

Constructing the Relational Mind

John G. Taylor

Centre for Neural Networks
Department of Mathematics
King's College, Strand
London, WC2R 2LS
U.K.

also at

Institute for Medicine
Research Centre Juelich
Juelich, D-52425
GERMANY

john.g.taylor@kcl.ac.uk

Copyright (c) John Taylor 1998

PSYCHE, 4(10), June 1998

<http://psyche.cs.monash.edu.au/v4/psyche-4-10-taylor.html>

KEYWORDS: Consciousness, qualia, relational mind, neural networks, semantic memory, cortical bubbles of activity, word awareness.

ABSTRACT: The "relational mind" approach to the inner content of consciousness is developed in terms of various control structures and processing strategies and their possible neurobiological identifications in brain sites. This leads naturally to a division of consciousness into a passive and an active part. A global control structure for the "single strand" aspect of consciousness is proposed as the thalamo-nucleus reticularis thalami-cortex coupled system, which is related to experimental data on the electrical stimulation of awareness. Local control, in terms of excitatory transfer from pre-processing sites to posterior working memory regions, is supported by data on subliminal perception timing and disambiguation of poorly defined percepts. The inner content of consciousness is understood as arising from the resulting relational features between inputs and stored pre-processing and episodic memories. Strong analogies are drawn between emergent properties of the model and suggested properties of "raw feels", supporting the thesis that working memories are the initial sites for the emergence of phenomenal awareness, and the frontal lobes for its further adumbration in terms of higher cognitive processing, including the creation of self.

1. Introduction

There is presently considerable activity in trying to understand the nature of consciousness. One of the deepest questions is that of determining "What is it like to be X?", where X is any being, other than oneself, having the potential for consciousness (Nagel, 1974). It is this question which has been claimed to be impossible for science to answer (Searle, 1991), and has led to the useful distinction between "hard" and "easy" questions about consciousness (Chalmers, 1996). The latter are those about the brain mechanisms which appear to support consciousness and which are being ever more clearly discerned via the recent developments in psychology, neuroscience and pharmacology. Non-invasive brain imaging, lesion studies and single and multi-electrode studies on animals and humans have yielded an enormous wealth of knowledge about such mechanisms. The harder sciences of engineering, physics and mathematics are now making important contributions to a theoretical framework for the various information strategies used at both the local (single neuron) and global scales.

In particular, the non-invasive instruments of PET, fMRI, MEG and EEG are now producing results of deep import for the networks of neural modules supporting a broad range of psychological tasks. Such coupled networks have been discovered for visual word processing, attention and motor sequence control, to name but a few tasks (Posner and Raichle, 1994, 1995). The increasing information about these networks raises the prospects for constructing a global model of the brain as a multi-modular system (Taylor, 1995a). The philosophy behind such an approach is to start with simple neurons and modules, and develop increasing complexity thereafter. The task itself clearly requires a multi-disciplinary approach. It should lead to a framework from which to tackle the easy problems of consciousness from a global viewpoint.

It is my purpose here to present a possible physical underpinning for consciousness. We will consider how this suggests a solution to the hard problem of consciousness after a presentation of the physical basis of the model. That will initially be at a level of specifying a "boxes"-type of global information flow in the brain. Parts of the model will then be made more specific in terms of putative neural structures able to perform the functions specified in the boxes model. Some supporting evidence for these structures and their function, as well as for the more global program being carried out, will be brought forward from neurophysiology and psychology.

This general model will then be compared briefly with the recent ones of Gray (1995), Baars (1988), Edelman (1989), Block (1995), Dennett (1991) and Aleksander (1996), so as to clarify the nature of the proposed model in relation to those already known to the reader. On the basis of the model, and supporting evidence for it, we will consider a tentative solution to the hard problem of describing "what it is like to be X" at the end of the paper, together with a reason for the emergence of consciousness as an essential and irreducible part of the system.

The main thrust of the answer to be developed here is that mind emerges through the relations between brain activities. After sensory input is initially coded by pre-processing

networks at a pre-conscious level it emerges into phenomenal awareness by way of a competitive process in regions of cortex in which there is a longer persistence of neural activity. This persistent activity, termed "bubbles" of cortical activity, can provide a context to later inputs and so guide their interpretation in terms of which higher level encoding will occur. The contextual information provided thereby gives the initial relational character to phenomenal awareness. As will be shown later, the detailed form of the competition and later use of the winners (which are decaying away) provide properties seen to be closely analogous to those of qualia, such as ineffability, transparency and presence among others. The possibility of bridging the mental/physical gap thus appears through competitive processes involving the creation and suitable later deployment of cortical bubbles of neural activity. Further relational structure involves the use of long-term memory, producing the "filling out" of initially low-content phenomenal experience.

2. The Relational Mind

2.1 Introduction to the Model

Relational structures have long been recognized as an integral part of brain and mind. Aristotle proposed over two millennia ago that thinking proceeds by the basic relations of contiguity, similarity and opposites, an idea developed strongly by the associationist schools of psychology in the 18th and 19th centuries. The manner in which ideas led one into another was considered seriously by the empirical philosophers Locke, Berkeley and Hume in the 17th century. We will try to put more detail into the phrase of Hume's that "mind is nothing but a heap or collection of different perceptions, unified together by certain relations..." (Hume, 1896). In particular, we will try to specify those relations in greater detail, so as to move towards justifying the basic thesis of the Relational Mind model (Taylor, 1973, 1991) that: the conscious content of a mental experience is determined by the evocation and intermingling of suitable past memories evoked (usually unconsciously) by the input giving rise to that experience.

This model goes beyond that of Hume in that the Humean relations between perceptions are now extended to include a range of past experiences entering into relation with a present one; these past experiences need not have been conscious originally nor as they are evoked to enter the relation.

The basic idea of the Relational Mind Model is that consciousness arises due to the activation by on-going brain activity, stemming from external inputs in various modalities, of somewhat similar past activity stored in pre-processing/semantic, episodic and value memories. The mental content of an experience is postulated to consist of the set of these stimulated, related activations along with the input; in this way consciousness can be said to consist of relations of an input to stored memories of relevant past experiences. Thus, the conscious experience of the blue of the sky, as seen now, is determined by the stored memories of one's past experiences of blue skies, say on

hillsides, at the seaside stretched flat on the sands, or in one's garden sitting in a chair looking at the sky.

It is these past memories which are supposed to "fill out", and make sense of, the input, whatever it consists of. They allow for more efficient response to the input, since by means of the related activities it has excited, the input has acquired a degree of familiarity. These past memories need not surface as conscious experiences in themselves, but are regarded as acting as constraints or guides to further experience. On the other hand, they may become conscious in their own right, so that one is reminded of past experiences as one lives through the stream of one's own conscious thoughts. These are constantly surfacing into consciousness and then disappearing again, as if continually flowing along just below the surface, only to erupt occasionally when the memory becomes salient. They are also in constant competition with each other, new inputs and their related memories fighting to become the new sources of action or of higher cognitive processing such as reasoning or planning.

But why is this Relational Structure able to create mind at all? How does it ever cross the "explanatory gap" of Levine (1983) and others? That question will be addressed towards the end of the paper. In particular, it will be shown that activity emerging from pre-processing/semantic modules onto so-called working memory nets, that can hold activity over a sufficiently long time for there to be traces of earlier activity to be involved with the new arrival, has properties very similar to those of phenomenal consciousness. However before we can reach that point we will first make the Relational Mind model more specific, in the form of a set of principles. Some general experimental support will then be given to the model in the following subsection before more detailed analysis is developed in the following sections.

2.2 Basics of the Relational Mind Model

The basic assumptions of the model are:

Assumption 1. Consciousness needs time to emerge, so that buffering of input activity in working memory modules is essential.

Assumption 2. The final emergence of consciousness occurs by means of two sorts of competitive process:

- a) local
- b) global

Assumption 3. The contents of consciousness arise from the activation of past memories (from long-term memory) or by means of continuing activity traces (on working memories).

This leads to the need for three different sorts of processors:

- (a) The Relational Processing system, in which preprocessed/semantically coded inputs activate related episodic memories.
- (b) The "Conscious I" global command system, in which global competition is carried out between activities on the various working memory systems.
- (c) "Local Breakthrough" systems, in which local intramodular competition occurs between nodes coding for different interpretations of the same input from earlier semantically coded input, as described in the preceding sections. Bubbles of neural activity are created as part of this emergence; these bubbles possess the (non-personal) properties of qualia outlined in Table 2.

From the above assumptions and the related processing systems a set of more specific principles arise. I present these principles at this early stage to make the structure of the model apparent from the beginning and to help motivate some of the considerations over the succeeding sections. These will develop neural systems to implement various of the principles, and they will then be collected together, along with further specific experimental supporting data, in the detailed discussion shortly.

The Relational Mind Principles are as follows:

- P1. For each input code A there exists a related pair of (possibly, sets of) pre-processing and working memory modules P_Ma, W_Ma.
- P2. There is a local competition, on a given W_Ma, between neural activities representing different interpretations of inputs in the preceding second or so.
- P3. A global competition is run between the activities on different working memories, the winner gaining access to consciousness.
- P4. Feedback from episodic memory is involved in the competition between the different PM/W_M pairs.
- P5. Upgrading of episodic memories occurs from the output of the winning working memory, denoted W_Mwin.
- P6. The activity in W_Mwin is stored in hippocampus and nearby sites as a buffer, to be usable (possibly non-consciously) at later times.
- P7. PMwin/W_Mwin activity guides thinking and planning sequences.
- P8. The activities in the non-winning PM's can still be used for automatic, non-conscious level processing, so this non-winning activity itself is not vetoed completely. On its own, the winning W_M activity is not recallable from long term memory.
- P9. Upgrading of memory, either of declarative form in the hippocampal complex or in semantic memories or non-declarative form in frontal-lobe-basal ganglia thalamic feedback circuits, can occur continuously as well as more specifically off-line (during sleep).

- P10: Report arises in different modalities from the output of WMwin and may have increasingly imperfect transmission as the level of the signal decreases.

These are the detailed principles of the model for which supporting neural modules, and related experimental data, will be discussed, at various levels of specificity, after some initial general comments.

2.3 Exploring the Model

The Relational Mind model was developed in a formal manner initially in (Taylor, 1973), then (Taylor, 1991), and more explicit neural underpinning has been given to it in more recent papers (Taylor, 1992a, 1992b, 1992c, 1993a, 1993b, 1994, 1995a, 1995b, 1996a, 1996b; Taylor and Villa, 1997). There are other models of the mind with considerable similarity to the Relational Mind; the detailed correspondence between them will be discussed later.

What then does the Relational Mind model suggest as the neural underpinning of consciousness? There are two operations being performed on inputs in the model:

- a) *evocation*
- b) *intermingling*

Evocation is the activation of earlier memories, which might themselves be of a variety of forms. For words, semantic memories are crucially involved. For vision, preprocessing is by primary and later visual areas: feature analysis occurs in the temporal ("what") and parietal ("where") streams, with the codes of shape, texture, colour, motion arising from the various areas involved.

Episodic or autobiographic memories (with the presence of "I") are also expected to be of importance, being used to give a more personal and social aspect to conscious content. Affective aspects of inputs are also expected to be relevant, imparting an emotional addition to an otherwise purely cognitive conscious content and arising from stored memories in limbic circuits, more especially the amygdala and mesial frontal cortex.

