The JANUS Architecture for an Artificial Brain: Medical Perspectives

F.J. Śmieja smieja@gmdzi.uucp

H. Mühlenbein muehlen@gmdzi.uucp

German National Research Centre for Computer Science (GMD),
Schloß Birlinghoven, 5205 St. Augustin 1, Germany.

March 22, 1991

Abstract

In the evolution of the mammalian brain the two-sided nature of the neural motor control and sensory input has been preserved in the architecture of the cerebral matter. We believe this fundamental two-sided nature of the mammalian brain to hold a very important clue to the problem-solving capability and method of the higher mammals. The JANUS idea for the design of a robot's brain is to have a two-hemisphere architecture on the macro-level and self-assessing modules of neural networks on the micro-level. At the highest architectural level there is the possibility of conflicts between the opinions of the two hemispheres, which can independently process the same data and generate motor decisions. The halves are connected by a simplified analogue of the Corpus Callosum, through which they can exchange information at various architectural levels. In this paper the JANUS architecture is motivated and described, and we explain how we expect to use this medically-inspired model to contribute to our understanding of neurophysiological rehabilitation, through construction of a *visuo-manual* prototype, which has two eyes and two arms.

1 Motivation

One of the higher-level aims of neural network research (or *connectionism*) is to develop systems that will be able to perform a wide range of tasks, and survive in a changing and perhaps hostile environment. Most neural network research to date has been involved in developing low-level models of network architectures and dynamics. Each of the individual networks alone cannot however satisfy all the requirements of an adaptive and at least partially autonomous system. It has been shown on a number of occasions that the fundamental scaling properties of such networks make the quest for a scalable general-function neural net rather naïve [8, 4, 5, 10]. What is required to solve large, complicated tasks is not large, complicated networks but a large number of small network modules, each of which may be specialized in the performance of a particular aspect of the complicated task, with perhaps suggestions made by these modules to be combined further up the hierarchy of the system.

But how might these modules be constructed, and how might they interact with each other, in order to keep chaos from making the system unusable? We propose an answer to such questions with our basic JANUS idea: the top level has a symmetrical dual-sided form, modelled on the gross features of the mammalian brain, and at the lowest level of the modular hierarchy is found a (MINOS neural network module with limited abilities but the vital property that it is *self-assessing*. Further hierarchical levels build their own self-assessment ability through combining those of lower levels. With such a handle, system knowledge and capability, although distributed, may be managed and organized, through use of various feedback loops that run upwards through the hierarchy.

With JANUS we intend to bridge the gap between small-scale neural network success and real-world applicability. Our prototype application is a visuo-manual system that has two eyes and two arms, with which it observes and manipulates its environment. The prototype learns from experience and self-supervision, initialized only with a few essential preprogrammed properties. Having successfully trained the system, we progress from mainly engineeering considerations to the medical aspects of our work: JANUS incorporates structural and functional medical knowledge of the brain, and our first experiments on the trained brain will be to investigate its plasticity after damage.

2 Introduction to the ideas of JANUS

The choice of the architectural constraints of JANUS was directly influenced by simple observations of the top-level structure of the human brain and its hemispherical functional lateralization [1, 9]. However, the similarities end at that level and a great deal of freedom is permitted in lower-level neural network module development, within the framework of the essential constraints. We chose the name JANUS after the Roman god for a specific reason: not only does our brain architecture have a double nature physically, but, most importantly, it has a double nature conceptionally. The brain not only looks out and observes and weighs up its environment, but it looks *inwardly* and is aware of and estimates its own processes. It is a *reflective* architecture. More will be said about this in section 4.

A JANUS brain is intended to control a robot that may exist in a changing environment. It can be affected by the environment and likewise can affect the environment itself. The robot can possess various sensory attributes in order to perceive the environment in which it exists. It can be affected by the environment either directly, through physical contact with objects, or indirectly in the development of thought and learning processes within the brain. The robot itself can affect (its perception of) the environment either non-physically through controlling the sensory inputs or directly through effecting physical alterations in the environment.

The highest-level description of the JANUS architecture is illustrated in figure 1. The sensory representation of the environment provides the input to the system (we show the visual input in figure 1), and the output from the system is in the form of motor connection to parts of the robot which can change its or the environment's state. The brain is divided into two halves laterally and two blocks vertically. All sensory and reinforcement signals from the left side of the robot pass directly to the right half of the brain, while those from the right side of the robot pass directly to the left half of the brain. The left half of the brain controls directly the motor outputs affecting the right side of the robot, and similarly the right half controls the left motor side.

