

A computational model for the primate neocortex based on its functional architecture

Alan H. Bond*

California Institute of Technology, Mailstop 256-80, 1201, California Boulevard, Pasadena, CA 91125, USA

Received 20 March 2001; received in revised form 25 June 2002; accepted 22 October 2003

Abstract

Experimental evidence has shown that the primate neocortex consists in the main of a set of cortical regions which form a perception hierarchy, an action hierarchy and connections between them.

By using a computer science analysis, we develop a computational architecture for the brain in which each cortical region is represented by a computational module with processing and storage abilities. Modules are interconnected according to the connectivity of the corresponding cortical regions.

We develop computational principles for designing such a hierarchical and parallel computing system. We demonstrate this approach by proposing a causal functioning model of the brain. We report on results obtained with an implementation of this model.

We conclude with a brief discussion of some consequences and predictions of our work.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Brain architecture; Perception-action hierarchy; Computational model; Logic programming; Primate social behavior

1. Introduction

1.1. Motivation

In our companion paper (Bond, 2004), we reviewed experimental data on neuroanatomical connectivity and neurophysiological activity of the neurons comprising the primate neocortex. The data were based largely on the macaque monkey, with additional data from humans. There was sound evidence for the widely held belief that the neocortex is made up of discrete cortical regions with specialized functional involvements. Our information-processing analysis of these findings concluded that each region processes certain types of data specific to that region. We also introduced information-processing concepts of goal, plan, sequence, event, and context as data types processed by certain regions. Furthermore, from an analysis of connectivity, we concluded that these regions are connected together in a particular architectural scheme, namely a perception-

action hierarchy. Our paper described these cortical regions, the types of data processed by each region, and the connections among regions.

What it did not do was explain how such a set of cortical regions provides the neural basis for complex organized primate behavior. Our current paper provides this explanation, by presenting a system-level theory of brain function, using as a basis the conclusions of our companion paper.

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.

1.2. System models

A *system model* treats 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 interactions among subsystems.

The use of a systems level of thinking in neuroscience, where more than one neural area are conceived as

*Tel.: +1-626-395-6430; fax: +1-626-792-4257.

E-mail address: bond@cs.caltech.edu (A.H. Bond).

URL: <http://www.cs.caltech.edu/~bond>.

working together, has a venerable history going back to Wernicke's (1874) book. There is, for neuroscientists, a natural systems level of explanation of experimental data, see for example Gazzaniga (1989). Current imaging evidence is showing coactivation of distributed areas in many tasks. From this type of evidence, McIntosh et al. (1994) have developed influence graphs, which give a measure of dynamic influence among neural areas. Mesulam (1990) has developed ideas on a distributed system mediating attention. Kosslyn (1994) has developed a system model of visual imagery and perception. Goldman-Rakic (1988) has investigated distributed systems for working memory, which involve areas in the frontal lobes and in the parietal and temporal lobes. Modular explanations of language processing have progressed (Geschwind, 1965) (Deacon, 1989) and now have support from imaging experiments. Petersen et al. (1988), for example, have produced a modular description of language processing. In this paper, we will model the primate neocortex as a system.

1.3. Natural science and causal models

From the hierarchy of functional involvement alone, we cannot construct a model of brain functioning. The experimental results demonstrate the involvement of some parts of the brain in some given behavior, but they do not demonstrate a causal functioning model of the brain actually operating to produce the behavior. By *causal* we mean that the model has a dynamics of changing in time from one state to another, each next state being determined from the current state. Very little experimental information is available on how different parts of the brain work together, what information flows, what is computed, or how the activities of different parts are coordinated. Most of natural science to date has been concerned with matter and energy and transformations among its various forms. What we will need for the brain is a natural science of information processing.

1.4. Computer science

In order to produce a working causal model of information processing in the primate neocortex, we shall turn to a computer science analysis, where we shall draw on knowledge of information-processing systems from several different specializations within the field of computer science—parallel architectures, distributed systems, formal description languages, and artificial intelligence planning.

Neuroscientific information-processing requirements can be used to constrain the design of a scientific model. We know that the brain does indeed work in a coordinated manner to produce behavior, we know that it is stable under a number of disturbances, we think it is

organized for real-time responsiveness, we believe it is distributed, and we have some idea of timings of components.

Computer science brings to our research notions of processing architecture, of data and process representation, and of control. A good standard treatment of system level abstraction and layers of description can be found in (Siewiorek et al., 1982). Further, computer science brings techniques for describing, specifying and implementing models. Description languages have been developed for the high-level description of complete computer systems. Standard treatments can be found for example for requirements in Davis (1993) and for real-time embedded system methods in Calvez (1993). In our case, it will be of particular value for understanding how control in the brain could be organized, and for understanding how to create a model by specifying and implementing one using logic programming.

We will determine computational principles by which such a hierarchical system of information processing regions can function to produce behavior.

1.5. The perception and action hierarchies of the primate neocortex

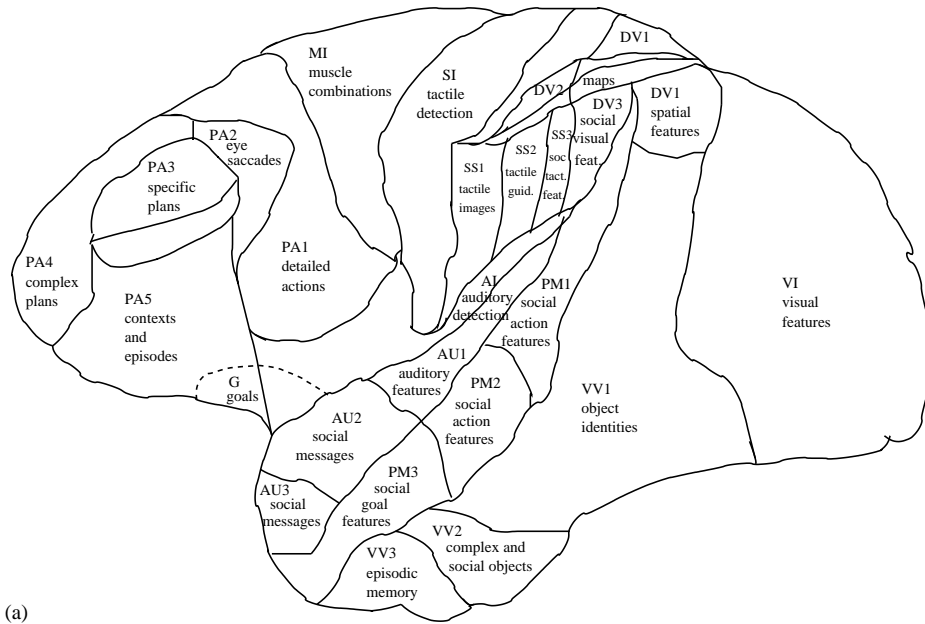
As a basis for our system model, we now summarize the findings of our companion paper (Bond, 2004), showing a hierarchy of function and data types in the cortex. We work with *regions* made up of several neural areas. We list the neural areas comprising each region and summarize their functional involvements in Fig. 1. We show the regions on a lateral view of the cortex in Fig. 2(a) together with an indication of their functional involvements. The region G is shown with a dotted line boundary to indicate that it is interior, being on the medial surface. In Fig. 2(b), we give the connectivity of the set of cortical regions. The hierarchy is diagrammed in (b) on its side with its top at the left, in order to make it correspond to the usual lateral view of the cortex in (a). The positioning of a perception region on a vertical line indicates connection to the corresponding action region.

1.6. Our model

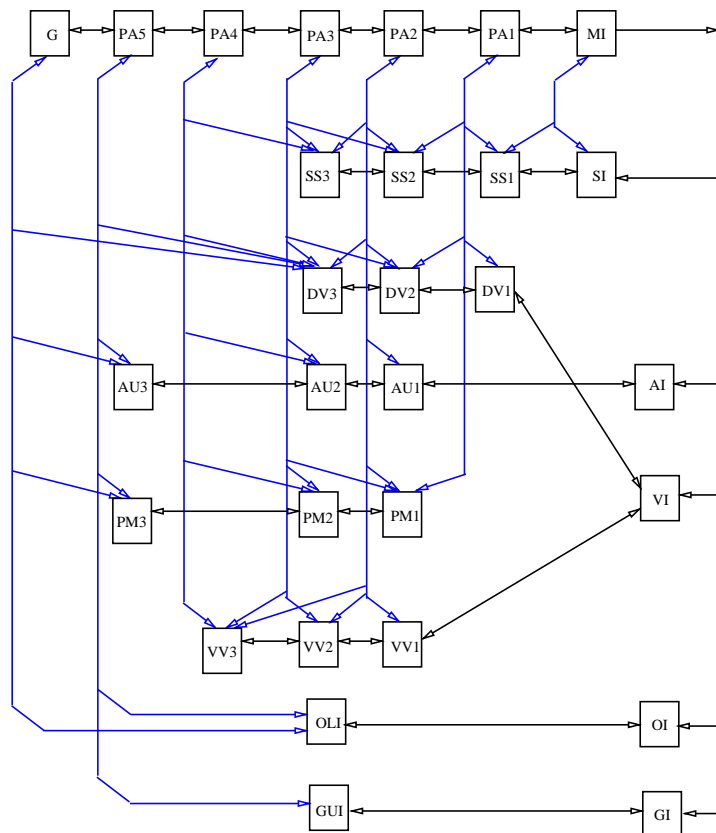
Our system model is diagrammed in Fig. 3 showing the set of implemented modules with approximately corresponding cortical locations. This approximate correspondence locates the perception hierarchy along the superior temporal sulcus (STS) following Perrett's findings, and with episodic memory for social relations in the anterior temporal lobe. Goals are 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