The process of intermingling requires some inputs to be singled out over others (for further, higher level processing) and requires the resolution of ambiguity; an important proportion of words and visual scenes contain ambiguities. Thus, the word "palm" has at least two meanings; which is to be chosen in a sentence is usually determined by the context. Similarly, in vision a Necker cube presents two possible percepts, with the lower side either facing toward or away from the viewer; which is chosen may depend on eye movements and adaptation.

A natural process for the reduction of ambiguity and of total input for later processing is competition between alternatives. Such a competition takes time to be completed and

would employ context to help reduce ambiguity. Thus, a buffer or "working memory" structure would appear to be needed, to hold traces of earlier inputs and also to allow time for the competition to be run; such competition would require inhibitory connections between the competitors, which might be in distant regions of the cortex.

Inputs are posited to be encoded at suitably high pre-processed level in a given modality or code. This pre-processed input is then sent to its appropriate working memory, where it competes with other activities already there as well as with those on other working memories in other codes and modalities. There may also be augmentation of the activities on these working memories by suitable episodic and value memories. Consciousness then arises as the continued activation of a winner, on a particular working memory, for the time during which it has won the competition against other such memory sites.

2.4 Support for the Model

In this section it is our purpose to flesh out more fully the Relational Mind model of passive or posterior consciousness (Cp) outlined briefly in the second section. Let us turn to the principles enunciated in section 2.2 for the Relational Mind approach to Cp, and discuss the evidence for them. They are:

P1. For each input code A there exists a related pair of (possibly sets of) pre-processing and working memory modules PMA, WMA.

By "code" is meant here the output of possibly several levels of non-conscious analysers acting on primary sensory input in a given modality. The highest levels of this hierarchy are denoted PMA. For vision the codes are: texture, shape, motion, position, colour, and possibly others. The coding level may be of primitives, such as for colour and motion. It may also be of a high-level alphabet, as suggested by the "geons" of Biederman (1987). These may arise from pre-processing by successive higher-order feature detectors created by the learning of AND gates. In addition, there is processing up to the level of phonemes and even few-syllable words on the semantic module.

There is experimental support of two kinds for the above principle. First, there is psychological evidence of a broad range (Baddeley, 1986, 1993) for working memories in various codes: the phonological store for phonemes, the visuo-spatial sketch-pad for spatial and shape inputs, and the "body-matrix" for somato-sensory inputs (Jeannerod, 1994). Second, there is neurological evidence from PET studies (Paulesu, Frith and Frackowiak, 1993) for the existence of localized cortical auditory working, and also from other PET studies of visual memories. There are numerous modules involved, as is to be expected if both semantic and continued working memory activity is being measured.

P2. There is competition, on a given WMA, between neural activities representing different interpretations of inputs in the preceding second or so.

Support for this principle has been presented in the previous sections. The intra-modular inhibition posited there may indeed be augmented by, or even entirely replaced by, sub-cortical effects arising from lateral inhibition on the nucleus reticularis thalami (NRT), acting over sufficient distances to explain the Libet data to be discussed later. Such sub-cortical inhibition has been used recently (May et al., 1998) to obtain good agreement with experimental data on the mis-match negativity observed in processing deviant frequencies in a train of standards, interpreted as due to the competition between the traces of activity on nodes activated for the standards in opposition to those for the deviant.

Further, less direct evidence for intra-modular competition can be gleaned from the phenomena of perceptual alteration, such as for the Necker cube. An analysis, by means of two neural modules, one coding for each of the alternative interpretations, has been performed by Riani and Masulli (1990), achieving good agreement with observed alternation times. The model fits P2 if the two mutually competing modules of Riani and Masulli (1990) are regarded as parts of the visuo-spatial sketch-pad of Baddeley (1986). This has been extended to the model of various recent data on perceptual rivalry in (Taylor, 1996c).

P3 A competition is run between the activities on different working memories, the winner gaining access to consciousness.

This principle will be given some support using the data of Libet and his colleagues (1964, 1979, 1991). There is also support for it from recent experiments by Baddeley and his colleagues (Baddeley, 1993) on intrusive thoughts. The understanding and control of these is important for depressed patients, with such thoughts often being of the patient's inadequacy, forming one of the principle complaints of the patient. There are two sorts of such stimulus-independent thoughts: sequential and coherent, and those of a more fragmented sort. Irregularly asking subjects, sitting silently in a quiet room, what their thoughts consisted of, led to the conclusion that about 80% of the time they did have an intrusive thought, with about 65% involving sequential thoughts and the other 15% fragmented ones. Performing distracting tasks, such as listening to and repeating back a 5 digit sequence presented one a second, led to considerable reduction of the coherent thoughts (to about 15%), but little change in the level of fragmentary thought production. The relevance of these experiments to the present discussion derives from an additional fact reported in Baddeley (1993). Subjects who were repeating back the 5 digits were divided into two groups: those who reported being aware of the digits during the digit span task and those who were unaware. The aware subjects had intrusive thoughts in less than 10% of the instances they were asked. On the other hand, the subjects able to perform the digit span task without being aware of so doing reported having stimulus-independent thoughts almost 60% of the time probed. This is in comparison to the 80% or so times such thoughts were reported by silent control subjects.

The important feature of this result is the crucial role played by awareness of the digits in considerably reducing the level of intrusive thoughts. Once consciousness of the digits occurred, there was a six-fold reduction in intrusive thoughts compared to subjects when

no other task was performed, or when the reporting was automatic. These data support the model of this process composed of two PM/WM pairs, each coupled to an episodic memory store E and to the TH-NRT-C system (i.e., thalamic-NTR-cortical system) for the global competition. The model suggests that the intrusive thoughts are produced from E, based in part on past unpleasant memories and on low self-esteem (where a value memory is included in E). The process of generation may then be due to incorrect levels of neuromodulators, such as may be at the source of the voices heard in schizophrenia. There may also be strong self-excitatory circuitry involved.

The crucial question here is how the intrusive thoughts emerge into consciousness. With reduced stimulus load entering various pre-processing memories PMa, PMb (where "a" might denote auditory and "b" visual codes), the internally generated fragments or sequences from the episodic memory E will have more importance in gaining control of the nodes on WMa or WMb. There is expected to be a reduction of the level of intrusive thoughts when one of the WMs, say WMa, has won the suggested competition, by P3, as in the case of awareness of the digits of the distracter task. Such reduction is not expected to occur when the digits are processed automatically, as can occur for input to PMA passing directly to the output without arousing awareness of it. This is consistent with the results in Baddeley (1993). The reduction by about 20% of intrusive thoughts in the automatic processing case may be explained by a general level of inhibition by feedback from the digit response outputs to the episodic memory and related systems. A further relevant experimental result was that simultaneous use of the visuo-spatial sketchpad by tapping a set of keys in a specific pattern led to a reduction of the intrusive thoughts from 80% to 30%. This result also supports P3 and the occurrence of global competition.

P4. Feedback from episodic memory is involved in the competition between the different PM/WM pairs.

This principle is basic to the Relational Mind model. There is some support for P4 from the analysis of intrusive thoughts discussed above. Stronger and more direct support for P4 comes from recent work of Roediger and McDermott (1994). They investigated the creation of false memories, in which previously experienced material was used to cause a subject to report falsely exposure to related, but unexperienced, material. This can be seen as involving reconstructive memory, involving active filling-in of missing elements whilst remembering, in contrast to accurate rote reproduction of material from memory.

In the experiments of Roediger and McDermott (1994) subjects heard and recalled lists of 12 words, each related to the same "lure". A 42-item recognition list was then used, consisting of 12 studied and 30 non-studied items. This list was presented several minutes after a set of 6 such lists of 12 items (each list entirely related to different words) had been read aloud to the subjects, and each list tested for free recall immediately after its presentation. The most important result was that 40% of the time the associated non-presented word (the lure) for a list was recalled. This is to be compared with the intrusion of any other word occurring only 14% of the time, so the subjects were not guessing wildly. On the recognition list the subjects were "sure" that the critical non-studied items had been on the originally presented lists over half of the time.

The basic question here concerns the phenomenal experience of the subject, in which conscious "recollection" of a supposed episodic memory occurred (likely brought about by activation of implicit associations between the list items and the associated word). There was a higher chance of the critical item being remembered if it had been produced on the immediate free recall test compared to no such production (probabilities of 0.58 versus 0.38 respectively). This result can be interpreted as indicating that there is a 20% chance of the critical non-studied word being produced at free recall in an implicit manner and then being laid down in episodic memory so as to be used at the recognition test as a basis for remembering. The experiences of both the conscious and the implicit incorrect recollections of associated words, the lures, either in the preprocessing module or in episodic memory, were then used to modify the response, and thereby the conscious experience, of the subjects during testing on the recognition lists. The above interpretation of the cases of conscious remembering are in support of P4, through the involvement of episodic memory which had only very recently been created (at free recall). Furthermore, this phenomenon is in support of the principle:

P5 Upgrading of episodic memories occurs from the output of the winning working memory, denoted WMwin.

This principle underlines the manner in which the episodic and working memories are proposed as having on-line excitatory feedback interaction. Further principles can be introduced, such as:

P6. The activity in WMwin is stored in hippocampus and nearby sites as a buffer, to be usable (possibly non-consciously) at later times.

P7. PMwin/WMwin activity guides thinking and planning sequences.

P8. The activities in the non-winning PM's can still be used for automatic, non-conscious level processing, so this non-winning activity itself is not vetoed completely. On its own, the winning WM activity is not recallable from long term memory.

This feature is related to the occurrence of apparently automatic processing, such as while driving, with no later memory of the inputs being experienced earlier. However, there appears to be phenomenal awareness of the events and memory of them held for some 3 seconds later. It would appear that there was no suitable backup of the posterior winning working memory activity by its pre-processed level of related activity, which is needed to be stored in long term memory in order give recallable material at a significantly later time (well after the short term memory has decayed away).

P9. Upgrading of memory, either of declarative form in the hippocampal complex or in semantic memories or non-declarative form in frontal-lobe-basal ganglia thalamic feedback circuits, can occur continuously as well as more specifically off-line (during sleep).

A further set of experimental results supportive of the above principles has been obtained by Marcel (1993) in his probing of "slippage in the unity of consciousness". His experiments showed how very noisy or low level inputs may be reported quite differently by different modalities. The resulting dissociation of knowledge from awareness seems similar to that occurring in blindsight. In the experiments, Marcel (1993) asked his subjects to guess the presence or the absence of a dim light. The level of the illumination was such that for only about 50% of the trials when the light was on, in a prior trial, did the subjects guess that it was. The main tests involved sets of trials for which the light was on for only 50% of the time, and to which the subjects were asked to respond in three different ways if they "felt" that a light had indeed come on: by blinking with the right eye, by pressing the button beneath the right forefinger, and by saying "yes". When the subjects were urged to report a conscious sensation of the light "as fast as possible but as accurately as possible" results were obtained showing a simultaneous dissociation. That is, on the same trial there might be a "Yes" response with one modality of response, but not with the others. The number of successes in a later set of trials was 77.5%, 67.5%, and 57.5% respectively for the different modalities, showing significantly different success levels. Abandoning the assumption of a single unitary consciousness, Marcel (1993) concludes "we have to say that either the phenomenal experience of the target light is multiply represented or that separate reflexive consciousnesses have differential access to a single experience."

