How we might be able to Understand the Brain

Brian D. Josephson

Department of Physics
University of Cambridge

http://www.tcm.phy.cam.ac.uk/~bdj10

Collaborators, etc.

Hermann Hauser
David Blair
Nils Baas
Andrée Ehresmann

presentation created using OpenOffice.org 1.1.0 running on MacOS 10.2.8

The problem: explaining phenomena generated by brain

for example: language

(complexity is the issue)

Important point: *describing* phenomena (e.g. as in linguistics)

is *not the same* as modelling them with a mechanism

linguists do one, neuroscientists the other

and rarely the twain do meet!

We need an *integration of the phenomenal and mechanistic points of view* that does not ignore questions of complexity

- There exist complicated systems that we do (more or less) understand

e.g. computer programs, electronic circuits, machines, life

what is special about these is the way they possess structured analyses

Assume that the brain, is in some sense, similar.

Implicated here is the *design issue*, a point of contention in the neurosciences.

The ‘neural constructivists’, seemingly ignorant of the way in which, in AI, *General Problem Solver* was superseded by *Expert System*, appear to ascribe, on the basis of generalisation from successes with comparatively simple skills, corresponding miraculous powers to general-purpose neural networks. But experiment suggests there is such a thing as ‘domain-relevant mechanisms’ (Karmiloff-Smith). The consideration they leave out (with apologies to President Clinton) is this:

‘It’s the efficiency question, stupid’

It would be curious indeed if the brain were (uniquely) *not* specifically designed to be good at the things it is good at, and had to learn everything from scratch instead.

The design is hidden from immediate view: as with circuits and computer programs, when things get complicated "we cannot see the trees (functional components) for the wood (the entire system)"; but the design is there nevertheless and the effectiveness of the system crucially depends upon it.

Now we begin the details …

Comment concerning design and implementation:

In general, the physical device used in a mechanism (e.g. transistor, pump) is merely the means of realising the requirements of a design.

Similarly, we assume that the nervous system is merely the means of realising a design: its status is secondary; it is just something for achieving what the design requires (comparatively basic information processing, and relationship learning, no magical powers).

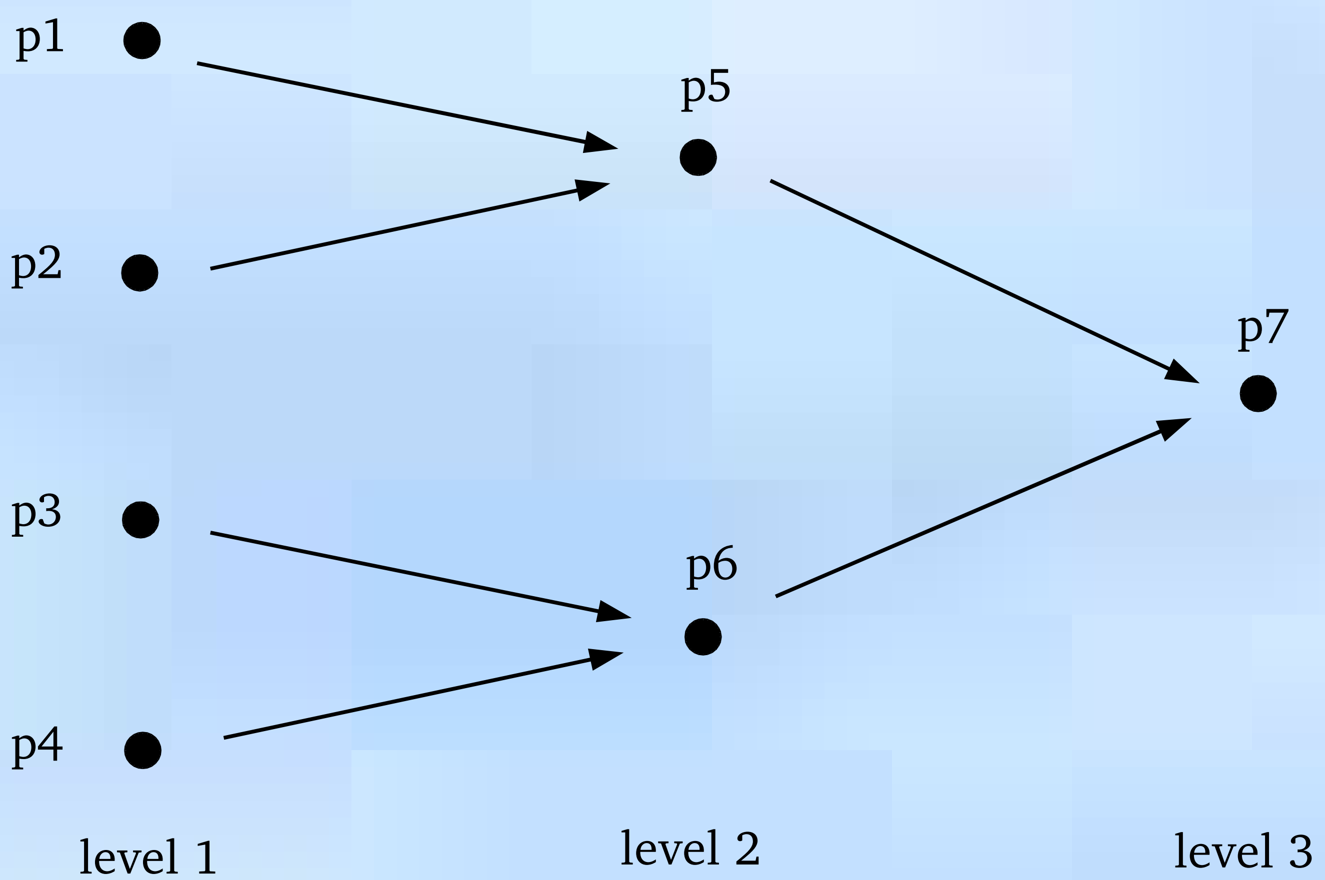
Models

Our understanding of a design is in terms of models, formal or informal.