hierarchy	region	corresponding areas	functional involvements	reference
somatosensory	SI	1,2,3	tactile detection	[Kaas and Huertas, 1988]
	SS1	PE,PEa,PF	tactile images	[Merzenich et al., 1981]
	SS2	PEc,rostral POa,PFPG	guidance of reaching and grasping	[Mountcastle, 1995]
	SS3	PGm,rostral PG	guidance of reaching and grasping	[Mountcastle, 1995]
	PM3	TPO-1,TPPro	socially significant tactile recognition	[Perrett et al., 1989]
auditory	AI	auditory cortex	auditory detection images	[Brugge and Reale, 1985]
	AU1	CB,RB,CPB (TS3)	auditory images, maps	[Juhani Hyvärinen, 1982]
	AU2	CSTG,CSTS,RPB (TS2)	socially significant auditory recognition	[Newman, 1978]
	AU3	RSTG and RSTS (TS1),CML,23b	socially significant auditory recognition	[Newman, 1978]
ventral visual	VI	V1, V2, V3, VP, V3A, V4t	visual features and images	[Essen et al., 1990]
	VV1	TE3, V4	object identity and motion	[Logothetis and Sheinberg, 1996]
	VV2	TE2	complex objects and long term memory	[Miyashita, 1993]
	VV3	TE1	complex objects and long term memory	[Miyashita, 1993]
dorsal visual	VI	V1, V2, V3, VP, V3A, V4t	visual features and images	[Essen et al., 1990]
	DV1	PIP,PO,MT	spatial features	[Essen et al., 1990]
	DV2	MIP, VIP, LIP,	eye saccades	[Andersen, 1995]
	DV2	MIP, VIP, LIP, DP, MDP, MST, FST	guidance of reaching and grasping	[Mountcastle, 1995]
	DV3	caudal 7a	spatial maps and spatial perception	[Andersen, 1995]
	PM3	TPO-1,TPPro	socially significant perception and guidance	[Perrett et al., 1989]
polymodal STS	PM1, PM2, PM3	TPO-4,TPO-3,TPO-2,TPO-1,TPPro	socially significant viewer-centered perception	[Perrett et al., 1990]
	PM3	TPO-1,TPPro	socially significant goal-centered perception	[Perrett et al., 1989]
	PM3	TPO-1,TPPro	episodic memory	[Perrett et al., 1989]
planning and action	G	24,25,32	goals and action selection	[Devinsky et al., 1995]
	PA5	11,12,13,14,FPPro	context and episode representation	[Petrides, 1994]
	PA4	9,10	complex plans, self-paced	[Petrides, 1994]
	PA3	46	specific plans with data in working memory	[Petrides, 1994]
	PA2, PA1	8,6	explicit detailed realtime action sequences	[Passingham, 1993]
	MI	4	muscle combinations	[Passingham, 1993]

Fig. 1. Summary of experimental findings for hierarchy of data abstraction.



(a)



(b)

Fig. 2. (a) Lateral view of the cortex showing regions and functional involvements, (b) connectivity of regions showing perception-action hierarchy.

social-spacing behaviors. This also used a simple low-level spatial navigation module which could be tentatively identified with PA1.

The model functions by continuously generating and selecting a goal, and elaborating and executing a corresponding plan via its action hierarchy, while

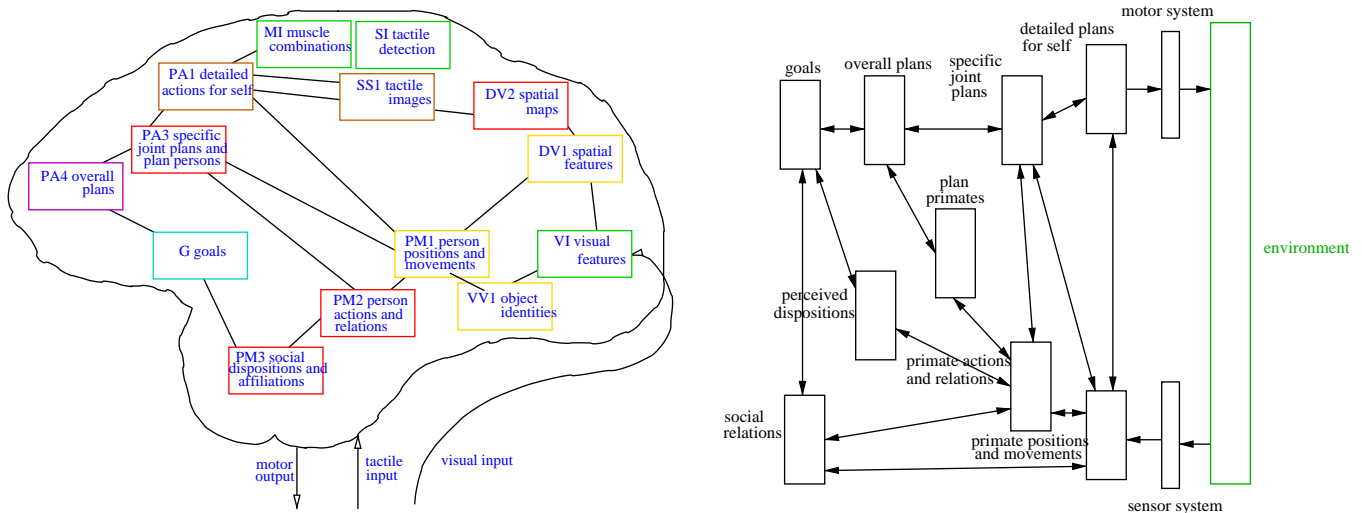


Fig. 3. Modules from neural areas of the primate neocortex, and our initial system model.

perceiving the world using its perception hierarchy, with continuous interactions between these hierarchies.

In summary, we can create a causal model of the brain at the system level if we model each cortical region by a continuously acting process which constructs, stores, and transmits data of the types specific to that region. Processes are connected by channels whose connectivity corresponds to cortico-cortical connectivity. This results in a system model whose dynamics include feedback, goal-direction, conditional plan elaboration, attention and situated action.

1.7. Predictions

From our model, we can obtain detailed predictions of temporal sequences of spatial distributions of cortical activation, for behaviors represented using the model. These predictions have a detailed time granularity of about 20 ms, and could be compared with fMRI or EEG data. 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.

1.8. Social interaction

The other main new advance is that our model shows how a perception-action model can result in a model of social interaction. This occurs because, in a situation with more than one animal, each animal continuously perceives the other, and continuously acts toward the other conditionally upon what it perceives. Further, the hierarchical organization of the perception-action system allows a hierarchical description of the social interaction with different levels of control and protocol. The model provides correspondences between measure-

ments of social interaction and the set of cortical regions and their activation patterns.

1.9. This paper

In Section 2, we derive computational principles from the biology of the cortex, and in Section 3 we describe how, guided by these principles, we can represent cortical data and processes using predicate logic. In Section 4, we describe the dynamics of our model. Section 5 gives a detailed description of a specific brain model which we have implemented on a computer. In Section 6, we report behaviors and results obtained with our implemented model. This model, and its implementation, therefore establish the correctness and feasibility of the approach. They exhibit an actual functioning brain model based on the available empirical evidence. Section 7 concludes with a brief discussion of some consequences and predictions of our work.

2. The biological basis of our computational approach

In this section, we examine what we know about the primate cortex, and we develop the basic computational elements upon which to design a system-level brain model.

2.1. Areas

The primate cortex is partitioned into distinct areas. We therefore structure our computational model into a set of corresponding modules.

2.2. Areas have specific interconnectivity

The connectivity among areas is the same, or similar, for all primates. We will connect our modules in the

same way. Areas are typically connected to a small number of other areas. Connections divide into long range and medium range. At medium range, an area is often connected to several neighboring areas that are contiguous with it. At long range, an area is usually connected to one, two or three areas that are further away, and not contiguous with it.

2.3. *Each area is involved in specific kinds of processing*

We will assume that each module processes only certain kinds of data, specific to that module.

2.4. *Processing is distributed*

Areas process data received and/or stored locally by them. There is no central manager or controller. This is a debatable issue. In our view, areas influence each other by data sent between modules. The set of modules works together in an integrated way, but by means of local processing and the exchange of data.

2.5. *There is a uniform process*

The primate cortex has a uniform structure over all of its area (Creutzfeldt, 1978; Ullman, 1991), having a six-layer organization comprising neurons from a small set of anatomical types. The numbers of these cells per unit volume are very uniform over the cortical surface, the main differences being in motor cortex which has more and larger pyramidal cells, and in visual cortex, which has a significantly, three times, greater density of cells. A canonical neocortical circuit can be described (Shepherd and Koch, 1998; Douglas and Martin, 1998) and regional variations from the canonical form characterized. Although long-range connectivity, as we have seen, tends to be clustered around cortical regions, short range (<3 mm) connectivity within the cortex is statistically quite uniform. It therefore appears that information processing within different cortical regions has a common basis or principle.

2.6. *Cortical processing proceeds at a uniform rate*

All modules do similar amounts of processing and run at about the same speed.

2.7. *Data parallelism in communication, storage and processing*

We assume that data are coded in parallel codes, such as population codes, so that a large set of parallel fibers carries a code for one message or one meaning. We assume that processing within a module is also highly parallel, operating on a large set of parallel fibers concurrently. Parallel coded data are transmitted,

stored, and trigger processing. Processing acts on parallel data to produce parallel data.

2.8. *The cortex works in real time*

The cortex's fastest reaction time to a stimulus is about 100 ms. The time to process information in one area and to pass it on to the next area is about 20 ms (Edmund Rolls, pers. comm.). Further, the path from incoming sensory stimulus to outgoing motor command runs through about five areas. Language is processed in real-time, both generation and recognition, and, as Goodwin has shown (Goodwin, 1981), even the co-construction of a sentence usually occurs in real-time including nonverbal signalling between participants during the generation of the sentence. Hence the action of an area is, at least some of the time, an immediate reaction to its incoming data. Cognition occurs by modules exchanging data and by repeatedly reacting to incoming data and newly computed data. The incoming data arrive at an area, a process occurs in about 20 ms, and output data are transmitted to other modules.

