Pergamon

Neural Networks, Vol. 10, No. 7, pp. 1207–1225, 1997

Printed in Great Britain
0893-6080/97 $17.00+.00

PII: S0893-6080(97)00064-6

1997 SPECIAL ISSUE

Neural Networks for Consciousness

J. G. TAYLOR

Department of Mathematics, King's College and Institute for Medicine, Research Centre Juelich

(*Received* 9 *March* 1997; *accepted* 28 *June* 1997)

Abstract—*The paper outlines a three-stage neural network model for (i) the emergence of consciousness at its lowest level of phenomenal experience, (ii) the development of actions on the emerged conscious activity so as to generate higher-order consciousness. In the model, the lower-first stage involves modules transforming inputs into various codes. It is only at the second level that further modules are activated in a manner which brings about the emergence of phenomenal awareness. The evidence in support of the initial two-stage model is briefly summarised. The two-stage model is then developed by means of a detailed neural framework which allows analysis of the extended duration of activity in posterior working memory sites.The third stage is then described in terms of known neural structures in the frontal lobes, and the manner they might support the creation of higher consciousness is briefly presented.* © *1997 Elsevier Science Ltd.*

1. INTRODUCTION

The race for consciousness has started (Taylor, 1998). There is a large amount of interest in the neuroscience and neural network community on the topic. Various groups are seriously trying to track it down by the use of non-invasive instruments (PET, fMRI, EEG and MEG) by observing the networks of connected modules which function so as to enable various high level cognitive tasks to be solved by the brain.

The current body of knowledge on consciousness is rapidly increasing. There is already a large reservoir of material on it discovered by psychologists over the last century, and this is now being combined with the results coming from non-invasive instruments, and at single cell level in monkeys, to attempt to indicate where and how consciousness arises, and is supported in the brain.

Various models of consciousness have been suggested in the past. These give answers to the two questions indicated in Figure 1 associated with the question marks inserted there. The figure shows schematically the preprocessing of input by the modules labeled 1, and its further analysis by the modules 2. It is the latter which are supposed to support the initial appearance of consciousness of the incoming stimuli. The two question marks attached to the flow of activity in Figure 1 are related to the following basic questions about consciousness:

1. Q1: how is content supplied to the conscious experience once it has been created?

   The second question mark is attached to the "hard" question of consciousness (Levine, 1983):

2. Q2: how does the transformation of activity from the modules 1 to those in 2 cause phenomenal awareness to arise in the first place?

These two questions are closely related to each other, but are logically distinct in that content may have many possibilities, but the underlying phenomenal experience itself appears to have a relatively constant set of attributes (which will be considered in due course). Question 1 attempts to explain how these variations of awareness occur, whilst question 2 explores the principles behind that emergence in the first place. It might be suggested that one should answer question 2 before question 1, but indeed the former is very hard to give a convincing explanation. That is why a number of reasonably acceptable proposals have been made to answer question 1, but with no acceptable solution to question 2 in sight.

The most promising of the answers to question 1 is in terms of some form of feedback from earlier memories, both of semantic and episodic form. Thus, there is the 'Remembered Present' (Edelman, 1989), the 'Relational

Acknowledgements: The author would like to thank Prof. H.-W. Mueller-Gaertner, Director of the Institute for Medicine, for hospitality at the Institute where this paper was written, and to his colleagues A. Ioannides, L. Jaencke, S. Posse and J. Shah for stimulating conversations. I would also like to thank S. Petersen for the simulation of Figure 10.

Requests for reprints should be sent to Dept of Maths, King's College, Strand, UK; E-mail: j.taylor@mth.kcl.ac.uk.

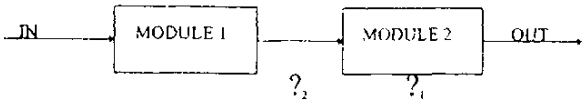

FIGURE 1. The basic two-stage model of awareness. Module 1 performs preprocessing in various modalities and codes; activity in module 2 is that on which phenomenal awareness is based. The two question marks denote the questions Q1 and Q2 stated in the text.

Mind' (Taylor, 1973, 1991) and the 'Predicted Present' (Gray, 1995) as candidate versions using, in one way or another, relational structures involving past memories in one form or another. However, these approaches do not seem to help solve question 2. It is the purpose of the first part of this paper to propose the bare bones of such a solution.

There is also the unresolved question as to how further higher level cognitive experience arises and is related to that of phenomenal experience. This leads to our third question:

Q3: Is there a dissociation between higher level cognitive processes (thinking, planning, self-awareness) and the lower level phenomenal experience, and if so what are the additional neural structures and mechanisms which support the higher level processes?

There has already been acceptance of some level of dissociation between passive awareness and controlled cognition in presenting the two-stage model of Figure 1, since otherwise it would not be possible to leave out the higher levels. The three stage model of Figure 2 does assume some dissociation but may simply be regarded as a summary of known anatomy, with the modules at stage 2 being in posterior cortex and stage-three modules in frontal lobe. However, the process of modeling the actions of the various parts of the more complete structures of Figure 2 would need to be performed with some knowledge of the level of any such dissociation. That will be considered at a later stage of the paper; for the moment we will assume that the methodology we are adopting, of modeling the lower two stages before attacking the third stage, will have some chance of being successful.

The detailed contents of the paper are as follows. It commences in Section 2 with an outline of the two-stage neural model for the emergence of consciousness. In this model the lower first stage involves modules transforming inputs into codes in the visual or other modalities. It is only at the second higher level that further modules are activated in a manner which brings about the emergence of phenomenal awareness. The evidence in support of such a two-stage model will be briefly summarised in Section 3.

The two-stage model is then developed in Section 4 by means of a detailed neural framework which allows analysis of the extended duration of activity in posterior working memory sites (which are posited as at the second

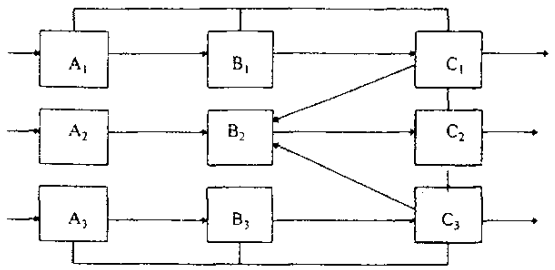

FIGURE 2. The three-stage model for the creation and control of awareness. The modules at the lowest or first stage are denoted A_1, A_2, ... and are involved only in low level feature analysis in the appropriate modality. A number of modules are involved in such processing. At the next stage are sites of 'buffer' memory in posterior cortex, denoted B_1, B_2, ... They are fed by preprocessed activity from first stage modules, and are coupled to each other so as to run a competition by lateral inhibition so as to select the most appropriate percept, given earlier context still present in activity traces on these modules, to enter phenomenal awareness. The final third stage is composed of modules which are frontal and possess attentional feedback control both to the second stage modules, so as to refresh activity there, and to the first stage modules to reduce or increase stimulus selectivity. Feedback is shown explicitly from the third to lower stages although it will also be present from the second to first stage (not all feedback is shown explicitly).

stage in the model) using the notion of 'bubbles' of activity which are formed in neural models with enough neurons and suitably strong recurrent excitatory and inhibitory connections (Amari, 1977; Taylor, 1997a). Such models have recently been suggested as explaining the development of observed orientation sensitivity in primary visual cortex (Douglas et al., 1995; Somers et al., 1995; Ben-Yishai et al., 1995). The purpose of Section 4 is to discuss the implications of the resulting dynamical activity for the two-stage model of the emergence of phenomenal consciousness.