Models allow us to calculate, or estimate behaviour.

We understand a *mechanism* in terms of a *collection* of models.

Structured analysis/hierarchical inference



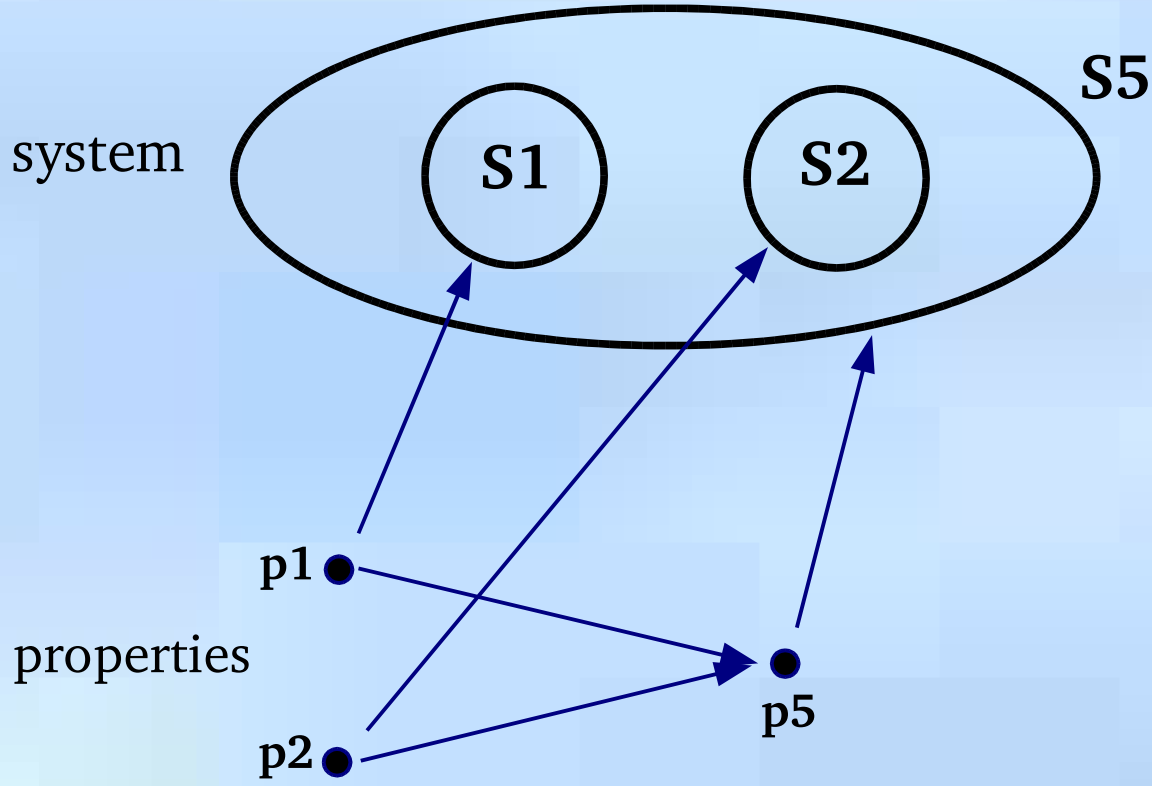
p_n are propositions re the various systems: we infer from each level to the level above.

The inference difficulty

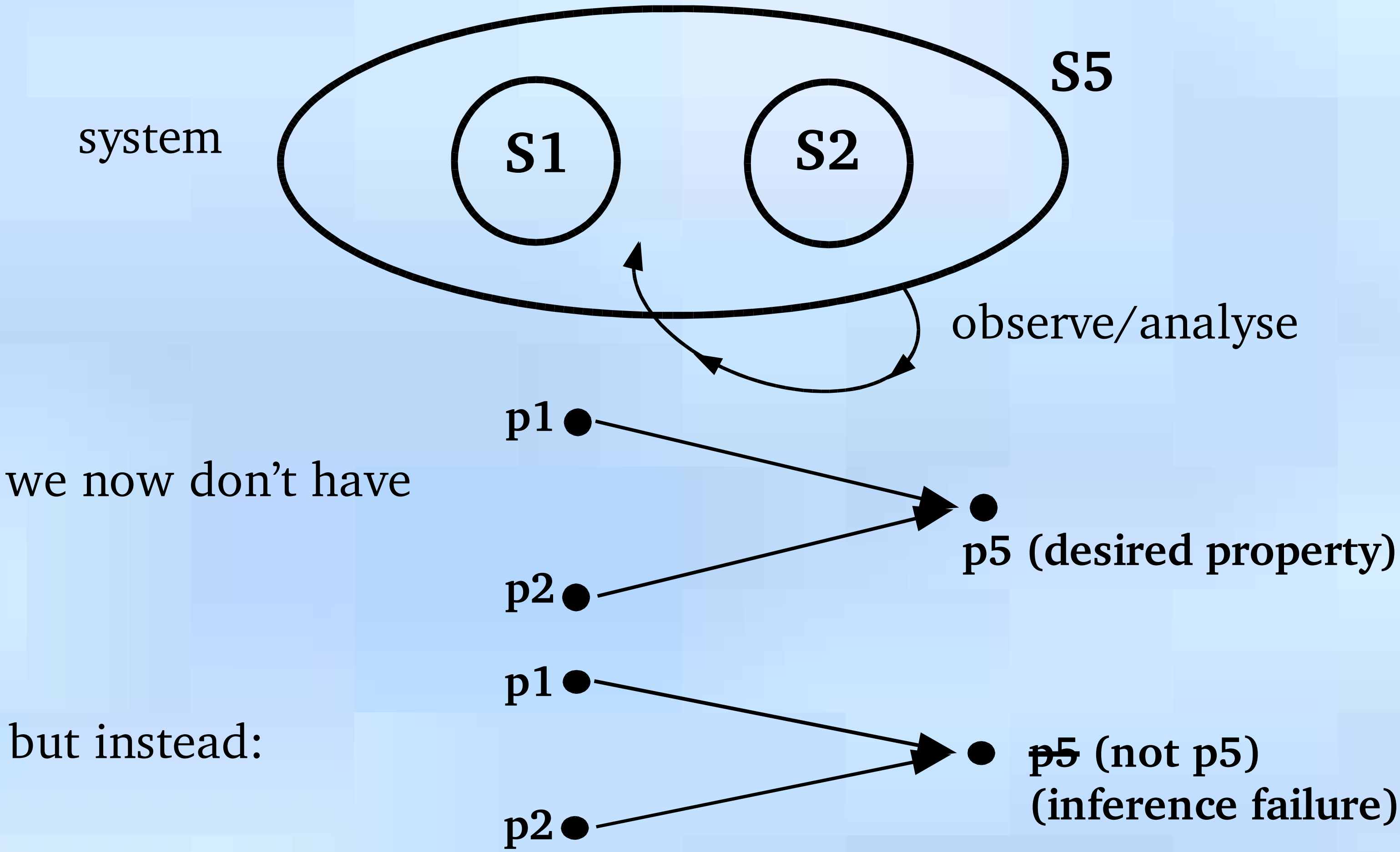
Such an inference scheme is applicable with ordinary mechanisms. With the brain, the difficulty is that the model-based inferences we require cannot be built in: indeed, the higher levels only manifest through development. The problem is resolved through the Baas hyperstructure concept/observer mechanism.