There exist important connections between the two hemispheres (the "Corpus Callosum"), through which information of various sorts can be exchanged. Thus it is possible for both sides to receive sensory information about the whole environment and robot's body. Finally, an approximate symmetry is imposed between the halves, with regard to all hierarchical functional levels of the vertical modular structure they possess.

3 Functional description of the brain

We have identified the various regions of our architecture using the biological names for regions of the human brain, not because we claim they have comparable processing power, but because the jobs they undertake are respectively taken care of loosely in the implied areas in the human brain. In figure 1 we identify the areas Corpus Callosum, Cerebellum and Cerebral hemispheres.

3.1 Cerebral hemisphere processing

In general these areas are concerned with high-level programs like planning, goal creation, conscious supervision of the learning of new tasks, and building new skills involving linking together various primitive motor programs. We find also sensory recognition and classification here. More particularly, and significantly for our prototype system (section 5), all the *visual processing* occurs here, involving basic shape recognition and binocular fusion. The processing is carried out by modules of neural networks.

Consciousness—the highest level self-supervision process going on in the brain—happens here, although its location tends to change, depending on the current area of highest attention on either hemisphere. The same inputs are accessible to both sides of the brain and the data is analysed by modules of neural networks. The same *information* is available to each hemisphere, but different *aspects* or global/local *representations* are considered. Likewise, in learning about the environment different aspects are considered, and in such a way two different interpretations of the environment may develop. The imposition of such differences on our architecture were inspired through observation of the many split-brain experiments [9] and the evidence they provide, in a nutshell, for the greater degree of analytical interpretation of events on the left side of the human brain, and more of a non-critical passive acceptance and modelling of the environment in a synthetical way on the right side. We hope to include such differences in our model eventually in a general way through imposing greater *self-supervision* of the processing occurring on the left side [1].

3.2 Cerebellum processing

In both sides of the cerebellum we find the basic control modules, such as those controlling limb movement and coordination. The cerebellum of our prototype is concerned with arm movement specifically. Reflex behaviour is also located here, by which we mean the automatic unintentional reactions to "pain" stimuli. For our prototype such pain stimuli will be in the form of direct knowledge of collisions of the arms with objects in the environment and with each other. Building-block motor programs are evolved in the cerebellum. These are programs that use the basic control procedures recurrently in order to produce some kind of fluent motion learnt by temporal neural network modules. In our prototype they will be concerned with learning smooth motions of the arms, both singly and in unison. Unconscious learning occurs here. As opposed to the conscious learning directed from the hemispheres, in

