categories, object and property.

A more psychologically plausible argument against my position might claim that *any* property of an object that one could give a name to could in principle be an attention-

drawing trait. This would potentially attribute to the dorsal stream any information conveyed by a predicate, thus destroying the hypothesis that it is the ventral stream that delivers predicates. I emphasize that such issues should be addressed with empirical (neuro-) psychological evidence, rather than purely philosophical argumentation. Some relevant evidence exists, pointed out by O'Brien and Opie (1999), in connection with blindsight, as follows.

Consider the comments made by Weiskrantz' subject D.B., after performing well above chance in a test that involved distinguishing between Xs and Os presented in his scotoma. While D.B. maintained that he performed the task merely by guessing:

If pressed, he might say that he perhaps had a "feeling" that the stimulus was either pointing this or that way, or was "smooth" (the O) or "jagged" (the X). On one occasion in which "blanks" were randomly inserted in a series of stimuli . . . he afterwards spontaneously commented he had a feeling that maybe there was no stimulus present on some trials. But always he was at a loss for words to describe any conscious perception, and repeatedly stressed that he saw nothing at all in the sense of "seeing," and that he was merely guessing. (Weiskrantz et al. 1974, p. 721)

Throughout D.B.'s verbal commentaries there are similar remarks. Although he steadfastly denies "seeing" in the usual way when presented with visual stimuli, he frequently describes some kind of concurrent awareness. He talks of things "popping out a couple of inches" and of "moving waves," in response to single point stimuli (Weiskrantz 1986, p. 45). He also refers to "kinds of pulsation" and of "feeling some movement" in response to moving line stimuli. (Weiskrantz 1986, p. 67)

Consequently, while blindsight subjects clearly do not have normal visual experience in the "blind" regions of their visual fields, this is not to say that they don't have any phenomenal experience whatsoever associated with stimuli presented in these regions. What is more, it is not unreasonable to suggest that what little experience they do have in this regard explains their residual discriminative abilities. D.B., for example, does not see Xs or Os (in the conventional sense). But in order to perform this task he doesn't need to. All he requires is some way of discriminating between the two stimulus conditions *q* some broad phenomenal criterion to distinguish "Xness" from "Oness." And as we've seen, he does possess such a criterion: one stimulus condition feels "jagged" while the other feels "smooth." Thus, it is natural to suppose that he is able to perform as well as he does (above chance) because of the (limited) amount of information that is consciously available to him. (O'Brien & Opie 1999, p. 131)

Unlike O'Brien and Opie, I am not mainly concerned with consciousness. I am content to concede that O'Brien and Opie have a point, and to fall back on the reservation that a formula as simple as *PREDICATE*(*x*) cannot be expected to mirror exactly all the processes of such a complex organ as the brain. The stark contrast between the blindsight patient's experience and his performance is evidence that the brain separates sub- or semiconscious awareness of the bare presence of an object from the vast array of judgments that can be made by a normal person about the properties of an object. Perhaps training can boost the set of properties which can act as attention-drawing traits. But I would predict that only a tiny subset of properties are natural attention-drawing properties, and that any properties added to this set by practice or training are likely to swing into action significantly more slowly than the primal attention-drawing properties.

This prediction conflicts with a prediction of Milner and Goodale's in their final chapter addressing further research

questions prompted by the dorsal/ventral distinction. They write “It is unlikely that the dorsal stream plays the major role in mediating this initial [attention] selection process, since object recognition and ‘semantic’ knowledge may have to be taken into account.” (Milner & Goodale 1995, p. 202) With due deference to Milner and Goodale, I suggest that their implicit premise that *all* “semantic” recognition takes place in the ventral stream may be too strong, and that a very limited set of primal properties can be accessed by the dorsal stream. I would further claim that access to these primal attention-drawing properties is highly encapsulated, unlike access to properties delivered by the ventral stream. It is an intuition of this difference that gives rise to the logician’s postulate that the fundamental logical structure is an asymmetric relation between two distinct logical types, predicate and argument.