Baas distinguishes between the deducible emergence just discussed, and observational emergence.

Deducible emergence: we can infer the behaviour of a given system (e.g. system for driving walking) from that of its component subsystems:

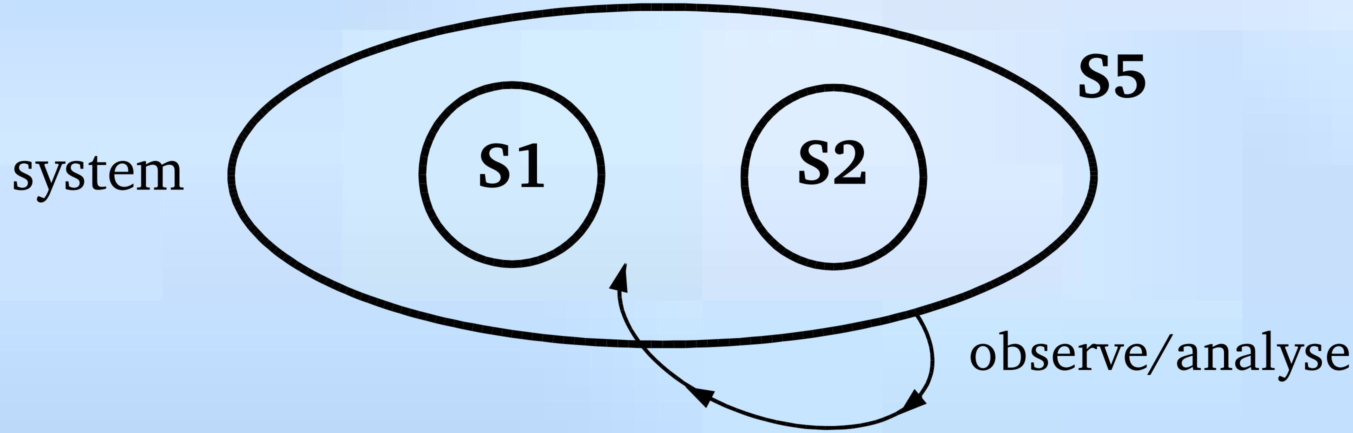


observational emergence

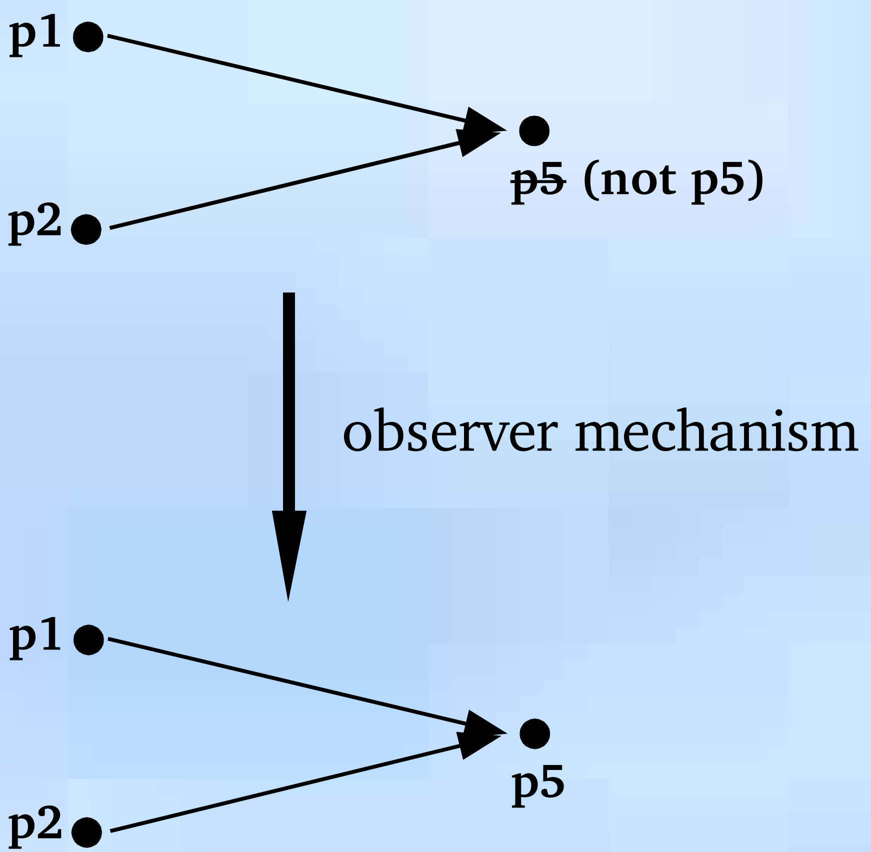


so we add an *observer mechanism*