The Relational Mind model would suggest the need for broadcasting the output of WMwin, in this case expected to be the visuo-spatial sketch pad, to the other WM sites. For example, the input may need to be communicated to the phonological loop. It may be that activity on WMwin, as also buffered in the hippocampus according to P7, may also be used to help that process. The model allows for noise in the transmission between WM modules, and indeed such noise is to be expected; this fits in well with the results of Marcel. Moreover, the different reaction times also fit the model well. Thus, the fastest response is in the visual with the eye at 315 msec, with the finger response (at 580 msec) and the spoken response (850 msec) lagging behind. Such a staggering of responses is also expected, with transmission to the body matrix and then to the phonological loop, say, from the frontal eye field. It is to be expected that if the input were to be in another modality, such as a touch or a sound, then the response error and timing patterns would be altered accordingly. These data give special support to a final principle, as well as to other aspects of the Relational Mind model:

P10: Report arises in different modalities from the output of WMwin, and may have increasingly imperfect transmission as the level of the signal decreases.

Blindsight has been mentioned above, as has support for the Relational Mind from other forms of deficit. I will conclude this section by remarking on two recent reports highly relevant to the Relational Mind model. The first of these is from Weiskrantz and his colleagues (Weiskrantz, Barbur and Sahraie, 1995) on the patient GY, who has lost part of his occipital cortex, with ensuing blindness in the right half of his visual field. In the report cited it was shown that GY possessed awareness of a moving spot of light in his

blind field if it moved fast enough, whereas he only possessed true blindsight for slow speeds. Due to loss of relevant V1 cortex it can be seen that GY's awareness can only be supported by a higher visual region. This is consistent with and supportive of the principle P3. The other report is from Goodale and colleagues (Goodale et al., 1991; Goodale and Milner, 1992), involving, among others, the patient DF suffering from visual form agnosia due to carbon monoxide poisoning. The patient showed strikingly accurate guidance of hand and finger movements towards an object which, however, she was unable to recognize. This, and other cases cited in the report, support the dissociation of posterior consciousness into components supported by different buffer sites of working memory. Thus patient DF had lost the object WM, and so awareness of object form, whilst another patient cited had destruction of spatial WM and found extreme difficulty in picking up objects which she could recognize.

We may conclude that there is a broad range of support for the principles P1 - P10. More detailed models of the various processes noted in section 2.2 will now be analysed in order to give more detailed and quantitative support.

3. The Global Competitive System

3.1 The Global Gate

Consciousness is a unified process, at least in normal experience. The working memory modules known to be sited in parieto-temporal, infero-temporal and frontal cortices (Salmon et al., 1996) are fairly distant from each other on the cortical surface. How can their activities be combined so as to produce the experienced unity of consciousness?

Some form of competition is needed to give this unity. An effective competition between distant modules in the cortex might be achieved by distant excitation of local inhibitory neurons; however, there is little evidence for such long-range lateral inhibition and other mechanisms have been proposed to achieve the required competitive action. Where should we look for suitable neural structures to support it?

One particularly relevant nucleus is the nucleus reticularis thalami (or, NRT for short); it consists of a thin sheet of mutually inhibitory neurons draped over the thalamus. Any axon from cortex to thalamus or vice-versa gives off collaterals to the NRT, so it occupies a strategic position to sample all the traffic between the thalamus and the cortex, and so act as a controlling gate of this traffic. It has even been suggested that "the reticularis gate becomes a mosaic of gatelets, each tied to some specific receptive field zone or species of input" (Schiebel, 1980). Schiebel goes on to suggest that "perhaps here resides the structurofunctional substrate for selective awareness and in the delicacy and complexity of its connections, our source of knowing and of knowing that we know."

Here I posit that the NRT functions, through its intrinsic global connectivity, to allow a long-range competition between various specialized cortical modules. Such a competition

arises from the combination of activity in various thalamic, cortical and NRT regions, and so the total system forms what I will term the TH-NRT-C complex. Loss of NRT action would lead to loss of global features of this competition; if NRT were dissected out, rendered only locally connected or even totally non-functional, then the system would at best function in the local mode suggested by Schiebel, and there would be no unity of consciousness. How serious that would be for a rat, for example, is unclear, but for a human it would be expected to be disastrous. Until such a direct test of the the global competitive function of NRT is made, it is necessary to find more indirect support for the thesis. Such support can be described briefly as: 1) the important position NRT occupies in its interposition between thalamus and cortex (as noted by Schiebel); 2) lesion studies of thalamus show deficits in reaction time (Posner and Petersen, 1990), although the damage to NRT in such lesions is not known; 3) enhanced pulvinar (a thalamic nucleus well-connected to the parietal lobe) activity is observed in attentional tasks (Posner and Petersen, 1990; LaBerge, 1990), and so would involve the NRT from its strong connectivity to the pulvinar (Jones, 1975); 4) a global wave running over the cortex and suggested as being preceded by a thalamo-NRT wave has been observed by MEG techniques in auditory experiments on humans (Llinas and Ribary, 1992). This has been noted to have a strong 40 Hz component, which may be related to the 20-40 Hz membrane oscillation of single nerve cells (Steriade et al., 1991); there has been similar 40 Hz signals observed in visual processing in cats and monkeys (Gray and Singer, 1989).

3.2 The "Conscious I"

A certain amount of spade-work has already been done to tackle the problem of constructing a viable model for the NRT as a competitive net. That lateral competition could be effective in enhancing the maximum of a set of local inputs has been shown by simulation of a simple model (LaBerge et al., 1992), based on the mode of NRT inhibitory action suggested by Steriade, Domitch and Oakson (1986). This explicitly used axon collateral feedback from different cortical areas to achieve a competitive relationship between activities in the different cortical areas. An independent approach (Taylor, 1992a, 1992b) was based on the existence of local inhibitory connectivity on the NRT sheet; this allowed the use of the analogy with the outer plexiform layer (OPL) of the retina, which has dendro-dendritic synapses which are gap junctions between the horizontal cells (Dowling, 1987). These synapses give the OPL a very well connected structure, allowing the efficient spread of local input to its whole surface; NRT should have a similar property. Hence, the epithet "Conscious I" for the model.

A mathematical model of the OPL was developed by Taylor (1990). The model was analysed particularly under the assumption that the neurons of the OPL formed a continuous sheet, so in what is termed the "continuum limit"; in this limit analytic expressions can be obtained for the OPL response. A similar model for the NRT sheet has a similar continuum limit, but now with inhibitory action of neighbouring NRT cells on each other. This action causes oscillatory waves to be produced over the whole NRT

sheet by local input, due to the manner in which local lateral inhibition suppresses activity over a certain distance but allows it to build up again far enough away on the sheet. In this manner a wave pattern is formed on the sheet. Such oscillatory patterns are observed in many systems (Cohen and Murray, 1981; Ermentrout and Cowan, 1978), and there is considerable understanding of them from a mathematical point of view.

Since the NRT is receiving activity from both thalamus and cortex and feeds back inhibitorily to the thalamus, we can see that the NRT can exert global control by means of a global wave of activity on itself over the cortical sheet. This has been shown by simulations of a number of cases (Taylor and Alavi, 1993a, 1993b); the conclusion is that a global wave of activity over NRT can be created by local thalamic input and can exert global control over activity accessing cortex in other, possibly globally distant, sites.

To fit with the MEG data reported earlier the whole anterior pole of NRT must exercise its authority over the more posterior activity, in keeping with the known neuroanatomy of the NRT. Indeed, it is known from neuroanatomy that control over the anterior part of NRT is exercised by external input from the limbic cortex (itself being outside direct NRT control), with arousal from signals coming from the intralaminar nucleus of the thalamus (a set of small nuclei in the thalamus which are now being appreciated as playing an important role in visuospatial awareness, attention and memory by means of their spreading a general arousal signal throughout the cortex, (Bogen, 1995a, 1995b; Purpura and Schiff, 1996)). The limbic cortex involves nervous tissue relevant to emotional drives as well as stored memories, so the general nature of the control on NRT activity will be from those sources.

3.3 Winning Global Control

Interesting results, now thirty years old (Libet et al., 1964) determined the threshold current for conscious experience when a 1 mm diameter stimulating electrode was placed on the post-central gyrus and just conscious experience of what seemed like a localized skin stimulus was reported (as a touch on the back of the hand). The stimulus was delivered as a series of short pulses (around 0.1 or so msec duration). There are two features of the data which stand out, and which can be summarized in two quantitative laws (Taylor, 1995b):

- (a) For threshold current to be consciously experienced over a short duration (less than 0.5 sec), the applied electrical energy (frequency times duration times square of current) must be greater than a critical value;
- (b) For a duration longer than about 0.5 sec the applied electrical power must be large enough to allow the conscious experience to continue.

The requirement of enough applied electrical energy to capture, or turn on, conscious awareness in the short term would seem to fit well with the NRT control structure model

above; such a circuit would be expected to function in terms of electrical energy requirements for winning the dominant mode.

The second result (b) leads to the need for enough injected power to keep the control system going; there will be a certain amount dissipated, and so power above the critical level will have to be injected to hold the control of consciousness achieved by the earlier injected electrical energy. A more detailed mathematical analysis of how the TH-NRT-C system can support the above laws has been given by Taylor (1996b).

In summary, the results of Libet give support to this system as underpinning the emergence of consciousness. Further details of the similarity between the simulated results and the experiences of Libet's patients are reported in the above references. Moreover, the results also support the proposition that consciousness arises from local sites of activity - posterior working memory modules - competing with each other by way of the lateral inhibition acting globally over the NRT sheet. This is especially notable given the need for the electrode activity to persist at least for 0.5 seconds before consciousness was caused in any of Libet's patients.

4. The Local Competitive System: Awareness of Words

We will consider only awareness of words here, due to space considerations, although it is possible to extend this discussion to other modalities (Taylor, 1998a, 1998b).

Psychological (Baddeley, 1986) and neurophysiological (Paulesu et al., 1993; Salmon et al., 1996) support has been given for the existence in Brodmann's area 40 of a localized site for a phonological working memory. This evidence does not prove that consciousness emerges in that (or any other) site, although it is supportive of such a claim when combined with the earlier postulate that consciousness first emerges in sites of working memory. I have previously suggested (Taylor 1992a, 1992b, 1992c, 1993a, 1993b, 1995a, 1996a) that the crucial component needed for persistent input activity to support the emergence of consciousness is short term memory, or more precisely a working memory (WM) of the sort of Baddeley and Hitch (Baddeley and Hitch, 1974; Baddeley, 1986). In a WM there is continued activity of a period of about 1 to 2 seconds before decay occurs without rehearsal. This identification of consciousness with short-term memory is not new, already having been stated by William James in the last century and more precisely in the papers of Atkinson and Shiffrin (1971) and Mandler (1975). It is clear from what has been said so far that the earlier accounts have been extended and remodelled so as to account for the existence of different WMs across modalities, such as the phonological store for phonemes and the visuospatial sketch pad for spatial vision.

