the environment that elicits sensory input *S*, the location of which is marked for attention by the dorsal ("where") stream of cortical processing. Presumably, neurons in the ventral stream would have two states, active or silent, corresponding to the "true" or "false" values of PREDICATE(*X*). By way of a concrete example, we might consider a hypothetical neuron in the ventral stream that evaluates APPLE(*X*), which would fire neural impulses at some fixed rate if the currently attended sensory input *S* corresponds to an apple, and would be silent otherwise. One problem with this scheme is that cortical neurons are not two-state elements but show graded responses to their inputs. Another, more serious, problem is that sensory inputs, being neural, are stochastic and therefore uncertain to some extent. It is hard to see how neurons in the ventral stream, or anywhere else in the brain, could ever be completely certain of exactly what has elicited their current pattern of input. Perhaps a better way to model the responses of such a neuron is $P(X = \text{APPLE}|S)$. The activity of the neuron would then vary from zero up to some maximal level of firing, which would be proportional to the probability that object *X,* eliciting sensory input *S,* is an apple. The computation of probabilities seems a more realistic basis for perception than the evaluation of statements in logic.

Prelinguistic agents will form only egocentric representations

Michael L. Anderson[a] and Tim Oates[b]

[a]*Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742;* [b]*Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, MD 21250.*
anderson@cs.umd.edu oates@cs.umbc.edu
http://www.cs.umd.edu/~anderson
http://www.csee.umbc.edu/~oates/

Abstract: The representations formed by the ventral and dorsal streams of a prelinguistic agent will tend to be too qualitatively similar to support the distinct roles required by *PREDICATE(x)* structure. We suggest that the attachment of qualities to objects is not a product of the combination of these separate processing streams, but is instead a part of the processing required in each. In addition, we suggest that the formation of objective predicates is inextricably bound up with the emergence of language itself, and so cannot be cleanly identified with any prelinguistic cognitive capacities.

In his search for the neural basis of the simple logical structure *PREDICATE(x),* Hurford focuses on the basic cognitive capacities of advanced, but prelinguistic agents – for he follows Batali (2002) in believing that language can be developed out of such simple logical structures. Not having any such agents to study directly, Hurford instead considers the capacities of higher primates, abstracting away from their linguistic abilities to uncover two basic perceptual processing systems – the dorsal and ventral streams (henceforth DS and VS) – which he suggests provide the basic components of *PREDICATE(x)* structure. However, he does not carry this thought experiment through consistently; and, failing to consider matters from the perspective of such a prelinguistic agent, he is led to interpret the information delivered by these perceptual processing streams in postlinguistic terms, thereby in a subtle way assuming what he is trying to prove. When things are instead considered from the standpoint of such an agent, it no longer looks as if the DS and VS provide attractive candidates for the separate components of *PREDICATE(x)* structure.

Hurford rightly emphasizes that the deliverances of the DS – the "where" pathway that provides information about the location and size and shape of an object – are cast in "egocentric" terms. The DS is a specialized perceptual processing system that represents information in a form optimized for calculating and directing motor responses aimed at an object in virtue of its location, orientation, and spatial extent. This information is used to guide such things as the orientation of sense organs for optimal perception, perceptual tracking, reaching, and grasping. Thus, the natural way to characterize what the agent knows in virtue of DS representations (what the information means to the agent) is in terms of egocentric spatial coordinates: Where it is in relation to the agent, and what might be done to get the agent-object relation into a preferred state. One might say that the DS places objects in an egocentric visuomotor space, or an egocentric action field, and the object is thereby presented to the agent in these terms.

This characterization of the function of the DS is largely in accord with Hurford's – but what drives Hurford's account is the supposed *contrast* between the egocentric "motor-oriented" information given by the DS, with the "cognitive" – and therefore in some sense more objective – information said to be delivered by the VS. But in the individual, prelinguistic, and thus (one might say) *functionally solipsistic* agents Hurford describes, this contrast is untenable. Like the DS, the VS is a specialized perceptual processing system, but in this case it is optimized for representing information about the *look* rather than the *location* of an object. Just as with the DS, the VS representations are used to select and direct appropriate motor responses; and just as with the DS, the most natural way to characterize what the agent knows in representing this information is what the object means to it – to its utility, goals, survival – and what it might appropriately do in response. VS representations likewise place the object in a visuomotor action field, useful for calculating responses based on the differences between individual objects as opposed to individual locations. To imagine instead that encountered objects are represented in terms of objective features or abstract qualities is to import into the VS the kind of representational scheme appropriate for language and logic, but of no use to agents whose primary concerns are individual and behavioral but not communal or communicative.

