American Journal of Primatology 49:315–338 (1999)

Describing Behavioral States Using a System Model of the Primate Brain

ALAN H. BOND*
California Institute of Technology, Pasadena, California

A system model of the primate neocortex is presented, based mainly on the neuroanatomy of the rhesus macaque monkey and consisting of a set of processing modules arranged as a perception-action hierarchy. These modules correspond to regions of the neocortex and their connectivity to that of the neocortex. A computational approach based on predicate logic is explained, and the results of a computer implementation of the model are reported, which demonstrate social behaviors involving affiliation and social conflict. The behavioral states of primates involved in these behaviors can be represented by the states of the system model, which have a logical representation and a diagrammatic form. It is shown how the behavioral states in goal-directed behaviors can be represented and also their short term moment-to-moment development in time. It is then shown how the state of social interaction among two or more primates can be represented, using their individual behavioral states, with interindividual action and perception. The causal dynamics of behavioral states is explained and also a control mechanism, namely, the use of confirmation signals, which stabilizes behavioral states and their dynamics. Stabilized behavioral states are seen as corresponding to coherent activations of the system, resulting from successful selection of module activations and intermodule communication with confirmation. Am. J. Primatol. 49:315–338, 1999.

Key words: system model; grooming; social action; primate; joint action; macaque

INTRODUCTION

Lesion and neurophysiologic recording studies have been mainstays of behavioral neuroscience, offering links between behavior and specific brain regions. Neuroanatomical studies have shown the existence of discrete cortical regions with specialized functional involvements and have provided evidence for pathways through which neural activity is relayed from region to region. However, what is lacking is a system-level theory of brain function against which to test specific hypotheses concerning the neural basis of complex, organized primate behavior.

Contract grant sponsor: National Science Foundation.

*Correspondence to: Alan H. Bond, California Institute of Technology, Mailstop 136-93, 1201 California Boulevard, Pasadena, CA 91125. E-mail:bond@vision.caltech.edu

Received 22 April 1998; revision accepted 20 May 1999

An ideal model would be based on empirical evidence derived from the foregoing research programs. It would also incorporate concepts from computer science, such as control, distribution, and transformation of information. If sufficiently detailed and comprehensive, a system-level model could be of great utility for laboratory experimentation on brain-behavior relationships, allowing investigators to manipulate variables of interest and test outcomes in a computer model. If it adequately explains empirical observations and performs realistically, such a model could usefully replace the more informal, partial, and relatively unspecified models currently invoked to explain the links between behavior and brain function in primates. Previous attempts to develop causal models of behavior have, in the main, had inadequate descriptions of behavior and inadequate descriptions of neural mechanisms.

We present here a system model of the primate neocortex, which shows how the set of specialized cortical functions can be put together using the connectivity of the neocortex to produce real behavior. Our model was developed using computer science concepts and is based largely on the anatomy of the macaque monkey. We will explain its use for the representation of the behavioral states of primates and will demonstrate behavioral scenarios involving affiliation and social conflict.

A *system model* comprises a set of processing modules that are connected together to form a *system*. This does not necessarily imply that the model becomes mathematically tractable or that individual modules need be represented in some standard way. A system is simply a common sense idea of treating an object of study as a set of interacting subsystems, each of which is easier to understand and to describe than the complete system. It results in explanations of objects as due to the action of each subsystem and the interaction among the subsystems.

In this paper, we will analyze the primate neocortex as a system. We are concerned with the processing of information, not energy, nor concentrations of specific chemical substances. Thus system modules will process and store information, and interactions among modules will be based on the transmission of information.

System level explanations of brain activity are well-known in neurobiology [Wernicke, 1874; Geschwind, 1965; Goldman-Rakic, 1988; Petersen et al., 1988; Gazzaniga, 1989; Deacon, 1989; Mesulam, 1990; McIntosh et al., 1994; Kosslyn, 1994]. However, our work is the first application of a system approach to a computational model of the primate neocortex.

There are two key questions currently facing primatology: Can a system model of the primate brain be developed which has a correspondence of subsystems to brain elements? Can we represent behavioral states of primates using a system model of the brain? In this paper, we will answer these questions in the affirmative and by explicit construction. A behavioral state will be the state of a causal system model, which has dynamics based on rules that allow us to determine the behavioral state at the next moment in time.

What is new is that the model shows that we get a system model with all the advantages of goal-direction, feedback, and regulation if we simply model each cortical region by a continuously acting process that maintains (constructs, stores, and transmits) data of the types specific to that region. The set of processes gives us a system model of behavior whose elements correspond to identifiable neural regions. To get the model to work, we have had to carefully analyze the data types observed in the action hierarchy and to reinterpret them in a computer science framework, to give notions of goal, plan, etc. as cortical data types.

From this we obtain detailed fMRI predictions, of spatial distributions of cortical activation, for behaviors represented using the model.

The other main new advance is that the model shows how a perception-action model can result in a model of social interaction. This occurs because each animal continuously perceives the other and continuously acts toward the other conditionally upon what it perceives, and because we can represent joint plans in the model. The hierarchical organization of the perception-action system allows a hierarchical description of the social interaction, which is natural and which corresponds to a pattern of activation of a set of cortical regions.

PERCEPTION AND ACTION HIERARCHY IN THE PRIMATE BRAIN

It is well known that primate neocortical areas involved in perception and action form hierarchies of connection and processing. Examination of experimental evidence for *functional involvement* of different areas of the brain shows a hierarchy of function and data types [Bond, 1997].

To reduce complexity, and following Pandya and Yeterian [1990], we will work with *regions* made up of several neural areas. We show our regions on a lateral view of the cortex in Figure 1a, with indications of their functional involvements.

In Figure 1b, we diagram the known corticortical connections [Young, 1993] from the hierarchies of perception regions to frontal regions. We used reported strength of connections, i.e., percentage of observed neural connections, as well as clustering and functional involvements, to assign areas to regions and regions to levels in the hierarchies. For detailed analysis, please refer to Bond [1997]. As can be seen, the connections between the perception hierarchies and the action hierarchy are clustered between regions at the same hierarchical level. Usually a region is connected to the corresponding region on the same level and also to regions immediately above and below this. We have not shown additional connections between different perception hierarchies nor additional connections between different levels of the action hierarchy.

OUR SYSTEM MODEL

Our Computational Approach

Basic architectural assumptions. To obtain a computational architecture for the primate brain, we took the functional hierarchy composed of neural regions and interpreted each region as a computational process or module. Each module in general has processing and storage abilities. We interpreted connections between neural regions to be communication channels that data may be transmitted along.

We will assume that all control and all processing in the system is accounted for by the local processing in each module. This assumption of distribution is based on a general observation that the cortex seems to be distributed without a global controlling process. This local processing can only operate upon data received by the module or stored in the module. There is no global control process that does monitoring, caretaking, or control of transmission. The assumption of distribution of data and distribution of control is a major constraint on the structure and dynamics of possible models. The distribution of memory and control in our model differs greatly from a conventional artificial intelligence (AI) computer program, diagrammed in Figure 2.

In a conventional AI program (Fig. 2a), there is a program that is serially

(a)

(b)