Detailed neural models are then briefly summarised for

1. the initial emergence of consciousness in terms of semi-autonomous 'bubbles' of activity in a neural system of the form of a continuum neural field theory (CNFT), as developed in a two-dimensional context (Taylor, 1997a);
2. application of CNFT to give brief initial explanations of various visual phenomena: apparent motion (due to moving bubbles) and stabilised image effects (destruction of bubbles);
3. how working memory modules, with extended temporal activity, can arise as from CNFT for suitably high density neuronal content (together with suitable ionically-driven adaptation);

The initial two-stage model is then extended to a third stage, as shown in Figure 2, in Section 5. There is much investigation presently as to the extent of possible dissociations between posterior sites of working memory, the

so-called 'buffer sites' of short term memory, and the anterior 'active memory' sites (Fuster, 1993). The nature of this evidence will be summarised and a third-stage model suggested in Section 6 which appears to be consistent with presently know data.

The paper finishes with a discussion, particularly on how the anterior activations can create the experienced features of higher consciousness.

2. THE TWO-STAGE MODEL OF AWARENESS

The two-stage model of the emergence of awareness in the brain is shown in Figure 1. There is supposedly no awareness of activity in stage 1 modules, whilst it is in modules at stage 2 that activity can enter awareness. What sort of neural networks would be suitable for such a separation? Some features of them have been summarised in Table 1, derived from analysis of introspection by psychologists and philosophers (and summarised in Metzinger, 1995). As seen from the table, there are spatial and temporal features of the networks which are of great importance. Thus, there are spatial features which require localised representations in sets of well-connected modules which are also well coupled to those modules involved in higher cognitive processes (the modules C of the third stage of processing in Figure 2).

The temporal features are also very specific, requiring suitably long temporal duration of activity (of the order of 300–500 ms) for phenomenal experience (PE) to arise but with little time between one experience and the next. Finally, there are emergent features which are also non-trivial, involving one-way and rapid emergence of PE at the highest level of a processing hierarchy with no ability to gain awareness of the activity of lower modules of stage 1.

Coding at the highest level of activity emerging into PE is most simply obtained by assuming that it arises from those first stage modules which are also coded at the second-stage level. The other most important feature in Table 1, that of the temporal aspect of duration, is a crucial part of the general features of the buffer stores in the psychologically-based working memory model (Baddeley and Hitch, 1974; Baddeley, 1986). These buffers have input from semantic-level coding, where all possible interpretations (in the case of words) have already been accessed in the semantic memory, so at the end of the stage 1 processing, before the buffer store is attained (Marcel, 1980; Levelt et al., 1991).

TABLE 1
Criterial features on neural networks to support PE

Spatial features:
(a) localised representations in localised modules
(b) modules well-coupled together
(c) modules well-coupled to higher-level modules
Temporal features:
(a) temporal continuation of activity
(b) time required to achieve PE activation
(c) no gap between different PE activations
Emergent features:
(a) one-way creation of PE activation
(b) PE created at highest level of a hierarchy
(c) rapid emergence once begun
(d) no ability to probe lower levels supporting PE but below PE creation level

The two-stage model of Figure 1 has been developed in a neural network implementation elsewhere (Taylor, 1996a). It uses semantic-level coding of activity at stage 1 and then buffered activity at stage 2, with lateral inhibition at this stage to achieve singling out of the contextually appropriate representation at the higher stage. It is this activity which then enters awareness by being broadcasting around the 'global workspace' of (Baars, 1988) or, in terms of the three stage model of Figure 3 gaining access to the anterior sites of working memory. Once at that third and highest level the material can be rehearsed or manipulated as desired. A neural network model of this higher level processing will also be described later in the paper.

Besides the construction of a model of the two-stage process, and due to the plethora of buffer stores now observed in posterior cortex in various modalities and codes (Salmon et al., 1996; Paulesu et al., 1993; Smith and Jonides, 1995) there must be some way of combining their various activities so as to achieve a unified experience of consciousness. Control structures in the brain are thereby needed which produce global correlations between the various component working memories. In particular, these control structures should support competition between various inputs, when suitably encoded, and only allow certain, most relevant, memories to be activated and related to the corresponding winning input. Such structures may also be involved with assessing the level of discrepancy of new incoming input with that predicted from later parts of activated stored pattern sequences. These aspects were considered (Taylor, 1992, 1993, 1994; Taylor and Villa, 1997), in terms of possible networks which could perform pattern matching and sustain competition. One of these involved the nucleus reticularis thalami (NRT), a sheet of mutually inhibitory neurons interposed between thalamus and cortex (see Figure 3).

It was suggested that the thalamus-nucleus reticularis thalami(NRT)-cortex complex may support such activities. This is due to the fact that the NRT is composed almost entirely of inhibitory neurons which are fed by activity coming from thalamus up to cortex and also by reciprocal cortico-thalamic connections. Since NRT also sends inhibitory inputs down onto thalamus, it is clear that such inhibition (which could also function as a release from inhibition if the NRT targets inhibitory interneurons in thalamus more effectively than the

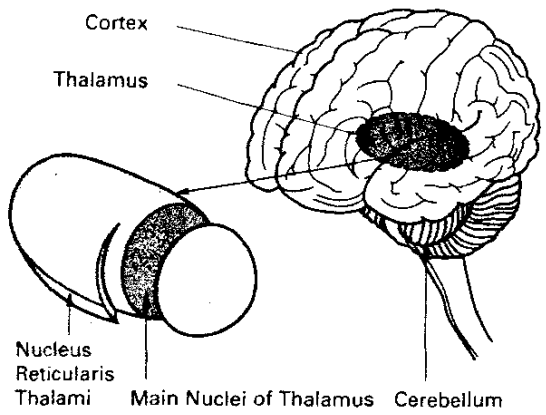

FIGURE 3. The Nucleus Reticularis Thalami (NRT) is a sheet of inhibitory neurons surrounding the upper and lateral parts of the thalamus. It only gives output to the thalamus as well as by lateral connections to other parts of itself, whilst receiving collaterals from both thalamo-cortical and cortico-thalamic axons.

excitatory relay cells to cortex) could exert a powerful control influence on cortical activity. This has been shown experimentally to be the case, with global effects of the NRT sheet especially being observed in the manner in which NRT controls the nature of cortical patterns of activity in sleep. There is also some evidence for a similar global form of control by NRT on allowed cortical activity in the non-sleep states (Skinner and Yingling, 1977; Villa, 1988).

It is possible, in a general manner, to understand this global control achieved by NRT of cortical activity in the following manner. Since any localised activity on

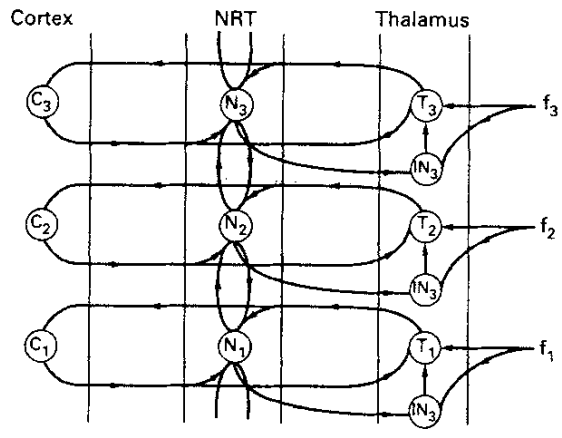

FIGURE 4. Circuit used to simulate the global competitive control action of the NRT over thalamic inputs to cortex, and hence to cortical activity itself. Input I enters both the main relay cells, denoted T, and the associated inhibitory interneurons IN in thalamus, and then feeds topographically to cortical cells C. There are collaterals from the cortico-thalamic axons and the thalamo-cortical ones feeding the NRT cells N, whcih are also laterally connected. These latter also feed back to both the inhibitory and relay cells in the thalamus, having greatest inhibitory effect on the former (from Alavi and Taylor, 1993).