In order to explain the model for word awareness, it is necessary to consider experiments which explore the manner in which earlier inputs may affect the response to later ones; manipulation of these earlier inputs, so that they could be either perceived or not, allows for various aspects of the emergence of awareness of the word inputs to be uncovered.

An indication of how such awareness can emerge can be gleaned from the analysis of experiments on subliminally processed lexical decision data (Marcel, 1980). A subject had to make a decision as to the word or non-word character of a string of letters. A sequence of three such strings constituted a given trial; the second letter string of the three strings was manipulated, by masking, so as to be at either a conscious level or a subliminal level or to be completely invisible to the subject.

Table 1
Examples of the different associative relationships
in all-word Triples used in the experiments of Marcel (1980)

Relationship	Examples
Congruent	Hand - Palm - Wrist
Incongruent	Tree - Palm - Wrist
Unbiased	Clock - Palm - Wrist
Separated	Hand - Race - Wrist
Initial	Speed - Race - Wrist
Unassociated	Clock - Race - Wrist

Typical 3-word strings used in the experiment are given in Table 1. When the second word was presented subliminally it was found that the reaction time to the decision task for the third word was affected by the semantic relationship to the earlier words in such a manner that apparently all the possible semantic codes were accessed by the second word. However, these results did not occur at conscious level. Only one meaning of a polysemous word (with several meanings) was accessed when presented at conscious level; it was then found to delay response to a third word of opposite meaning.

A model of this processing was suggested (Taylor, 1996a) in which all possible meanings of a word are accessed in a semantic module at an unconscious level (as emphasized by Marcel, 1980); only in a later working memory site is there competition, so as to prevent activation of contextually incongruous word meanings and their emergence into awareness. Previous activity still held on the working memory site is used to give constraints on the present winner, a competition arising from a small amount of lateral inhibition between nodes coding for words with opposite meanings, giving one or other of these words a competitive edge.

The model was able to give a simple explanation of some of the numerical features of the data of Marcel, in particular of the changes in reaction time to the lexical decision about the third word brought about by the subliminal or conscious activation of the second word. In this interpretation conscious awareness initially arises, at its most elementary

level, from a process of disambiguation using the activation traces of earlier inputs (still present) on the site of working memory to give appropriate context to disambiguate the most recent input.

The success of the model led to the question of how the working memory buffers achieved the assumed long lifetime of its activation trace compared to the much shorter-lived activations in pre-processing modules. That will be considered in the next section, leading to a suggestion of how to bridge the explanatory gap.

5. The Emergence of Qualia

5.1 Identification with Qualia

Let us restate the model presented so far for the emergence of phenomenal consciousness, supposedly sited solely in posterior cortex. The primary awareness of various modalities is suggested to occur after coding at a lower order feature-analysis level. There may be a number of such stages, both serial and in parallel. Consciousness only then arises in those modules with sufficiently long decay of activity (from recurrent excitation) and with lateral inhibitory and excitatory connections for intra-module competition, allowing a single winner to emerge when its activity rises above some criterial level.

How could the neural computations outlined above provide an underpinning for qualitatively different phenomenal awareness? That is the problem of the explanatory gap, to which we will give a tentative answer here. In particular, we will consider how the properties supposedly possessed by qualia, and well summarized in Metzinger (1995), could possibly be assigned to neural activity in the model we are considering.

Table 2
Comparison of properties of qualia with those of working memory activities

Property	Qualia	Working Memory Activity
1. Presence	*persistence *latency *seamlessness	*persistence *latency *rapid change-over
2. Transparency	*can 'look through them' *fully interpreted	*smooth broadcasting to *each other at highest coding level
3. Ineffability	*infinitely distant *intrinsic	*one-way creation *at highest level, no visible back-up

4. Uniqueness	*one winner		
5. Boundedness	*correlation between different codes at different levels		
6. Perspectivalness	*involvement of self modules		

The properties of interest for qualia are presented in Table 2 and will be discussed in turn. The first of these is "presence", involving various temporal features of phenomenal experience. The temporal component of purely phenomenal awareness (with no self component) has three characteristics, as delineated in Table 2: persistency, latency and seamlessness.

The first of these, that of temporal extension of phenomenal experience, is a very important component of the sense of the "subjective now". Such persistence is expected to be an intrinsic part of the process of winning a competition. This is the opposite side of the coin to latency (the delay of the onset of consciousness of an input after it occurs); until another winner arises the earlier one is still "king". In the competitive model latency will arise as the time needed for the competition to be won on a working memory site. There is also the further property that we will call seamlessness: that the transition from one content of awareness to another happens rapidly enough to provide a sense of continuity in awareness.

Both latency and seamlessness were probed by Libet (1964, 1994). In the former paper it was shown that about 500 msec is needed to cause the "artificial" arousal of consciousness by direct electrical stimulation of the somatosensory cortex. Libet (1994) also notes that the change-over process of consciousness was brief, occurring in less than a tenth of the time it took for the total process. That this could occur in a neural system was shown by the stimulation of a simplified version of the TH-NRT-C system, as I reported earlier (Taylor and Alavi, 1995; Taylor, 1996a); seamlessness arose in the competitive model in agreement with the report of Libet.

The next notion of transparency (being able to "look right through phenomenal experience") may be seen as arising from the well-connected system of working memories, for which there must be easy transfer of correlated information once a winner has occurred. Such transfer would also carry with it correlated preconscious material across different codes, if the activity were still available, since there are good connections between modules at the same level of complexity (Mesulam, 1981). Thus, transparency is achievable by sufficiently strong connections between the various components of the heteromodal cortex, giving a "transparent feel" to the activity.

There is also the "fully interpreted" character of qualia (they do not need further processing to be conceptually meaningful). Such a property could arise from the fact that

the working memory sites are at the highest level of the posterior processing hierarchy. Thus, all coding and competition has been performed and all incompatibilities ironed out. No more interpretation needs to be performed; all this has been accomplished as awareness is achieved.

The third property of qualia listed in Table 2 is ineffability (that qualia are impossible to probe and "get behind"). They are atomic or intrinsic and not apparently related to any other aspects of existence. This property is well mirrored in the one-way character of activations on working memory sites, achieved by the use of sufficiently local feedback of neural activity so as to cause the observed long-lifetime of working memory activity in terms of so-called "bubble" formation. Such a bubble is a semi-autonomous state of neural activity (Amari, 1977) which may be used to explain various features of visual and auditory processing (Taylor, 1996b; May et al., 1998). Bubble creation occurs in a manner in which the details of the particular input creating it are lost to the output neural network. This can be seen as a typical feature of the recurrent network style of processing, but is made more effective by bubbles being created due to their autonomy. These bubbles would lose their input identity in detail, and so gain an intrinsic character as experienced for qualia.

The properties of uniqueness and boundedness have been included in Table 2 to make it more complete, but as with perspectivalness they will not be considered here due to lack of space.

5.2 The Explanatory Gap Reconsidered

The results of the comparison between the properties of qualia and working memory activity are clear from Table 2. There is a remarkable correspondence between the two, and it appears reasonable to suppose that there is a valid identification, at least at the level of approximation being used here. More detailed analysis will have to be made to properly justify the proposed identification.

It is relevant to point out that there is a new factor here: we are attempting to make an explicit identification across the brain/mind divide, using specific properties of the activities of each side. Properties of awareness, as seen from the outside or by a third person, are interpreted in a manner allowing them to be identified with those as experienced in the first person. Any serious attempt to bridge the mind/brain gap must propose such an identification; but given the nature of the correspondences argued for, perhaps the gap does not appear as unbridgeable as some have claimed.

The above discussion helps provide an answer to the question as to why consciousness is a necessary part of the activity of the Relational Mind model. For the right hand column of Table 2 summarizes the most important properties of the activity emerging onto the well-connected sites of working memory. These properties are then seen to be very similar to the main properties of qualia. Hence, a system constructed according to the

principles of the Relational Mind Model possesses the attributes of phenomenal consciousness as part of its activity. The function being performed by the model cannot be other than conscious ones. They have been so built as to result in that identification.

6. Active Consciousness

6.1 Dividing Consciousness

The aspect of consciousness described so far, which I denote passive or posterior consciousness, Cp, involves purely the "raw feels" or phenomenal experience. This is to be contrasted with a more anterior and active conscious component, which I denote Ca, involving more strongly the components of introspection and intentionality. The anterior siting of Ca is to be expected since the frontal lobes are strongly supportive of actions and action-set; self and social knowledge also have a strong frontal basis, as is known from deficits in frontal patients.

This leads to the analysis of the total conscious experience as

$$C = (Cp, Ca) \quad (1)$$

Such a separation has already been made by Davidson (1970), who divided conscious experience into a phenomenal part and that associated with beliefs and emotions; the former can roughly be identified with Cp and the latter with Ca. Such a division has also been made by Block (1995), where the labels p and a denote phenomenal and access respectively (although his division does not include self-consciousness and metacognition). The method being developed here, that of attempting to discern the neural regions supporting various informational aspects of consciousness and thereby give such aspects further support, has led to a similar division independently (Taylor, 1995b).

The Relational Mind approach can also be extended to Ca, so that the necessary relations involved in introspection and intentionality are emphasized. Some of these will be carried over from the lower level of conscious content of Cp, gained through its detailed competitive emergence by added semantic and other memory activations. But the extra content in Ca will be expected to arise from the further structures in the frontal lobe. These support self, social memory, prediction systems, drive and goal information. There are also further buffers, termed "active memories", which can hold activations for up to 30 seconds or more. This extended time window increases enormously the relevant relations that may be activated to give content to Ca and allows far greater flexibility in response. Yet the conscious content of Ca can still be considered in terms of the Relational Mind model.