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 egocentric 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)

There is an interesting parallel (more than merely coincidental) in the uses of the term “binding” in logic and neuroscience. The existence of a blue dot can be represented in FOPL as  $\exists x [BLUE(x) \& DOT(x)]$ . (The ordering of the conjuncts is immaterial.) Here the existential quantifier is said to “bind” the variable *x* immediately after it, and, also importantly, all further instances of this variable must fall within the scope, indicated by brackets, of the quantifier. The variable and its binding quantifier thus serve to unite the various predicates in the formula, indicating that they denote properties of the *same* object. Logical binding is not a relationship between a predicate and its argument, but a relationship between all predicates in the scope of a particular quantifier which take the bound variable as argument. In neuroscience, “binding is the problem of representing conjunctions of properties. . . . For example, to visually detect a vertical red line among vertical blue lines and diagonal red lines, one must visually bind each line’s color to its orientation” (Hummel 1999). Detection of properties is generally achieved via the ventral stream. The dorsal stream directs attention to an object. Once attention is focussed on a particular object, the ventral stream can deliver a multitude of different judgements about it, which can be represented logically by a conjunction of 1-place predications. The bare drawing of attention to an object, with no category judgements (yet) made about it, corresponds to the “ $\exists x$ ” part of the logical formula.

Evidently, the brain does solve the binding problem, although we are not yet certain exactly how it does it. The claim advanced here for a connection between predicate-argument structure and the ventral/dorsal separation does not depend on what, in detail, the brain’s solution to the binding problem turns out to be.

### 2.3. Related proposals

#### 2.3.1. Landau and Jackendoff: Nouns and prepositions.

Jackendoff and Landau (1992) and Landau and Jackendoff (1993)<sup>4</sup> 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. Landau and Jackendoff 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 they 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. Landau and Jackendoff 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. Landau and Jackendoff 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 Landau and Jackendoff's proposed dorsal/preposition correlation. Prepositions express predicates, many of which give spatial information, both egocentric and allocentric. Their 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" (p. 118) is located in the posterior parietal region, while many other kinds of visual judgement, including computation of allocentric spatial information, 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 & Goodale 1995, p. 89). 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, p. 242) 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 express 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 logical distinction is between predicates and individual variables, not between different syntactic subclasses of words which express predicates. Thus, the logical 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 Landau and Jackendoff focus. This idea is close to what I believe Bridgeman (1993), in his commentary on them, states: "cognitive and [motor-oriented] spatial systems can be distinguished on a lower level than that of Landau and Jackendoff, a level that differentiates linguistic from nonlinguistic coding" (p. 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 indefinite pronouns, such as *something* and *anything* 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 *Something 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.

**2.3.2. Givon: Lexical concepts and propositions.** Givon (1995, pp. 408–10), in a brief but pioneering discussion, relates the dorsal and ventral visual pathways to linguistic information in a way which is partly similar to my proposed correlation. In particular, Givon correlates information accessed via the ventral stream with lexical concepts. This is very close to my correlation of this information with prelinguistic predicates. Prelinguistic predicates are concepts (or what Bickerton calls "protoconcepts"), and they can become *lexical* concepts by association with phonological forms, once language gets established. My proposal differs from Givon's in the information that we correlate with the dorsal stream, which he correlates with "spatial relation/motion – propositional information about states or events" (p. 409). Givon, writing before 1995, relied on several of the same sources as Landau and Jackendoff, and, like them, assumes that "the dorsal (upper) visual processing stream analyzed the *spatial relations* between specific objects and *spatial motion* of specific objects. This processing track is thus responsible for analyzing specific visual *states* and *events*" (p. 409, emphasis in original). As mentioned above, Milner and Goodale (1995) subsequently presented evidence that such allocentric spatial information is not processed in the dorsal stream. Elsewhere in Givon's account, there is an acknowledgement of the role of the stream to the temporal lobe in accessing information about spatial motion:

Further, even in non-human primates, the object recognition (ventral) stream analyzes more than visually perceived objects and their attributes. Thus Perrett et al. (1989) in their study of single-cell activation in monkeys have been able to differentiate between single cortical cells that respond to objects (nouns), and those that are activated by *actions* (verbs). Such differentiation occurs within the object recognition stream itself, in the *superior temporal sulcus* of the left-temporal lobe. And while the verbs involved – e.g., moving an object by hand towards mouth – are concrete and spatio-visual, they involve more abstract computations of purpose and causation. (Givon 1995, p. 410, emphasis in original)