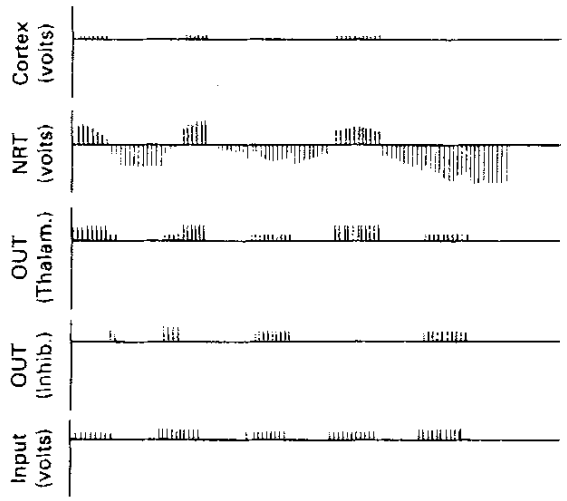

FIGURE 5. Results of a simulation of the circuit of Figure 4 for a line of 100 cells, for an input given on the bottom line. There is a global wave of activity set up over the NRT, as seen in the activation curve iin the second line of the figure. This causes a reduced level of input to cortical cells, as their activity indicate, shown on the top line.

the NRT will try to damp down activity on it elsewhere by lateral inhibition, the NRT can sustain 'bunched' spatially inhomogeneous activity, in which competition between neighbouring thalamic or cortical inputs onto it is occurring. This spatially structured activity may occur globally over the whole NRT sheet if it is well enough connected laterally, as is seen to be the case in spindle generation in sleep (Steriade et al., 1990). In this manner the NRT may function as a global controller of cortical activity. As such it appears of great relevance to include in models of the control circuitry for consciousness (Taylor, 1992, 1993, 1996b; Alavi and Taylor, 1992, 1993, 1995; Baars and Newman, 1993, 1994; Harth, 1995; Kilmer, 1996). Known local circuitry (Steriade et al., 1990) was used to construct a simulation of the coupled thalamo-NRT-cortical system following the circuit of Figure 4; the resulting activity arising from a simulation of 100 cortical neurons is shown in Figure 5. The resulting extended two-stage model is shown in Figure 6.

In summary the proposed two-stage model has the following features:

1. coding occurs up to semantic level (denoted SM) in the first (preprocessing) stage in Figure 6,
2. there is feedforward transfer of activity from the first to the second stage at approximately the same level of coding, with the feedforward map being
3. SM → WM
4. in a given code (as shown in Figure 6),
5. duration of activity is longest of all posterior modules in the WM of the second stage,
6. there is a process of global competition between

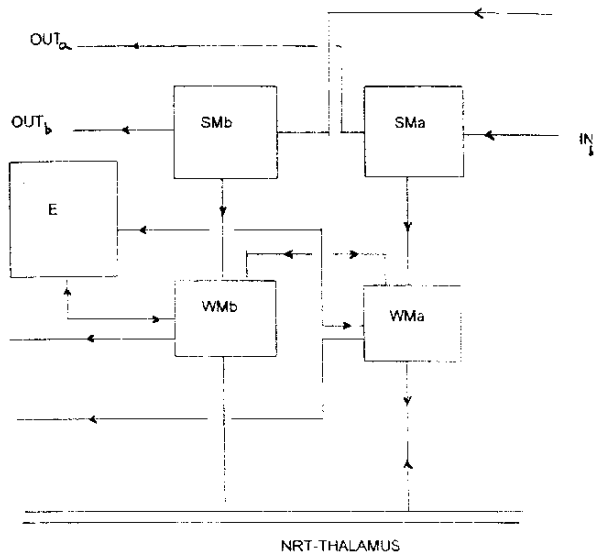

FIGURE 6. The resulting more detailed first two stages of the networks of the 3-stage model. The various features are discussed in the text.

activities on different WMs, supported by the TH/NRT/C system,

7. there is a strong reciprocal connectivity between the posterior WM sites and frontal areas at the highest level of coding (to allow rapid and effective access to the frontal 'active' working memory sites). Such connections are observed neuroanatomically.

The above model of the two-stages of phenomenal awareness leaves unanswered many detailed points. In particular:

1. what is the dynamical origin of the temporally extended activity traces on the working memory sites?
2. how are the codes in the WMs built up?
3. how do features become bound across codes and modalities?
4. what is the temporal dynamics of the emergence of awareness?

Answers to some of these questions (especially the 1st, 2nd and 4th) will be developed in Section 4. Before that we turn in the next section to describe some of the experimental basis for the separation of the emergence of awareness in two parts, as the 2-stage model claims.

3. EXPERIMENTAL SUPPORT FOR THE TWO-STAGE MODEL.

The two-stage model supposes that there are two sorts of cortical processing: firstly at the level of the stage-1 modules in Figure 1 and supporting neural activity of which there is no direct awareness, and secondly that at the level of stage-2 modules of Figure 1, of which there is direct awareness. This is not a new idea although it has been severely challenged in the past.

The reason for such a challenge is not too difficult to discern. If the two-stage model is true then there must be some properties possessed by the stage-2 modules which are not by the stage-1 modules. Such properties should be open to direct analysis by experiment and modeling, and perhaps allow the manner in which awareness is created to be determined. On the other hand, if there were no special modules in the brain supporting awareness then consciousness could preserve its monolithic mystery; it would still be some unknown feature possessed by a complex enough set of nerve cells but not one we can yet (if ever) see how to probe. Of course the two-stage model shades into the one-stage, monolithic one as the preprocessing stages of Figure 1 reduce in size and importance. However the existence of any modules at the non-aware level would still destroy the monolithic approach and make the task of probing consciousness more hopeful.

There are two main sources of support for the two-stage model: altered experience brought about by deficits, and that caused by suitably subtle paradigms for normal people. The first class contains modification of many forms of experience due to brain damage brought about by accident or disease. All of these are germane to the general program of explaining the mind in terms of the brain but some are more crucial to the task at hand, that of supporting the two-stage model, so will be described briefly in this section.

The two most important deficits to be discussed here are those of neglect and of blindsight. Neglect involves the inability to notice inputs to one side of the body. Thus, a patient will be unable to notice one side of their field of view; if such a patient is shown a picture containing two houses, placed one above the other, with one of the houses having flames emerging from its left side, then the patient may well prefer the house without the flames if asked to choose between them. This and other tests (Bisiach, 1988) indicate that the patient has knowledge up to a high level about the neglected side of their field of view but it is not part of their phenomenal experience.

Blindsight is a similar phenomenon in which there is knowledge without awareness. It is a phenomenon which has been well documented (Weiskrantz, 1986), and occurs to those who have lost a part of the primary visual cortex and so cannot see things in that part of the visual field. Yet they have knowledge of moving spots of light in their blindfield, since they can guess above chance as to where a spot of light has moved across a screen in front of them. It has even been found, in some cases, that the patient possesses some form of awareness of the movement-termed 'contentless awareness' (Weiskrantz et al., 1995), if it is fast enough. Thus, it would appear as if there is still the ability of modules after V1 (the primary visual area) to create awareness provided there is a large enough input by some alternate route.

Besides the cases brought about by damage and disease there is also the phenomenon of 'normal blindsight' (Kolb and Braun, 1995). A subject is required to pick out from an array of short bars oriented in a particular direction a small subset oriented in the perpendicular one. It is found that, provided the field of oppositely oriented bars is small enough, the subject is not aware of them but can be correct as to where they are more often than by chance alone. Here again there is knowledge without awareness, so that the orientation analysis, performed very likely in V1, is producing knowledge which is below the level of awareness.