Figure 1 summarises the model being presented:

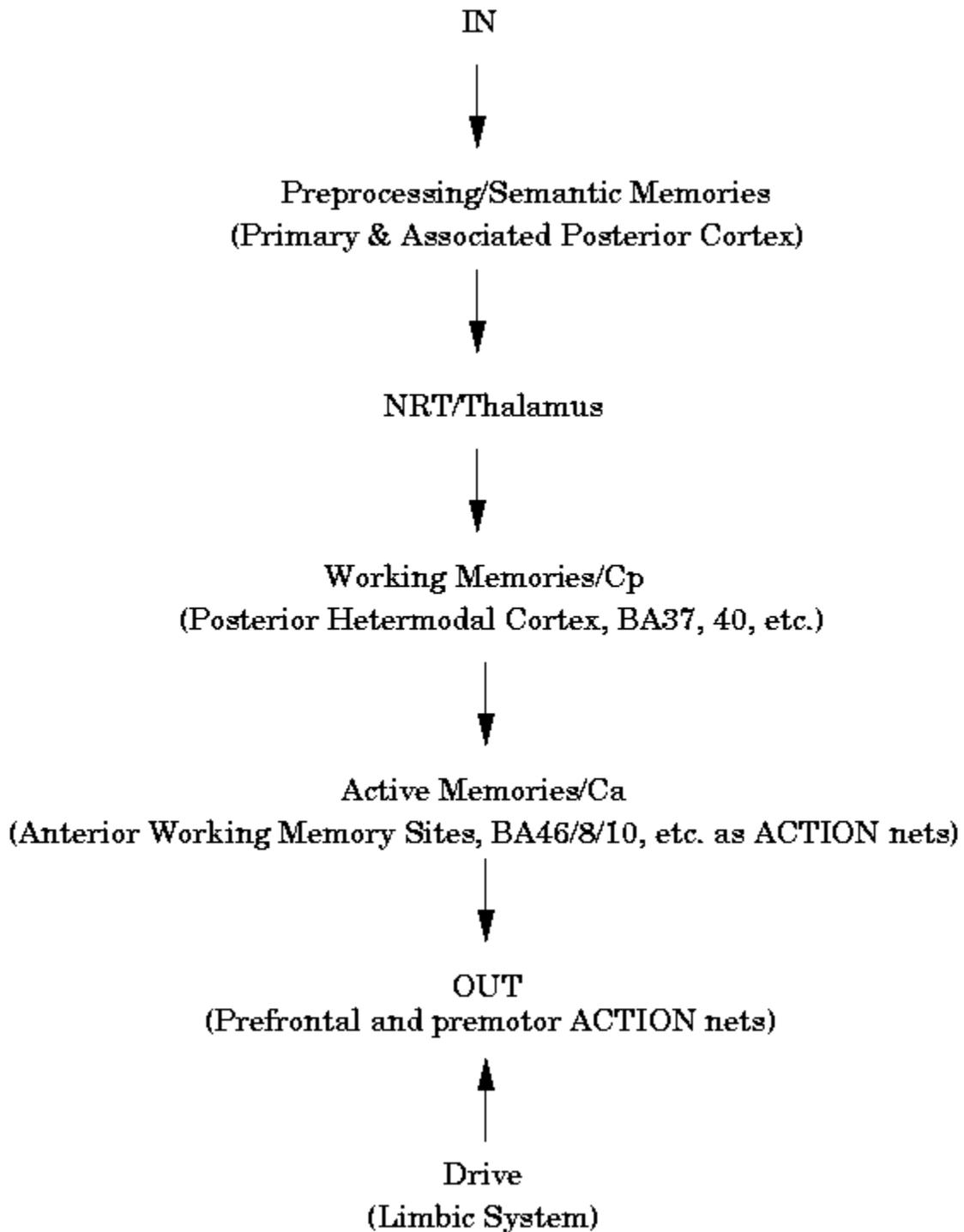


Figure 1

A summary of the structural pathways of the model presented in this paper. Where the symbols Cp and Ca are placed beside the relevant memory modules (working and active memory, respectively) as argued earlier. The arrows denote the direction of flow of activation, which are especially involved in the competition run by the TH-NRT-C system. Output is taken only from the relevant ACTION

network system corresponding to motor response, which could involve purely subcortical responses as well as those controlled by cortical representations. The drive system (specifically the hypothalamus and other portions of the limbic system) has not been discussed in the text, but has been included to indicate the manner in which reasons for actions arise.

6.2 Building Anterior Consciousness

We now attempt to apply the Relational Mind model to anterior or active consciousness, Ca. In order to do so it is necessary to consider the frontal versions of the processes of "evocation" and "intermingling" that would support Ca, and so allow us to build a similar set of principles for Ca as we did for Cp in the previous section.

There is presently no completely accepted indication of the nature of active memory representations. However, strong support for population vector coding in motor and pre/supplementary motor cortices for output responses (coding of the direction of motor actions) has been presented in Georgopolous (1994). This also gives a mechanism for allowing the active memory regions of frontal cortex to be used as "blackboards". Material can be written on these by input, causing the recurrent lateral cortical and recurrent thalamo-cortical loops to tend rapidly to the relevant attractors of the population coding; modulation by basal ganglia will act thereby to modify the thresholds and direct or change the attractors. A neural model of this has been presented in the ACTION network of Taylor (1995) and Alavi and Taylor (1996), which has some similarity to models of Houk and Wise (1993). There are also a further range of neural models of frontal components of working memory (Carpenter and Grossberg, 1993; Dominey and Arbib, 1992, 1994; Guigon et al., 1994; Kirillov et al., 1993; Zipser et al., 1993).

The ACTION Network is composed of the basal ganglia acting in a disinhibitory manner on the thalamo-cortical recurrent loops (with lateral connections as well), so as to cause an attractor to be set up by an external input in the (possibly high) dimensional space given by the input features. This generalizes the two-dimensional action-direction coding of the motor cortex to the other loops of the frontal system. The process of learning the correct connection weights for such an input set, so as to be able to write any material from posterior WM and other memory sites onto the frontal cortex, is presently under analysis (Taylor and Taylor, 1998) and has been considered in some detail by Arbib and Dominey (1994). These connections would allow the active memories to have written on them material from Cp, from autobiographical/episodic memory databases related to this material (as discussed in the previous section) and from goal or other memory buffered in hippocampus.

The process of "intermingling" to achieve Ca is also expected to require competitive processing. That such competition can be supported by the ACTION network was suggested by Alavi and Taylor (1996), where the inhibitory nuclei in basal ganglia (striatum and globus pallidus) may be able to function in a similar manner to that of NRT. The competition would then involve threshold-changing processes, as performed by the disinhibitory action of striatum on thalamus. There are 5 possible ACTION

networks, corresponding to the 5 great frontal loops of Alexander and colleagues (Alexander and Crutcher, 1990) involving motor, premotor and supplementary motor cortex (action sequences), limbic (affect), orbitofrontal (social), frontal eye fields (eye movements) and dorsolateral prefrontal cortex (cognitive/attentional). Each of the cortical regions is expected to have suitable connections for writing on them from posterior and limbic sites; this seems to be the case from neuro-anatomical knowledge.

6.3 Developing Active Consciousness

Having constructed some of the possible neural machinery to support anterior/active consciousness, it has now to be put to work. In particular, it is important to show, though in the space available only very briefly, how the frontal processes of attention, thinking, planning, prediction and schema learning and selection might be achieved. Also, the place of emotions in such a galaxy of functions needs to be clarified. More general aspects of Ca, such as intentionality and introspection, will be considered in the next section.

A number of these processes have already been hinted at in (Taylor, 1995a): attention can be decomposed into the component sub-processes of comparison, activation/inhibition and monitoring (Stuss et al., 1995); all of these can be performed by the ACTION network (Alavi and Taylor, 1996), as can sequence learning and generation, at the basis of schema learning and production. Prediction is also involved in sequence generation, in which hippocampal activity, as in Gray's model (Gray, 1995; Kubota and Gabriel, 1995), should be included. Thinking has been discussed (Taylor, 1995a), using action transformations on representations written on prefrontal cortex so as to make such representations closer, after comparison by the ACTION network, to a goal held in the Papez circuit. Planning can use similar techniques to discover transformations taking a given state to a goal state. Note the difference between thinking and planning; on this view, the former uses whatever transformations that can be constructed so as to reach a goal - it is the sequence of states that are important - whereas the latter emphasizes the set of transformations and not the states.

Emotions will be considered as global activations from the limbic system when concerns (Frijda and Moffat, 1993) - that is, differences between desired and actual states - become large. Such comparisons (as concerns) are related to those of the model of Gray, but are used to give a global signal of affect to the cortical sites of Ca and Cp so as to cause modification of on-going behaviour. This may be crucial in order to change schemata (Frijda and Moffat, 1993), as well as having an important effect in motivation and intelligence.

The decomposition $C=(C_p,C_a)$ would seem to imply that Cp and Ca are independent; that cannot be true, since Ca may well depend for its creation on the output of Cp. It was possible to develop a model for Cp without the need for Ca; in particular, it was earlier

possible to identify activity emerging into Cp on the appropriate WM as possessing most of the properties of qualia.

Ca is dependent upon Cp, since it will be produced, at least in passive situations, by the downloading of activity onto it from Cp. There may be independence when Ca is created through internally generated activity, as in thought or planning. Even then, however, content for Ca has to arise from structures also used by Cp, such as preprocessing/semantic memories, after processing by the associated WMs (and so by Cp) as part of the overall activity. However, strong activation of Cp may not always occur when PMs are being used, especially when creative processes are occurring at a subconscious level. More generally, the interdependence between Ca and Cp is also the reason why it does not seem possible simply to write down the analogues for Ca of the principles P1 to P10. Moreover, these latter will themselves have to be modified to account for this reciprocal relationship.

What is the more precise relationship between Cp and Ca? Are the active memory sites themselves part of the global competition run by the TH-NRT-C complex? Indeed, they seem to be, in so far as all non-limbic thalamic nuclei, including those involved with the recurrent frontal lobe loops, are coupled to NRT in a similar manner to the more posterior nuclei. The frontal cortex may have an edge on the posterior part in view of the waves travelling to the posterior, as observed by Llinas and Ribary (1991). Only the limbic nuclei are partially decoupled from NRT, in that they give collaterals to NRT, but the latter has no feedback to them.

Hence, Cp and Ca are coupled together as part of the TH-NRT-C complex, with emotional signals being downloaded onto them, but with no direct reciprocal coupling (and hence being less tightly cortically controlled). The exact coupling between anterior active memory sites and posterior working memory seems to allow for coupled sites of activity, with Ca being composed of conjoint anterior and posterior activity. Thus, the decomposition (1) denotes a complex interaction between Cp and Ca which could be written (to a first approximation) as

$$C = (Cp, Ca(Cp)) (3)$$

The second component shows dependence on Cp from its downloading directly onto frontal lobe. It will also have additional content from further processing, such as thinking or planning, carried out on frontal sites. Any phenomenal character of Ca can thus arise from that of Cp, and further content arising from the intrinsic character of the frontal transformations themselves. This latter added content will also have phenomenal content due to its arising as part of the global competition on the TH-NRT-C complex discussed earlier. There is added complexity in Ca due to the further ACTION network involvement. Due to reasons of space, that will have to be discussed elsewhere. (See also Bapi, Bugmann, Levine and Taylor, 1997.)

7. The Construction of the Self

7.1 General Comments on the Self

Before it is possible to attack the perspectival aspect of "what it is like to be," the manner in which the self might be created by brain activity must be considered. Here again the Relational Mind approach may be used for guidance, but now with respect to episodic memory, and in particular that part or extension of it termed "autobiographic memory" (Nelson, 1993). That involves issues related to the self, personally relevant goals and personal meanings.

The most influential model of the self is that of the central executive, which has an ancient history (Norman and Shallice, 1986; Shallice, 1988; Stuss and Benson, 1986). This has recently been expressed in information processing terms, which are appropriate to attempts to extend it to the neural arena. The executive is involved in tasks such as "making one's mind up" or in answering such questions as "Do I believe in ...?" or "Do I Like...?" or "How important is that?" The decisions, beliefs, goals, attention to inputs, and feelings and emotions involved in everyday conscious experience are all those concerned with the "I" of the inside world used in the above phrases rather than with the "Them" that is outside and serving as input. We can also consider the important model of the brain developed in Arbib and Hesse (1986), with its basis in schema theory, at both unconscious and conscious brain activity level, for embodying the significant aspect of a situation.