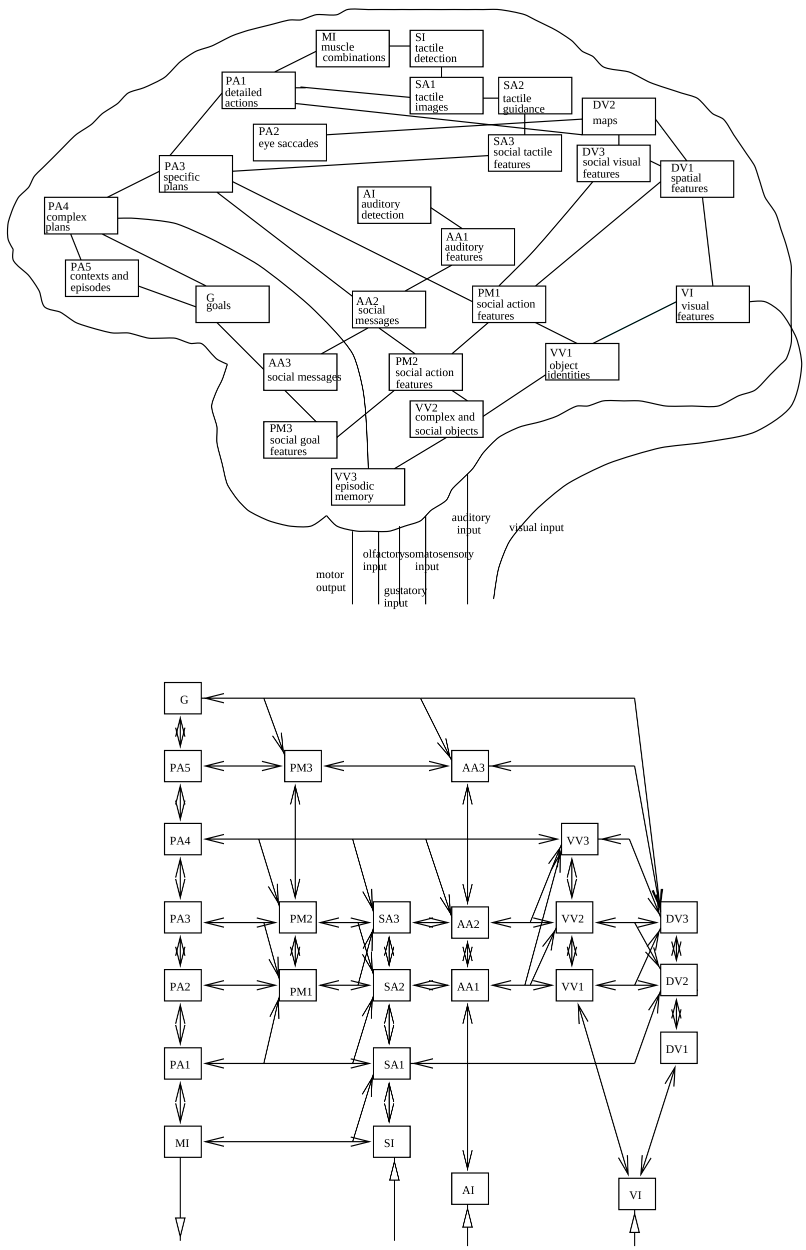


Fig. 1. **a:** Lateral view of the cortex showing regions and functions. b: Cortical regions and connections form a perception-action hierarchy.

executed one instruction at a time by a processor. During this execution, it reads and writes data from and to an addressable memory. The program can examine any part of the memory, one item at a time, and has a *global data perspective*. Also, there is a *single point of control*, that is, at a given moment in time there is one place in the program instructions where activity occurs.

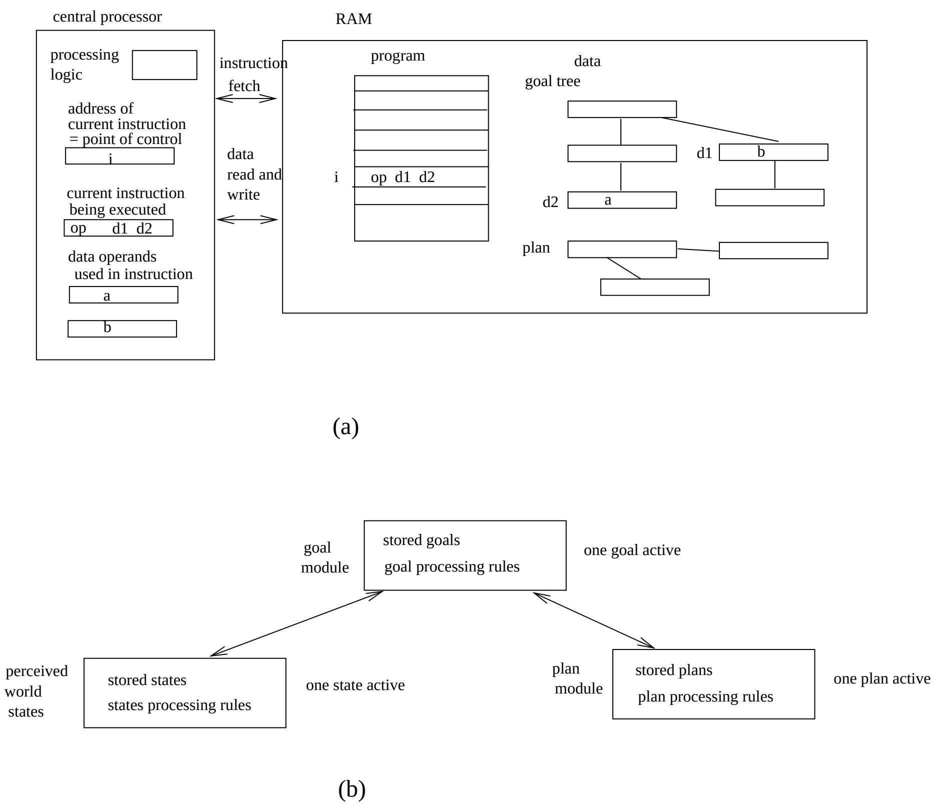


Fig. 2. Difference between Conventional AI Program and our Distributed System.

Our model uses distributed data and control (Fig. 2b). The data is not all in one memory but is in separate modules depending on data type. There is a process in each module that can access and modify only data in that module. There is therefore no process that has access to all data, and no global data perspective, and there are many distributed processes, not a single control focus.

Each cortical region will be represented by a continuously acting module, which is a process with stored data items. The main determiner of processing will be the type of data being processed (rather than the function being computed), different regions being specialized for different data types. Depending on the temporal characteristics of the module, these stored items may constitute volatile, short-term, or long-term memory. Items that are activated as a result of computation will have their activation sustained and will correspond to working memory. Thus, potentially, both long-term memory and working memory are distributed over the set of modules. Petrides [1994] has made a similar observation.

The modules operate concurrently, that is, they all operate at the same time in *parallel*. The characteristic time for a brain region to process data is about 20 msec (Edmund Rolls, personal communication).

Logical representations. The representation of system modules by neural nets, and the representation of module interaction by large parallel sets of axons carrying neural signals, seemed to us to be impractical for systems with ten or more modules. Instead, we turned to a computer science approach.

We adopted a very general and flexible method that uses predicate logic. Information in the brain will be represented as sets of *logical expressions*. Logical expressions are thought of as representations of assertions that something is true. Examples of the simplest logical expressions, called literals, are **position(adam,360,300,0)** and **goal(affiliation,affiliate(adam,alice)).** We will refer to these simple logical expressions in our model as *descriptions*, and these represent the information structures that are stored and processed in the brain. Thus, **position (adam,360,300,0)** could be a description stored in a brain module and representing the information that this module has concerning the position of a primate named adam.

Thus, a module will store descriptions, processing will take descriptions as input and generate descriptions as output, and information transmission will consist of transmitting descriptions from one module to another.

Thus, the action of a large set of parallel axons transmitting a firing pattern of neural spikes will be abstracted to the transmission of a stream of descriptions, which represent the information being communicated between the modules. For example, a set of 10,000 axons carrying a population coding of an angle giving the orientation of a perceived primate might be abstracted as an expression of the form **body_angle(Primate,Angle),** for example, **body_angle (adam,35),** or a form carrying breadth and height of a population code might be **body_angle(Primate,Angle, Height,Width),** for example **body_angle (adam,35,10,3)**.

We will call the outer name of a description, such as "**position**," a *predicate name*. We can capture the idea that a given module only processes certain kinds of data by limiting it to storing and processing only descriptions with a given limited set of predicate names. Thus an early visual module might store descriptions of the form **retinal_position(Feature,RX,RY),** and auditory module **sound(Frequency,Intensity),** a motor module **action(Motor_action,X,Y,Z),** and a planning module **plan_action(Action,Agent).** We will use the convention that names can have any number of characters; names denoting variables will start with a capital letter, and constants will start with a lowercase letter.

The processing within a module may actually be a sophisticated learned and learning neural net, but we will abstract and approximate information processing by a set of processing *rules*. To illustrate the idea, suppose we have a module which computes spatial relations among other primates and the subject. Then it might receive the positions and velocities of these primates from other modules and then compute which primates are approaching each other in space. A simplified rule to do this might look like the following:

```
position(M1,X1,Y1,Z1),velocity(M1,Mag1,Dir1),
position(M2,X2,Y2,Z2),velocity(M2,Mag2,Dir2)
→
direction(X1,Y1,Z1,X2,Y2,Z2,Dir),
Rvel is cos(Dir1-Dir)*Mag1 + cos(Dir2-Dir)*Mag2,
Rvel > 0,
assert((approaching(M1,M2,Rvel))).
```

This rule calculates the angle of the line connecting two perceived primates and then calculates their relative velocity along this line. If this is greater than zero then a description, indicating that these primates are approaching one another, is generated, which includes their velocity of approach.