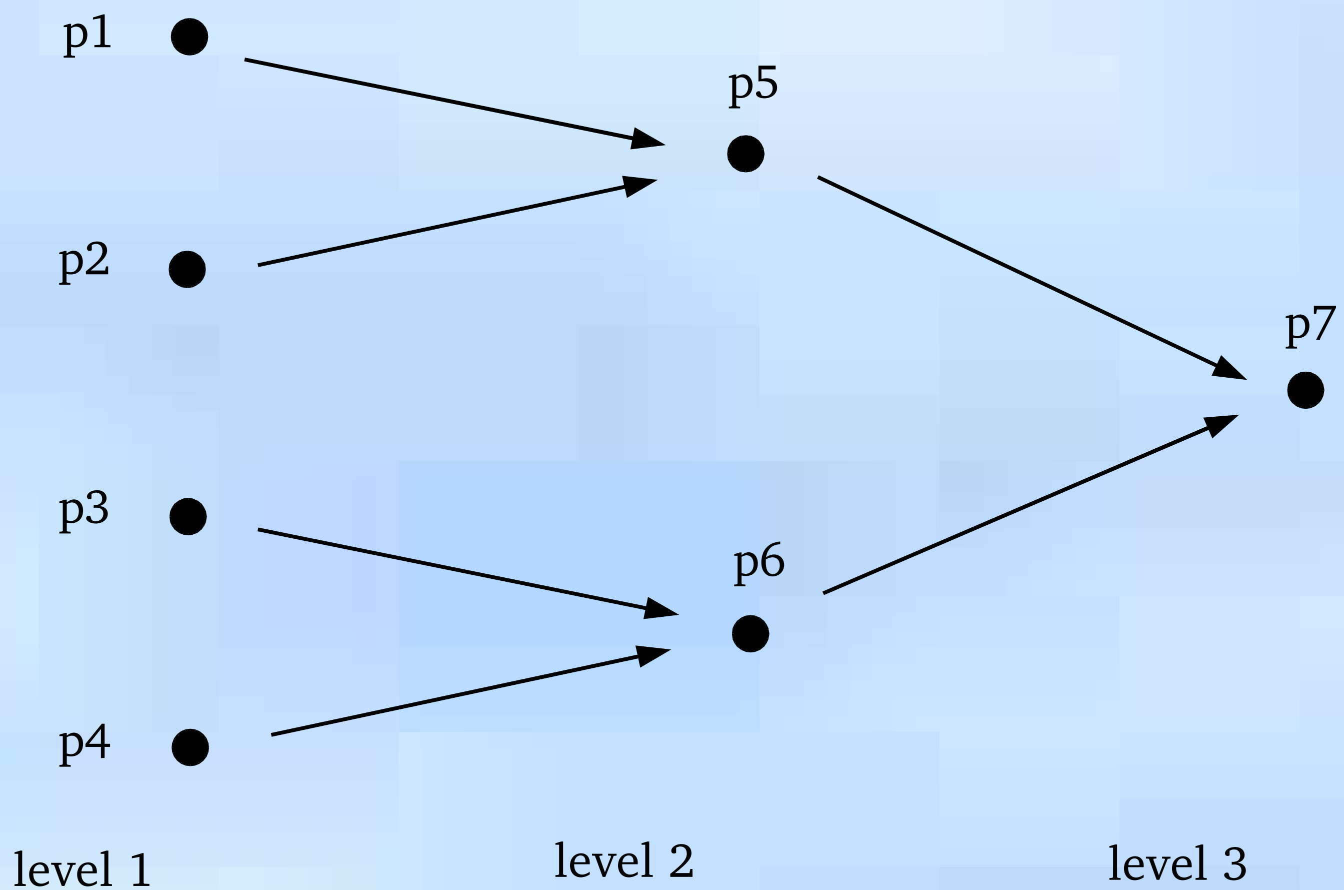
UNIVERSITY OF CAMBRIDGE



In the Baas approach, the various subsystems don't start off working together in a way associated with an inferential hierarchy. Instead, the behaviour of various combinations is observed and analysed, and adjustments are made in accord with the outcome of the analysis. This leads to the following progression:

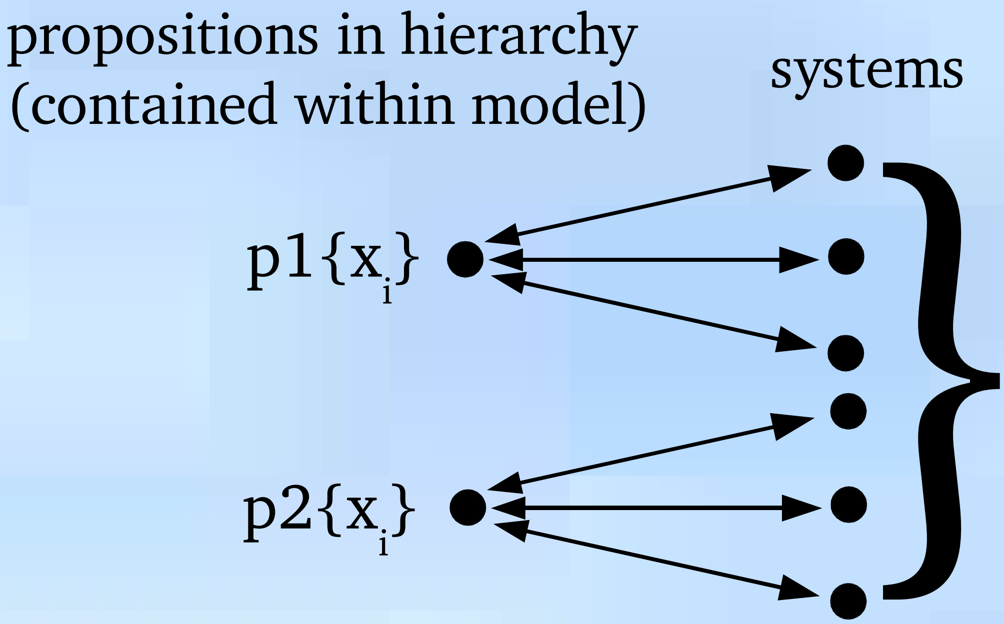

p1
p2
p5 (not p5)
observer mechanism
p1
p2
p5

Thus, piece by piece we might generate our hierarchy:

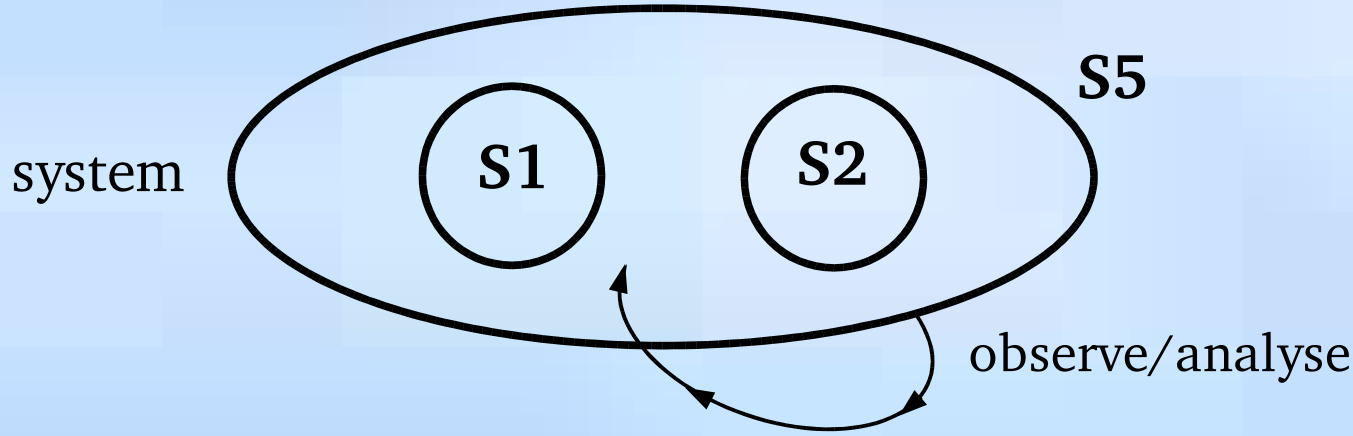


But obviously this is too schematic!

The hierarchy portrayed is actually a collection of localised inferences interrelated by the conclusion of one being the premiss of another, but each individual proposition can apply to many systems.



(x_i are parameters that can vary with time, and from system to system)



These processes of construction keep going on, involving many kinds of systems, each kind governed by its own model. The systems run themselves in accord with the various models, gradually achieving higher levels of the hierarchical scheme.

Note: Such constructions can involve abstract processes as well as physical ones.

With each skill, there is a specific ladder of development with the steps being levels in the hierarchy. Systems are established at one level and then used to form the next (e.g. rising to the vertical, standing in balance, taking a step, taking a number of steps), each transition being governed by its own model (similarly for steps in language, where there is a more complex structure).

Comment on top-down aspect

We have described system generation in a bottom-up manner, but there is also a top-down aspect. There is no incompatibility here, any more than a photograph of an object taken from above is incompatible with one taken from below! The observer mechanism is responsible for endowing higher level systems with special properties in this case, making it useful to think of them as wholes (cf. computer programs).

Systems

Models are defined within *systems*

The inferences concerning models are made on the basis of the assumptions of the system

Inferences may also be made concerning the effectiveness of any observer/analysis mechanisms

The systems concerned may be of an abstract or concrete character

Abstract levels

Cognition involves abstract activity as well as concrete (physical) actions.

Physical actions are modelled by physical models, abstract activity by abstract models, whose significance will be discussed later.

The picture we have developed is highly abstract, and about as far removed from the real world as is the M-theory of the high-energy physics community (joke!).

But, just as we flesh out the general proposition "life is based on chemistry" by finding out which molecules and chemical reactions do what, we can (hopefully) turn the present proposals into a realistic theory of brain and behaviour by filling in some details.

What details?

- Specify the systems and the observer mechanisms informally

- Create a corresponding more formal specification
- Integrate with the behavioural aspects

- Investigate the neural implementation

There is partial overlap between these aspects, but in this approach the neural aspects come late in the game rather than forming the starting point —t he nervous system is seen as a resource, a means to an end.