The processing of the outside world is achieved by a monitoring mechanism: the self arises from the brain involved in such outside monitoring. If the outside is developing in a manner agreeing with predictions made by related systems (supporting future prediction, such as automatic action sequences generated by frontal lobe and basal ganglia), then the self need not intervene. If there is a discrepancy, then the self has to become involved in the active processes outlined above of deciding, believing, feeling, realizing goals, etc. Such activity, involving transformations of neural activity from other modules (especially posterior working memories from autobiographic memory in hippocampal and mesial frontal regions), is known from lesion and non-invasive studies to involve the frontal lobes crucially. It is there that the great control loops of motor activity are sited, along with their companion loops for emotional and thinking processes (Alexander, DeLong and Strick, 1986). It is these neural underpinnings which allow a neural model of the self to be constructed. In order finally to approach the very difficult problem of the inner point of view, it might first be useful to consider other features of consciousness which may be understood in terms of the relational model. This will help give confidence in the model, as well as adumbrate certain of its features. Searle presents an interesting list of twelve aspects of consciousness (Searle, 1991), but we will only consider here subjective feeling.

7.2 What it is Like to Be

Subjective feeling presents us with the problem of the inner view. To echo Nagel (1974): "... every subjective phenomenon is essentially connected with a single point of view and it seems inevitable that an objective, physical theory will abandon that point of view." How can the relational mind model approach this supposedly scientifically unrealizable "inner point of view"? As Nagel puts it even more pointedly, the problem is "... I find the hypothesis that a certain brain state should necessarily have a certain subjective character incomprehensible without further explanation." To respond to this, the possible nature of such a point of view must be described from within the model, to see how closely the subjective point of view can be approached.

The point of view of the sentient being X will be taken to be the detailed content of conscious awareness of the being X. This detail, according to the Relational Mind model, is composed of the working and/or active memory activities which have just won the consciousness competition. To that is to be added all the feed-back and simultaneously activated semantic and episodic (autobiographic and value) memory related to the input. Moreover, there are also parallel activities in the other coding, working and active memories, which may be involved in linguistic report (as in the case of the phonological loop). But most especially, the "point of view" is determined by the semantic and episodic memories related to the input in the winning working and/or active memory.

It is possible to explore this view of the subjective character of experience in more detail by considering separately the contributions made by the pre-processing, semantic and the episodic memory feedbacks. The pre-processing and semantic memory content of consciousness will have general culture- and species-specific characteristics. In the semantic case there will be words of the natural language in which the human has been brought up, as well as words of other languages learnt by the person during their schooling or as part of their general life experiences. For animals, such as dogs, similar encoding of the few words with which they are familiar will occur, although there need be no corresponding phonological loop to give conscious experience to words, but only direct output to spatial working memories allowing the conscious experience of such words to acquire a visual form.

Similarly, other modalities will have culture-specific pre-processing memories, such as for shapes in the visuo-spatial sketch pad (Baddeley, 1992; Baddeley and Hitch, 1974). Person-specific encodings in the semantic memories will also occur, such as dialects or particular shapes, although these will usually still be shared with others in a local area. Thus, the pre-processing/semantic encodings will have a certain degree of objectivity associated with them, although that will not be absolute.

There is still an important aspect that has yet to be explored. Nagel (1974) indicates that the "point of view of X" we are after is critically tied to the concepts possessed by X and evoked by the inputs which cause that point of view. The pre-processing/semantic memory has been emphasized above as giving a major part of the "point of view." However, the more detailed nature of the concepts encoded thereby has not been stressed. But, according to Nagel, the nature of the concepts developed and used by a species

appears to be the main entree to the inner point of view. Thus, the nature of pre-processing/semantic memory requires more analysis.

The semantic memory developed by X is based on the sensorimotor response schemata used in experience of the object which evoked the concept (in addition to linguistic encoding, for species with such skills). It is possible on this approach to give an objective analysis of the neural activity excited by a particular object and to describe those sensorimotor acts which would be involved with it (where semantics is here equated with "relevant virtual actions"). It would then be possible, by means of imagining the making of such acts (assuming the species X were not too distant from our own musculature), to begin to imagine what it would be like to be an X experiencing the object in question.

In the case of the bat, it would be the manner in which obstacles were avoided, others of its own kind related to, and prey obtained, which would be the central points of action sequences. The actions in all such cases would be guided by one's radar returns, so that there would be a coupled set of returns and actions taken for the three different sorts of object. Even if no actual avoidance or approach actions were in fact being taken, the "virtual actions" coded, by means of threshold modifications in structures homologous to human basal ganglia, should be able to provide activity which would give a semantic structure, special to the bat, to the input. There is not direct experience of these virtual actions, but they would give a general set of constraints, on developing responses to the objects, which are proposed as corresponding to the "inner feel" that is part of Nagel's quest.

It is through episodic memory content that the particular subjective point of view arises. The personal "colouring" added to consciousness, the subjective "feel" of the experience, following the Relational Mind model, is given by the parallel activation and feedback of memories associated with the present input. Particular people, buildings, or more generally sight, smells, touches and sounds from relevant past events are excited (usually subliminally) and help guide the competition for consciousness of the present input in its working memory. The laying down of these episodic memories, even if they were only sub-consciously experienced, is an indication of the filtering character of the activity of the competition winner. The level of relevance and significance of these past experiences to ongoing activity will not be fixed, but will depend on mood and emotion. These stored memories are energized both by inputs and basic drives, and they guide the continuing cortico-thalamic activity in a top-down manner from limbic structures (Taylor, 1995b).

8. Comparisons with Past Models of Consciousness

As noted earlier it is appropriate to attempt to develop understanding of the Relational Mind model of (2) by comparing it to past models of consciousness. This comparison will be restricted to models which have proved influential in very recent thinking on the subject and which have the greatest relevance to the Relational Mind model. The first of these models is Dennett's "Multiple Drafts model" (Dennett, 1991). This model (in which

there are a set of specialized processors, each of which can be accessed for its contribution to conscious experience at any one time) can be identified with the set of competing working memories in the Relational Mind model, the added value of the model being the manner in which semantic, episodic and other memories are crucial to the ongoing amplification of conscious experience by their activation and feedback, as well as the specific manner in which phenomenal experience is suggested as coming from the creation of semi-autonomous "bubbles" of activity by input.

There is also the separation of consciousness into the two separate components Cp and Ca, which gives further complexity to the nature of such experience. In particular, broadcasting from Cp to Ca provides externally driven content to Ca, in addition to the self and value memory sources of such content arising within Ca. Such an important binary division of consciousness is absent from Dennett's model.

It should not be thought that the dreaded homunculus has smuggled its way back into the system. There is no supreme executive assumed to exist, although certain sites perform so as to allow decisions to be made between competing alternatives.

The Relational Mind model in its simplest form may be seen as an extension of the "global workspace" model of Baars (1988), in which input sensors compete to broadcast their activity to other such and to effectors by the medium of the "global workspace" (GW). The Relational Mind model identifies the global workspace with the interacting set of working/active memory modules, but goes beyond this by emphasizing the need for relational memory structures to amplify and constrain incoming signals.

It is these which are claimed here to give an important component of the inner view, especially the properties seemingly akin to qualia, as reported by Table 2. At the same time, some resemblance to neurophysiological structures in the brain is required to effect the various modules of operation of the GW model. This is provided in the Relational Mind model by the thalamo-NRT-cortex complex (for Cp) and the frontal lobe-basal ganglia-thalamic loops (for Ca). The former was hinted at by Baars (1988), and elsewhere other thalamic structures, as well as layer 1 of cortex, have been considered relevant to the GW model (Newman and Baars, 1993; Newman, Baars and Cho, 1995). However, the specificity of the local buffer sites of emergence of consciousness in our model are an advantage relative to the Global Workspace model.

There is similarity also between the Relational Model and that of Edelman (1989). His emphasis on refference is very similar to the need for feedback from memory structures for guidance and prediction. These themes, however, were not developed by Edelman (1989) using the working/active memory structures discussed earlier; nor using the crucial and unique experimental data of Libet and colleagues (Libet et al., 1964; Libet et al., 1979) and of Marcel (1980), as discussed earlier. Edelman does not consider the themes of global and local competition, nor the resultant features of consciousness, nor again any distinction between Cp and Ca.

The supervisory attentional system (SAS) of Norman and Shallice (1986) has been very influential, especially in guiding experiments in working memory (Baddeley, 1992; Baddeley and Hitch, 1974). However, it has been admitted by one of the proponents that the model has a problem underpinning conscious experience (Shallice, 1988). In particular, a question raised by Horne (1993) is "Who does the contention scheduling for the SAS?" It would seem the buck stops there, whilst in the Relational Mind model the buck is constantly handed from winner to winner.

One important point of contact between the Relational Mind model and the SAS system arises from the distinction between Ca and Cp in the former, for thereby the relational approach to Ca fills in the missing element in the SAS system; this leads to a program for frontal lobe analysis (Bapi, Bugmann, Levine and Taylor, 1997).

There are three recent models of consciousness with which I must conclude this comparative discussion. First is the model of Gray (1995), in which the contents of consciousness "correspond to the outputs of a comparator that, on a moment by moment basis, compares the current state of the organism's perceptual world with a predicted state." This is an attractive thesis, and has already been noted above as part of the mechanisms for the production of Ca. However, there is no reason given by Gray for there to be any conscious content to the resulting comparator, as is supplied by the emphasis here on the relational structures involved (see also Taylor, 1996a, 1998a). The lack of such inner content may explain why it was that Gray concluded "... neither this nor any existing comparable hypothesis is yet able to explain why the brain should generate conscious experience of any kind at all."

The approach of Block (1995) fits the Relational Model quite well. He builds on the original model of Schachter (1989) (which has a set of specialized modules inputting to a module for "phenomenal consciousness," coupled to an episodic memory system, that in turn feeds to the executive and response systems). On the basis of this and other models, Block subdivides consciousness into a phenomenal part, denoted P, and an access part A (together with further components which are not considered in much detail). The paradigm P-conscious states for Block are sensations, whilst those for A-consciousness are thoughts, beliefs and desires, "states with representational content expressed by 'that' clauses" (Block, 1995). He argues strongly for avoidance of the conflation of P with A.

It might be suggested that there is a rough identification of P with Cp and A with Ca. This is supported by the manner in which Block separates P and A consciousness along the divide of awareness versus consciousness without attention. However, there is an important difference between the two divisions of consciousness. Block assumes that P and A have the same information content; "A is the information processing image of P." That is certainly not the case with Cp and Ca; the first has purely phenomenal content, the latter has that (transferred to the frontal active memory sites) plus whatever is created by active transformations on this frontal activity. It may be that any such transformed activity always sends its content back to posterior WM sites. However, it is unlikely that the total content can be transferred, since the posterior WM sites are purely to create phenomenal consciousness, as discussed in detail in section 7.

Finally, there is the interesting automaton approach of Aleksander (1996), in which it is claimed that there are special neurons whose activation leads to the "personal sensation of consciousness," these being a subset of all neurons. The mechanism by which the special neurons are chosen is that they are involved in "iconic memory", which consists of sculpting suitable basins of attraction by external inputs and actions taken by the neural system. Such a memory is activated by the same input as it had already learnt.

The siting of these attractor states in any neural modules in the human brain is not attempted, although there is the claim that such states must occur in the brain since it itself is a neural state machine. The existence of attractor states in the brain has been argued strongly on neurophysiological grounds (Amit, 1990, 1995), the best candidate for that being the hippocampus or nearby temporal regions. However, the hippocampus is not the site of consciousness, as noted earlier in respect to Gray's model of consciousness; to simply identify persistent attractor states in the brain with consciousness is too simple (as noted by several other critics of Aleksander's program).