A rule has then three parts: i) input patterns matching to stored descriptions and which bind variables such as **X1, Y1,** etc. to explicit values; ii) the rule body in which some calculation involving the values of these variables is carried out (The complexity of this calculation is intended to be that achievable by a simple neural net, which we imagine to allow arithmetic operations, simple geometric functions and inequality tests. The body is a sequence of operations; if any fail, then the rule fails for the particular values of the variables being used); and iii) result patterns which give descriptions computed if the rule succeeds. Results may be output to other modules and/or stored in the processing module.

We use names such as **body_angle, position,** etc. using English words, which suggest the meaning of the expression. However, the precise scientific meaning of expressions is determined by their effect upon the action of the system model. We can call this their *computational semantics*. Thus, if information in descriptions like **position(M,X,Y,Z)** behaves like position information to the system, then we could take it to correspond to a position. It may have some properties of positions but not all, for it may be limited in accuracy or it may not be able represent all possible positions such as very distant ones, etc. In predicate logic, predicate names are arbitrary tokens and the system will work exactly the same with other names. For example, instead of the name "**position**," we could use the name "**xx**." Since the meanings of expressions are only determined by the dynamics of the model, this abstract logical approach enables us to work with different kinds of information, which may be processed by the brain but which do not correspond to sensory or motor information or even to any current English words.

The big advantage of a logical approach is that it allows arbitrary information structures and arbitrary information processing to be represented. It includes real variables and functions of real variables, but it also allows us to represent discrete structures such as plans as sequences of actions, and the conditional temporal sequential processing of plans. Continuous representations such as fields and differential equations cannot do this easily. In addition, logical expressions are linguistic forms, which can be made intuitively expressive by an appropriate choice of names. This allows us to bring a precisely specified model close to the language that practicing primate neurobiologists might use in discussing information and its storage, processing, and transmission in the brain.

Each stored description has a *weight*, which is a real number in the range [0.0,1.0] representing its strength at the current time. Weights attenuate with time but are refreshed by incoming matching descriptions. Transformations use description weights to compute the weights of the descriptions they create, using a weighted linear sum. Each module can specify conditions of competition among rule activations. Typically, only the results of the strongest rule activation are transmitted to other modules. By the use of logical representations, we avoid all use of random access addresses; instead, basing all computation in the model only on associative matching and competition.

The system model uses a discrete time, with one unit of time corresponding to 20 msec. In each time unit, all matching instances of all rules in all modules fire and then all transmissions of information among modules occur. This is intended as an abstract representation of all the activity that occurs in the brain during that period of one time unit.

States and causal dynamics The state comprises the current set of stored descriptions in each module, the set of rule instances firing, and the set of descriptions being transmitted along the channels. The dynamics of the system is given by the action of the rules in each module, together with the action of storage management functions such as store updating and attenuation. The description of states and causal dynamics constitutes a scientific theory of the action of the brain.

Perception-Action Hierarchy

Having established a computational approach and the use of logical representations, we now turn to the architecture of the neocortex, which as we saw is a perception-action hierarchy. Figure 3 shows how a perception-action computational architecture could support the functioning of the brain in behavior. A plan is selected and elaborated, receiving input from the perception hierarchy to allow it to elaborate appropriately.

Within a given level, the component of the action hierarchy at that level is elaborated down to the next lower level, and evaluations are assessed and transmitted back up to the next higher level. By *elaboration*, we mean taking data that describe action at one level and generating data that describe that action in more detail. More detail includes 1) exactly how to act (which detailed action components), 2) in what order, 3) exactly at what times, 4) exactly where in space, and/or 5) who will do which actions. By an *evaluation* we mean, for example, a value indicating progress, success, or failure; such a value can also be associated with a particular description, for example, one representing a particular action or a particular goal. The perception hierarchy and action hierarchy cooperate

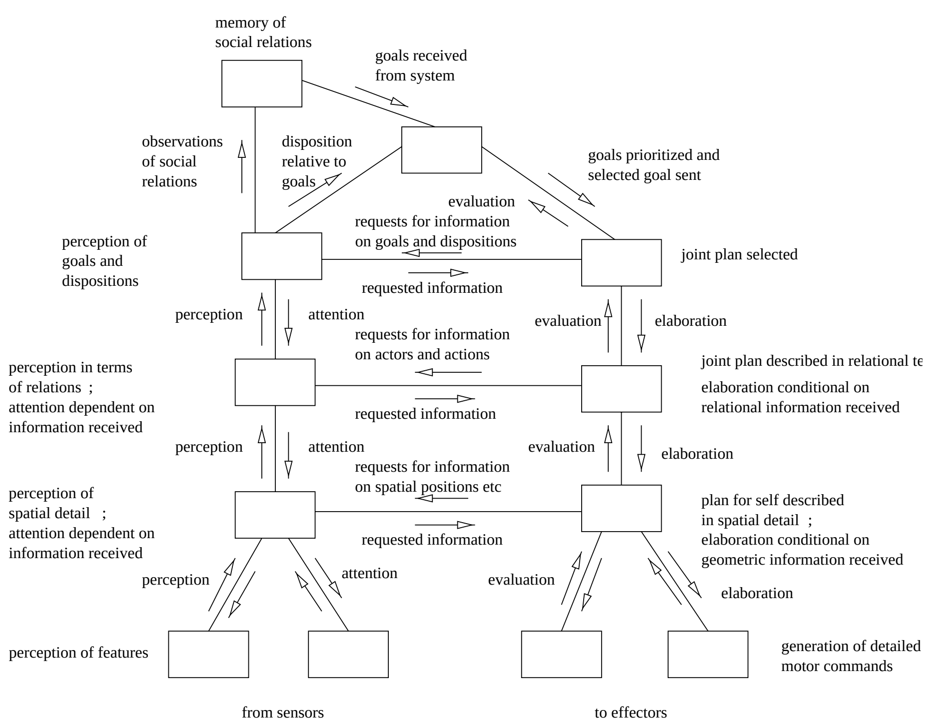


Fig. 3. Functioning of interacting perception and action hierarchies in behavior.

closely. The action hierarchy must elaborate the currently selected plan conditionally upon the perceived environment. The modules of the perception hierarchy at a given level derive information required for successful action elaboration at that level.

The perception hierarchy receives descriptions representing tuning information and direct requests, attention information, and prediction information from the action hierarchy. This information provides a context for perception and enables the optimal use of processing and communication resources by the perception hierarchy in supporting the realtime action. Thus, our perception-action architecture provides a framework for attention mechanisms.

Plans. We need to represent plans that the system can execute, and for this we generalized the standard artificial intelligence representation of plan to one suitable for action by more than one collaborating agent [Bond and Gasser, 1988]. A *plan* will be a description that represents a sequence of joint actions and that causes the system to attempt to carry out such actions. Each step in the sequence specifies an action for every agent collaborating in the joint action, including the subject agent, each such action being conditional upon specified tests on the perceived situation. For example, a primate's plan to take offered food from another's hand cannot proceed unless it determines that the other's hand is also on a correct corresponding trajectory.

The way a plan is executed is to attempt each step in turn, and during a step, to verify that every collaborating primate is performing its corresponding action and that any tests are satisfied, and to attempt to execute the corresponding individual action for the subject primate.

Goals and plans as data. Our model is a set of active data stores, each of which is specialized for a different kind of data. Data are patterns of neural activity in sets of neurons and we represent these by symbolic expressions. The meanings of these data are not derived from arbitrary names used in them but from their function and role in the system. Goals and plans are also treated by us as particular types of data, and there are stores specialized for storing and processing them. A goal expression represents a goal for the system in that it tends to make the system seek to attain that goal. Similarly, a plan tends to make the system carry out that plan.

Flexible hierarchical control. In computer science, the term control concerns the organization of activity, what gets executed, when, and under what conditions. The interacting perception-action hierarchical arrangement can support the flexible control of resources, since it can act differently and conditionally depending upon the availability and strength of data, and upon the computational resources and time available. Where the plan is uncertain, the lower levels of the hierarchy can function using stored defaults, but when the plan is generated from a goal, it can control and modify activity in the lower levels of the hierarchy by sending information to them. Where the plan elaboration needs more, different, or more accurate perceptual data, it can send such requests to the perceptual hierarchy for finer tuning and attention.