How the models work

1. They involve model of world + model of network (e.g. balance, edges in visual field)

2. The network model is 'mathematised', i.e. described in terms of signals and numbers

3. The models take into account the context-dependence of effective performance, in 2 ways:

(a) using parameters
(b) using distinct systems for significantly different situations

What the neural architecture does

1. does necessary computations on the basis of representational schemes

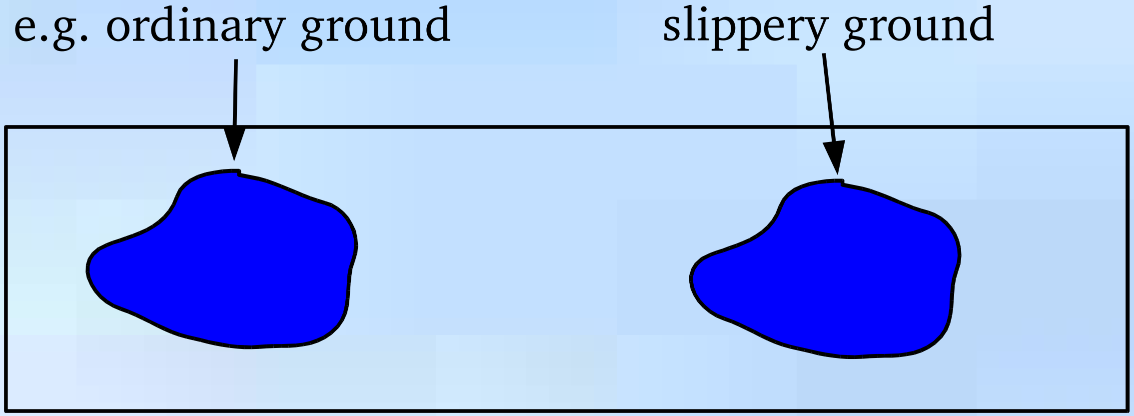
2. builds circuits in the ways dictated by models, i.e. so as to perform the necessary operations

(consider the way electronic circuits are dictated by the requirement that their processes produce the desired outcome)

The relevance of context

Different subsystems are highlighted in different contexts, to be 'where the action is' in the nervous system in that context.

Different circuits are built in different situations, and invoked again when the situations repeat.



UNIVERSITY OF CAMBRIDGE

Our whole approach stresses the importance of abstract systems and models in design.

In this respect it is consistent with standard neuroscience, but conflicts with constructivist dogmas which hypothesise general learning abilities of neural networks.

Unlike standard neuroscience, it has the capacity to cope with complexity, on the basis of the observer mechanism and its ability to construct hierarchies.

We now discuss the relevance of these ideas to two concepts discussed in the literature:

(a) abstraction via 'representational redescription' (Karmiloff-Smith)
(b) Jackendoff's theories of language

RR says basically that when we have developed a skill we 'redescribe' our successful behaviour, representing it by another mechanism. We learn to apply this new representation in various ways, affording a generality not available with the original mechanism, where behaviour tends to be frozen to preserve it.

The new feature of our approach is the idea that very specific, model-based, mechanisms (which should be accessible to developmental studies), underlie such processes. These models typically involve abstract relationships, such as membership of a class, or generalisation.

AI algorithms, for which justification can in principle be given on logical or mathematical grounds, give some insight into what may be involved.

Now we come to language, which has been extensively studied by linguists, who have proposed many descriptive models.

Jackendoff's models are of particular interest as they do not extract a part such as syntax, and claim to get a proper account on that basis. The proposed approach offers the possibility of extending his ideas and integrating them into neuroscience.