Further support for V1 being below the level of awareness has also been given by an experiment which uses the adaptation of subjects to oriented bars, which takes place over about 5 s of viewing the bars (He et al., 1996). Once such adaptation has taken place, and before recovery, the contrast threshold for detecting oriented bars of the same orientation is found to have increased considerably. If the bars were 'crowded' by having bars of opposite orientation surrounding them in the field of view, loss of awareness of the particular bars now occurs but there is still an increase of the threshold for detecting bars presented at a variable contrast thresholds. The phenomenon of crowding is also found to be asymmetric between the upper and lower halves of the field of view, an asymmetry not possessed by V1. This means that V1 is involved in the adaptation to the orientation analysis but not in the final production of awareness. As the authors concluded 'activation of neurons in V1 is insufficient for conscious perception'.

This is further supported by the so-called waterfall effect or 'motion after-effect' (MAE). This occurs as an adaptation after about 30 seconds of viewing a moving scene; on cessation of the motion there is an apparent motion of the now static scene in the opposite direction. The time course of this has been investigated under fMRI (Tootell et al., 1996a) where they detected in subjects continued activity in the motion-sensitive area V5 for about the same length of time that the MAE persists in humans (about 9 s). even more interestingly they observed (Tootell, 1996b) a longer time course of the effect (by about 3 s) in higher visual areas beyond V5. This supports the notion that awareness arises in sites with about 3 seconds of duration of neural activations.

Single cell data from monkeys are also relevant to the search for the modules involved in the creation of consciousness. Binocular rivalry (in which different images are presented to the two eyes and the percept will switch every few seconds between the two possibilities) has been used in monkeys to determine in which area there is a predominance of neurons signaling the percept when it is experienced (which they can be trained to do faithfully). The use of this technique, over several years, has led Logethitis and colleagues (Logothetis and Schall, 1989; Leopold and Logothetis, 1996) to be able to show that, in vision, certain earlier areas (V1, V2, V4)

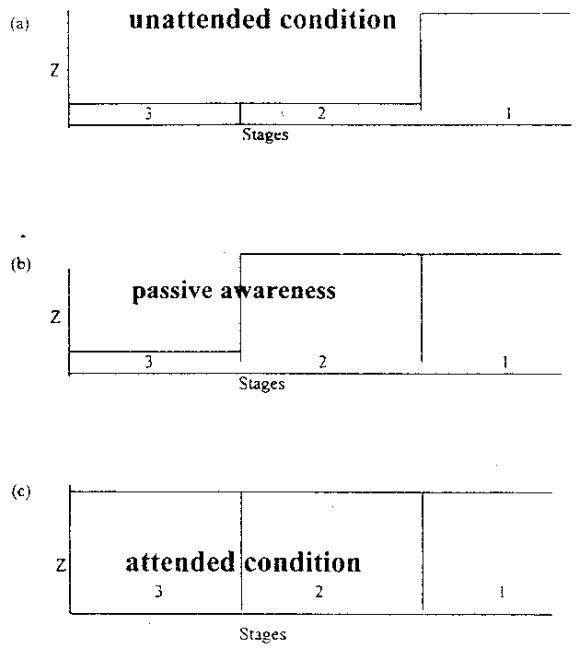

FIGURE 7. The levels of activation of the various modules of the 3-stage model for a given input under the three conditions of a) ignore b) passive awareness c) directed attention. Under condition (b) both the first two stage modules are activated, whilst under (c) all are on. There will be differences expected between the total levels of activation across the conditions due to feedback modulation, so the levels in the different conditions are normalized.

have about equal numbers of neurons firing both for the percept they support (in terms of direction of motion, for the paradigm in which the rivalry is between two sets of moving bars going vertically upwards versus moving downwards; a neuron supports a movement direction if it fires preponderantly when that direction of movement is presented) and the opposite one. As they wrote, 'The majority of cells in all areas continue to respond even when it is perceptually suppressed'. However their most recent analysis in the inferotemporal area (Sheinberg and Logothetis, 1997) has shown that in areas TE and TEO there is now a majority of active neurons in favor of the percept being experienced. Where the change-over area is sited is unclear and very important to discover.

There is also support from psychophysical experiments on word processing, in which the processing of a word presented visually is probed by a further auditorily-presented probe word (Levelt et al., 1991). The conclusion of the study was that a two-stage process occurs in which at the first stage there is automatic processing of all meanings of the word, which at a later stage are reduced to a suitably unique phonological representation. This is similar to the earlier two-stage 'logogon' model of (Morton, 1969, 1979) both supporting the proposal of the two-stage model of Section 2 that awareness arises at the second stage.

There is also increasing evidence in support of the

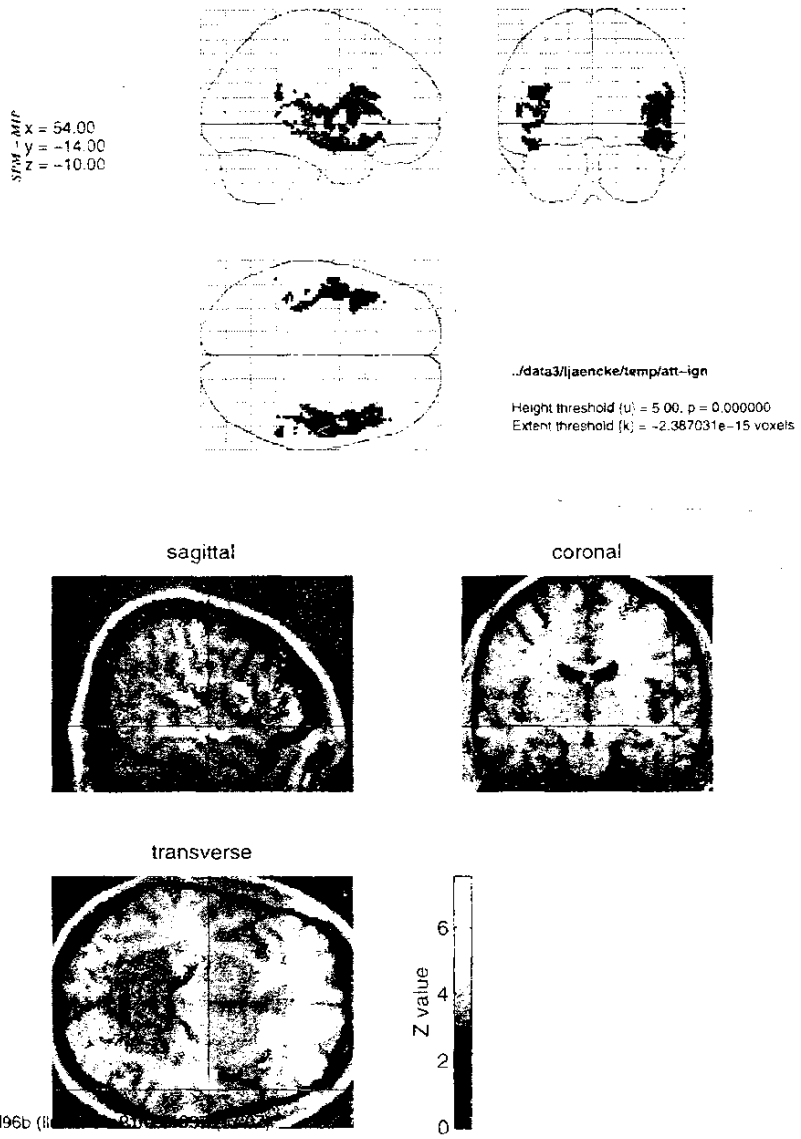

FIGURE 8. Differences in activation at various levels of the brain from an fMRI experiment (Taylor, Jaenke et al, 1997) in which the passive levels had subtracted from them levels due to ignore conditions for listening to a sequence of syllables. Note the non-zero levels of activity in various sites (threshold set at 5 leads to significance at $p < 0.01$).