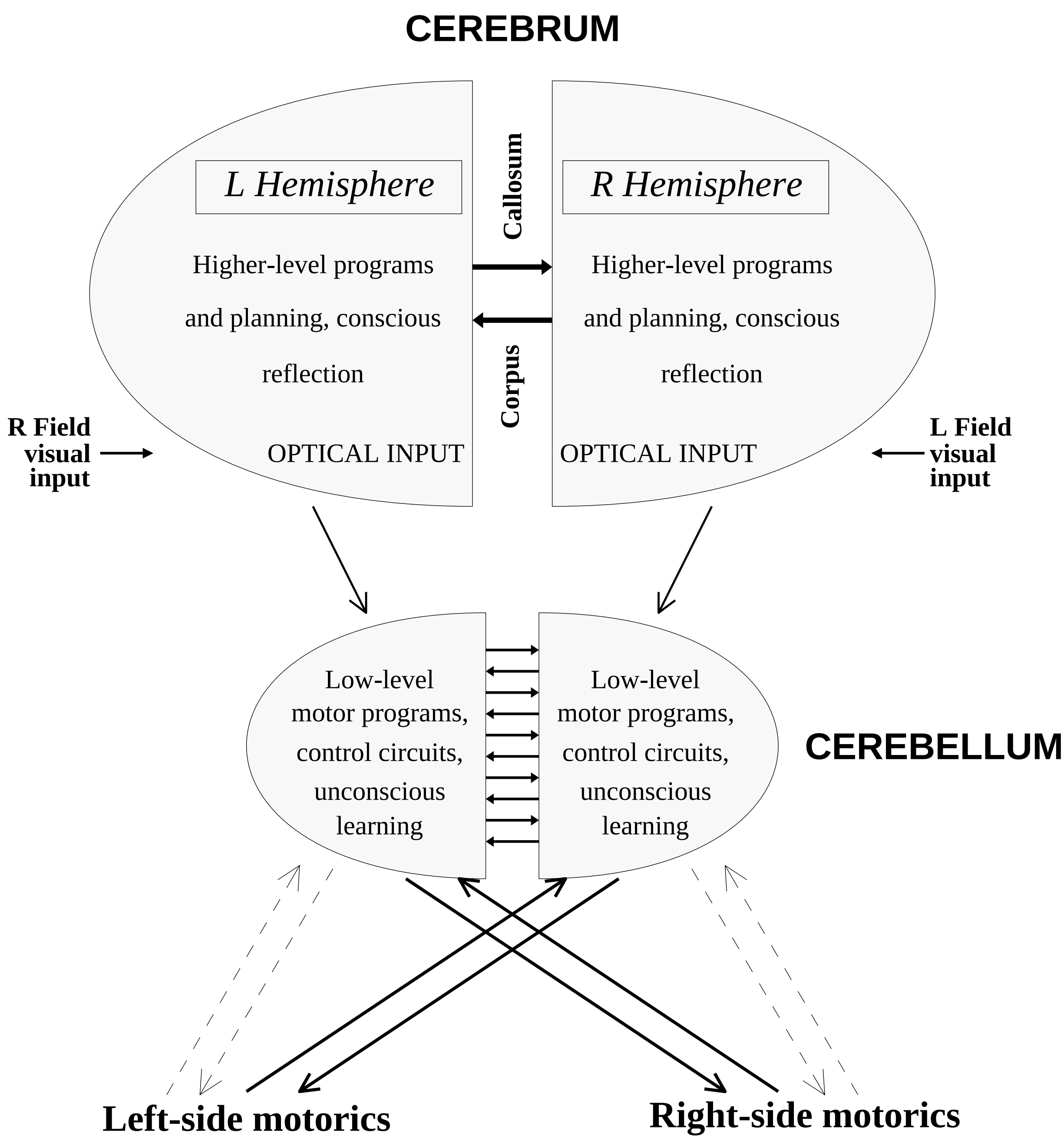

Figure 1: JANUS: Top-level illustration of the architecture. Identified are the three biologically inspired regions: Cerebral Hemispheres, Cerebellum and Corpus Callosum, together with the sensory (here visual) input areas and motor control outputs.

this region we find modules that are passive and continuously (although variably) receptive to transient signals, and may learn current associations of sensory signals with observed actions unconsciously. The degree of unconscious/passive learning depends on the persistence of the exercise. In the prototype they allow for the possiblility of passive learning of one-side control of the robot through experiencing other-side control, which is important for some rehabilitation programs [6]. Attentional processes that constantly monitor the sensory inputs in order to discover interesting changes in the environment or avoid unpleasant ones, are also located here. For the prototype they are necessary for changing the concentration area of the eyes during a scanning exercise and pulling new objects into focus. They are also necessary for enabling the robot to learn fast reaction behaviour, that it will without conscious processing carry out when necessary.

3.3 Lateral connections

The Corpus Callosum is the name given to the complete set of connections between the two cerebral hemispheres of the brain. These are simple fixed strength connections, that can be excitatory or inhibitory. In our prototype they allow the information from the left and right visual fields to be combined in both hemispheres, and for conflicts in higher decisions to arise, through incompatible desires on each side.

It is neurophysiologically unclear how it is that the left hemisphere seems to have almost total control of the right motoric system, and the right hemisphere of the left motoric system. We have modelled this observation here through a partition of the Cerebellum into two halves, connected by many nerves that do not possess the same 1–1 interhemispherical associative property as the Corpus Callosum. In addition, the left side of the Cerebellum is connected 90% directly to the right motoric system, and 10% to the left motoric system, and similarly for the right side of the Cerebellum.

4 Self-assessment and self-supervision

Probably the central theme to our JANUS concept is the notion of *self-assessment*, or self-supervision, within a hierarchy of adaptable modules. The modules can modify themselves, and higher levels can act on other levels, extending them, improving them, repairing them, etc. In order that this might be possible, each module, and each level, requires to nullify its essential limitations through **looking in on itself and deciding when it is unsure of its ability**. In our prototype, self-assessment reveals itself as the mechanism through which the robot may get itself to learn about the world, requiring absolutely minimal external interference.