Our Initial Model of the Primate Neocortex

A primate minisociety. In order to set up a brain model, we needed to decide what environment the brain would have and what behaviors to consider. We chose to consider the case of social interaction. Our strategy was to build in social interaction into our brain design from its inception. Social interaction is arguably the most general type of behavior and leads us to construct a general model. Social behavior involves perceiving dynamically changing environments of primates who have complex dynamics. It involves generating social behavior, which is joint and requires real-time coordination of action. We used a "minisociety," in which a group of primates interact socially, in a naturalistic three-dimensional environment, with each model primate controlled by a brain model. Thus the instantaneous state of the environment is mainly the positions, orientations and configurations of these primates. Motivation in the system is achieved by defining a social memory module, which stores knowledge of affiliative relations, and which, among other things, generates affiliation goals, since affiliative behavior is a known driving force in primate groups [Kling and Steklis, 1976]. We will refine this in a later section.

The initial model. In order to construct this model and to determine its properties, we developed an initial brain model consisting of data and process representations, with eight memory modules. Figure 4a shows our initial model.

An outline of each of these memories, the descriptions they store and the processes they include, is as follows:

i) the *social relations* module contains all memory of social information, including kinship, affiliative, and dominance relationship information. It generates affiliation goals (represented by descriptions) and sends them to the *goals* module.
ii) The *goals* module contains all goals currently held. It activates the most important goals and sends this information to the *overall plans* module.
iii) The *overall plans* module receives goals and instantiates suitable joint-plans, sending them to the *specific joint plans* module.
iv) The *specific joint plans* module receives a joint-plan and generates a detailed action based on descriptions received from the perceptual hierarchy. For the others involved in the joint plan, the detailed action or state is verified, and for the self, its detailed action is sent to the *detailed actions for self* module.
v) The *detailed actions for self* module receives the detailed self action from the *specific joint plans* module, receives object and location information from lower levels of the perceptual hierarchy, mainly from the *primate positions and movements* module, and outputs a detailed motor action for this to the *motor* system.
vi) The *primate positions and movements* module receives sensory descriptions of the state of the external world and provides information on requested primates to the *primate actions* and relations and *detailed actions for* self modules.
vii) The *primate actions and relations* module computes higher-level descriptions of the action of each primate involved in the current joint action. It requests information on particular primates from the *primate positions and movements* module.
viii) The *plan primates* module receives information from the *overall plans* module as to which other primates are involved in the joint action and passes this on to the *primate actions and relations* module.
ix) The *motor* system does some processing to generate the external action given the direct action received from the *detailed actions for self* module

Note that we have very much simplified the perceptual and motor hierarchies in this initial model. The perceptual hierarchy is simply the *primate posi-*

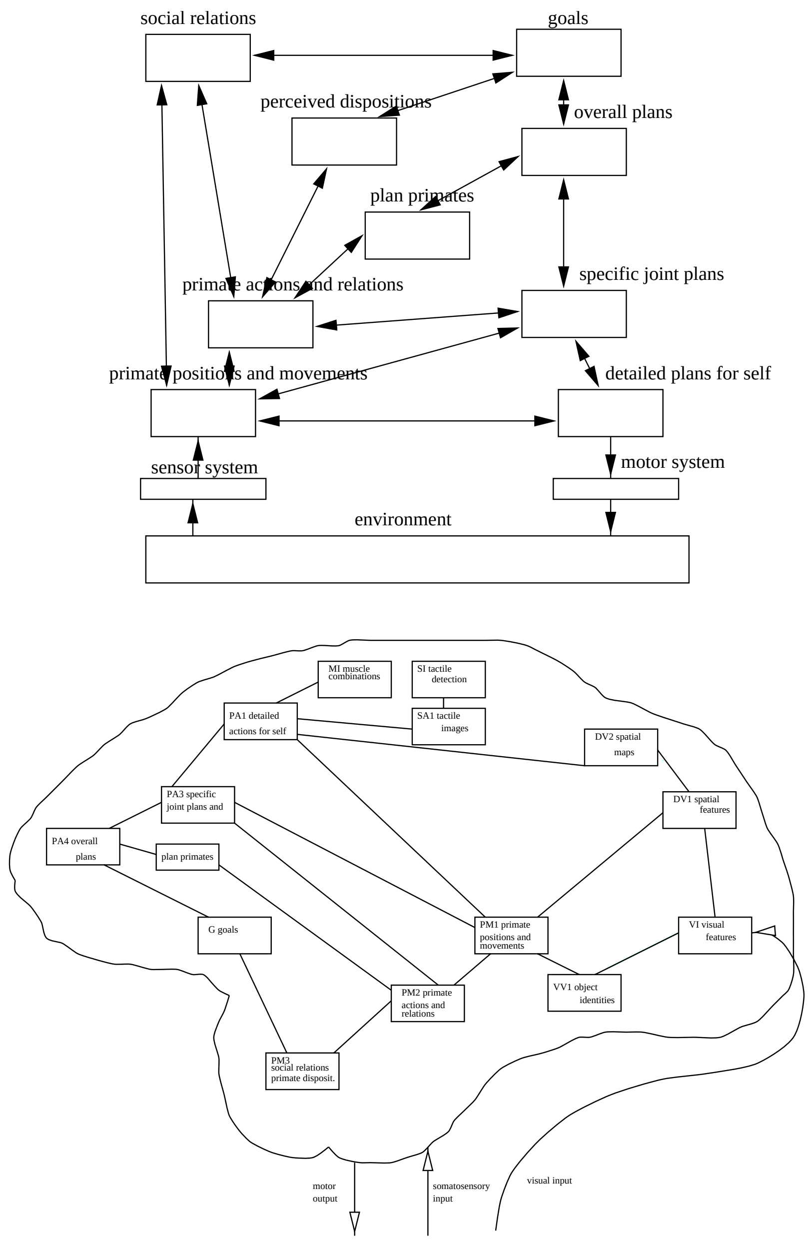


Fig. 4. **a:** Initial system model of the primate brain. **b:** The approximate correspondence of the system model to the neocortex.

tions and movements and *primate actions and relations* modules, and the action hierarchy is the *overall plans*, *specific joint plans* and *detailed actions for self* modules.

The environment is modeled by a set of descriptions of what is the case at the current time. Sensing consists of sending sensing goals to the environment which returns descriptions satisfying those goals, which are then put on input channels. Effectors send proposed description changes to the environment, which determines what changes actually take place.

Correspondence to the neocortex. Figure 4b shows the set of implemented modules with their potentially corresponding cortical locations. This approximate correspondence locates the perception hierarchy along the superior temporal sulcus, following Perrett's findings, and with episodic memory for social relations in the anterior of the temporal lobe. Goal processing is predominantly in anterior cingulate. Specific joint plans and detailed plans for self are in dorsal prefrontal. Tactile sensing in somatosensory regions and spatial maps in dorsal visual regions were used in our extension of the model for social spacing behaviors. This also used a simple low-level spatial navigation module which could be tentatively identified with premotor cortex.

Motivation. Kraemer [1992] has reviewed mechanisms of attachment. Innate affiliation schemas may be subcortical and form part of a social attachment feedback control system. During development, these schemas are probably developed into working models, and particular instances of affiliation relations will probably be stored cortically. Thus, a better account of the generation of affiliative goals would be that both subcortical and cortical representations of specific affiliation relations would generate signals that would be propagated to the anterior cingulate gyrus. The cortical component would indicate the specific affiliation involved and the subcortical component its intensity and other more generic qualities.

The Action of the Model

How one module works. Figure 5a shows the basic idea of how a module works. Descriptions continuously come in along channels and are stored, and rules continuously match to the store, generating new descriptions, which are usually output along channels but may also be stored in the same module.

Initiation of action. From a quiescent state, goals are generated by various subsystems and sent to the goal module. The goal module selects the most important goal and, if this is sufficiently important, sends it to the overall plans module. The overall plans module selects a suitable joint plan and sends it to the specific joint plans module. That module verifies that the collaborating others are in a state compatible with the joint plan and extracts values of variables from this perceived state, for example, the particular posture and spatial relationship to the subject. These values allow the module to select and elaborate the appropriate detailed action for the self, which it sends to the detailed plans for self module. This module, from detailed positional information received, constructs the detailed movement patterns and sends them to the motor system. This process of goal generation, goal selection, plan selection, plan elaboration, action specification, and motion specification proceeds continuously, renewing the information every time cycle.

Figure 5b shows how goals are continuously generated, prioritized, plans selected and instantiated, detailed joint plans constructed, and self plans and self actions constructed in a cascade of specialized processes. This corresponds to the sequence of frontal regions.