2.9. *Data is "wide"*

The data items being transmitted, stored and processed can involve a lot of information; they can be complex. Thus, if we have a parallel set of 1 million neurons, then the code for one choice or component of a data item might involve 10 000 neurons, and then the set of neurons might transmit 100 such components or choices simultaneously as one data item. Scientists describing this situation to each other use natural language which is less "wide", so we tend to unconsciously assume that a data item in the brain is of similar complexity to a natural language word or phrase. However, a single data item, we suggest, may convey as much information as a whole paragraph.

3. **Representing data and processes using logic**

3.1. *Logical modeling*

Our intention is to develop a very general and abstract kind of model that can be changed and specialized in the light of results obtained with it. For this purpose, we use predicate logic expressions to represent data and predicate logic inference rules to represent transformations of data. Given an abstract model, we will then later be able to consider more specific implementations of it, in particular, how it can be implemented as a neural net. However, the abstract model is self-contained, it can be run on a computer and its behavior found, and it can generate falsifiable predictions that can be tested against experiment.

3.2. Data items, and their storage and transmission

We will assume that we can view transmission and storage of data in the brain as codes which represent some information with a specific meaning. We therefore assume that, for the purposes of modeling at this level of analysis and abstraction, we can view all data streams and storage as made up of discrete data items. We will represent each data item by a logical literal which indicates the meaning of the information contained in the data item. In order to allow for ramping up and attenuation effects, we give every literal an associated weight, or strength, which is a real number. An example data item is `position(0.8, [adam, 300, 200, 0])` which might mean that the perceived position of a given other animal, identified by the name “adam”, is given by (x, y, z) coordinates (300, 200, 0). This might be a data item that is transmitted from one brain module to another. In the brain this would actually be implemented by a set of parallel neurons firing in a spatial pattern at certain firing rates. Its effect however is that the receiving module now has the information about the position of adam.

3.3. Memory

Each cortical region will be represented by a continuously acting module which is a process with storage. 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. Every module may in general have stored data items. Depending on the time 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; compare (Petrides, 1994).

3.4. Processing within a module

We represent the processing within a module by a set of rules. A rule matches to incoming transmitted data items and to locally stored data items. All the processing by a module is described by a set of left-to-right rules which are executed in parallel.

The patterns on the left-hand side of rules also have weights, and the strength of a rule instance is the product of the matching data item weights and the rule weights, multiplied by an overall rule weight.

A rule may do some computation which we represent by arithmetic. This should not be more complex than can be expected of a neural net. This arithmetic is

represented in the body part of the rule, written as a “provided” expression, for example:

```
if position(W1, [M1, X1, Y1, Z1]), position(W2, [M2, X2, Y2, Z2]), then too_near(W, [M1, M2, D]), provided(distance(X1, Y1, Z1, X2, Y2, Z2, D), D < 10.0).
```

This rule is intended to determine whether one animal is too near another.

The results are then filtered competitively. Typically, only the one strongest rule instance is allowed to “express itself”, by sending its constructed data items to other modules and/or to be stored locally. In some cases however all the computed data are allowed through.

3.5. Uniform process

The uniform process is then the mechanism for storage and transmission of data and the mechanism for execution of rules.

3.6. Uniform rate

We achieve uniformity of rate by describing time by a discrete time scale; the model runs in discrete time cycles. In one processing cycle, all the rules in all the modules are executed once, that is, all rule instances, and then all selected data are communicated between modules and/or stored locally. The events of a processing cycle represent and abstract all the changes occurring during that time interval.

3.7. Perception-action hierarchy

Modules are organized as a *perception-action hierarchy*. We diagram our concept in Fig. 5.

This is an abstraction hierarchy, so that modules higher in the hierarchy process data of more abstract data types. We use a fixed number of levels of abstraction.

The perception hierarchy receives sensory data items at the bottom and derives higher level descriptions to form a percept. The action hierarchy generates more and more detailed descriptions of action, that is, it elaborates the plan to the point where motor actions are generated at the bottom of the action hierarchy.

An example of a perceptual rule is:

```
if position(M, X, Y, Z) and orientation(M, A) and self_position(X1, Y1, Z1) then oriented_towards(M), provided(angle_towards(X, Y, Z, X1, Y1, Z1, A1) and app_equal(A, A1)),
```

i.e., from data giving another primate’s position and orientation, and from the subject’s own position, calculate the angle from the primate to the subject and test if that angle is the same as the primate’s orientation, if so, create a new datum representing the fact that the other primate is oriented towards the subject.

An example of an action elaboration rule is:

```
if plan_self_action(walk_towards(M)) and position(M,X,Y,Z) then plan_self_act(walk_towards(X,Y,Z)),
```

i.e., if the planned action for the self in terms of relations is to walk towards some primate, and if this primate's position is X,Y,Z then generate a new datum, which represents planned action for the self in terms of detailed position, to walk towards the position X,Y,Z.

The goal module has rules causing it to prioritize the set of goals that it has received, and to select the strongest one, which is then sent to the highest level plan module.

We also use a long-term memory module, which perceives social action and maintains the memory of affiliative relations. It generates goals to affiliate and sends them to the goal module.

3.8. The external world

Primates operate in an *external environment* which is a 3D spatial world. The environment is everything external to the brain, so it includes the body. A primate has *sensors* which interrogate the environment and generate sensed feature descriptions which are represented as literals. These input data items are sent to specified modules each cycle. Some modules act as *effectors* in that they send motor commands, represented as literals, to the environment. The environment receives motor commands from all the primates and computes what changes to make. Clearly, primates can only communicate with each other via the environment.

3.9. Symbols

We hasten to point out that the use of a discrete representation for specifying data and processing does not imply that we are assuming that the brain processes symbols. First, any processing mechanism will be expressed in some language, which is necessarily a discrete representation, and will use mathematical symbols. For example, a neural net is usually expressed in mathematical symbols or in a programming language such as C. Another example is neural transmission mechanisms, such as Hodgkin–Huxley equations which are represented as mathematical symbols. Second, a symbol used in cognition is a particular type of data whose representation would have to be defined in the model and whose behavior and properties would be given by sets of rules. One can explicitly postulate that the brain uses symbols (Newell, 1990). There is quite a lot of psychological work on clarifying if and when the brain uses symbols, and if so how to model it, see Holyoak and Hummel (2000) and Pylyshyn (1984).

4. Dynamics of our model

4.1. Perception-action hierarchy

Figs. 4 and 5 show 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.

4.2. Conditional elaboration—situation

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 which describe action at one level and generating data which 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 (5) who will do which actions. We diagram an example of this in Fig. 6(a). By an *evaluation* we mean, for example, a value indicating progress, success or failure; such a value can also be associated with a particular datum, for example, one representing an action or goal.

4.3. Conditional perception—attention

The perception hierarchy and action hierarchy cooperate 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. We diagram an example of this in Fig. 6(b). 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.

4.4. Continuous action

Action is continuous with a small time granularity. The primate brain runs at about 20 ms, and our implementation runs at about 100 ms on a 300 MHz processor. Thus, stored data are updated every cycle, the selection of rule instances is updated every cycle, and updated motor commands are output to the environment every cycle. The process of goal generation, goal selection, plan selection, plan elaboration, action

Level	Perception	Action	Types of Information	Example
Level 5: goals and context	perception of goals perception of context	prioritization and selection of goals maintenance of current context	desired states described abstractly, priorities and urgencies of such states. contexts - classes of events and episodes, themes, general plans applying to classes of episode. objects, actions and relations described generally.	goal - affiliate with X. current resting situation, family foraging summer afternoon, X is aunt
Level 4: joint plans plans in social relationship	perception of social features, situations social actions and intentions	generation of social plans	plans for a well-defined situation class, objects, actions, relations corresponding, involving others.	groom with X
Level 3: joint plan in relational form	perception of features features that indicate spatial relations actions and intentions	construction, execution and monitoring of explicit joint plan in relational form	joint plans with assigned roles and including defined actions specified in terms of relations	groomee is X, groomer is self, (approach, prelude, groom)
Level 2: self action in detail	perception of position, orientation movement, velocity	construction and execution of detailed plan for self	concrete actions for the self, including detailed spatial and temporal characteristics. detailed motor programs - to allow realtime performance of the actions without immediate feedback	approach to X at position (300,360,0) get up, turn and walk
Level 1: motor actions	perception of somatosensory, tactile, features for muscle selection and	activation of muscle groups	individual actions by sets of muscle groups.	front right leg(), front left leg()

Fig. 4. Computational hierarchical levels used in our model.