In spite of that, the program of Aleksander has considerable similarity with that being pursued here, especially the requirement that the iconic state activity persists over time. But the program has a different emphasis: my aim is to build a model of human consciousness which is founded on, and is consistent with, the present knowledge of brain science; Aleksander's program is to build a machine with such an internal structure that we would be supportive of the claim that it is conscious in our sense. However, unless we know what human consciousness is in scientific terms, and so have a well-tested theory of it, we could not prove that such a machine was indeed conscious in our sense of the word.

The various models are compared in Table 3.

Table 3
Comparison of the different models of phenomenal consciousness. "C" denotes phenomenal consciousness; MRAS denotes Mesocephalic Reticular Activating System;
? denotes not discussed or insufficient clarity to site in any place in the brain

Author	Model Name	Site of "C"	Properties of "C"
Dennett	Multiple drafts	Whole brain	No qualia
Baars	Global Workspace	Layer 1 or MRAS	Not explained
Edelman	Remembered present	Limbic (?)	Not explained
Norman & Shallice	Supervisory Attentional System	Frontal	Not explained
Gray	Comparator	Hippocampus	Not explained

Block	Two "C"s	?	Not explained
Aleksander	Iconic machine	?	Not explained
Taylor	Relational Mind	Working memories	"bubbles"

9. Discussion

A model has been presented which is claimed to enable one to discover what it is like to be the conscious being X. It is assumed that X has suitable neural network structure to enable us to recognize separate modules in X performing the activities of global competition, and episodic memory, and also sets of sensory modality-specific working, active and semantic memories. Then the inner view that X has of its experiences is given by the relations set up by the memory activities accompanying the winning working/active memory in its global competition with other such memories. The various features of X's consciousness, of uniqueness, intentionality, familiarity and subjective character are then derived from the model, as it is supposed other features of Searle's list (1991) could be. Moreover, identification of neural structures involved with the nucleus and fringe of consciousness may be made (James, 1950; Mangan, 1993).

The answer to the question as to what makes neural activity conscious, as given by the Relational Mind model, is:

- (a) the holding of input on a suitable buffer to allow for its further processing;
- (b) the activation by input of suitable memory representations from similar past activity (evocation);
- (c) the intermingling of those representations with the buffered input so as to remove ambiguity, reduce information throughput, and perform possible transformations on the buffered activity.

The Relational Mind model is testable in many ways, with predictions of observable thalamic timing systems and specific timings for the turning on of appropriate working memories, differentially for peripheral, thalamic and cortical stimulation. It is also testable in the phenomena of neglect, as indicated above for blind-sight. Non-invasive MEG and rCBF measurements should allow details of the essential circuitry for consciousness ultimately to be ascertained. A program, using non-invasive instruments, to observe details of the two-stage emergence of phenomenal awareness has been proposed elsewhere (Taylor and Muller-Gartner, 1997).

This paper has made extensive reference to the neural underpinnings of human consciousness. It is possible to attempt to remove as much of the living neural circuits as possible without loss of consciousness. That is an important endeavour but one which seems most appropriate to commence once the model presented here, or a successor, is accepted as explaining human consciousness; it will have passed its main tests. Only then will such a program to build artificial consciousness have a firm start, based on at least one known example.

Acknowledgements

The author would like to thank Dr. D. Gorse, Professor, S. Grossberg and Professor W. Freeman for trenchant comments, Dr. P. Fenwick of the Institute of Psychiatry, London and Dr. A. Villa of the Institute of Physiology at Lausanne for useful conversations, Professor B. Libet for useful discussions on his work, Dr. D. Watt on neuropsychological implications, and Professor Mueller-Gaertner and his colleagues at the Institute of Medicine, Research Centre Juelich for sharing their extensive knowledge on non-invasive instruments and for support in the search for awareness.

References

- Alavi, F. & Taylor, J.G. (1996). A basis for long range inhibition across cortex. In J. Sirosh, R. Miikkulainen & Y. Choe (Eds.) *Lateral interactions in the cortex: Structure and function*. Hypertext book published at <http://www.cs.utexas.edu/users/nm/lateral-interaction-book/cover.html>.
- Alexander, G.E. & Crutcher, M.D. (1990). Substrates of parallel processing. *Trends in Neuroscience*, 13, 266-271.
- Alexander, G.E., DeLong, R. & Strick, P.L. (1986). Parallel organisation of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9, 357-381.
- Amari, S-I. (1977). Dynamics of pattern formation in lateral-inhibitory type neural fields. *Biological Cybernetics*, 27, 77-87.
- Amit, D. (1995). The Hebbian paradigm reintegrated: Local reverberations as internal representations. *Behavioural and Brain Sciences*, 18, 617-625.
- Amit, D. (1990). *Modelling brain function*. Cambridge University Press.

- Atkinson, R.C. & Shiffrin, R.M. (1971). The control of short-term memory. *Scientific American*, 225, 82-91.
- Arbib, M.A. & Hesse, M.B. (1986). *The construction of reality*. Cambridge University Press.
- Arbib, M.A. & Dominy, P.F. (1994). Modelling the role of the basal ganglia in timing and sequencing of saccadic eye movements. In J.C. Houk, J.L. Davis & D.G. Beiser (Eds.) *Models of information processing in the basal ganglia* (pp. 149-162). MIT Press.
- Baars, B.J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.
- Baddeley, A. (1993). Working memory and conscious awareness. In M.A. Conway & P.E. Morris (Eds.) *Theories of memory* (pp. 11-28). Hillsdale NJ: Erlbaum Associates.
- Baddeley, A. (1986). *Working memory*. Oxford University Press.
- Baddeley, A. (1992). Is working memory working? *Quarterly Journal of Experimental Psychology*, 44, 1-31.
- Baddeley, A. & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.) *The Psychology of Learning and Motivation*. Academic Press.
- Bapi, R., Bugmann, G., Levine, D. & Taylor, J.G. (1997). Neural models of frontal function. *Behaviour and Brain Sciences* (submitted).
- Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, 94, 115-147.
- Block, N. (1995). On a confusion about a function of consciousness. *Behaviour and Brain Sciences*, 18, 227-287.
- Bogen, J.E. (1995a). On the neurophysiology of consciousness: I. An overview. *Consciousness and Cognition*, 4, 52-62.
- Bogen, J.E. (1995b). On the neurophysiology of consciousness: Part II. constraining the semantic problem. *Consciousness and Cognition*, 4, 137-158.
- Carpenter, G. & Grossberg, S. (1993). Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions. *Trends in neuroscience*, 16, 131-140.
- Chalmers, D. (1996). *The Conscious Mind*. Oxford University Press.
- Cohen, D S & Murray, J (1981). A generalised diffusion model for growth and spersal in a population. *Mathematical Biology*, 12, 237-249.

Conway, M.A. & Rubin, D.C. (1993). The structure of autobiographical memory. In A.F. Collins, S.E. Gathercole, M.A. Conway & P.E. Morris (Eds.) *Theories of memory*. Hillsdale, NJ: Lawrence Erlbaum.

Crick, F. H. C., Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2, 237-249.

Crick, F.H.C. & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, 375, 121-3.

Davies, M. & Humphreys, G.W. (Eds.) (1993). *Consciousness*. Oxford: Blackwell

Davidson, D. (1970). Mental events. In L. Foster & J. Swanson (Eds.) *Experience and theory*. Duckworth.

Dempsey, E.W. & Morison, R.S. (1942). The electrical activity of a thalmo-cortical relay system. *American Journal of Physiology*, 138, 283-289

Dennett, D.C. (1988). Quining qualia. In A.J. Marcel & E. Bisiach (Eds.) *Consciousness in Contemporary Science* (pp. 42-77). Oxford: Oxford University Press.

Dennett, D.C. (1991). *Consciousness explained*. Allen and Lane.

Dominey, P. & Arbib, M.A. (1992). A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex*, 2, 153-173.

Dowling, J. (1987). *The retina*. Cambridge, MA: Harvard Univ. Press.

Edelman, G. J. (1989). *The remembered present*. New York: Basic Books.

Ermentrout, G. B. & Cowan, J. D. (1978). Studies in mathematics. *The Mathematical Association of America*, 15, 67-117.

Frijda, N.H. & Moffat, D. (1993). A model of emotions and emotion communication. In *Proceedings of the 2nd IEEE International Workshop on Robot and Human Communication*, pp. 29-34.

Fuster, J.M. (1994). Frontal lobes. *Current Opinion in Neurobiology*, 3, 160-165.

Georgopolous, A.P. (1994). New concepts in generation of movement. *Neuron*, 13, 257-268.

Goodale, M.A. & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, 15, 20-25.

Goodale, M.A., Milner, A.D., Jakobson, L.S. & Carey, D.P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature* 349, 154-156.

Gray, C. M. & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences US* 86, 1698-1702.

Gray, J. A. (1995). The contents of consciousness: a neurophysiological conjecture. *Behavioural and Brain Sciences*, 18, 659-676.

Guigon, E., Dorizzi, B., Burnod, Y. & Schultz, W. (1994). Neural correlates of learning in the prefrontal cortex of the monkey. *Cerebral Cortex* (in press).

Hilgard, E. R. (1977). *Divided consciousness*. John Wiley, New York.

Houk, J.C. & Wise, S.P. (1993). Outline for a theory of motor behaviour: involving cooperative actions of the cerebellum, basal ganglia and cerebral cortex. In Rudomin, P., Arbib, M.A. & Cervantes, F. (Eds.) *From neural networks to artificial intelligence* (pp. 452-470). Springer Verlag.

Horne, P.V. (1993). The Nature of Imagery. *Consciousness and Cognition*, 2, 58-82.

Hume, D. (1896). *A treatise on human nature*, edited by E. A. Selby-Briggs. Oxford: Clarendon Press.

Jackson, S.R., Marrocco R. & Posner M.I. (1994). Networks of anatomical areas controlling visuospatial attention. *Neural Networks*, 7, 925-944.

James, W. (1950). *The principles of psychology*. Dover Books.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and mental imagery. *Behavioral and Brain Sciences*, 17, 187-245.

Jones, E. G. (1983). The thalamus. In Emson, P.C. (Ed.) *Chemical neuroanatomy* (pp. 257-293). New York: Raven Press.

Jones, E. G. (1975). Some aspects of the organisation of the reticular complex. *Journal of Comparative Neurobiology*, 162, 285-308.

Kirillov, A.B., Myre, C.D. & Woodward, D.J. (1993). Bistability switches and working memory in a two-neuron inhibitory-feedback model. *Biological Cybernetics*, 68, 441-449.

Kubota, Y. & Gabriel, M. (1995). Studies of the limbic comparator: Limbic circuit training-induced unit activity and avoidance behaviour in rabbits with anterior dorsal thalamic lesions. *Behavioral Neuroscience*, 109, 258-277.

La Berge, D., Carter, M. & Brown, V. (1992). Thalamic and cortical mechanisms of attention suggested by recent PET experiments. *Neural Computation*, 4, 318-333.

La Berge, D. (1990). A Network Simulation of Thalamic Circuit Operations in Selective Attention. *Journal of Cognitive Neuroscience*, 2, 358-373.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64, 354-61.