Cortical processes and confirmation signals. We developed a notion of *cortical process* that is a distributed set of rule activations in different modules which are dynamically linked together by transmission of a type of descriptions that we call confirmation signals. The way confirmation signals work is as follows. If a module receives a data item that causes activity, i.e., some

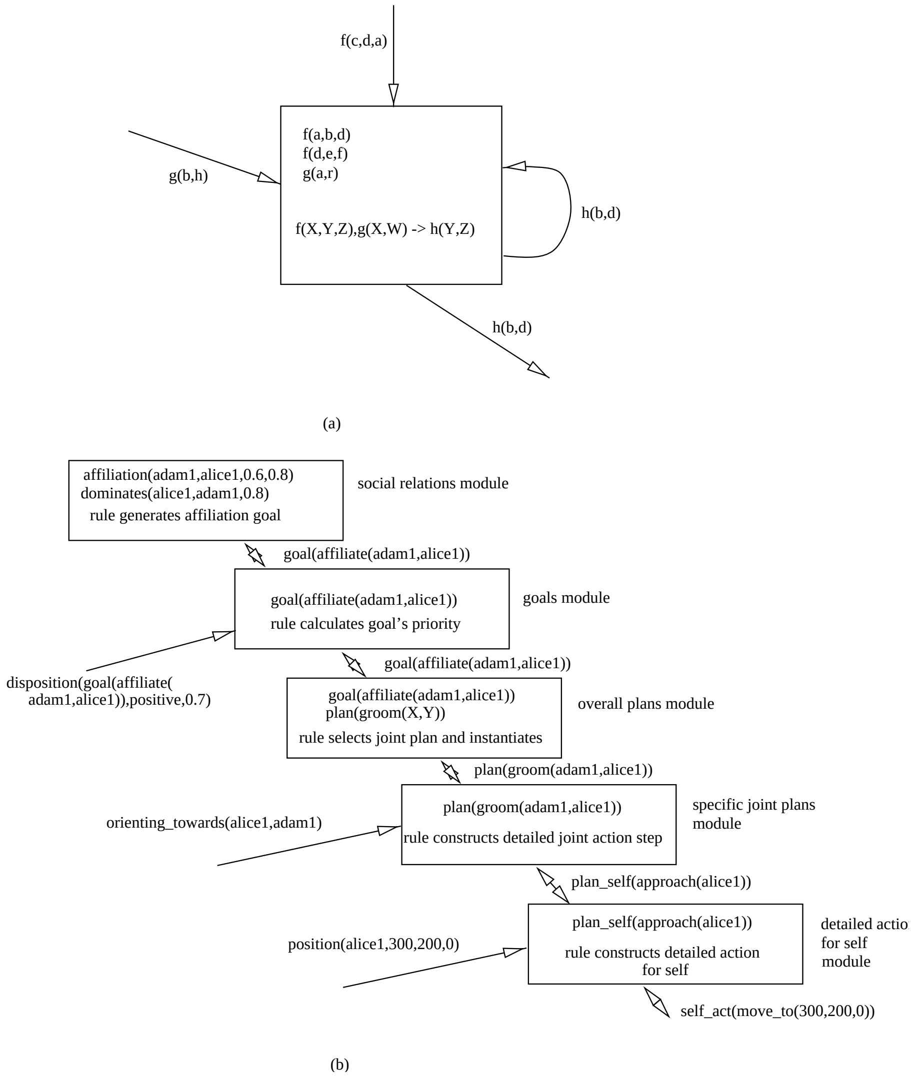


Fig. 5. How the model works.

rule to fire, then its sends a positive confirmation message back to the sender, in the form of a confirmation description, evaluating that data item and boosting the rule activation that sent that data item. Confirmation signals tend to stabilize distributed activity. If, on the other hand, received data does not cause any execution, a negative confirmatory signal is sent back, which tends to attenuate the sending rule activation and thereby to allow competing choices to be tried.

The basic action of the brain model is to try to establish a plan consistent with the response of the environment and with its own goals. It does this by trying different alternatives at each level on a competitive basis, and subject to confirmation of successful elaboration.

Continuation of action. Provided the internally generated goals and the external environment do not change greatly, the continuous process of plan elaboration, perception, and action will continue. A change in spatial positions will simply result in different positions being perceived, this positional information being passed to the self action module and a different and more appropriate detailed action being generated using the updated position. The other levels will continue as before. Thus, the system will track changes in position.

Greater changes in position, posture, and action may result in different spatial and action relations being perceived at the next higher level. The relational information passed to the action module at this level may cause a different type of self action to be generated, but one that is still consistent with, and an elaboration of, the more generally specified plan received from the level above this.

Thus, the levels of the hierarchy of perception and action correspond to a hierarchy of control concerning variations of 1) new positions and/or orientations, 2) new spatial relations, action types, or action phases, 3) new plans, 4) new goals, and 5) new social situations, respectively.

GROOMING BEHAVIOR AND JOINT ACTION

Our first attempt at a joint action demonstration scenario was to examine a prototypical situation in which two primates groom. In order for this, or any other, joint action to occur, each primate would have to develop joint goals and joint plans; it would have to be guided in its actions by feedback from perceiving the other's actions. In this way, the initiation, development, and achievement of the joint action could be modeled. We developed a four-phase plan for a groomer (orientation, approach, grooming-prelude, and then grooming), and a groomee (waiting, orientation, grooming-prelude-response, and then grooming-response), and we developed suitable rules for activity in each module in each phase. We ran our computer realization and the primates did indeed carry out the four phases described, leading to a primate named adam1 grooming a primate named alice1, which involves joint planning and action. For two primates, each has a joint plan in which the other's action is represented and verified at each step. Furthermore, each primate's actions at each moment depend on what the other is doing. The joint action is established by each primate trying different plans and variants competitively until a match is achieved and a successful joint plan is established for each primate. This matching process occurs because of competition among data and rule activations, and because of feedback of success and failure in the processing hierarchy. These simple social behaviors were obtained using about 15 rules per module. The implemented system automatically generates animated movies.

Figure 6a shows our formulation of pairwise grooming. Figure 6b shows a movie frame from pairwise grooming behavior obtained with our system. We show in Figure 7 an instantaneous state of the model.

We show in Figure 8 an instantaneous state of the model, with the two interacting primates. At that moment, adam1 is walking toward alice1 as a result of selecting a goal to affiliate with her, and to do this by grooming her. He is perceiving that alice1 is in the process of orienting toward him and takes this into account in generating his own action of walking directly toward her. She perceives him walking toward her, and since she has selected a goal of affiliating with him by being groomed by him, she generates an action to orient toward him and therefore to turn her head toward him. We show the left-hand sides of dominant rules in each box and the transmitted right hand sides on the lines representing channels.

(a)

Phase	groomer's behavior	groomee's behavior
Orientation phase	groomer orients to and moves towards groomee groomer is far from groomee	groomee may remain passive
Approach phase	groomer is sufficiently near to groomee and is visible to groomee	groomee should at least acknowledge the presence of the groomer by some orienting - eyes, head, body, etc.
Grooming prelude phase	groomer is very near to groomee within touching distance orientation should be mutual groomer initiates grooming prelude	groomee is mutually oriented to groomer, accepts grooming prelude by giving grooming prelude response
Grooming phase	groomer moves to grooming position and grooms groomee	groomee responds to grooming by giving grooming response

(b)

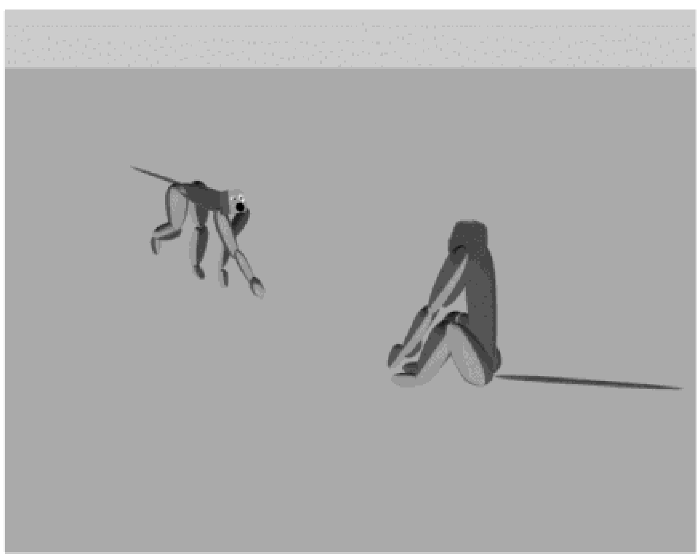


Fig. 6. **a:** Our formulation of grooming behavior. **b:** Grooming behavior obtained with our model.