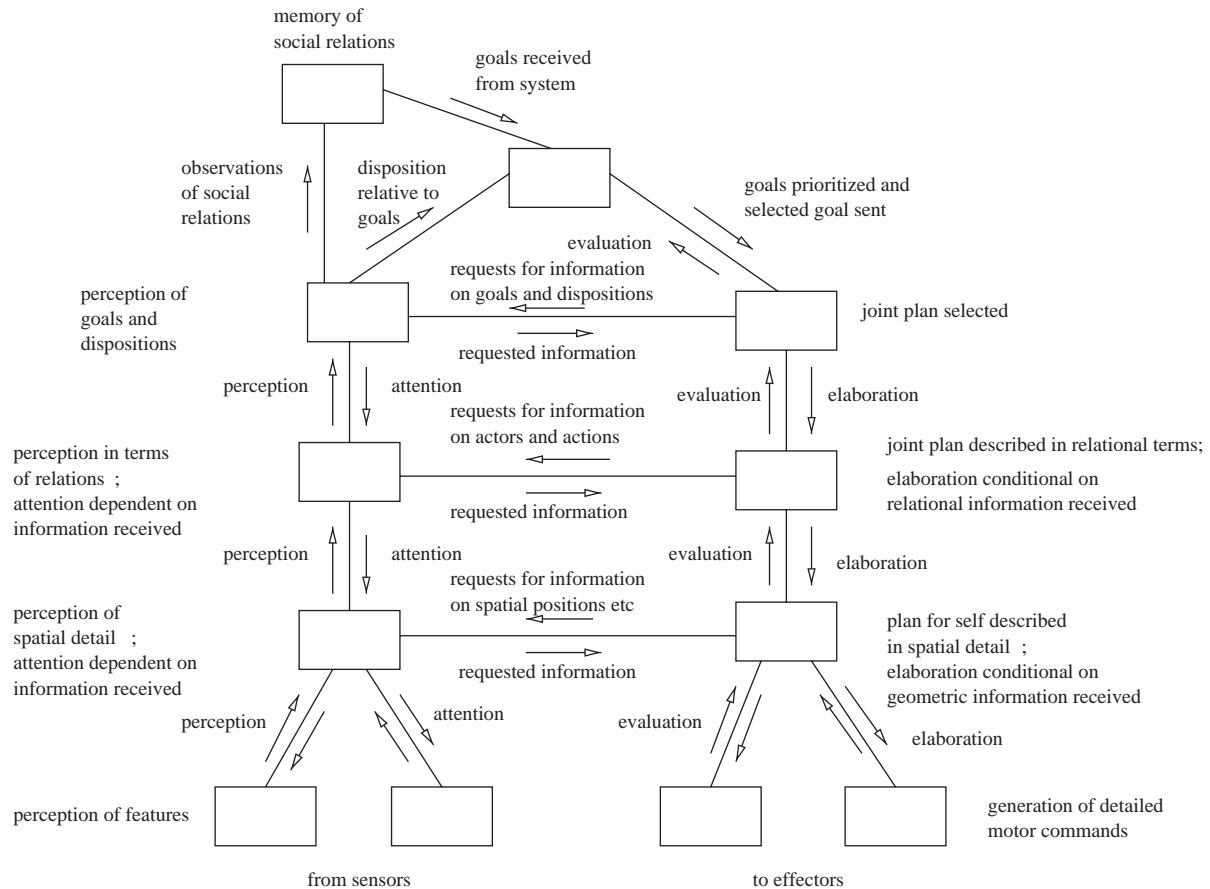


Fig. 5. Functioning of interacting perception and action hierarchies in behavior.

specification and motion specification proceeds continuously, renewing the information every cycle.

4.5. The stability of distributed activation

Each module selects a dominant rule which outputs data to other modules. However, this can lead to incoherence, modules can get into states with crossed purposes, and attempts to elaborate plans tend to collapse under challenge. We developed a simple, biologically plausible mechanism which stabilizes distributed activity. If a module receives a data item that causes successful activity, it sends a *confirmation* message back to the sender, evaluating that data item. Successful activity is defined as any rule firing, not necessarily a selected one. We diagram an example of this in Fig. 6(c). The confirmation message is specific to and contains the particular data item sent. When the sending module receives a confirmation message it boosts the level of the rules generating that data item. This therefore consolidates the strength of the selected rules. Further, if a selected rule does not receive confirmation messages, its strength will attenuate, thereby allowing competing choices to be tried.

4.6. Viable behavioral states

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 motivations. It does this by trying different alternatives at each level on a competitive basis, and subject to confirmation of successful elaboration. A state of the module in which such a plan is established, and is executing consistently with perception and confirmed elaboration, can be called a *viable state*.

4.7. Different levels of control

Provided the internally generated goals and the external environment do not change radically, 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

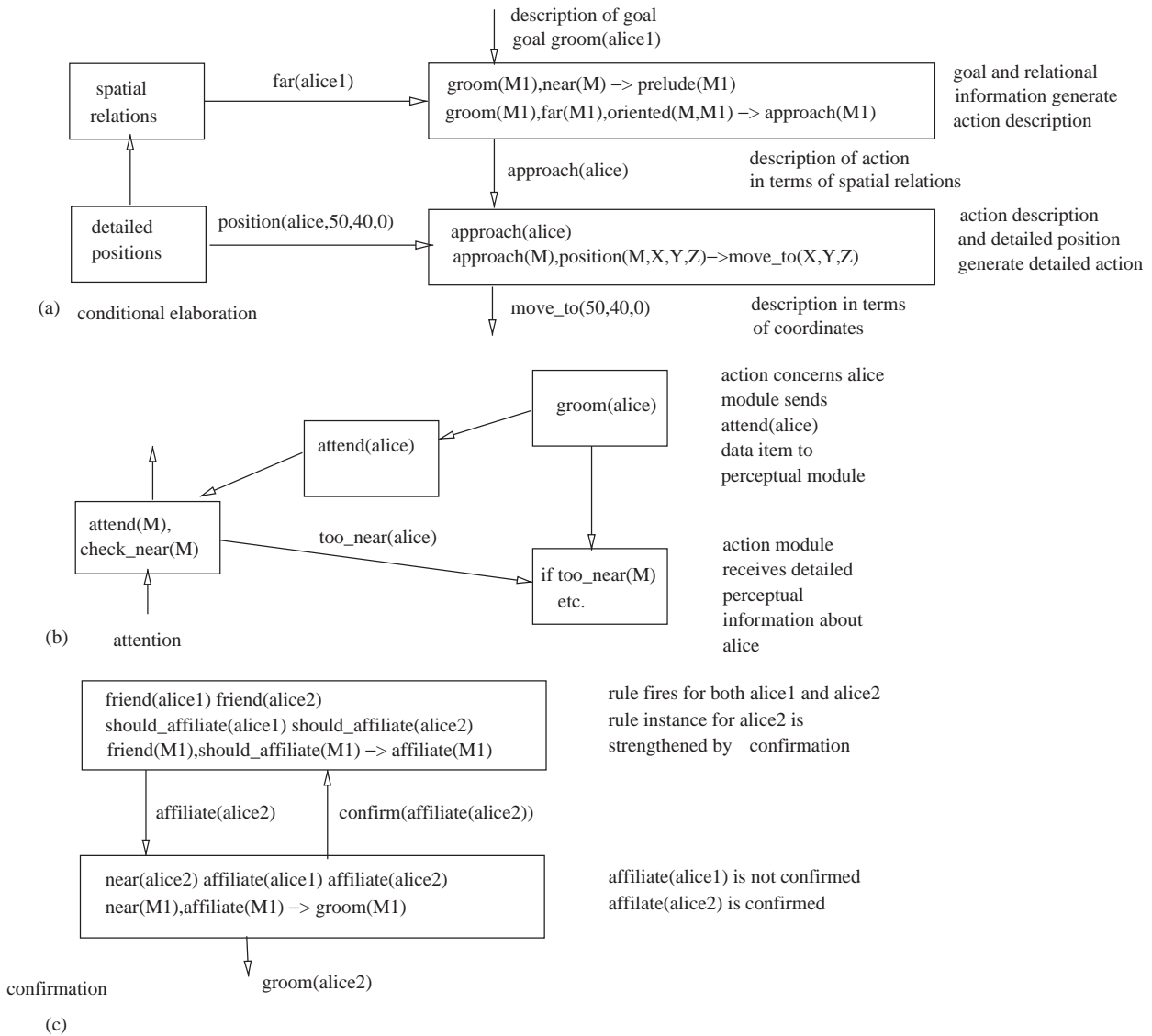


Fig. 6. How the model works.

perceived at level 3. 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 level 4.

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. This is depicted, using a cortical correspondence diagram, in Fig. 7.

4.8. Joint action

We developed a notion of plan suitable for social action. A *joint plan* is a set of joint steps, with temporal and causal ordering constraints, each step specifying an

action for every primate collaborating in the joint plan, including the subject primate. 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 to attempt to execute the corresponding individual action for the subject primate. We made most of the levels of the planning hierarchy work with joint plans, the next to lowest works with a “selfplan” which specifies action only for the subject primate, and the lowest with concrete motor actions. However, the action of these two lowest levels still depended on information received from the perception hierarchy.

4.9. Imagining brain activity

The modules operate concurrently, that is they all operate at the same time, in *parallel*. We can perhaps