This attribution undermines Givon's earlier identification of the dorsal stream as the stream providing information about spatial motion. Note that Givon begins to correlate neural structure with the specifically linguistic categories of noun and verb, a move which I avoid. I correlate information accessed by the ventral stream with predicates, regardless of whether these eventually get expressed as nouns, verbs, adjectives, or any other lexical category. The present proposed correlation of distinct neural pathways with *logical* predicates and individual variables differs from both Landau and Jackendoff's and Givon's proposals in claiming completely *prelinguistic* correlates for the ventral and dorsal pathways. The correlation that I propose for information delivered by the dorsal stream is developed in more detail in the next section.

**2.3.3. Rizzolatti and Arbib: A prelinguistic "grammar" of action.** Rizzolatti and Arbib's paper (1998) contains a section entitled "A pre-linguistic 'grammar' of action in the monkey brain." Like me, they are concerned with a neural precursor to language, found in monkey brains. There are superficial similarities between our proposals and differences, which are important to state.

Rizzolatti and Arbib use a kind of logical notation to convey an idea about the activity of "canonical" macaque F5 neurons in grasping small objects.



We view the activity of “canonical” F5 neurons as part of the code for an imperative case structure, for example, Command: grasp-A(raisin), as an instance of grasp-A(object), where grasp-A is a specific kind of grasp, to be applied to the raisin. Note that this case structure is an “action description,” not a linguistic representation. “Raisin” denotes the specific object towards which the grasp is directed, whereas grasp-A is a specific command directed towards an object with well specified physical properties. (Rizzolatti & Arbib 1998, p. 192)

The formula used here by Rizzolatti and Arbib is best taken as a shorthand for a sequence of separate processes; the compression into a single formula gives rise to several potentially misleading infelicities. Logically, a term like “raisin” is a predicate, and therefore (in FOPL) should not be used as an argument. This is not a merely pernicky point. Key to my own proposal is the idea that a *predicate* is the logical expression of a judgement about the category to which some attended-to object belongs. The process of perceiving something to be a raisin is, I claim, well represented by the formula *RAISIN(x)*. Allowing, for the moment, “GRASP-A” as a predicate, the sequence of events in the monkey’s brain with which R&A are here concerned would be better expressed as

<i>RAISIN(x)</i>	Equation 1
<i>GRASP-A(x)</i>	Equation 2

That is, the judgement that the attended-to object is a raisin precedes the motor instruction to grasp it in a certain way, if the animal is acting with any deliberation. If the animal does not make a deliberate categorical judgement, but simply reflexively grabs the object (with activation essentially limited to the dorsal stream), then, according to the correlation I propose, there is no question of the predicate *RAISIN*, or any other predicate, being involved. I have less to say about the use of predicate notation to cover motor instructions. Classical logic was devised as a way of objectively representing (inter alia) observable events and states of affairs, and the present proposal is to link logic to the neural processes involved in *perception* of stimuli from outside the animal, and not to the mechanisms involved in purposeful action by the animal. Rizzolatti and Arbib’s discussion, while appealing to a notation which is logic-like in that it apparently has predicate-argument structure, does not in fact deconstruct this formula and attribute the separate parts to different neural processes, as is proposed in the present article.

### 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, which 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 an object nothing more than a bundle of features?<sup>25</sup> The notion of an object, as opposed to its features, is important for the central claim of this article, that modern neuroscience 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 attention 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 that 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 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. (p. 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. (p. 501)

And,

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. (Duncan 1984, p. 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., Egly et al. 1994; Vecera & Farah 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: “In-

stead 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” (p. 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” (p. 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 features.<sup>6</sup>

A series of papers (Baylis 1994; Baylis & Driver 1993; Gibson 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.” (Baylis & Driver 1993, pp. 466–67)

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. (p. 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” (p. B13).

Each set moved in a distinct and randomly chosen direction. We found that simultaneous judgments of speed and direction were more accurate when they 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. (p. 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 (“Gabor”), 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 holistic 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 neither 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. Everything 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).