SOCIAL CONFLICT, CHANGE, AND TERMINATION OF ACTION

We next developed a more complex scenario involving four primates and social conflict. The idea was to set up a situation in which at least one primate would set up an initial goal to affiliate with another but then would find that it could not, since it would not be receiving cooperative feedback, and so it then would turn to another goal to affiliate with a different other primate.

Figure 9a shows the data used for the social conflict scenario. Figure 9b illustrates the social conflict scenario obtained with our system. Behavior was achieved in which conflict occurred and a change of cortical process was needed.

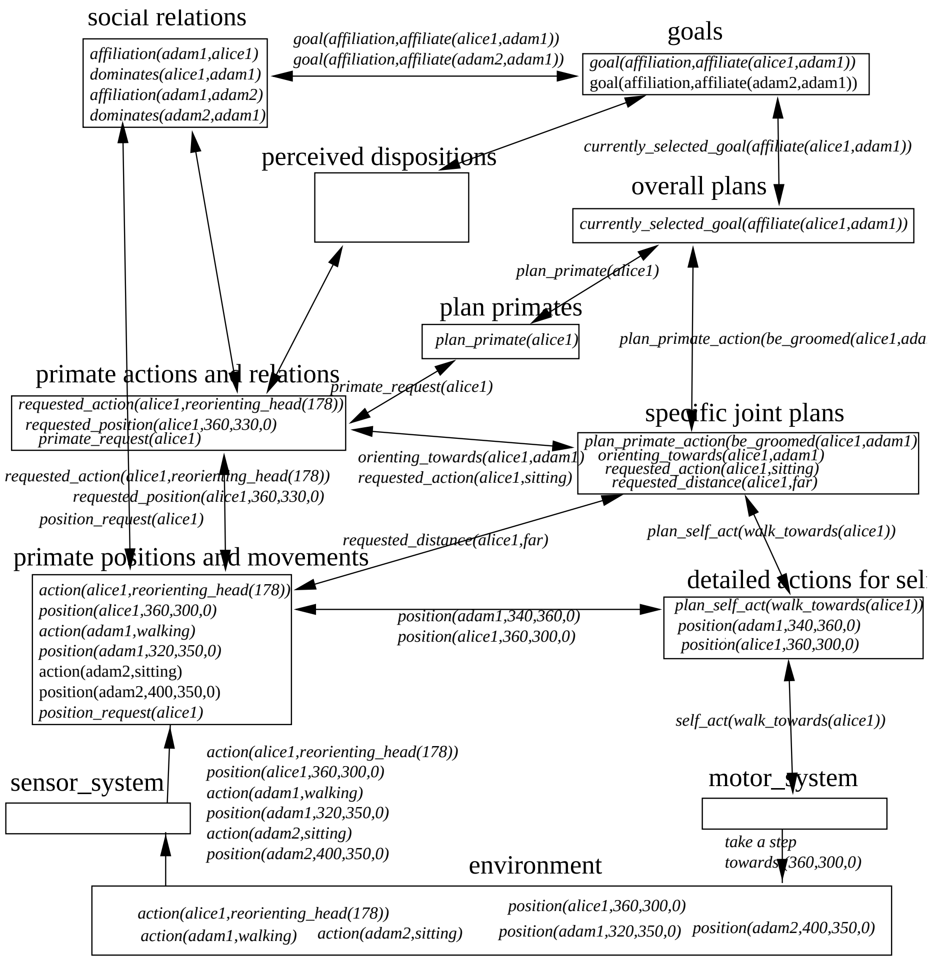


Fig. 7. Example of a behavioral state of the model.

The “moment of truth” is captured in Figure 9b where adam1 realized from adam2’s walking toward alice2 that adam2 did not wish to enter into joint activity with him. This caused disconfirmation of his elaborated plan to groom with adam2, and eventually disconfirmation of the corresponding goal. A new goal was competitively selected to groom with alice1, and this joint action was able to be completed. The existing distributed activity collapsed as the disconfirmation propagated upwards, producing a kind of “stunned silence’.” Then, a new goal was competitively selected and a new plan elaborated downwards.

We also ran the same scenario with a different brain model, which had an additional *perceived disposition* module for the perception of the dispositions of others. Dispositions were represented as positive or negative evaluations of certain goal types. A disposition represented the subject primate’s perception of the attitude of another primate toward a given goal. Perceived dispositions were stored in the perceived dispositions module and from there transmitted to the goals module. The weights of goals generated were made conditional upon this primate disposition feedback, The change of plan was then accomplished more

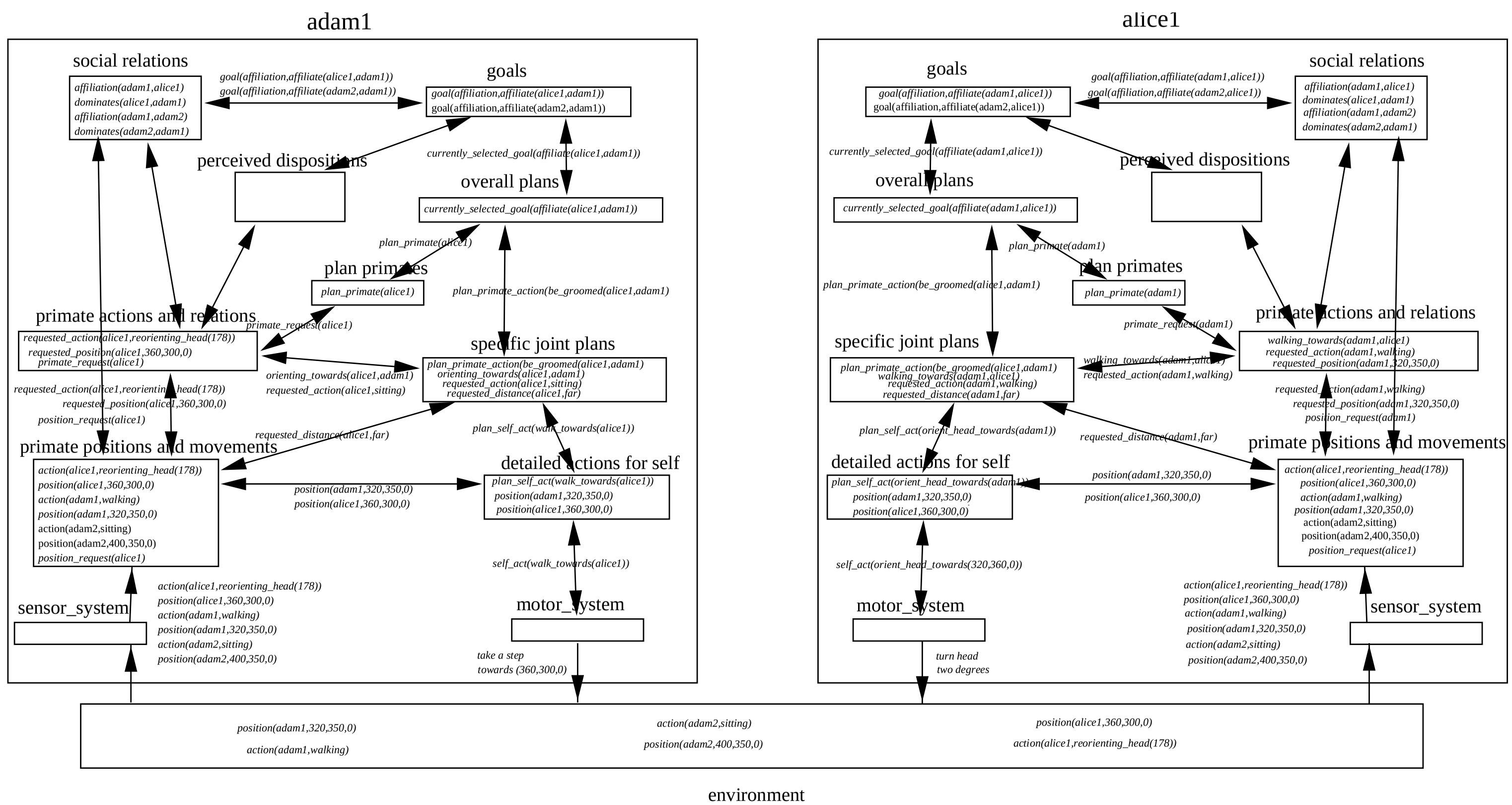
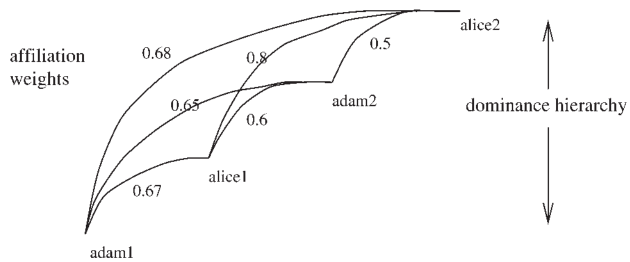


Fig. 8. Instantaneous behavioral states of two interacting primates.