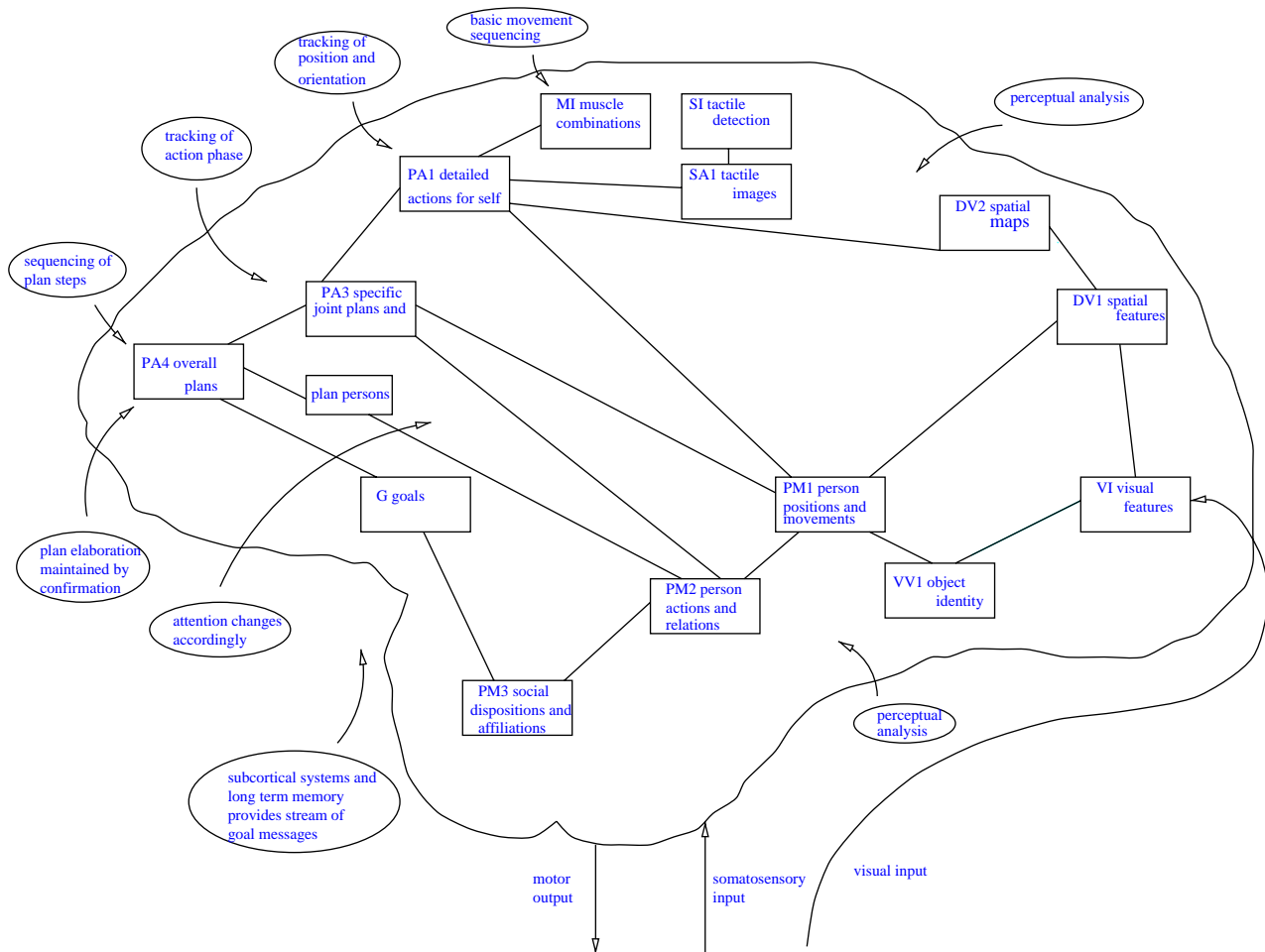


Fig. 7. Response to variation in environment.

imagine a cortical surface covered with an array of cortical regions like a patchwork quilt, each region lighting up by a different amount during each cycle. Each region also stores and processes different information in each time interval, so there are different sets of expressions in each module which are constantly changing, and expressions flowing between modules.

The usual image of information activity in the brain is that of information flowing through a set of cortical areas, forming a pathway. We conceive of data flowing through the brain as passing sequentially through a series of brain regions. During each cycle, incoming information may be used to generate new information and/or to store information. In general, the same information is not passed on; instead, new information, derived from all the inputs received at that moment and any information already in storage at that moment, is transmitted. Because of the high processing speed of modules, a high rate of data transmission is maintained through the brain.

To understand how the brain produces behavior, we need a more general concept of computation that allows information transformation activity, combination of

information, storage and retrieval of information, and activity conditional upon properties of the information. Our “array of regions” image allows us to think of each neural region as responding conditionally to information, as having time to compute new information and to store information, and as sending different information in different directions to different other areas, including sending information back to areas “upstream”.

5. Our implemented brain model

5.1. The choice of behaviors and external environment

In order to precisely define a brain model, we needed to decide what behaviors to consider and what external environment the brain would have. We chose to consider the case of social affiliation. We used a “minisociety”, in which a group of primates (monkeys) interact socially, in a naturalistic 3D 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. We motivated the system by defining long-term memory, which stores knowledge of affiliative relations, as, among other things, generating affiliation goals, since affiliative behavior is a known driving force in primate groups (Kling and Steklis, 1976).

Our impulse was to build in social interaction into our brain design from its inception. In the event, this has proved to be a fruitful decision; 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.

5.2. The initial model

We developed an initial brain model consisting of data and process representations, with eight memory modules, shown in Fig. 3. An outline of each of these memories, the descriptions they store and the processes they include, is as follows:

- (i) the *affiliations* module contains all affiliative information, including kinship and dominance relationship information. It generates affiliation

- goals 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.

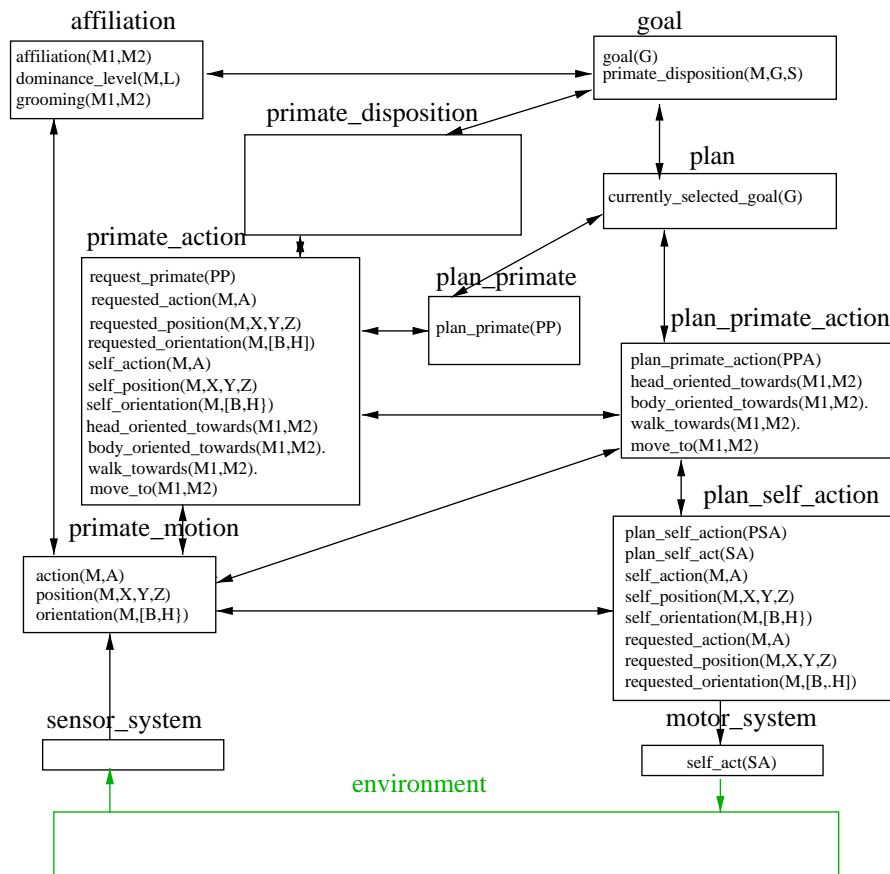


Fig. 8. Description types in each module.

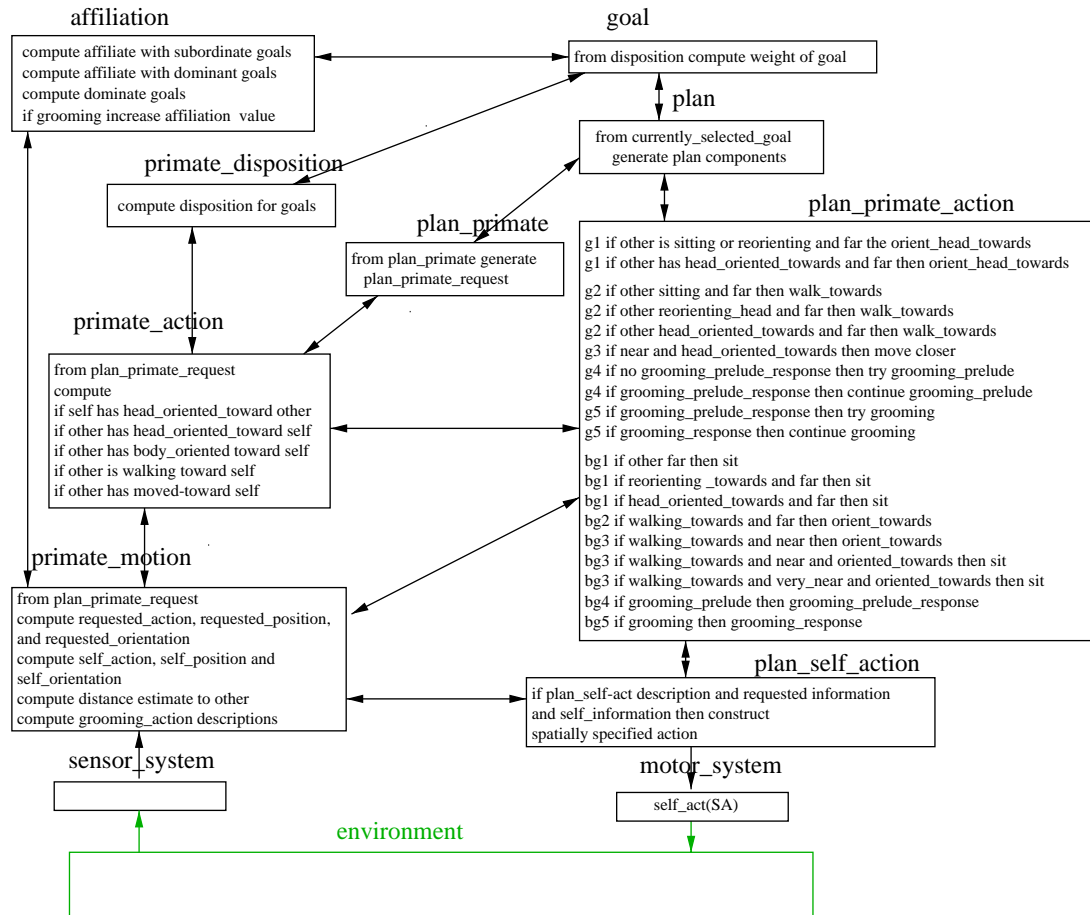


Fig. 9. Outline of description transformations in each module.

- (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 positions 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.

Fig. 8 indicates the data types processed in each module. Fig. 9 outlines the rules operating in each module, for the case of two-primate grooming.