#### 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 ob-

jects 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; 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, pp. 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, p. 82, emphasis 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, p. 150)

Since this passage was written, in the early 1980s, 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; 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” (p. 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” (p. 178). Ballard et al. (1997) 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,” p. 280). Pylyshyn assumes “a pool of four or five available indexes” (Pylyshyn 2000, p. 201). 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 handle a maximum of only 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,  $w, x, y, z, \dots$ . Cowan (2001) provides a very thorough and extensive survey of studies of short term memory, concluding that there is a remarkable degree of similarity in the capacity limit in working memory observed with a wide range of procedures. A restricted set of conditions is necessary to observe this limit. It can be observed only with procedures that allow assumptions about what the independent chunks are, and the limit the recursive use of the limited-capacity store . . . The preponderance of evidence from procedures fitting these conditions strongly suggests a mean memory capacity in adults of 3 to 5 chunks, whereas individual scores appear to range more widely from about 2 up to about 6 chunks. The evidence for this pure capacity limit is considerably more extensive than that for the somewhat higher limit of  $7 + 2$  stimuli (Cowan 2001).

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 Antell & Keating 1983; Gelman & Gallistel 1978; Mandler & Shebo 1982; Russac 1983; Schaeffer et al. 1974; Starkey & Cooper 1980 for some relevant studies). Both Ullman (1984, p. 151) and Pylyshyn (2000, pp. 201–202) make the connection between subitizing (which Ullman calls “visual counting”) and the marking or indexing of locations in a scene. Trick and Pylyshyn (1993; 1994) explain 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 (1987; 2000a).

The simple clauses of human languages are constrained to a maximum of about four or five 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).<sup>7</sup>

Ballard et al. (1997) 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, for example, “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 nonhuman 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 arguments, derives from the considerations of computational efficiency advanced by Ballard et al. (1997).<sup>8</sup>

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).<sup>9</sup> (Woll & Kyle 1994, p. 3905)

See also Liddell (1990), McDonald (1994), and 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. (1997) 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. There are a few languages with four-way contrasts. Tlingit is one such language. In Tlingit,

*yáa* “this (one) right here” is clearly “close to *Sp*”; *héi* “this (one) nearby” is characterized by a moderate distance from *Sp* without reference to the *Adr*; *wéé* “that (one) over there” is again not identified by the location of the *Adr*; and *yóo* “that (one) far off (in space or time),” the fourth term, is simply remote from the speech situation. (Anderson & Keenan 1985, p. 286)

Anderson and Keenan mention two other languages, Sre and Quileute, as also having four-way deictic contrasts. They mention one language, CiBemba, with a five-way system, and one, Malagasy, with a seven-way system; frankly, I am skeptical of the claim for seven degrees of contrast along a single dimension in Malagasy. “Systems with more than five terms along the basic deictic dimension are exceedingly rare” (Anderson & Keenan 1985, p. 288).

The extreme rarity of languages providing more than five contrasting deictic terms in any subsystem corresponds nicely to the “pool of four or five available indexes,” or visual deictic variables, postulated by Pylyshyn. 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 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. It is important to note the highly elastic size of the domains appealed to in deixis. Within deictic systems, “near” and “far” are typically relative, not absolute. Hence, within a domain which is all in some sense near the speaker, there nevertheless will still be a distinction between “near” and “far.”

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 convenience.<sup>10</sup> So far, we have made no move to suggest how such nonlinguistic 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 develop.<sup>11</sup>

I have suggested certain parallels between the prelin-

guistic representation of events (restriction to three to five participants, location of the participants in egocentric space) and features of modern human languages (clause size, limits of deictic systems, anaphora in sign languages). I believe that these features of language can ultimately be traced back to evolutionary precursors in the prelinguistic representations. But it also seems very likely that in the evolution of the language capacity, the human brain has liberated itself from certain of the most concrete associations of the prelinguistic representations. Thus, when a modern human processes a sentence describing some abstract relation, such as *Ambition is more forgivable than greed*, it is unlikely that any specifically egocentric space-processing (parietal) areas are activated. The relation between ancient egocentric visuo-spatial maps and modern features of language is, I would claim, rather like the relationship between ancient thermoregulation panels and wings, a relationship of homology or exaptation. If the ancient structures had never existed, the modern descendants would not have the particular features that they do, but the modern descendants are just that, *descendants*, with the kind of modifications one expects from evolution.