We will briefly describe the two fundamental building-block modules of our systems, which tie together neural networks in such a way that they possess this self-assessing ability. The lowest level is called MINOS and consists of a neural network that can learn and a *monitor*, which is also adaptive. As the neural network adapts the monitor also adapts, and according to the progress of the neural network and/or the frequency of its being exposed to a particular learning sequence, generates a varying *confidence* that this module has in itself with respect to the current task. In order to extend the limited knowledge that a single MINOSmay confidently learn, we collect together a number of them, each of which is confident in different aspects of a task, although unconfident in many, and string them together in an AUTHORITY. The AUTHORITY selects the MINOSthat is most confident (and, incidentally, gets the most appropriate MINOSto learn new knowledge). Note that the AUTHORITY inherits a confidence from the MINOSs it is comprised of. In a sense, our modules are like **holons** in the sense used by Koestler [3], in that they too are small JANUS's that form part of the whole, but function independently. The same applies for higher levels in the module hierarchy. For more details see [7].

5 The Dual Arm, Dual Eye prototype system

Our prototype system has two arms and two eyes. Initially we constrain its arms to move on a 2-D plane, known as the "table". The eyes are a fixed distance raised above and back from the table, so that the arms may be seen the whole time. Each of the (initially 2: shoulder and elbow) arm joints emits a ray of light, and they are the only evidence for the robot that it has arms.

There is no communication of muscle stresses from the arms back to the brain, nor is there tactile information from the arms. Thus the only clues the robot has of the existence of these arms is through observation of the effect they have on the (visual) environment, when certain muscle changing commands are enacted. The prototype must *discover* for itself, through self-observation and learning, what the dots on its retina mean and how they represent controllable tools in the environment.

The left side of the brain directly controls the right arm and right eye motion, and the right side the left arm and left eye motion. Most of the left visual field input comes via the right side of the brain, and most of the right visual input via the left side of the brain.

At birth the eyes already move together in the large non-focusing motions, but not in the stereoscopic fusion movements. The eyes may move up/down and right/left (simultaneously).

The system must also learn about the extensive nature of its arms, and thus avoid collisions between them, and also with objects that may be in the environment.

6 Training the prototype

The prototype should reach a competent standard through stepwise learning. The learning experiments are based on the classical experiments by Piaget on infants and by Held on baby monkeys. Such experiments confirmed the hypothesis that learning of motor control requires sensoric (in our case visual) guidance. Assuming the necessary visual learning has been mastered (we do not go into this here), we plan the following steps for the building-up of the motor abilities.

1. *Random movement*
Muscular excitations are generated in the **random excitation generator** and led into the L/R arm direction predictor, generator and arm movement effector modules. The current configuration of the arm system is observed visually. On the completion of the random arm movement the new configuration of the arm is determined visually and the actual **direction** moved by the arm is determined (in the hemispheres) and this information is directed back to

the prediction modules. Here it is learnt to associate the initial configuration and the random excitation with the actual observed direction moved, and the initial configuration and the observed direction with the randomly generated excitations.

2. *Randomly generated directional movement*
Similar to stage 1, but a random goal direction is chosen, and motion is not merely predicted, but intentionally directed, using the prediction ability to "imagine" the result of a movement.

3. *Reaching a goal—Fitts' law*
When the above two stages have been well learnt, the robot will confidently be able to move in particular directions, given a particular random starting arm configuration. The process need not be learnt perfectly in all possible starting configurations. In the third stage there is a randomly generated goal on the table, and the arm starts off in a randomly chosen, but comfortable, position. The arm must reach the object. The visual system and the hemisphere involved in the reasoning are used to determine the direction of the goal from the end of the arm at any point in the process. Initially the process takes a while to be completed. On repetition however, starting from exactly the same initial arm configuration, the process should take a shorter and shorter time. With a number of such tasks it is hoped to produce a Fitts' law-like learning curve.

4. *Motor program learning*
Next are learnt typical complete arm motions such as circular or figure-of-eight. In this stage we develop the fundamental motor program modules in the Cerebellum.

5. *Movement of two arms together*
At this point we need to communicate from the environment the reinforcement signal: collision. The robot should learn to recognize the extensive nature of its arms, and move them (randomly) such that they do not collide.

6. *Goal-directed movement of the two arms*
Now the arms should be controlled by both hemispheres working together to achieve coordinated (but simple) movement. For instance, with the end of the arms touching, a point should be reached.

7. *Manipulation of objects on the table*
Now it is time for the robot to recognize foreign objects in its environment: initially point objects, it will learn to move them around the table and observe the change to the environment. Later extensive obstacles will be present, and the arms will have to work (together) to manipulate objects around them in order to reach a goal.