Results were obtained with grooming, social conflict and social spacing behaviors. These simple social behaviors were obtained using about 15 rules per module.

6. Behaviors and results obtained

6.1. Two primate grooming behavior and joint action

We experimented with a prototypical situation in which two primates groom. We developed a four-phase plan for a groomer (orientation, approach, grooming-prelude, then grooming), and a groomee (waiting, orientation, grooming-prelude-response, then grooming-response), and we developed suitable rules for activity in each module in each phase. We ran our computer implementation and the primates did indeed carry out the four phases described, leading to a primate named adam1 grooming a primate named alice1. We show in Fig. 10 images from a visualization generated by our system, showing a frame from each of the four phases of grooming.

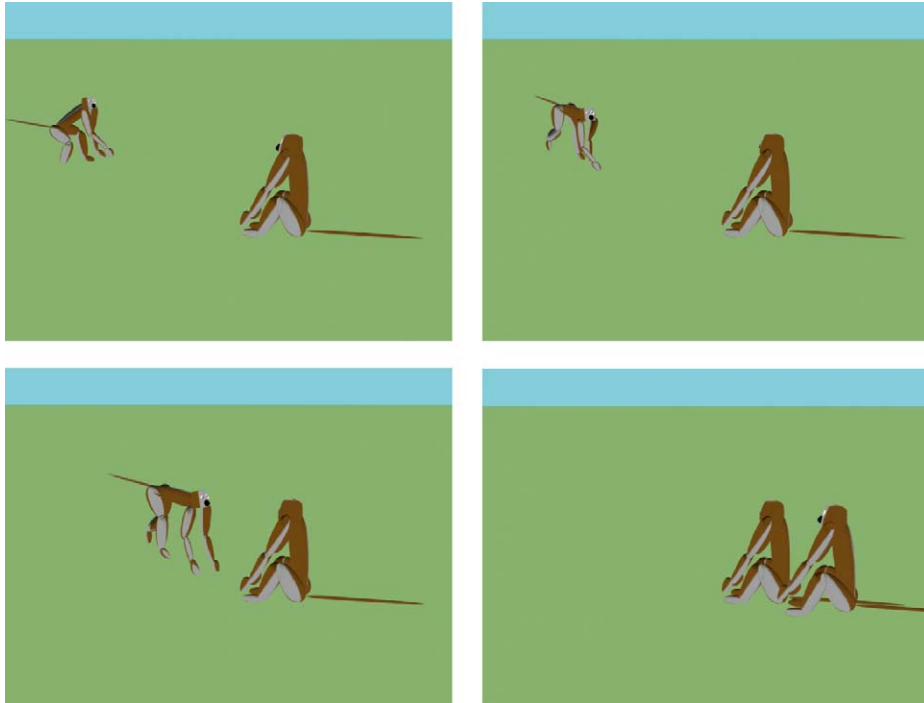


Fig. 10. Grooming sequence.

We show in Fig. 11 an instantaneous state of adam1's brain at a point in time when he is walking toward alicel, as a result of selecting a goal to affiliate with her, and to do this by grooming her. He is perceiving that alicel is in the process of orienting toward him and takes this into account in generating his own action of walking directly toward her. We show the left-hand sides of dominant rules in each box, and the transmitted right-hand sides on the channels.

Fig. 12 shows the states of adam1 and alicel, with complementary perceptions of each other and complementary plans being elaborated.

6.2. Social conflict, change and termination of action

We wanted to investigate failure of action, and dynamic change of selected goals and selected plans, so we developed a more complex scenario involving four primates and social conflict. This involved a situation in which a 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. Behavior was achieved in which conflict occurred and a change of cortical process was needed.

The strongest goals of alicel and adam1 were to groom adam2, but adam2's strongest goal was to groom alicel2. Fig. 13 shows the sequence obtained. Alicel and adam1, on the left and top of the pictures, first oriented to adam2, on the right. When they perceived it orienting

and moving toward alicel2, at the bottom of the picture, their plans failed. The "moment of truth" is captured in Fig. 13 (top right) where adam1, and alicel1, realize from adam2's walking toward alicel2 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 alicel1, and this joint action is able to be completed.

6.3. The direct perception of disposition

We also implemented this social conflict scenario using an additional 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 computed in the new module and from there transmitted to the goals module. It was relatively straightforward to develop rules for perceiving other's dispositions in a limited context. For example, if an animal is moving away, it is probably negatively disposed to grooming. We represented dispositions as evaluations relative to a given goal. So, a disposition data item in alicel's brain might be `person_disposition(adam1,0.9,negative,goal(groom(alice1,adam2)))`. The weights of goals generated in the goal module were made conditional upon this primate disposition feedback, The change of plan was

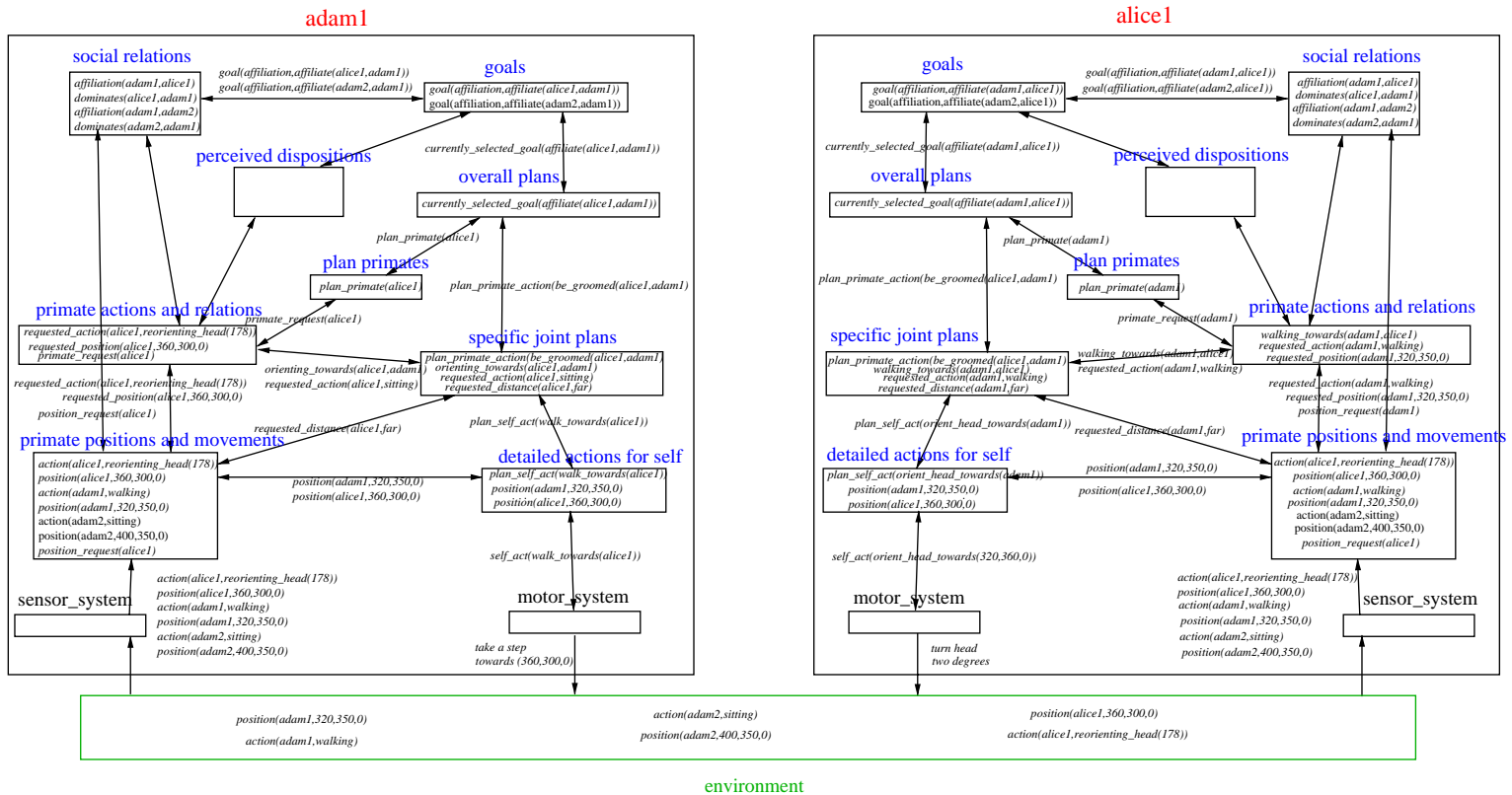


Fig. 12. Instantaneous behavioral states of two interacting primates.

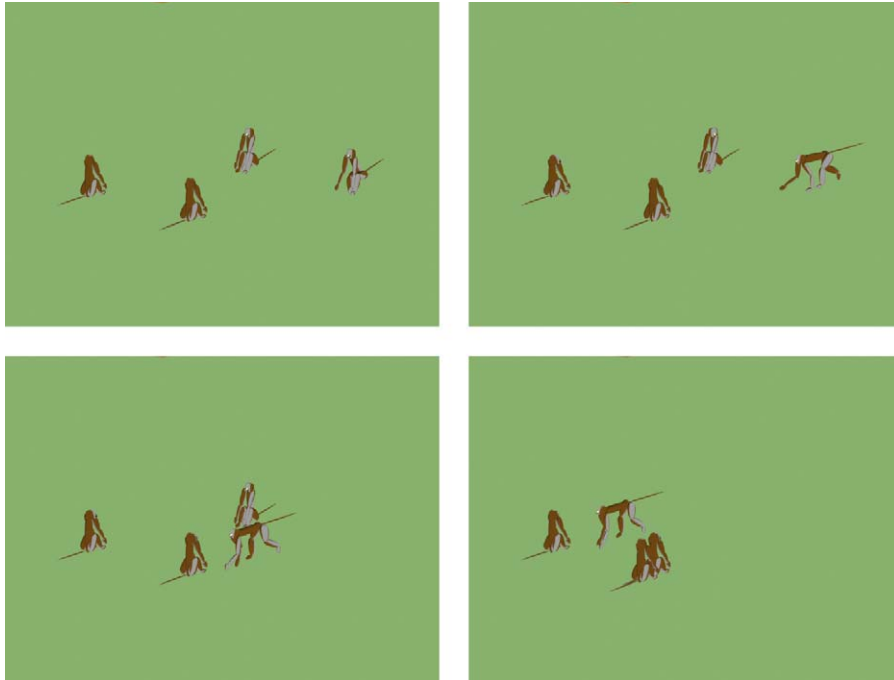


Fig. 13. Social conflict sequence.