## 5. Common ground of neuroscience, 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. The 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 “deiktikos”; see Lyons 1977, p. 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 & Chapman 1987, p. 270).<sup>12</sup> 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; 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 languages; for 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 indexical.<sup>13</sup> (Pylyshyn 2000, p. 205)

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 twentieth 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; 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 (though perhaps “praxics” might have been a better term) 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, p. 314)

Likewise Milner and Goodale (1995, p. 88) write of the “content or semantics” of nonverbal 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 & Goodale 1995, p. 88). 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 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. This history is sketched in Figure 2.

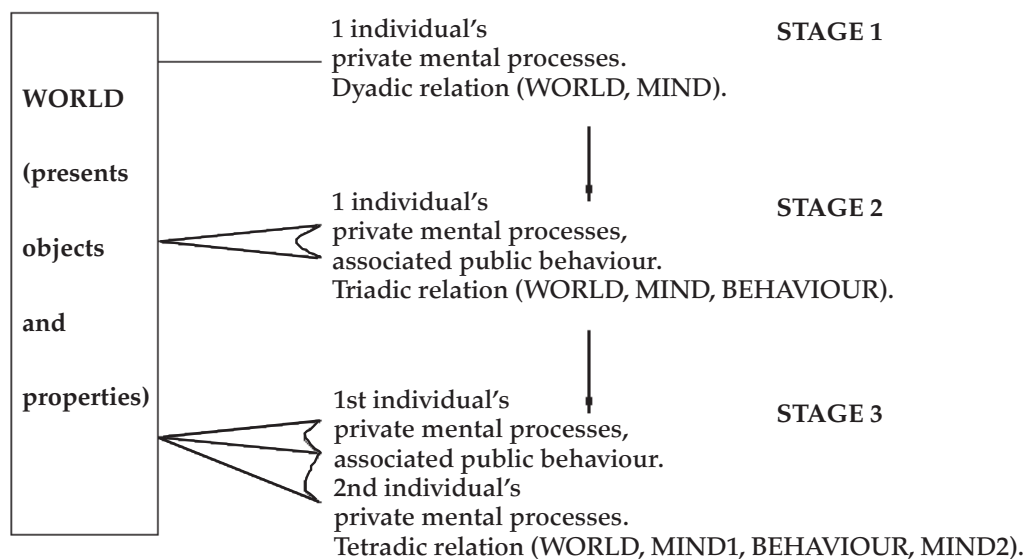


Figure 2. The evolution of reference. The relationship between mental processes and the world is the original and enduring factor. The last stage is successful reference as understood by linguists, and as manifested by people speaking natural languages. The stages may overlap, in that further evolution 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

### 6.1. *It could have been otherwise*

It could conceivably have been otherwise, both from a logical and a biological point of view. Consider, first, alternative biologies. We can conceive of a world in which organisms sense the ambient temperature of their surroundings by a single sensory organ which doesn't distinguish any source of radiant heat. Further, such a creature might have a keen sense of smell, and be able to discriminate between thousands of categorically different smells assailing its smell organ. And the creature might have arrays of light detectors evenly spaced all over its body, all feeding into a single internal organ activated by an unweighted average of the inputs. Such a creature would have no internal representation of objects, but only a set of “zero-place predicates.” It could sense “The world outside is in such-and-such a state.” Certainly, the higher animals on planet Earth are not like this, but I would be surprised if some lower animals were not somewhat like it. It just happens to be the case that the laws of physics, chemistry, and biology conspire to produce a world containing discrete categorizable objects, and so, not surprisingly, but *not logically necessarily*, advanced creatures have evolved ways of dealing with them.

An alternative logic is also easily conceivable, in which there is no predicate-argument structure. It already exists in the form of the propositional calculus, typically introduced in logic textbooks as a simple step towards the more “advanced” predicate calculus. A propositional calculus, with no predicate-argument structure, would be all that is needed by the creature described in the previous paragraph.