proposed evolutionary steps

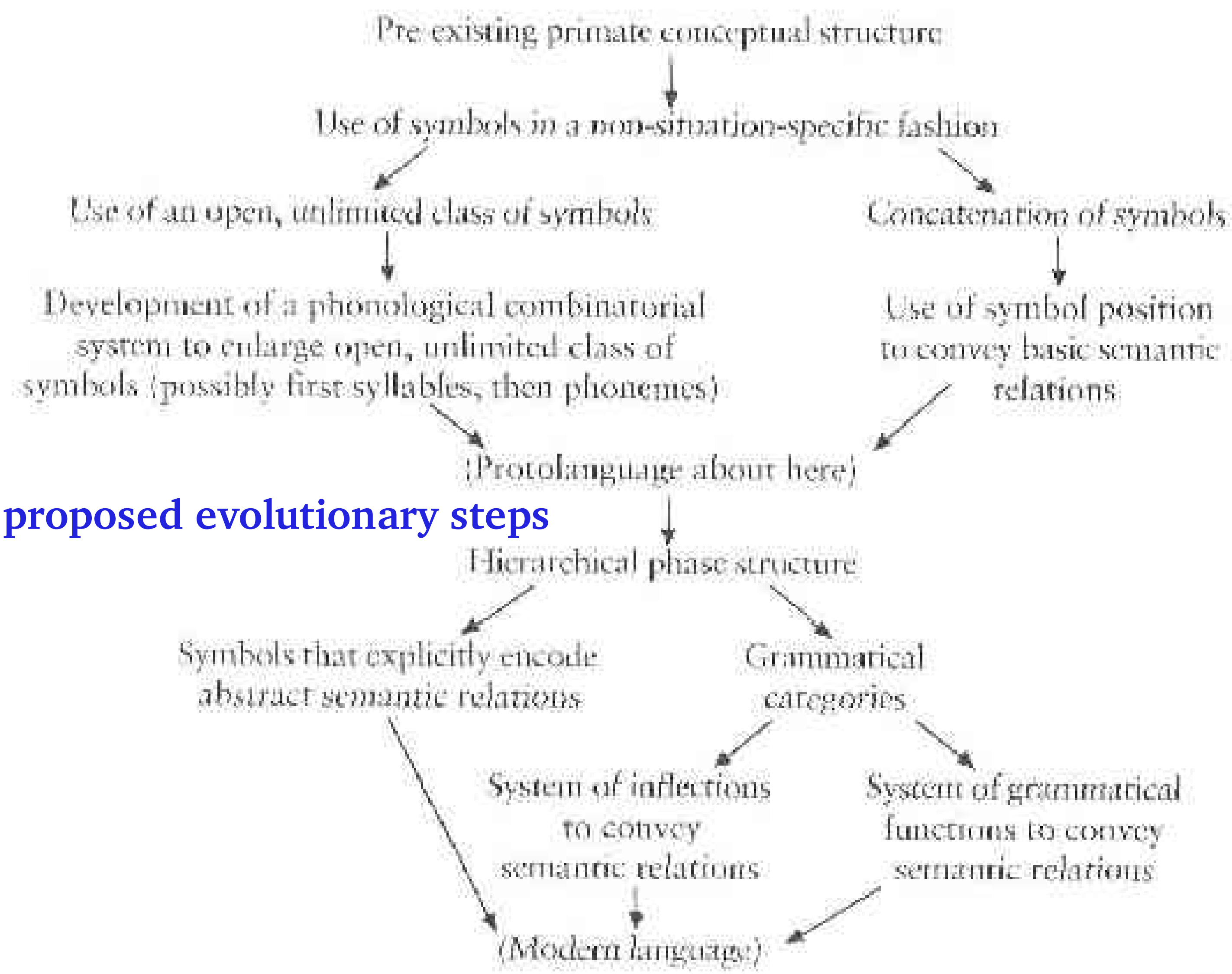


Fig. 8.1. Summary of incremental evolutionary steps

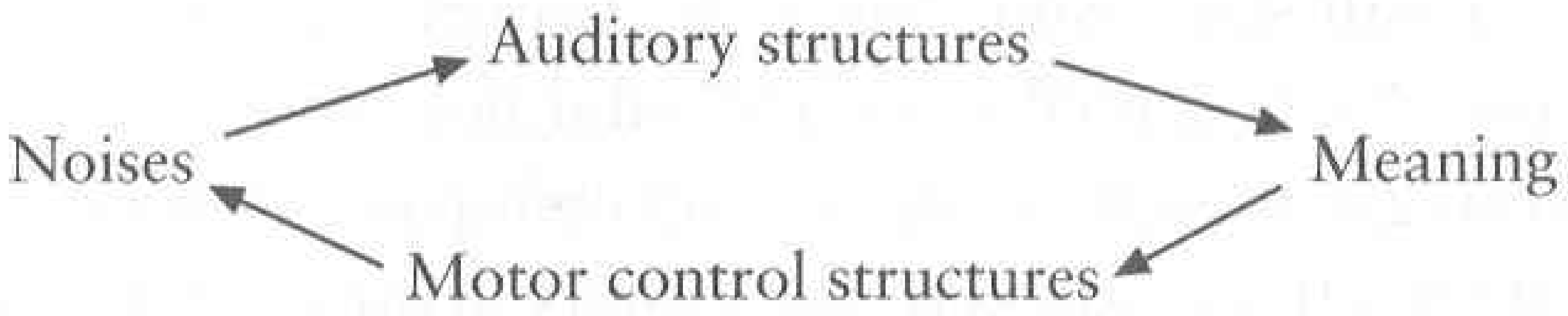


Fig 8.2. Architecture of early single-symbol stage

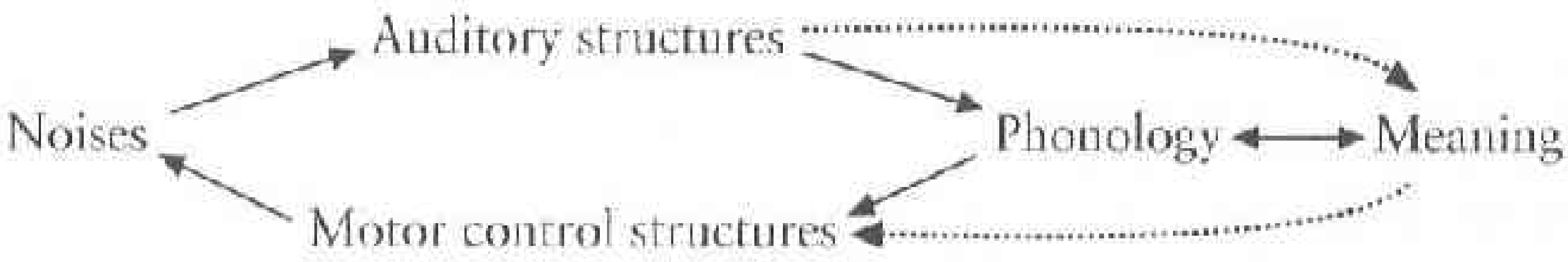


Fig. 8.3. Architecture of protolanguage/Basic Variety

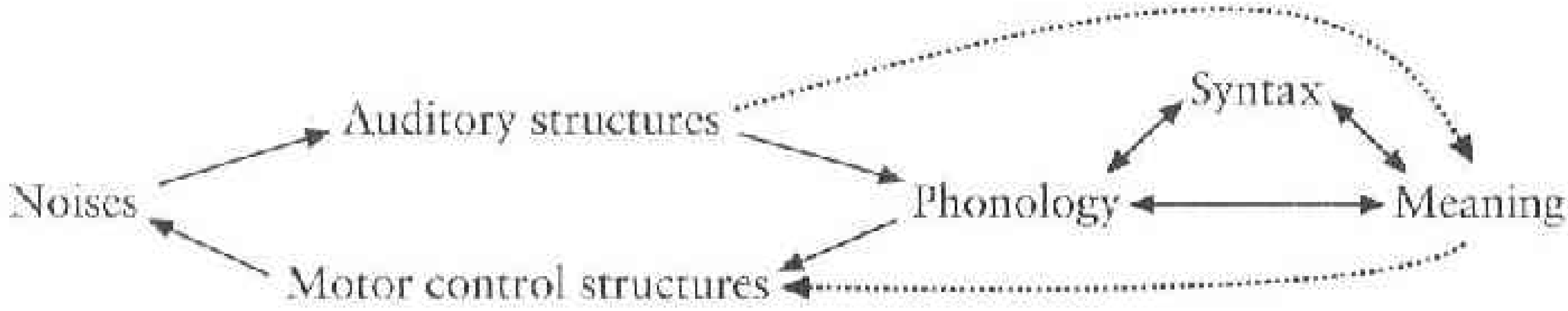


Fig. 8.4. 'Early modern' tripartite architecture

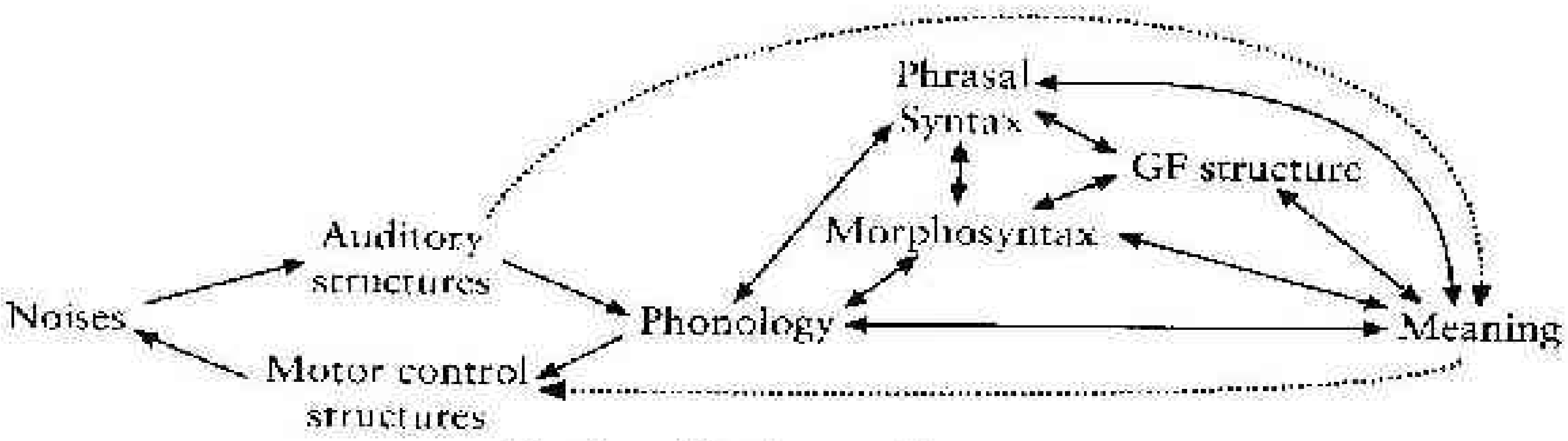


Fig. 8.5. Modern architecture

Our approach, since it links together skills, models, and mechanisms, should have much to say about these details. What is going on in all of these processes is that various connections are being explored and exploited, all on the basis of models applicable to particular things in particular situations.

e.g. ‘use of symbols to convey basic semantic relations.’ This is already an informal model, which can be refined into a formal model. Then can come circuit analysis, and finally linkage to the real nervous system.

Representational redescription and syntax

At every stage, circuits are being constructed to accomplish particular things on the basis of particular models; each such circuit is used to do more things in more contexts at the next level.