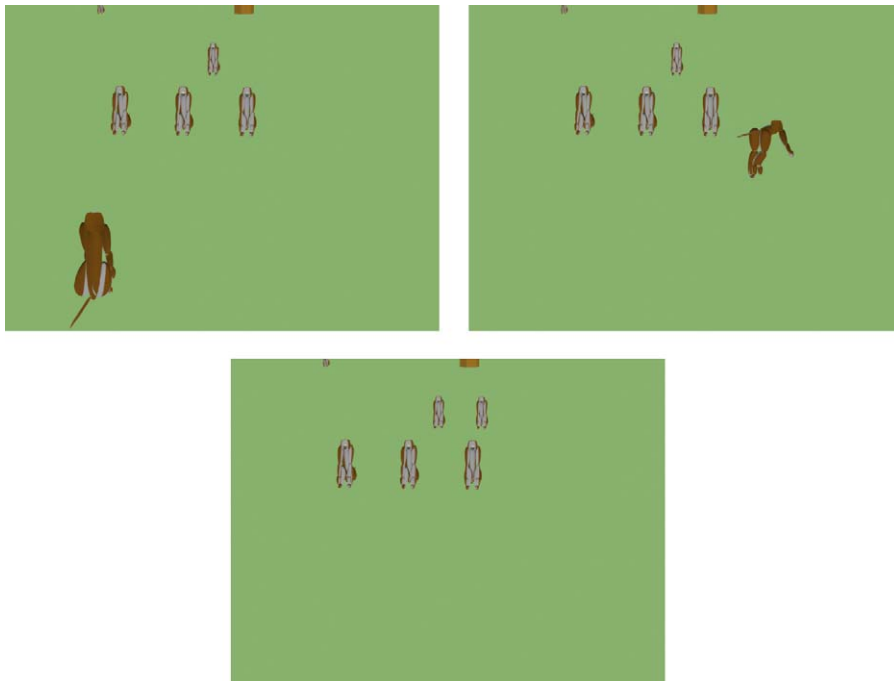


Fig. 14. Avoidance sequence.

or modelling framework is available. Fodor's argument for central processes is that at some stage all of the different input information has to be integrated to produce behavior and that this cannot, for unstated reasons, occur in a modular system. Our model shows how integration can occur through the use of abstracted

types of data and that integration can occur in a modular fashion. Further, the use of higher-level data types allows the use of modules with uniform processing power. This is not only in consonance with the uniformity of the neocortex, it also allows the cortical model to use energy resources optimally by allowing

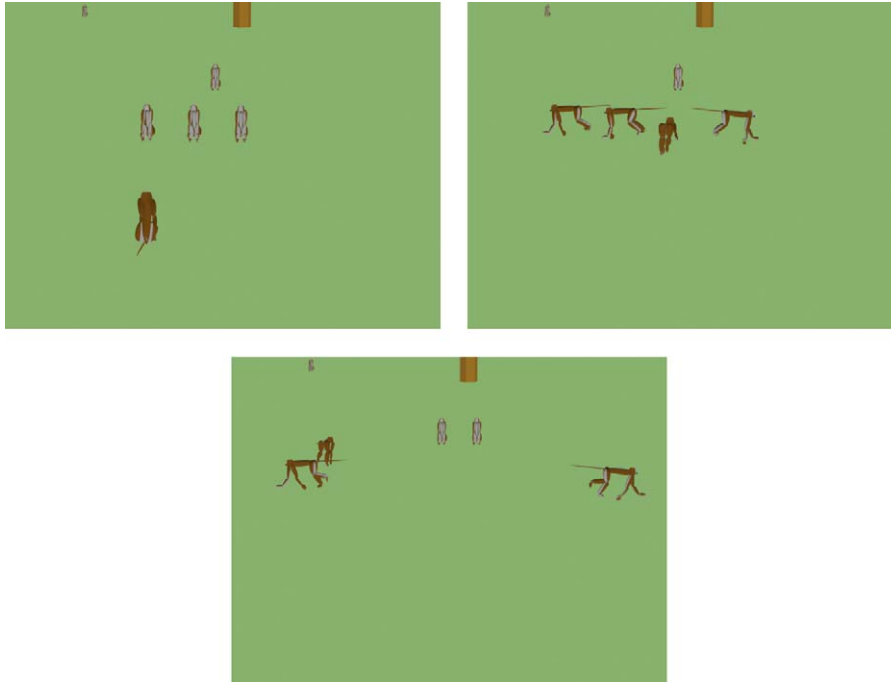


Fig. 15. Displacement sequence.

behavior to be generated by concurrent activity of the set of modules.

7.2. States, causal dynamics and scientific theory

The image of an array of lit up processing modules visualizes the instantaneous state of the system. 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.

7.3. Computational semantics

Although we can use names for descriptions and modules which suggest their significance by association to English words, their precise meaning is given by the action of the model. For example, the semantics or meaning of a goal description in the goal module is neither more nor less than that on being selected it is passed to the plans module and causes a plan to be selected and elaborated. The precise meaning of a module is a continuous process given by its description types and description transformation rules. Thus, planning and action modules are defined as continuously generating elaborations of plan elements they

receive, perception modules are defined as continuously analysing incoming data and derive more abstract data. Goal modules are defined as storing, prioritizing and selecting from goal descriptions they receive.

7.4. Motivation

Kraemer (1992) has reviewed mechanisms of attachment. Innate affiliation schemas may be subcortical and 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 subcortical and cortical representations of specific affiliation relations would generate signals which 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 qualities.

7.5. Properties of our model

Our work shows how to construct a working model of the brain at an architectural level. By using logical abstraction of data types, and processes as sets of parallel rules, a tractable computational model can be created, demonstrated and experimented with.

By representing control explicitly and in a distributed manner, we can see how a hierarchy of modules with storage and processing abilities can operate, in a flexible

way depending on the sensed environment and current stored contents. We have shown that (a) control loops can form for controlling the environment at different levels of disturbance, (b) plans can be generated and elaborated into explicit action, (c) action can focus perception, (d) perceived information can influence action directly at the appropriate level, and (e) distributed processes can form and cohere, and can restructure depending upon eventualities.

We have modeled control in the brain as well as data, and have shown how distributed control can work. Initiative and intention are distributed. In a set of interdependent modules, many modules are essential for overall action and the contribution of any one of them can be controlling.

Our model is relatively “vanilla”,¹ i.e., non-idiosyncratic. It is designed as a ground-level representation which can be further developed in the light of future knowledge.

Our model is of course an approximation. We anticipate corrections, but our scheme and principles will probably not be strongly affected by the particular assignments of areas to regions or regions to levels, for example.

For modules with memory, the action of the module is typically to update the memory according to the incoming descriptions. Thus memories track in memory the stream of descriptions they receive. Each module operates in a processing and communication environment of a set of other modules. There is a current state of the system whose characteristics usually change slowly compared to the reaction time of one module (cf. Arieli et al., 1996).

Modules can be characterized by their ability to *combine* information of more than one given types from more than one input, and to *create* and to *store*, data of given types. The development of modules with the abilities to store certain types of data would probably be aided by an existing flow of data from sensors and from subcortical areas to each module. Learning to process data by particular data transformations similarly would be aided by the co-occurrence of particular data on input and output channels of each module.

7.5.1. Consequences for experimental design and the interpretation of results

Characterizing the type of data processed by a given area is seen as key. From this a process model can be developed, as we have shown. A given area however receives data from more than one input and from downstream as well as upstream. If measurements cannot be easily made, our model can give some idea of the kinds of data that will be received from downstream.

7.5.2. A complete computer

Our model is a complete specification of brain activity, including control as well as data. In order to use this approach to describe brain activity to compare with the results of a particular experiment, it is therefore necessary to hypothesize the entire computational mechanism, both as regards data being processed and plans being executed by the brain. In a conventional boxes and arrow model, only the data and data paths are specified, the rest being described imprecisely using natural language.

7.6. The binding problem

Our model embodies a solution to the binding problem. Data coherence and control synchronization occur because specific pattern matchings and associations occur between modules. In addition, specific confirmatory messages, positive and negative, control the coherent activity of distributed processes. The demonstration of this approach to the binding problem is only now possible because we have an explicit brain model and environment so that explicit storage contents and associations can be modeled.

7.7. Predictions

Our main contribution is to have produced a computational approach and explicit working architecture for the primate brain. The change of thinking thereby involved produces many opportunities for falsification and verification:

7.7.1. Information roles of cortical areas

(i) attention requirements and requests for information are generated by the action hierarchy and cause focusing and changes in processing in the perception hierarchy. (ii) During initiation and change of action, we should get more processing in higher levels of the hierarchy. (iii) PA5 should be involved in the contextual control of action.

7.7.2. Falsifiable predictions of brain area activation

For the two grooming strategies, we can now generate detailed predictions of brain area activation sequences that should be observed during different types of processing situation. Using our computer realization, we can generate detailed predictions of activation levels for each time step. Since there are many adjustable parameters and detailed assumptions in the model, it is difficult to find clearly falsifiable predictions. However, we can also make a simplified and more practical form of prediction by classifying brain states into four types, shown in Fig. 16. Let us call these types of states G, E, P and M respectively. During grooming, we expect orientation would involve sequences such as G,P,E,M,

¹My apologies to Vanilla planifolia.