Here is a final thought experiment. A “Turing robot” is entirely conceivable as a working automaton, capable of navigating and surviving in a complex world. Instead of reading a character on a tape, the Turing robot “reads” a patch of the world in front of it, matching the input to some monadic symbol occurring in the quadruples of its instruction set. Instead of shifting the tape to right or left, it shifts itself to an adjacent patch of world, and it can act, one monadic action at a time, on the patch of world it is looking at. Given a complex enough instruction set, such a robot could replicate any of the complex computations carried out by an advanced real live creature successfully negotiating the world. The Turing robot's hardware, and the individual elements of its software instruction set, the basic quadruples, contain nothing corresponding to predicate-argument structure, though it is probable that we could interpret some higher-level pattern or subroutine in the whole instruction set as somehow corresponding to predicate-argument structure. The dorsal/ventral separation in higher mammals is, I argue, an evolved hardware implementation of predicate-argument structure.

### 6.2. *Falsifiability*

This article is an instance of reductionism. It takes two previously unrelated fields, logic and neuroscience, and argues that what logicians are really dealing with, whenever they appeal to predicate-argument structure, has a basis in neural processing. This in no way minimizes the validity of studies in logic; rather it enhances their validity. Biologists working with Mendelian genes without knowledge of DNA were doing valid work. “Abstract” work on the structure of human thought, and its relationship to language, must continue. But as long as we recognize that the object of study, both in logic and in linguistics, has a psychological basis, one of us should also work on bridging the gap between theoretical studies couched in logico/linguistic terminology and empirical studies in psychology and neuroscience. Only those who view logical and linguistic structure as Platonic, in some way existing independently of human minds, can ignore psychology and neuroscience.

Can a reductionist argument be falsified? Yes. Some proposed reductions are just plain wrong, some are well justified, and some are partly right. What justifies a reductionist argument is the goodness of fit between the two independently established theories. The present argument would be invalidated if it could be shown that any of the following apply:

The canonical arguments of predicates in logic do not denote individual objects.

Canonical predicates in logic do not denote properties.

The dorsal stream processes properties at least as much as the ventral stream.

The ventral stream plays a large role in drawing attention to objects.

I concede that an extreme version of my reductionist proposal *is* falsified in many ways, because, on the logical side, for example, formal semanticists often use nonobject-denoting terms as arguments of predicates, and on the neurological side, for example, some detection of properties is achieved by the dorsal stream. So the fit between the practices of logicians and formal semanticists with predicate-argument structure and the neural facts is not quite perfect. But, I claim, there is enough of a clear parallelism between the two domains to indicate that neuroscience has revealed facts which significantly inform the domain that logicians and formal semanticists traditionally deal with. Here again I mention that the brain is vastly more complex than even the most baroque of logical formalisms, and that one should expect complexities arising from brain studies that logical studies simply do not relate to. A logical formalism relates to the brain in the same way as a road map relates to a real place.

### 6.3. *Then, now, and next*

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 chutter could signify that the animal is having a *SNAKE(x)* experience, that is, 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.

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, translation of Kant 1783).<sup>14</sup> 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.

NOTES

1. The logical formula is simplified for convenience here.
2. A complication to this picture arises from work on the recog-

nition of facial expressions by blindsight patients (de Gelder et al. 1999; 2000; Heywood & Kentridge 2000; Morris et al. 1999). 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.

3. Belin 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.

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

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

6. Egly 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. (p. 173)

This again conflates attention-shifting, a preattentive (and post-attentive) process, with attention itself. These experiments relate only to attention-shifting, as the title of Egly et al.’s (1994) 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.)

7. There 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.

8. The 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 prelinguistic creature could have.

9. Both third and fourth person space in BSL are like available pronouns for entities being signed about, other than the speaker or hearer. It is not that BSL has four grammatical persons in the sense that English has three (1st – speaker, 2nd – hearer, 3rd – all other entities).

10. This formula, like any FOPL formula, conveys no temporal transitions. Tense logic is more complicated than FOPL.

11. See Batali (2002) for a computer simulation of the emergence of public language from representations of exactly this form.

12. Agre and Chapman (1987) do not, as stated by Ballard et al. (1995), use the term “deictic.”

13. Indeed, 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.