So what we, in fact, have in the case of the individual agents Hurford describes are two specialized processing mechanisms that, although optimized for representing different aspects of perceptual information, are both nevertheless engaged in interpreting that information in conceptual terms suited to the selection and direction of appropriate motor responses. It does not seem that the products of these two visuomotor control systems lend themselves to natural combination in the form Hurford needs. Indeed, there seem to be substantial gaps between what these pathways deliver – the egocentrically presented features of objects – and *PREDICATE(x)* structure. We next describe two of these gaps and identify possible approaches to bridging them.

Consider first the fact that the representations formed by the VS and DS are *egocentric.* Language is useful only to the extent that it enables agents to share meaning with one another, but for meaning to be shared, it must be objective, not subjective or idiosyncratic. An isolated agent capable of forming prelinguistic concepts, or *predicates,* from the information delivered by its VS will form just those concepts that help this one agent survive in whatever environment it finds itself. These concepts will be cast within a single, agent-centered frame of reference, meaningful only from its own individual perspective. But suppose this agent discovers others of its kind. Through repeated interactions with these new agents, objective features of the world – those features commonly available and salient to others – can be identified and thus used to form the concepts (predicates) that serve as the semantic basis of language. The computational model of language evolution developed by Luc Steels (Steels 1997) leverages this idea of repeated interactions to separate the objective from the subjective and thereby evolve a stable, shared lexicon. Genuinely objective predicates, and the representational schemes that support them, arise only as the result of the formation of such shared, stable, intersubjective representation systems.

Second, it's clear that in *PREDICATE(x)* the two components – *PREDICATE()* and *x* – are qualitatively different. But if both processing streams are delivering the egocentrically presented features of objects, then neither stream seems properly specialized

for producing *objects* over *properties.* Rather, if *both* pathways can be said to return information about objects – such as the distance to an object in the case of the dorsal pathway or the color of an object in the case of the ventral pathway – then it seems that *both* pathways must have access to (or contain) a neural mechanism that individuates objects in the visual field, making it possible to bind deictic markers and extract information about, for example, distances or colors. Given this mechanism, the dorsal pathway could produce representations to underlie predicates like *REACHABLE(x),* and the ventral pathway could produce representations to underlie predicates like *RED(x).* Indeed, it seems that the data presented in (Goodale et al. 1994) support this view, for the patient R.V. was able to discriminate objects despite apparent damage to DS processing.

Predicates: External description or neural reality?

Michael A. Arbib

Computer Science Department, Neuroscience Program and USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520.
arbib@pollux.usc.edu **http://www-hbp.usc.edu/**

Abstract: Hurford argues that propositions of the form PREDICATE(*x*) represent conceptual structures that predate language and that can be explicated in terms of neural structure. I disagree, arguing that such predicates are *descriptions* of limited aspects of brain function, not available as representations in the brain to be exploited in the frog or monkey brain and turned into language in the human.

Note: The numbered paragraphs relate to the corresponding sections of the target article; unnumbered paragraphs contain my comments.

(Section 1.2.) The basic ontological elements are whole events or situations and the participants of these events. The event described by *A man bites a dog* could be represented as

$$\exists e, x, y, bite(e), man(x), dog(y), agent(x), patient(y) \quad (1)$$

I don't think this works. We need to replace *agent x* by *agent(x, e)* to indicate in which event *x* plays the stipulated role; similarly for *y.* For Hurford, the discussion of episodes is an aside to his concentration on 1-place predicates, but I suggest that the crux for a prelinguistic representation is the event and the "action-object frame" *A(x,y)* – agent *x* is doing *A* to object *y* – and its variations. Rizzolatti and Arbib (1998) examined "whether a 'prelinguistic grammar' can be assigned to the control and observation of actions. If this is so, the notion that evolution could yield a language system 'atop' the action system becomes much more plausible" (p. 191).