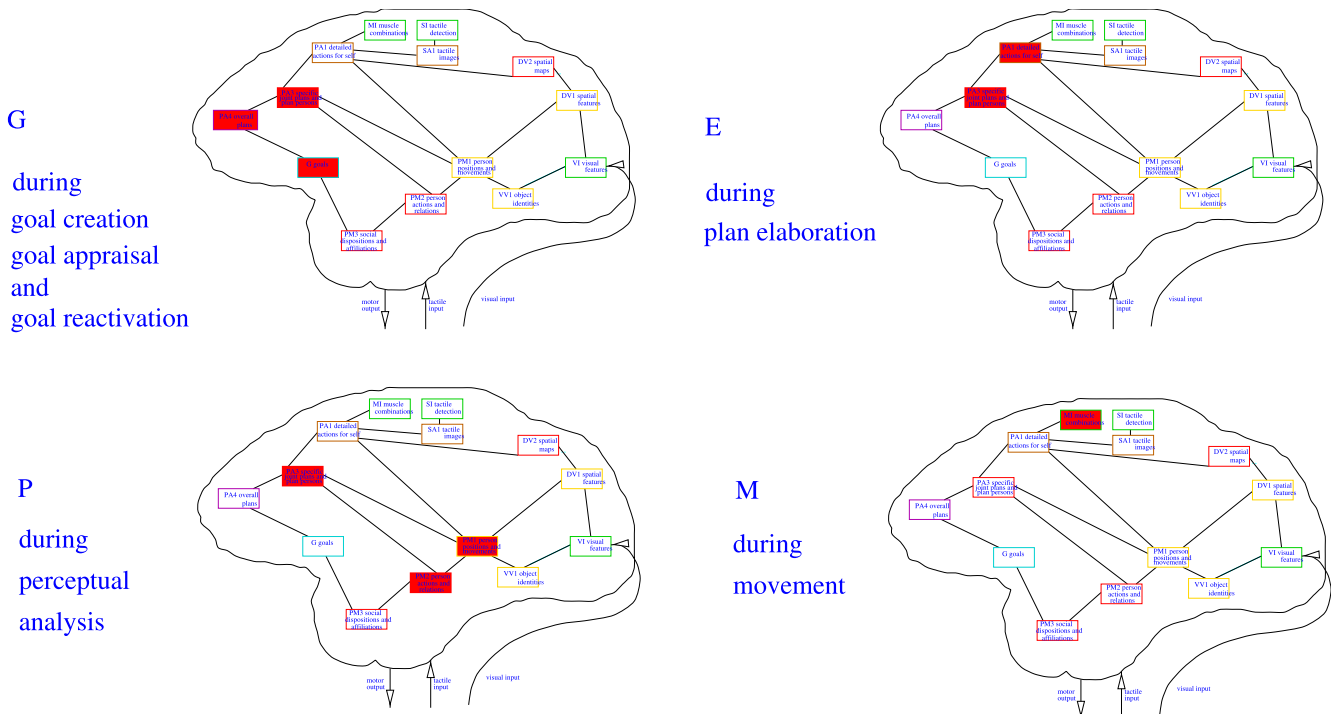


Fig. 16. Predicted brain area activation for different kinds of processing.

during approach G,E,M,E,M..., grooming prelude P,E,M,E,M,..., grooming P,E,M,E,M,M,M.. During social conflict, orientation G,P,E,M,E,M and during plan failure G,G,G,...P,E,M,E,M.

It should be noted that there is some redundancy in the model, so that, if a mismatch to experiment is found, it would be possible to make some changes to the model to bring it into better correspondence with the data. For example, the assignment of modules to particular brain areas is tentative and may need to be changed. However, there is a limit to the changes that can be made, and mismatches with data could falsify the model in its present form.

7.7.3. Correspondence to lesion studies

As regards lesioning, 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.

7.7.4. Individual differences

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).

8. Summary and conclusion

From the functional architecture of the primate brain, we were able to design a computational architecture using a computer science analysis. We defined a computational approach representing neural regions as distributed modules whose data contents were represented by logical expressions and whose processing was represented by the action of sets of rules. A particular explicit brain model was defined and implemented. Its demonstration of social behaviors in a minisociety showed the correctness and completeness of the model and the strength and feasibility of the approach.

We have developed a method of brain modeling where we do not need to make postulates concerning the encoding of information by neurons or the specific neural mechanisms used for processing information. Our method of modeling takes into account processing and storage constraints of the brain and yet does not require modeling the behavior of individual neurons.

The model has causal dynamics and makes falsifiable predictions concerning information processing in the brain. Conversely, of course, the model makes no statement concerning detailed neural activity or neural encoding of information.

Instead, we can use abstract descriptions of information and of the processing of information. This level of modeling allows one to make scientific theories concerning the types of data being transmitted, stored, and processed, transformations of data, distribution of data in the brain, specializations of processing at different locations. We can also represent issues of processing resources and timing of processes.

Acknowledgements

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 would like to thank Professor Stott Parker of UCLA for his steadfast friendship and support. This work has been supported by the National Science Foundation, Information Technology and Organizations Program managed by Dr. Les Gasser, and is currently supported by the Cooperation and Social Systems program managed by Dr. Susan Iacono. The author would also like to thank Professor Pietro Perona and Mani Chandy for their support and encouragement at Caltech, where this research is currently being pursued. He would finally also like to thank Professor Steve Mayo at Caltech for his generous provision of computing resources.

References

- Arieli, A., Sterkin, A., Grinvald, A., Aertsen, A., 1996. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* 273 (5283), 1868–1871.
- Bond, A.H., 2004. An information-processing analysis of the functional architecture of the primate neocortex, in this issue.
- Calvez, J.P., 1993. *Embedded Real-Time Systems*. Wiley, New York.
- Creutzfeldt, O.D., 1978. The neocortical link: thoughts in the generality of structure and function of the neocortex. In: Brazier, M.A.B., Petsche, H. (Eds.), *Architectonics of the Cerebral Cortex*. Raven Press, New York, pp. 357–383.
- Daigneault, S., Braun, C.M.J., Whitaker, H.A., 1992. An empirical test of two opposing theoretical models of prefrontal function. *Brain Cognition* 19, 48–71.
- Davis, A.M., 1993. *Software Requirements*. Prentice-Hall, Englewood Cliffs, NJ.
- Deacon, T.W., 1989. The neural circuitry underlying primate calls and human language. *Hum. Evol.* 4, 367–401.
- Douglas, R., Martin, K., 1998. Neocortex. In: Shepherd, G.M. (Ed.), *The Synaptic Organization of the Brain*, 4th ed. Oxford University Press, Oxford, pp. 459–510.
- Fodor, J.A., 1983. *The Modularity of Mind: an Essay on Faculty Psychology*. MIT Press, Cambridge, MA.
- Gazzaniga, M.S., 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 and 585–644.
- Goldman-Rakic, P.S., 1988. Topography of cognition: parallel distributed networks in primate association cortex. *Ann. Rev. of Neurosci.* 11, 137–156.
- Goodwin, C., 1981. *Conversational organization: Interaction between Speakers and Hearers*. Academic Press, New York and London.
- Karmiloff-Smith, A., 1993. *Beyond Modularity*. Oxford University Press, Oxford.
- Holyoak, K.J., Hummel, J.E., 2000. The proper treatment of symbols in a connectionist architecture. In: Dietrich, E., Markman, A.B. (Eds.), *Cognitive dynamics: Conceptual and Representational Change in Humans and Machines*. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 229–263.
- Kling, A., Steklis, H.D., 1976. A neural substrate for affiliative behavior in nonhuman primates. *Brain Behav. Evol.* 13, 216–238.
- Kosslyn, S., 1994. *Image and Brain*. MIT Press, Cambridge, MA.
- Kraemer, G.W., 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, A.R., Grady, C.L., Ungerleider, L.G., Haxby, J.V., Rapoport, S.I., Horwitz, B., 1994. Network analysis of cortical visual pathways mapped by PET. *J. Neurosci.* 14, 655–666.
- Mega, M.S., Cummings, J.L., 1994. Frontal-subcortical circuits and neuropsychiatric disorders. *J. Neuropsychiatry Clin. Neurosci.* 6, 358–370.
- Mesulam, M.-M., 1990. Large-scale neurocognitive networks and distributed processing for attention, language and memory. *Ann. Neurol.* 28, 597–613.
- Newell, A., 1990. *Unified Theories of Cognition*. Harvard University Press, Cambridge.
- Petersen, S.E., Fox, P.T., Posner, M.I., Minton, M., Raichle, M.E., 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. In: Boller, F., Grafman, J. (Eds.), *Handbook of Neuropsychology*, Vol. 9. Elsevier, Amsterdam, pp. 59–82.
- Pylyshyn, Z.W., 1984. *Computation and Cognition*. MIT Press, Cambridge, MA.
- Shallice, T., 1988. *From Neuropsychology to Mental Structure*. Cambridge University Press, Cambridge, England.
- Shallice, T., Burgess, P.W., Schon, F., Baxter, D., 1989. The origin of utilization behaviour. *Brain* 112, 1587–1598.
- Shepherd, G.M., Koch, C., 1998. Introduction to synaptic circuits. In: Shepherd, G.M. (Ed.), *The Synaptic Organization of the Brain*, 4th ed. Oxford University Press, Oxford, pp. 1–36.
- Siewiorek, D.P., Bell, C.G., Newell, A., 1982. *Computer Structures: Principles and Examples*. McGraw-Hill, New York.
- Ullman, S., 1991. Sequence-seeking and counter streams: a model for information processing in the cortex. AI memo 1311, MIT, Artificial Intelligence Laboratory.
- Wernicke, C., 1874. *Der aphasische Symptomencomplex*. Frank and Weigert, Breslau.