(a)



(b)

goals generated

alice1 wishes to	groom adam2 be_groomed by adam1	0.75 0.4875
adam1 wishes to	groom adam2 groom alice1	0.5625 0.4875
adam2 wishes to	groom alice2 be_groomed by alice1 be_groomed by adam1	1.0 0.75 0.5625
alice2 wishes to	be_groomed by adam2	1.0

(c)

Fig. 9. **a:** Data involved in social conflict scenario. **b:** Goals generated by the social relations module at the beginning of the scenario. **c:** Moment of social conflict.

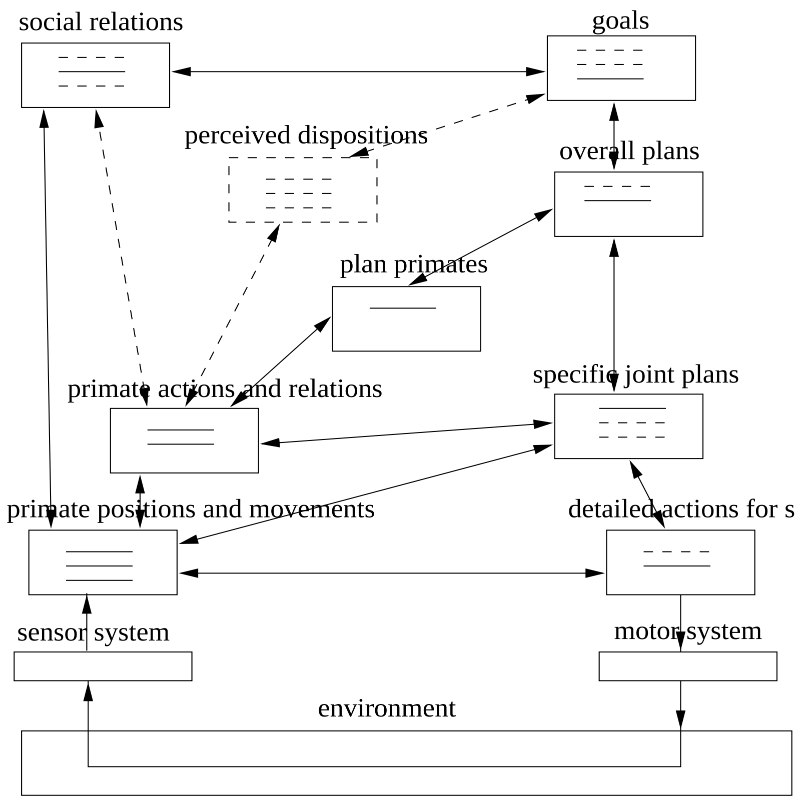


Fig. 10. Behavioral state as coherent distributed neural process, solid lines, descriptions which are components of dominant active rules; dashed lines, other descriptions.

smoothly and quickly. As soon as the negative disposition of adam2 was perceived by adam1, his goal to groom him were rapidly reduced in weight, allowing the alternative goal to become selected. This was accomplished without prior dismantling of the existing process, the new process simply displaced the old one competitively as it elaborated to each level.

COHERENT BEHAVIORAL STATES

Primate behavioral states are the instantaneous states of our system model of the primate neocortex. We have explained above that a state is the set of active rules and descriptions in each module together with the descriptions currently being transmitted between modules. We will use the notion of *coherence* to indicate the degree to which this distributed state holds together and is relatively stable over time. Figure 10 attempts to depict a behavioral state as a *coherent* dynamical process that is distributed over active modules. This state is maintained and changed by descriptions and confirmation information sent between modules.

We have already described different kinds of changes and adjustment of behavioral states that occur in our model, during the initiation of action, its adjustment to environmental change, and in major changes due to a change of dominant goal. We found that the time to establish or to change between coherent states was of the order of 10–20 cycles, which corresponds to approximately 200–400 msec.

DISCUSSION OF SOME ABSTRACT THEORETICAL ISSUES

A *hierarchy* is a partial ordering of some elements, and when we consider an information-processing system. there can be many different kinds of ordering relation applying to different aspects of the information processing. Data representation is perhaps the clearest example, where the things represented are progressively more general, representing greater amounts of space and time, more stimulus dimensions, more objects, more actions, and so on. In a hierarchy of control, what is controlled from one level is more general than the level below. The properties and dynamics of hierarchical control are quite rich and subtle. Hierarchical systems work well since they allow a large complex system to be realized using a set of processes of limited processing power, and the hierarchical architecture provides a framework for organization, control, and coordination for the set of processes.

We need models that are capable of a wide variety of behaviors. Ideally, we would like any model of behavior to be able to represent any function from stimulus to response. It has been known for a long time that all infinite computational machines are equally powerful, in that they can all compute the same large class of functions, called the computable functions. Such machines can be called *universal* machines. It is also well known [Minsky, 1967] that in order to be universal the main requirement is an unbounded data store. Our model is universal since it provides unbounded storage in its modules. It can compute any computable function, with infinite domain and range, provided it is given a suitable (finite) set of rules. This observation is of course academic, but it is relevant in discussing biological models. Any stimulus-response model has a finite number of elements and is thus a finite machine. Hence it cannot compute arbitrary behaviors; in fact any such model is limited to a fixed set of behaviors. In modeling complex adaptive behaviors and cognitive processes, we submit that it is necessary to have a model which is computationally universal.

This argument is analogous to the general argument given by Chomsky [1957, 1959] in which he showed quite generally that many syntactic description systems were equivalent to finite state grammars and that natural language was richer even than the context-free grammars and required machines more powerful than finite machines. Although it is true that a set of behaviors limited to a finite set of combinations of elements is finite and therefore could be represented by a finite model, this would not give the scientific insights that an infinite model can give. Chomsky used the notion of *creativity*, that language users always produce novel sentences, and we can perhaps use it in connection with behavior.

Control is trivial in simple models; the system selects an item from a table and outputs the behavior. We would like more complex control, where different mental states occur in a course of computation that depends conditionally on what is currently stored and what is currently being perceived.

Another big difference between a stimulus-response model and our model is that ours does *continuous monitoring and goal setting*. Indeed, the perception-action architecture, the way we have implemented it, has great properties. It provides for continuous top-down activity driven by goals but also allows ongoing perception to condition and modify that activity. Furthermore, it combines these two modes of operation dynamically. These properties also give advantages for describing joint action where two or more animals are motivated by a combination of goals and also perceptions of each other.

Our model may help in avoiding the problems of *indexing* in stimulus-response models. The indexing problem is solved to some extent since we are now selecting from a whole set of modules not just one, and then these modules compute consequences of the combined set of choices, so there is not a bottleneck of a single process or table that is supposed to discriminate all behaviors. In addition, stimulus-response models use single real numbers, but our model uses logical expressions, so these are richer descriptions allowing richer indexing. Our model uses a *logical representation of data and process* and this can represent a wide variety of complex types of data and process. This representation also allows us to have a universal computational model. It is related to the natural language used by biologists in describing biological data and processes. The deep relationship between logic and computation was first explored by Gödel, who showed that logical inference was a universal computational method. We suggest that the logical approach to computation is quite appropriate for the description of computational processes in biological systems. Each module could be described logically by giving what kinds of data it can store and the logical rules by which this data is transformed by the action of the module. Such a logical description would constitute both a theory of the module and a computational model of it.

DISCUSSION OF BEHAVIORAL ISSUES AND PREDICTIONS

There are some subtle and difficult issues in the description of information processing that can derail attempts at scientific description. Typically there are many different information processes with the same input-output performance. Thus for any process involving data describing high-level plans, etc., there will always also be a table-lookup or (finite-state) automaton process with the same performance. Of course the table or automaton will usually have an astronomical number of states. One can also distinguish between “high-level“ form and “table” form by the high-level form’s flexibility and adaptation to wider classes of situation. Biological systems will often develop very optimized and routinized forms of a given process. Furthermore, given current experimental methodologies, animals usually are trained to overlearn and routinize behaviors. We reach the situation where heavily overlearned routinized behaviors are indeed reduced to automatic sequences involving a minimum of feedback and control. Finally, scientific descriptions are encouraged, if not coerced, to be parsimonious, postulating a minimum of information processing apparatus.