2-stage model from non-invasive measurements. Thus, if a stimulus is ignored (by means of a suitably distracting task in another modality) the activity in cortical areas can be compared with that arising from passive awareness of the stimulus itself without any distractor. The 2-stage model would lead to the expectation of differences in activation along the lines of that shown in Figure 7. These differences in activation level have been seen in an auditory listening task with an added attentional load, as shown in Figure 8 (Taylor et al., 1997). This is taken from an fMRI study of five subjects in which there is comparison between subjects ignoring a sequence of syllables and listening passively to them. There is clear activation of regions in BA44/45 (Broca's area) and auditory areas when one passes from the ignore condition to the passive listening condition; these are candidates for the emergence of phenomenal awareness of syllables. The activation in Broca's area may be part of attetional processing and not due to phenomenal awareness.

Similar support arises from a recent MEG experiment (Vanni et al., 1996) in which subjects were presented with line drawings of objects or of non-objects, with a visual mask then delivered 30, 45 or 106 ms later. The level of the peak amplitude in only one visual area, the right lateral occipital (LO) area, was found to be significantly correlated with the level of awareness of the object drawings. This result indicates that the right cortical visual area LO is importantly involved in the creation of awareness of the objects described by the line drawings.

There is also a set of recent fMRI studies determining areas activated by various working memory tasks, such as the n-back task (Cohen et al., 1996). In this task a subject observes a sequence of letters or numbers presented one per second and has to respond when the same symbol occurs after n presentations (where n has been told to the subject by the experimenter before the particular run, with values of $n = 1$ to 4). Various areas were found to be significantly active compared to a background task, in particular posterior sites in BA 40 and BA 18/19, as well as prefrontal sites when n became 2 or greater. This was especially so for the dorsolateral prefrontal cortex (DLPFC), a fact important for our analysis of the third stage modules of Figure 2. However this area was not active for $n = 1$, so corresponding to posteriorly sited buffer memory (holding an object in memory over 10 s), and supporting the dissociation between frontal (stage 3) and posterior (stage 2) cortical areas under low-load conditions.

The conclusion of this section is that there is strong experimental support for

1. the existence of two separate stages at the lowest level in Figure 1,
2. the localisation of activity associated with the second stage of Figure 1 to specialised sites in posterior cortex, usually termed sites of 'buffer memory',
3. the determination of sites also involved in the stage-3 modules of Figure 3, involved in the creation of higher -order consciousness.
4. dissociation of the third from the lower stages under low attentional load.

4. A NEURAL NETWORK FOR THE TWO-STAGE MODEL

4.1. The Basics of Bubbles

We now turn to the development of a neural model to help support the two-stage model of Figure 1 and use it to give tentative answers to the questions raised at the end of Section 2. In particular we will consider the questions:

1. what is the dynamical origin of the temporally extended memory traces residing in the short-term working memory sites considered in this section?
2. how are the codes in these sites constructed?
3. what are the neural principles behind the creation of phenomenal awareness in these sites?

As stated in the Introduction we will employ continuum neural field theory (CNFT) whcih has been used over many years as a first approximation to cortical structure (Amari, 1977; Beurle, 1956; Ermentrout and Cowan, 1978). One of the important results of that period is the discovery of cortical 'bubbles' of activity which are initially created by input but which persist after stimulus offset (Amari, 1977). A brief introduction to bubbles and some of their properties in CNFT is given in this subsection, and relevant applications developed in the following one.

CNFT is based on a model of cortex as a continuous two-dimensional sheet of neurons with a lateral connection with the shape of a 'Mexican hat'. This allows the creation of localised activity, the bubbles, which can persist, and remain localised in spite of the absence of input due to the recurrent activity produced initially causing repeated firing of the neurons initially activated. The basic CNFT equation is constructed in terms of the membrane potential of a neuron and denoted by $u(x,t)$, at the point x and time t. It will be assumed that there is lateral connectivity other neural sheet defined by the lateral connection weight function $w(x - x')$ between the two neurons at the relevant points x. The connection weight will be usually taken to be of Mexican hat form as a function of the Euclidean distance $|x - x'|$. There is also an afferent connection weight function $s(x, y)$ from the thalamic position y to the cortical point x. The response function of a neuron will be taken to be determined by its mean firing rate, which is given as some function f of the membrane potential u of the relevant cell.

The membrane potential $u(x, t)$ will satisfy the CNFT equation (Amari, 1977)

$$\tau \partial u(x,t)/\partial t = -u(x,t) + \int dx' w(x-x') f[\,u(x',t)\,] + \int dy\, s(x,y)\, I(y,t) + h \qquad (1)$$

where $I(y, t)$ is the input to the thalamic position y at time t, h is the neuron threshold and the integration over the lateral connection weight is over the manifold M of neurons.

There are well-known autonomous solutions to (1) in the case when M is one-dimensional (Amari, 1977). In that case equation (1), for a static solution and with no input, becomes:

$$u(x) = \int w(x-x')\, 1[u(x')]\, dx' + h \qquad (2)$$

where the sharp threshold response function $f = 1$ (the step function) has been assumed in (2). A 'bubble' is defined to have a positive membrane potential over an interval, independent of input. This is formalised as

4.1.1. *Definition.* A 'bubble' of neural activity is a localised persistent solution to the CNFT equations. Its size is in general dependent on input, but its continued existence is not (to within effects of adaptation).

Let us consider the bubble extending from $x = 0$ to $x = a$:

$$u(x) > 0\,,\ 0 < x < a;\ u(0) = u(a) = 0 \qquad (3)$$

and otherwise $u < 0$. Then from (2) and (3), u is obtained

explicitly as

$$u(x) = \int_0^a w(x-x')\,dx' + h = W(x) - W(x-a) \quad (4)$$

where the function W is defined by

$$w(x) = \int_0^x w(x')\,dx' \quad (5)$$

Necessary conditions for the bubble to exist are that the membrane potential vanishes at the ends of the interval $[0, a]$, so

$$u(0) = u(a) = 0 = W(a)' + h \quad (6)$$

It is then possible to show that $u(x) > 0$ for $0 < x < a$ if $h < 0$; $u(x) < 0$ otherwise.

Stability of the resulting solution then requires

$$dW(a)/da < 0, \ \vee \ w(a) < 0 \quad (7)$$

Thus, the one-dimensional bubble exists under the conditions (6) and (7).

There are a number of further important results derived in Amari (1977) concerning the nature of bubble solutions and their extension to input dependence which will be briefly summarised here:

1. the parameter ranges for h and for the parameters in W can be determined so as to allow for autonomous solutions of various types ($\varnothing$ or the trivial one, ∞ or the constant non-zero one, an a-solution as the bubble of finite length a described above, and a spatially periodic solution),
2. complete determination of those patterns which are stable and those which are unstable, from amongst the stationary solutions described above,
3. response to input stimulus patterns: a bubble of finite length moves to a position of maximum of the input,
4. two bubbles interact, if close, with attraction (from the Mexican hat connection weight function), if more distant with repulsion, and if very distant with no effect on each other,
5. there can occur spatially homogeneous temporal oscillations (between a layer of excitatory and one of inhibitory cells)
6. traveling waves can persist.

Returning to the full two-dimensional bubble in the region D (which will be called an $R[2]$-solution here), we define

$$W_\infty = \lim_{R\to\infty} W(R) \quad (8)$$

where

$$W(R) = \int_D w(x-x')dx' \quad (9)$$

where x is only allowed to be on the boundary of D, which has radius $|x| = R$.