Libet, B., Alberts, W. W., Wright Jr E. W., Delattre, D. L., Levin, G. & Feinstein, B. (1964). Production of threshold levels of conscious sensation by electrical stimulation of human somato-sensory cortex. *Journal of Neurophysiology*, 27, 546-578.

Libet, B., Wright Jr E. W., Feinstein, B. & Pearl, D. K. (1979). Subjective referral of the timing for a conscious experience. *Brain*, 102, 193-224.

Libet, B. (1982). Brain stimulation in the study of neuronal functions for conscious sensory experience. *Human Neurobiology*, 1, 235-242.

Libet, B., Pearl, D.K., Morledge, D.E., Gleason, C.A., Hosobuchi, Y. & Barbaro, N.M. (1991). Control of the transition from sensory detection to sensory awareness in man by the duration of a thalamic stimulus. *Brain*. 114, 1731-1757.

Libet, B. (1994). Private communication.

Llinas, R., Ribary, U. (1992). Chapter 7 of Basar, E. & Bullock, T. (Eds.) *Induced rhythms in the brain*.. Boston: Birkhauser.

Lopes da Silva, F. H., Witter, M. P., Boeijinga, P. H., Lohman, A. H. M. (1990). Anatomic organization and physiology of the limbic cortex. *Physiology Review*, 70, 453-511.

Mandler, G. (1975). Consciousness: Respectable, useful and probably necessary. In Solso R. (ed). *Information processing and cognition: The Loyola symposium*. Lawrence Erlbaum.

Mangan, B. (1993). Taking phenomenology seriously. *Consciousness and Cognition* 2, 89-108.

Marcel, A. J. (1993). Slippage in the unity of consciousness. In *Ciba Foundation Symposium no 174, Experimental and Theoretical Studies in Consciousness*. John Wiley and Sons.

Marcel, A. J., (1980). Consciousness and preconscious recognition of polysemous words. In R. S. Nickerson (Ed.) *Attention and performance, VIII* (pp. 435-467). Lawrence Erlbaum.

- Marcel, A.J. (1983a). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, 15, 197-237
- Marcel, A.J. (1983b). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processing. *Cognitive Psychology*, 15, 283-300
- Marcel, A.J. & Bisiach, E., (Eds.) (1988). *Consciousness in contemporary science*. Oxford: Oxford University Press.
- May P., Tiitinen, H., Ilmoniemi, M., Taylor, J.G. & Naatanen, R. (1998). Frequency change detection in human auditory cortex. *Journal of Computational Neuroscience* (submitted).
- Mesulam, M.M.(1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology* 10, 309-325.
- Metzinger, T. (1995). The problem of consciousness. In T. Metzinger (Ed.) *Conscious experience* (pp. 3-40). Thorverton, UK: Imprint Academic, Schoningh.
- Monchi, O. & Taylor, J.G. (1995). A model of the prefrontal loop that includes the basal ganglia in solving the recency task. In *Proceedings of the World Congress on Neural Networks 1995* (pp. 3-51). INNS Press and Lawrence Erlbaum.
- Monchi, O. & Taylor, J.G.(1996). Simulations of three prefrontal loops with respect to working memory and attention. *Proceedings, Computational Neuroscience Conference, 1997*. (in press).
- Nagel, T. (1974). What is it like to be a bat?. *Philosophical Review*, 83, 435-450.
- Nelson, K. (1993). Explaining the emergence of autobiographic memory in early childhood. Chapter 13 in A.F. Collins, S.E. Gathercole, M.A. Conway & P.E. Morris (Eds.) *Theories of memory*. Hillsdale NJ: Lawrence Erlbaum.
- Newman, J. & Baars, B.J. (1993). A neural attentional model for access to consciousness: A global workspace perspective. *Concepts in Neuroscience*, 4, 255-290
- Newman, J., Baars, B.J. & Cho, S-B. (1995). A neurocognitive model for attention and consciousness. In Globus G.G. & Stamenof M.I. (Eds.) *Advances in consciousness research*. Benjamins (in press).
- Norman, D.A. & Shallice, T. (1986). Attention to action. Reprinted in Davidson, P.J., Schwartz, G.E. & Shapiro, D. (Eds.) *Consciousness and self regulation, vol 4*. Plenum.
- Paulesu, E., Frith, C.D. & Frackowiak, R.S.J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342-5

Posner, M.I. & Petersen, S. (1990). The attentional system of human beings. *Annual Review of Neuroscience*, 13, 25-42.

Posner, M.I. & Raichle, M.E. (1994). *Images of mind*. Scientific American Library, Freeman.

Posner, M.I. & Raichle, M.E. (1995). Precis of images of mind. *Behavioural and Brain Sciences* 189, 327-383.

Purpura, K.P. & Schiff, N.D. (1996). The thalamic intralaminar nuclei: a role in visual awareness. *The Neuroscientist* 3 (in press).

Riani, M. & Masulli, F. (1990). Modelling perceptual alternation by using anns. In Caianiello, E.R. (Ed) *The Second Italian Workshop on Parallel Architectures and Neural Networks*. World Scientific.

Roediger, H. & McDermott, P. (1994). Creating false memories: remembering words not presented in lists (unpublished manuscript).

Romo, R. & Schultz, W. (1992). Role of primate basal ganglia and frontal cortex in the internal generation of movements. *Experimental Brain Resesearch*, 91, 396-407.

Rusinov, V.S. (1973). *The dominant focus*, translated by Luria A.S. and Doty R.B. New York: Consultants Bureau.

Salmon, E., Van der Linden, M., Collette, F., Delfiore, G., Maquet, P., Degueldre, C., Luxen, A. & Franck, G. (1996). Regional brain activity during working memory tasks. *Brain*, 119, 1617-1625.

Schachter, D. (1989). On the relation between memory and consciousness. In Roediger, H. & Craik, F., (Eds.) *Varieties of consciousness: essays in honour of E. Tulving*. Lawrence Erlbaum.

Schiebel, A. B. (1980). Anatomical and physiological substrates of arousal: A view from the bridge. In Hobson, J. A. & Brazier, B. A. (Eds.) *The reticular formation revisited* (pp. 55-66). New York: Raven Press.

Searle, J. (1991). *The rediscovery of mind*. Cambridge: University Press.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press.

Skinner, J.E. & Yingling, C.D. (1977). Central gating mechanisms that regulate event-related potentials and behaviour. In Desmedt, J.E. (Ed) *Progress in clinical neurophysiology: Attention, voluntary contraction and event-related potentials* (pp. 30-69). Klinger.

Steriade, M., Domich, L., Oakson, G. (1986). Reticular thalamic neurons revisited: Activity changes during shifts in states of vigilance. *Journal of Neuroscience*, 6, 68-81.

Steriade, M., Curro-Dossi, R. & Oakson, G. (1991). Fast oscillations (20-40 Hz) in the thalamo-cortical systems and their potentiation by Mesopontine Cholinergic Nuclei in the cat. *Proceedings of the National Academy of Science*. 88, 4396-4400.

Steriade, M., Ropet, N., Kitsikus, A., Oakson, G., (1980). Ascending activation neuronal networks in midbrain reticular core and related rostral systems. In Hobson, J. A. & Brazier, B. A. (Eds.) *The reticular formation revisited* (pp. 125-167). New York: Raven Press.

Stuss, D.T., Shallice, T., & Alexander, M.P. (1995). A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences*, 769 191-211.

Tanji, J. & Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. *Nature*, 371, 413-415.

Taylor, J. G. (1973). A model of thinking neural networks. Seminar, Institute for Cybernetics, University of Tübingen (unpublished).

Taylor, J. G. (1990). A silicon model of the retina. *Neural Networks*, 3, 171-178.

Taylor, J. G. (1991). Can neural networks ever be made to think?. *Neural Network World*, 1, 4-11.

Taylor, J. G. (1992a). Towards a neural network model of mind. *Neural Network World*, 2, 797-812.

Taylor, J. G. (1992b). From single neuron to cognition. In Aleksander, I. & Taylor, J. G. (Eds.) *Artificial neural networks 2, vol 1*, (pp. 11-15). Amsterdam: North-Holland.

Taylor, J. G. (1992c). Temporal processing in brain activity. In Taylor, J. G. et al. (Eds.) *Complex neurodynamics, Proceedings of the 1991 Vietri Conference* (pp. 258-292). Springer-Verlag.

Taylor, J. G. (1993a). A global gating model of attention and consciousness. In Oaksford, M & Brown, G (Eds.) *Neurodynamics and psychology* (pp. 157-180). New York: Academic Press.

Taylor, J. G. (1993b). Neuronal network models of the mind. *Verhandlung der Deutsche Zoologische Gessellschaft*, 86.2, 159-163.

Taylor, J. G. (1994). Relational neurocomputing. Invited talk, Special Interest Group Meeting, World Congress on Neural Networks 1994, San Diego CA, and talk at the IEE Conference on Symbolic Computing, IEE Savoy Place, London.

Taylor, J.G. (1995a). Modelling the mind by psyche. In Fogelman-Sulie F. & Gallinari P. (Eds.) *Proceedings of the international conference on artificial neural networks 1995* (pp. 543-548).

Taylor, J. G. (1995b). A competition for consciousness? *Neurocomputing 11*, 271-296.

Taylor, J.G. (1996a). Breakthrough to awareness. *Biological Cybernetics*, 75, 59-72.

Taylor, J.G. (1996). Where and how does consciousness emerge? King's College preprint (unpublished).

Taylor, J.G. (1998). *The race for consciousness*. MIT Press.

Taylor, J.G. (1998a). Cortical activity and the explanatory gap. *Consciousness and Cognition*. (in press)

Taylor, J. G. & Alavi, F. (1993a). A global competitive network for attention. *Neural Network World*, 5, 477-502.

Taylor, J. G. & Alavi, F. (1993b). Mathematical analysis of a competitive network for attention. In J. G. Taylor (Ed.) *Mathematical approaches to neural networks* (pp. 341-382). Elsevier.

Taylor, J.G. & Alavi, F. (1995). A global competitive neural network. *Biological Cybernetics*, 72, 233-248.

Taylor, J.G. & Taylor, N. (1998). Experiments in modelling frontal lobes for temporal sequence learning. *Neuroscience Abstracts* (to appear).

Taylor, J.G. & Mueller-Gaertner, H-W. (1997). Non-invasive analysis of awareness. *Neural Networks*, 10, pp. 1185-1194.

Taylor, J.G. & Villa, A.E.P. (1997). The conscious I - a neuro-heuristic approach to the mind. To appear in *The brain of homo sapiens*. Rome: Istituto della Enciclopedia Italiana.

Villa, A. (1988) *Influence de l'ecorce cerebrale sur l'activite spontanee et evoque du thalamus auditif du chat*. Thesis, University of Lausanne.

Weiskrantz, L., Barbur, J.L. & Sahraie, A. (1995). Parameters affecting conscious versus unconscious visual discrimination with damage to the visual cortex. *Proceedings of the National Academy of Science*, 92, 6122-6126.

Zipser, D., Kehoe, B., Littlewort, G. & Fuster, J. (1993). A spiking network model of short-term active memory. *Journal of Neuroscience*, 13, 3406-3420.