Further empirical laws of motor behaviour that we intend to employ, like Dejong's log-log linear learning law, and motor learning aspects that are in general relevant to our work, may be found in [2].

7 Specific medical aims with the prototype system

We have two basic aims with this system. First, we intend to develop the prototype visuo-manual system so that it performs as suggested: finding out about the environment and supervising its own learning, reproducing the appropriate motor learning laws observed in mammals. Having succeeding in this respect, we have in our possession a skillful neural network based modular robot, with a two-hemisphere adaptable brain. This we will then use to investigate various aspects of brain research. Our first example is that of rehabilitation. Rehabilitation studies [6] have shown how, after damage of the cortex on one side of the brain makes conscious limb movement for the affected side impossible, although the motor cortex is undamaged, it is possible slowly to relearn the control movements for this side through running the limbs controlled by the other side through the motions. Such information is interhemispherical training, conducted either over the Corpus Callosum connection or through the Cerebellum. We intend to perform such studies using the prototype's arm movements capability.

We will also investigate the process of recovery from various forms of localized damage that we can impose on the trained brain. Since, as mentioned, each level of our system is such that it possesses the JANUS inward-looking nature, and additionally consists of *distributed knowledge*, in the form of individually (but not globally) expendible MINOSs, we expect to be able to recover from localized removal or destruction of brain matter, although the resulting system will have a lower performance than before. This lower performance, however, will be recognized by the system itself, and is, significantly, a result of deficiencies in less important knowledge, rather than complete fallout of knowledge seen to by the damaged modules. We hope to determine whether adaption of remaining ability in combination with other usable parts of other modules proves itself to be better than complete relearning with the smaller number of usable modules now available.

Starting with the reproduction of the various human motor laws mentioned in the last section for single arm movement, we will also progress to such experiments touching the question of when conscious learning can be transferred to an automatic unconscious skill, and indeed when learning can proceed unconsciously with the higher supervision ("consciousness") of the system not imposing learning instructions.

8 Conclusion

JANUS is an attempt at modelling biological brains not at the lowest neuronal level, but at the module level. The modules are constructed in such a way that they allow complex problem-solving strategies to arise through harmonizing many conflicting actions and sub-strategies. In JANUS the system uses actively the interaction between more analytical problem solving and more synthetical problem solving. The system tries to improve continuously using its "reflective" architecture.

The JANUS implementation and simulation is made possible by a very powerful tool: the object-oriented modular simulator SESAMe [11]. Our initial experiments with such systems of neural modules have been very promising.

References

[1] J.C. Eccles. *The Human Mystery.* Springer Press, 1979.

[2] W. Iba. Human motor behaviour: A short review of phenomena, theories and systems. Technical Report 89-34, Department of Information and Computer Science, University of California, October 1989.

[3] A. Koestler. *The Ghost in the Machine.* Picador, London, 1976.

[4] M. Minsky and S. Papert. *Perceptrons.* MIT Press, 1988. See particularly the Epilogue.

[5] H. Mühlenbein. Limitations of multilayer perceptrons – steps towards genetic neural networks. *Parallel Computing*, 14(3):249–260, 1990.

[6] H.W. Schumacher, H. Wassmann, and D.B. Linke. Plasticity of the brain after extensive cerebral infarction: Report of a case and discussion. In R. Wüllenweber, M. Klinger, and M. Brock, editors, *Advances in Neurosurgery, Vol. 15*, pages 302–305. Springer-Verlag (Berlin–Heidelberg), 1987.

[7] F. J. Śmieja. Multiple network systems (MINOS) modules: Task division and module discrimination. In *Proceedings of the 8th AISB conference on Artificial Intelligence, Leeds, 16–19 April, 1991*, 1991. Also available as GMD technical report 638.

[8] F. J. Śmieja and H. Mühlenbein. The geometry of multilayer perceptron solutions. *Parallel Computing*, 14:261–275, 1990.

[9] S. Springer and G. Deutsch. *Left brain, right brain.* Freeman (New York), 1985.

[10] G. Tesauro and B. Janssens. Scaling relationships in backpropagation learning. *Complex Systems*, 2:39–44, 1988.

[11] C. Tietz, P. Hendricks, A. Linden, and H. Mühlenbein. Object-oriented simulation of complex neural architectures on parallel computers. In T. Kohonen, editor, *Proceedings of COGNITIVA 90*, pages 387–400, Paris, 1990. AFCET.