It is now possible to extend the methods 1977 to deduce the same results as in the one dimensional case for the questions (a) to (f) raised above. Using the same enumeration:

1. Theorem 1. In the absence of input:

   There exists a $\varnothing$ solution iff $h < 0$.
   There exists an ∞-solution iff $W_\infty > -h$.
   There exists an $R[2]$-solution iff $h < 0$ and $R > 0$ satisfies

   $$W(R) + h = 0 \quad (10)$$

   It is possible to extend the classification of the solutions for varying levels of the stimulus h. Let $W_m = \max_x W(x)$.

2. Theorem 2. The nature of the various solutions for different parameter ranges is as in Figure 9.

   To determine which of these solutions is stable it is necessary to extend the one-dimensional discussion of Amari (1977) to two (or higher) dimensions. From the two-dimensional extension of equation (6) the boundary of D, defined by the radius $R(t)$ at time t, satisfies the constraint

   $$u(R(t), t) = 0 \quad (11)$$

   On differentiation of equation (11) with respect to t and use of equation (1) there results

   $$dR/dt = -[G(R(t)) + h]/\tau v \quad (12)$$

   where v is gradient of u normal to ∂D and is negative. The equilibrium case results on setting the right

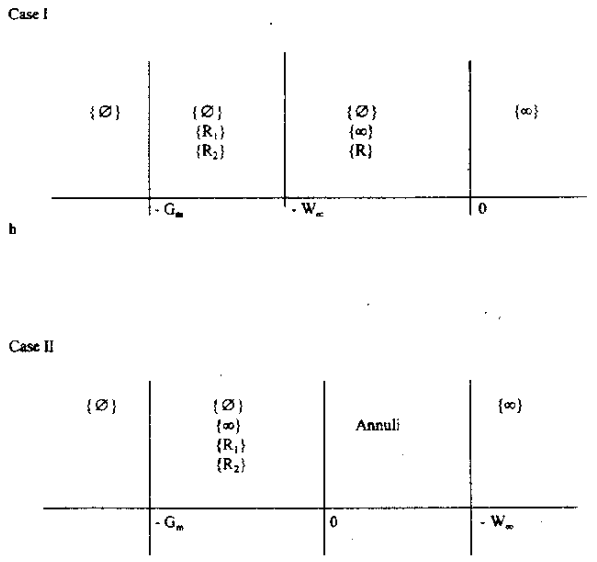

FIGURE 9. Parameter ranges for the existence of stable bubbles in two-dimensional CNFT (from Taylor, 1997a). The meaning of the paramteres is given in the text.

hand side of equation (12) to zero. The stability of this solution is determined by the sign of dG/dR: $dG/dR < 0 \Leftrightarrow$ stability. This leads to the stability classification of the solutions as given in theorem 2.

3. The response to stationary inputs of an $R[2]$-solution can also be treated as in Amari (1977). Consider a small stationary input $\varepsilon I(x)$, which is not assumed to be circularly symmetric so that the asymptotic bubble will not be circularly symmetric either. The equation of constraint is, following equation (12), that

$$dx/dt.\ \nabla u + \partial u/\partial t = 0 \text{ on } \partial D(t) \qquad (13)$$

Replacing the time derivative of u on the left hand side of equation (13) by equation (1) it is now possible to derive the condition for $x(t)$ on the boundary of $D(t)$,

$$dx/dt.\ \nabla u = (1\tau)[h + \varepsilon I + G(|x(t)|)] \qquad (14)$$

On expanding in a perturbation series in the small quantity ε, with

$$||x(t)|| = R_0 + \varepsilon(s,t) \qquad (15)$$

(where s denotes a parameter specifying which point on the boundary of the unperturbed $R[2]$ is being considered in the perturbation (15)) there results the constraint

$$d\varepsilon/dt.\ \nabla u = (1/\tau)[\varepsilon + \varepsilon.\ \nabla G(|x(t)|)\ ], \qquad (16)$$

where $\varepsilon = \varepsilon(s,t)$ is a vector describing the direction as well as the magnitude $\varepsilon(s,t)$ of the perturbation of the boundary point x, and the derivatives in (16) are evaluated at $\varepsilon = 0$, so at $x(t) = R_0$.
The result from the constraint (16) is that the net radial movement of the boundary of $D(t)$ is towards the region of largest input. There will be a movement of regions of $\partial D(t)$ towards lower values of the input, if these are positive, but there will be a larger velocity of movement towards those regions of the boundary nearer the maxima of I.

4. The one-dimensional result (d) above can also be extended in the same way, where the effect of one region (say $D1$) on another (say $D2$) is given, in terms of the lateral interaction term in equation (2), as the effective input to a neuron in $D2$ at the point x of amount

$$s(x) = \int_{D1} w(|x - x'|)dx' \qquad (17)$$

This will have the same effect as in the one-dimensional case, with attraction between the bubbles at $D1$ and $D2$ if they are close enough (as determined by $s(x)$), repulsion if the two regions are further separated, and ultimately no interaction between the bubbles at all if they are beyond the range of the lateral interaction term w (if that is finite).
5. The case of spatially homogeneous oscillations extends immediately to the two-dimensional case, since only ∞-solutions are being considered.
6. This case involves temporal structure and is considered more fully in Taylor (1997a).

4.2. Applications of Bubbles

We will apply the bubble solutions only to the cases of apparent motion, the fading of stabilised images and the lifetime of bubbles in cortex. Bubbles have also been applied to a variety of other cortical processes: the development of topographic maps, both in the one dimensional case (Takeuchi and Amari, 1979) and for two dimensions (Taylor, 1997a), control of saccades by the superior colliculus, on which the bubbles are supposed to form (Kopecz and Schoner, 1995), the modifications of the somatosensory constant topographic map by rapid unmasking due to removal of portion of an input (Petersen and Taylor, 1996a) or by relearning (Petersen and Taylor, 1996b), the guidance of head-directed cells (Zhang, 1996) and in explaining pre-attentive auditory memory (May et al., 1995). The two cases to be considered are closely related to features associated with the possible manner in which bubbles could enter directly into perception, and so are most relevant for discussion here.

4.2.1. *Apparent Motion.* This is the very well-known phenomenon in which a spot of light appearing on a screen in one place within 200 or so ms after the disappearance of one within 5 degrees of the second leads to the experience to a viewer of the motion of the first spot to the second. There has been considerable research on this phenomenon (see, for example, Cavanagh and Mather, 1990 and Ullman, 1979) and interesting attempts to correlate and explain it (Dawson, 1991). Here CNFT will be used to give an explanation at the level of principle, without attempting to explain all of the details of the phenomenon by the model (which is discussed more fully elsewhere).

The basic idea is the visual input from the initial spot creates a bubble in a cortical area (very likely MT) which is then dragged across the area by the appearance of the new spot in the manner arising from the result (d) above. The disappearance of the first spot and the appearance of the second causes a change in the background in which the bubble created by the first spot resides. It is thus attracted to the central region of activity brought about