Syntactic structure can be viewed as the mechanism underlying a modification of representational redescription:

The basic formulation of RR states that after we have developed a skill we ‘redescribe’ our successful behaviour, representing it by another mechanism, and then learn to apply this new representation in various ways, affording a generality (power) not available with the original mechanism. Syntactic structures can be seen as derived structures offering ways of improving access to the semantic intent of a speaker. Specialised forms of observation and analysis are involved in developing a capacity.

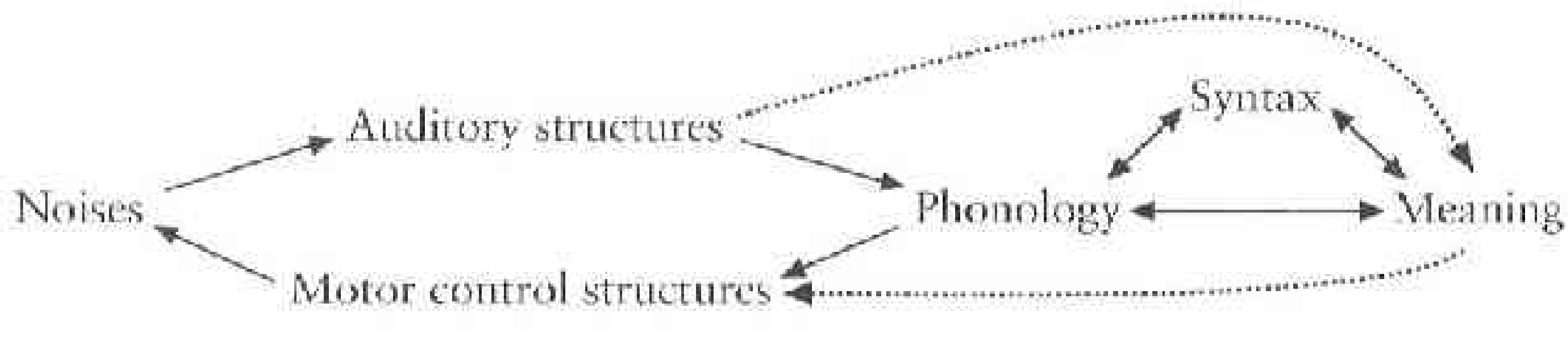


Fig. 8.4. ‘Early modern’ tripartite architecture

Conclusion

It has been argued that

• models are the basis on which the nervous system is designed

• observer mechanisms make use of model-related behaviour to generate hierarchies

On the basis of the picture described in this paper, the brain can be seen to 'make sense', and the neurosciences can now proceed to unravel the complexties of the brain in a way that has formerly been denied to workers in the field.

How we might be able to Understand the Brain

THE END