In order to observe planned and goal-directed behaviors in a systematic way, allowing clear quantitative and statistically well-founded results, it may be necessary to develop new experimental paradigms involving the presentation of novel problems to the subject. This will involve developing sets of novel problems with controlled, or at least understood, interactions among problems. It will probably also require different training regimens. It will require understanding how to mathematically characterize different novel problems and solution strategies.

The model makes *predictions* and gives insights into primate behavior that are of interest to primatologists. The most direct predictions are patterns of brain activity, observed by brain imaging, showing more frontal involvement higher in the hierarchy when situations of greater novelty or degree of variation are being dealt with. These predictions have a detailed time granularity of about 20 msec. In order to obtain fMRI data for social behaviors, one could perhaps use a visual display showing video-clips of social interactions or an interactive video game where the subject makes moves in a social interaction game.

However, there are more directly behavioral predictions; for example, the model suggests that behavior is structured hierarchically, with different perceptual and action mechanisms at different levels. The model predicts temporal latencies when the subject has to set a new goal [Anderson et al., 1993]. The model also suggests that detailed temporal ordering of behavior may be subjected to hypotheses, rather than just frequencies where temporal order information is thrown away.

As regards comparison with lesion studies, the model has very few modules at the moment and therefore little redundancy. Completely knocking out a module would produce major disruption; however, one could get some phenomena such as utilization behavior [Lhermitte, 1983; Shallice et al., 1989] by lesioning the planning module, for example. If lesioning simply weakens a module then we would get other phenomena. There is some correspondence to a broad classification of frontal lesioning effects due to Cummings [Mega and Cummings, 1994] where medial frontal lesions lead to apathy, dorsal frontal lesions lead to executive dysfunction, and orbital frontal lesions lead to impulsivity. This would correspond respectively in our model to lesioning the goal module, lesioning the planning module and lesioning the interface to subcortical perception-action systems.

As regards individual differences in response tendencies or dispositions, the model would in the simplest approach postulate that such individual differences result from individual differences in the performance of brain modules. Thus, one would try from a set of observations to derive a set of brain module characteristics and a set of individual parametric values, perhaps something similar to Daigneault et al. [1992].

The model provides a representation and a computational architecture. The details of data types and rules in the model will come from primatological observation. Thus, detailed predictions of grooming rates, dominance versus fertility patterns, etc. really flow from rules developed by primatologists. The architecture and representation allows these rules to be organized into a cortically plausible system and to be connected with behavior. Byrne and Whiten [1991] have tried defining observed behavior by rule sets. Shallice [1988] has proposed a rule model for the frontal lobe.

In general, the model provides a system language whereby primatologists can express their insights in a precise scientific form and can compute predictions of their resulting theory. I have been working on defining this brain modeling language, and my computer system will be able to take theories expressed in this language and to create the corresponding computational model. What I have done is give some initial theories of grooming and spatial behaviors using my architecture.

Finally, our system model of the primate neocortex allows us to contemplate future models which will include cortical and subcortical processes and the interactions between them.

SUMMARY AND CONCLUSION

To summarize in the form of achievements and results, i) a system model of the primate brain was successfully designed and implemented; ii) a practical modeling approach was used based on predicate logic expressions and inference rules; iii) grooming behavior was formulated, implemented and successfully demonstrated; iv) a behavioral state was defined as the instantaneous state of the system model, which has a logical form and a diagrammatic representation; v) the behavioral states of two, or more, interacting primates represent social interactions; vi) the behavioral state is the state of a causal system model, which propagates in time according to the given rules; and vii) stable behavior results from coherent behavioral states, which comprise a distributed set of local rule activations linked by stable patterns of messages and corresponding confirmations; coherent behavioral states change in times of the order of a few hundred milliseconds.

ACKNOWLEDGMENTS

This research was carried out at the Sony Computer Science Laboratory in Tokyo, the Computer Science Research Department at The Aerospace Corporation, El Segundo, California, and the Computer Science Department at the University of California at Los Angeles. The author thanks Professor Stott Parker of UCLA for his steadfast friendship and support. This work has been partially supported by the National Science Foundation, Information Technology and Organizations Program managed by Dr. Les Gasser. The author thanks Professor Pietro Perona for his generous support and encouragement, and Professor Steven Mayo for providing invaluable computer resources, at Caltech, where this research is currently being pursued. He also thanks Professor Nicholas Blurton-Jones of UCLA for very valuable criticisms of earlier versions of the paper and Leslie Brothers, M.D., for valuable suggestions on writing style.

REFERENCES

Anderson JR, Kushmerick N, Lebiere C. 1993. The Tower of Hanoi and goal structures. In: Anderson JR, editor. Rules of the mind. Hillsdale, New Jersey: Lawrence Erlbaum Associates. p 121–142.

Bond AH. 1997. The current status of the perception-action hierarchy concept for the primate neocortex. submitted for publication.

Bond AH, Gasser L. 1988. An analysis of problems and research in distributed artificial intelligence: readings in distributed artificial intelligence. San Mateo, CA: Morgan Kaufmann Publishers.

Byrne RW, Whiten A. 1991. Computation and mindreading in primate tactical deception. In: Whiten A, editor. Natural theories of man. Oxford: Basil Blackwell. p 127–141.

Chomsky N. 1957. Syntactic structures. The Hague: Mouthon.

Chomsky N. 1959. Review of B.F. Skinner's book "verbal behavior." Language 35:26–58.

Daigneault S., Braun CM.J, Whitaker HA. 1992. An empirical test of two opposing theoretical models of prefrontal function. Brain Cognition 19:48–71.

Deacon TW. 1989. The neural circuitry underlying primate calls and human language. Hum Evol 4:367–401.

Gazzaniga MS. 1989. Organization of the human brain. Science 245:947–952.

Geschwind N. 1965. Disconnexion syndromes in animals and man. I. and II. Brain,88:237–294, 585–644.

Goldman-Rakic PS. 1988. Topography of cognition: parallel distributed networks in primate association cortex. Annu Rev Neurosci 11:137–156.

Kling A, Steklis HD. 1976. A neural substrate for affiliative behavior in nonhuman primates. Brain Behav Evol 13:216–238.

Kosslyn S. 1994. Image and brain. Cambridge, MA: M.I.T. Press.

Kraemer GW. 1992. A psychobiological theory of attachment. Behav Brain Sci 15:493–541.

Lhermitte F. 1983. Utilization behaviour and its relation to lesions of the frontal lobe. Brain 106:237–255.

McIntosh AR, Grady CL, Ungeleider LG, Haxby JV, Rapoport SI, Horwitz B. 1994. Network analysis of cortical visual pathways mapped by PET. J Neurosci 14:655–666.

Mega MS Cummings JL. 1994. Frontal-subcortical circuits and neuropsychiatric disorders. J Neuropsychiatr Clin Neurosci 6: 358–370.

Mesulam MM. 1990. Large-scale neurocognitive networks and distributed processing for attention, language and memory. 28:597–613.

Minsky M. 1967. Computation: finite and infinite machines. Cambridge, MA: M.I.T. Press.

Pandya DN, Yeterian EH. 1990. Prefrontal cortex in relation to other cortical areas in rhesus monkey: architecture and connections. Prog Brain Res 85:63–94.

Petersen SE, Fox PT, Posner MI, Minton M, Raichle ME. 1988. Positron emission tomographic studies of the cortical anatomy of single word processing. Nature 331:585–589.

Petrides M. 1994. Frontal lobes and working memory: evidence from investigations of the effects of cortical excisions in nonhuman primates. p 59–82. In: Boller F, Grafman J, editors. Handbook of neuropsychology, vol 9. New York: Elsevier Science Publishers B.V.

Shallice, T. 1988. From neuropsychology to mental structure. Cambridge, England: Cambridge University Press.

Shallice T, Burgess PW, Schon F, Baxter D. 1989. The origins of utilization behaviour. Brain 112:1587–1598.

Wernicke C. 1874. Der aphasische symptomencomplex. Breslau: Frank and Weigert.

Young MP. 1993. The organization of neural systems in the primate cerebral cortex. Proce R Soc Lond B252:13–18.