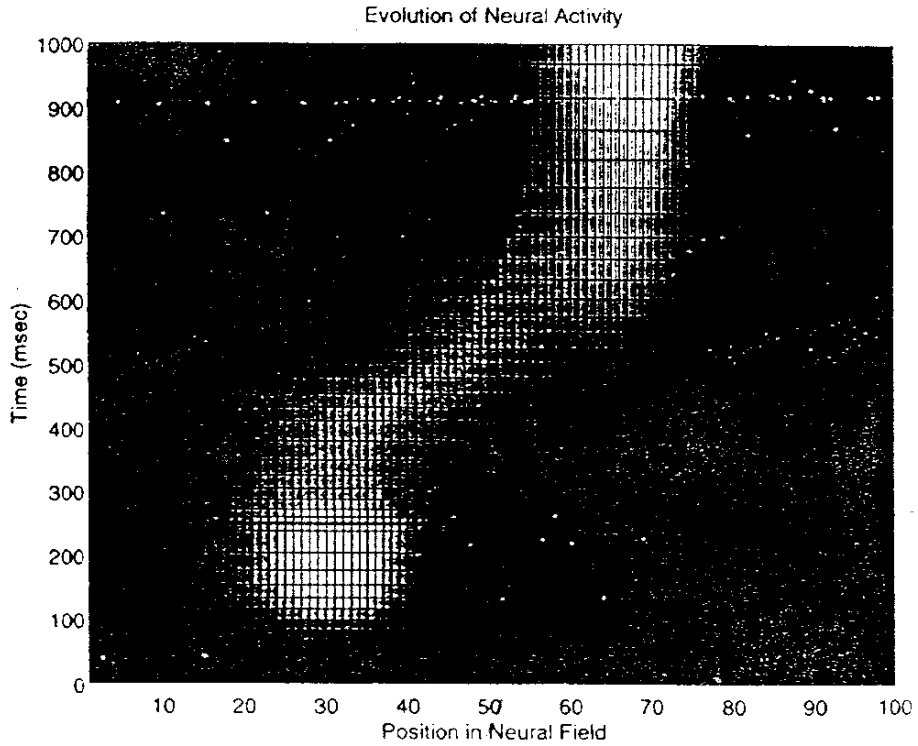

FIGURE 10. The dragging of a bubble, due to a localised input at the point 30 at time 0-300, formed in a line of neurons when the new input at the position 70 is switched on at the time 250. This is a proposed explanation of AM (see text).

by the second spot on its appearance, following the above argument. This process is difficult, although not impossible, to analyse in detail using the time-dependent equation (1), but a one-dimensional simulation is shown in Figure 10. In this, a bubble is created by a spot at time $t = 0$ at the position 30, using equation (1) with suitable parameters for the input to cause the creation of a bubble. This initial input is then removed at time 250 and a new input used at the position 70 at that time. As seen from the figure the bubble created by the first spot reduces in size when its input is removed but persists to move over to the position of the second input. It leaves a trail as it goes, causing the impression of a moving input.

4.2.2. *Stabilised Images.* These are experienced when the retinal image is kept fixed on the eye and the image ultimately fades from view (Pritchard, 1963). The initial loss of image may be due to adaptation in the retinal pathway, so reducing the image able to reach later stages. The process leads to very interesting patterns of image loss in which there are both Gestalt-type and ensemble effects observed. One particularly interesting effect was observed by (Crane and Piantanida, 1983) by stabilising the line between a rectangle of green and one of red. Normally it is not possible to observe both green and red together (they are opponent colours) but on stabilising the dividing line between them three [illegible] percepts were observed by subjects:

1. an overall 'red/green' colored region they had never experienced before,
2. dots of red and green interspersed throughout the visual field,
3. islands of red in a sea of green or vice versa.

It is possible to give a simple model possessing these three activations by means of two coupled CNFTs, one coding for red, the other for green. The red and green neurons at a given position are coupled inhibitorily to each other. The resulting model is shown in Figure 11. The coupled equations for the membrane potentials u and v for the two sets of colours are

$$du/dt = -u + w_{uu} * \theta(u) - w_{uv} * \theta(v) + h_u + I_u \quad (18a)$$

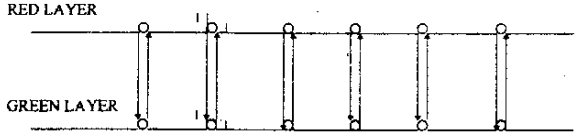

FIGURE 11. The architecture used for modelling red/green stabili[illegible]ge percepts. See text for details.

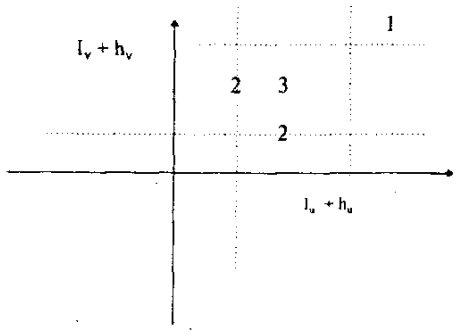

FIGURE 12. The regions of the ($I_u + h_u$, $I_v + h_v$)- plane associated with the possible solutions to the model of figure 11, as discussed in the text.

$$dv/dt = -v + w_{vv} * \theta(v) - w_{vu} * \theta(u) + h_v + I_v \quad (18b)$$

(where * denotes convolution over the cortical area, w_{uu} and w_{vv} are the usual Mexican hat functions and w_{uv}, w_{vu} are short range (excitatory) carriers of inhibition to their opposite colour). The inputs I_u and I_v are assumed to have been reduced on stabilisation of the image, so as to lead to a range of possible solutions to (18a) and (18b). It may be shown (Taylor, 1997b) that there are at least three forms of solution to (18):

1. $\infty x \infty$ (in the notation of earlier in the section, where '∞' denotes an infinitely extended bubble),
2. interdigitating spatial oscillations,
3. solutions of the type $\{(-\infty, a_1] \cup [a_2, \infty\} \times [a_1, a_2]$ (giving the support of the bubble solution in the separate coordinates).

It is seen that these solutions correspond respectively to the experiences of subjects listed above. The range of the parameter plane ($I_u + h_u$, $I_v + h_v$) for which these solutions (i)–(iii) exist is shown in Figure 12 and Figure 13.

4.2.3. *Bubble Lifetimes.* Having seen that bubbles can help explain some of the phenomena of perception (and other features of cortical processing as well, as noted in the earlier references) we now turn to the important question as to how bubbles might disappear It is very unlikely that they persist for ever, and if they did so then they would present an ever increasing background of 'noise' interfering with current ongoing processing. There is even some evidence for their possessing a finite lifetime from work of Lu et al. (1992). They exposed subjects to a sequence of sounds with the inter-stimulus interval (ISI) being gradually increased. They discovered that the amplitude of the N100 response (100 ms after stimulus onset) reached saturation at suitably long ISI in both primary and secondary auditory cortex, with the rise to saturation corresponding to a decaying trace with lifetime of about 3 s in primary and 5 in secondary auditory cortex.

What mechanism could cause the decay of such a trace or more generally of the bubbles which they might represent? The most likely answer is that of adaptation of the responses of the neurons in the CNFT. Spike adaptation is a well-studied phenomenon (Connors et al., 1982; McCormick, 1986) arising from slow after-hyperpolarizing currents I_{AHP} and I_M, which can last for several seconds (Schwindt et al., 1988). These after-currents essentially raise the threshold for neuronal response. There may also be an effect from previous traces of neural activity which shunts out later activation (May et al., 1997).

The results of Lu et al. (1992) and of Uusitalo and

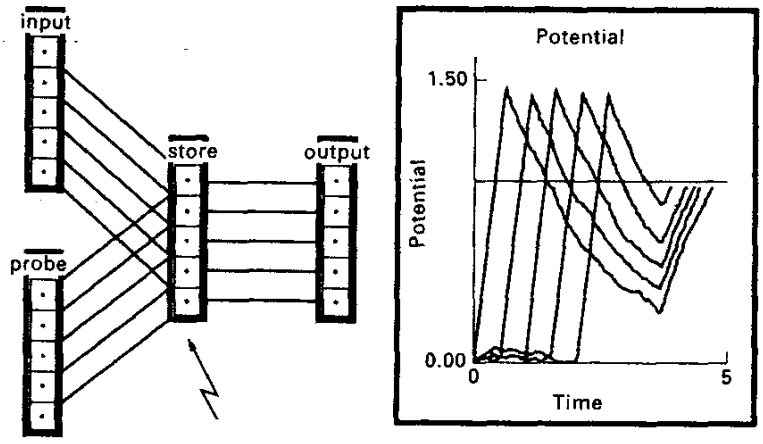

FIGURE 13. The model of recency in working memory (Hastings and Taylor, 1994). The nodes are dedicated to a given input item, and the resulting activity decays as corresponding to a leaky-integrator neuron. Probe re-activation has to lead to neuron activity above a criterial level before reponse. The resulting 'universal forgetting' formula for the reaction time to a new input as having been in the previously presented list fits known data well.