This talk of a prelinguistic grammar was not meant to imply that gestures may be a primitive form of grammar, for our approach was semantic rather than syntactic:

> We might say that the firing of "mirror" F5 neurons is part of the code for a declarative case structure, for example,
>
> Declaration: grasp-A(Luigi, raisin)
>
> which is a special case of grasp-A(agent, object), where grasp-A is a specific kind of grasp, applied to the raisin (the object) by Luigi (the agent . . . *this is an "action description," not a linguistic representation* (Rizzolatti & Arbib 1998, p. 192. Emphasis added).

Being able to grasp a raisin is different from being able to say, "I am grasping a raisin," and the neural mechanisms that underlie the doing and the saying are different. However, the case structure lets us see a commonality in the underlying representations, thus helping us understand how a mirror system for grasping might provide an evolutionary core for the development of brain mechanisms that support language.

(Section 2.) Representations of the form PREDICATE(x) 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.

More specifically, the notion is that a person may attend to a limited number of objects, and *x* then stands as an index for one of those objects. Thus, a scene might be represented by a conjunction

$$P1(X1) \,\&\, P2(X2) \,\&\, P3(X3) \,\&\, P4(X4) \quad (2)$$

where each *Xj* indexes some region of the scene, and *Pj(Xj)* indicates that the object at that location possesses property *Pj.* This leads to another point which (I think) weakens Hurford's critique of Rizzolatti and Arbib (1998):

(Section 4.) An example of a scene-description might be

$$APE(x) \,\&\, STICK(y) \,\&\, MOUND(z) \,\&\, HOLE(w) \,\&\, IN(w,z) \,\&\, PUT(x,y,w) \quad (3)$$

translating to *An ape puts a stick into a hole in a mound.*

The inclusion of PUT(*x,y,w*) in (3) reinforces the point that Hurford's focus on unary predicates does not do justice to describing animals which perceive to act, with acts dependent on relations between objects. The key question remains: "How do we go from predicates that we may use to describe internal behavior to neural representations that themselves abstract from the activity levels and parameterizations of schemas and their underlying neural networks, and instead provide abstractions that may in turn be refined to yield the cognitive and semantic forms that drive the production and perception of the phonological forms of language?"

In discussing the possible neural basis of (2), Hurford (sect. 4) cites papers from 1984 onward. However, I would claim some priority in this area with *the slide-box metaphor* (Arbib 1972; Didday & Arbib 1971): In the days before computer graphics, movie cartoons were drawn using *cels,* which I there called *slides.* Because the cartoon might run for seconds without the background changing, one may draw this background just once. In the middle ground, there might be a tree about which nothing changes for a while except its position relative to the background. It could thus be drawn on a separate slide and repositioned as needed. In the foreground, key details might change for each frame. The slides could then be photographed appropriately positioned in a slide-box for each frame, with only a few parameter changes (including minimal redrawing) required between successive frames. The slide-box metaphor suggested that a similar strategy might be used in the brain, with long-term memory (LTM) corresponding to a "slide file" and working or short-term memory (STM) corresponding to the "slide-box." The act of perception was compared to using sensory information to update slides already in the slide-box and to retrieve other slides as appropriate, experimenting to determine whether a newly retrieved slide fits sensory input "better" than one currently in the slide-box, which, in the brain, corresponds to a mass of neural tissue linking sensory and motor systems. A crucial point was that retrieval of a slide provided access to a wealth of information about the object it represented, including appropriate courses of action.

I cite this background to stress that (3) is a pale approximation of the slide-box metaphor, which is in turn a pale approximation to the multilevel modeling methodology that unifies the functional schemas of schema theory (Arbib 1981; Arbib et al. 1998) with the dynamics of detailed neural networks. For example, one *schema* in the visuomotor system of a frog (Arbib 1987a) might correspond to a pattern of neural activity signaling the likelihood of a small moving object in a region x_1 of the visual field, another schema might signal the likelihood that a large moving object is moving with velocity v in region x_2, while a third might indicate the likelihood that a barrier of extent w is located around region x_3. Thus, rather than being predicates that return 0 or 1, they are functions or likelihood distributions over a multidimensional parameter space. Moreover, the frog's actual course of action (the