Williamson (1996) will be interpreted here in terms of the decay of a lasting trace of activity in the particular region from which the magnetic field was being measured (auditory or primary visual cortex respectively). This will be shown by using a very simple model of the response of a neuron, that of the leaky integrator (Hodgkin–Huxley) neuron, but with an after-hyperpolarizing current as observed in Schwindt et al. (1988). The equation for the membrane potential u of this neuron will be

$$\tau\, du/dt = -u + I(t) - \lambda \int_0 \exp[-(t-t')/\tau'] f[u(t')]dt' \tag{19}$$

where $I(t)$ is the input current at time t, λ is the strength of an after-hyperpolarising current dependent on the response $f[u(t)]$ integrated over the time the neuron is active with an exponential decay of lifetime τ', and τ is the intrinsic lifetime of activity on the surface of the neuron (assumed to have no internal structure).

Under the experimental paradigm of Lu et al. (1992) the input is on for the time T and off for the time which we denote by the value ISI.

During the period that the input is on, u builds up its value driven by the input I. Let us suppose that the hyperpolarising lifetime τ' is much longer than the time T of duration of the input. This is valid for the long-lasting potassium-dependent hyperpolaristions mentioned by Schwindt et al. (1988), which we assume to be the one of equation (19). At the end of the input period the membrane potential begins to decay exponentially, so behaves as $\exp[-t/\tau]$ (multiplied by a suitable constant, and to within an additive constant).

We now assume that the term on the left-hand side of (19) can be neglected (in other words the time constant τ is relatively short) with the result that the time dependence of u at (or close to) the beginning of the next input, say at the N100 response, has the value (to within a constant initial value from $t = T$)

$$-\lambda \int_0 \exp[-(t-t')/\tau'] f[u(t')]dt' \tag{20}$$

where $t = T + \text{ISI}$. For f chosen to be the semi-linear function $f(x) = (x)_+$ (the positive part of x) and assuming that the membrane potential is positive through the ISI (so that f is linear) then the membrane potential at $t = T =$ ISI, is from (20) equal to

$$constant - \lambda exp[-(T + ISI)/\tau] \tag{21}$$

For an ISI = 0 there will be no N100 so that (21) must then be equal to zero. Thus (21) becomes (for fixed T)

$$u(T + ISI) = A(1 - exp[-ISI/\tau]) \tag{22}$$

Formula (22) fits the results of Williamson and colleagues (1992, 1996). This justifies interpreting the value of the time constant (as the duration of the neural trace of activity in the cortical site from which they were measuring. A more complete analysis can be given, under other assumptions on the relation between τ and τ', with a similar result to that presented above.

There is the further question as to how neurons can possess such large and variable time constants, both across areas and across subjects, as have been measured by Williamson and colleagues (1992, 1996). We will now show that these variations are to be seen as arising from variable levels of recurrence, giving an increase to the intrinsically identical time constants of each neuron by an amount dependent on the amount of recurrence; this level of recurrence can vary from one cortical region (and from one subject) to another, and is crucially dependant on the cell density in the short-term memory stores.

In order to consider bubble decay we will consider in more detail the effect of the long-lasting after-hyperpolarisation current used in equation (19) on bubble lifetime. The two-dimensional expression replacing equation (19) is

$$\tau \partial u(x,t)/\partial t = -u(x,t') + \int dx'\, w(x-x') f[\,u(x',t)\,] + \int dy\, s(x,y)\, I(y,t) + h - \lambda \int_0^a \times \exp[-(t-t')/\tau'] f[u(x,t')]\, dt' \tag{23}$$

where τ' is a measure of the lifetime of the adaptation current, and λ denotes its strength. It is now necessary to calculate the lifetime of a bubble created using equation (23). Let us first consider the one-dimensional case; that for two dimensions will follow straightforwardly.

A particular case of interest is when a bubble has initially been created by an input which is then removed. That could be due, for example, to the neural module acting as the source of the input having a shorter lifetime for the persistence of bubbles than the one under consideration. It would also occur if the bubble is created in a primary sensory module and the input itself has been modified.

To discuss this case it is appropriate to first reduce even further to a single recurrent neuron. For that case the membrane potential equation, from (23), is:

$$\tau \partial u(t)/\partial t = -u(t) + w\theta[\,ut)\,] + h - \lambda \int_0^a \exp[-(t-t')/\tau']\, \theta[u(t')]\, dt' \tag{24}$$

where a step function response has been taken for the neuron. From equation

$$u(t) = u(0) + [1 - \exp(-t/\tau)][h + w] \tag{25}$$

$$-\lambda \int_0^t exp[-(t-t')/\tau]dt' \int_0^{t'} \exp[-(t' - t')/\tau']dt''$$

From (25), with $u(0) > 0$, $u(t)$ will remain positive initially in time. Moreover (25) reduces to the expression

$$u(t) = u(0) + [1 - \exp(-t/\tau)][h + w - \lambda\tau\tau'] + \lambda\tau(\tau')^2[\exp(-t/\tau') - \exp(-t/\tau)]/(\tau' - \tau) \tag{26}$$

where the last term on the right hand side of (26) is replaced, for $\tau = \tau'$, by the expression $\lambda\tau\tau'\exp(-t/\tau)$. The last term in (26) may be neglected if $\tau' < \tau$, so that if

$$\lambda\tau\tau' > h + w + u(0) \tag{27}$$

then for suitably large t, $u(t)$ will become negative and the firing of the neuron will then cease. If no new input arrives then no further activity will ensue from the neuron.

The initial lifetime of the bubble is given by equating the right hand side of (26) to zero. Using the assumption that $\tau' \gg \tau$ in (26) gives the approximate value for the life-time T as

$$T = -\tau' ln\{1 - u(0)/\tau[\lambda\tau\tau' - h - w]\} \tag{28}$$

where the factor $[\lambda\tau\tau' - h - w]$ is positive by (27). Equation (28) is the formula we wish to extend to the case of a one- and then a two-dimensional CNFT.

Firstly the case of a bubble solution infinitely extended in either dimension reduces to the above analysis with the constant w in the single neuron case being replaced by the quantities $w = \int w(x)dx$, $w = \int w(x)dx$ in the one and two dimensional cases respectively.

The relevant equation in one dimensions for a finite-sized bubble solution (Taylor, 1997a) has the extra adaptation term

$$-\lambda\tau\tau' \tag{29}$$

(dropping the term of $O(\tau')$ in (26)) and the added initial value $u(l(0), 0)$, where $l(t)$ is the size of the bubble at time t; the input term involving S has also to be dropped. The bubble will have a finite lifetime if the adaptation term is so negative that there exists a solution to the resulting equation for the asymptotic size of the bubble Amari, 1977:

$$h - \lambda\tau\tau' + W(2l(\infty)) = 0 \tag{30}$$

where $W(x)$ is the first integral of the connection weight w over the bubble domain. Such a solution could arise if

$$\lambda\tau\tau' - h > W_m \tag{31}$$

where W_m is the maximum value of W. Thus if (31) is true then the bubble will have a finite lifetime given, under the same approximation as for the single neuron, by

$$T = \tau' ln[-h(\lambda\tau\tau' - h\,W_m)] \tag{32}$$

This approximation should hold for both the one and two dimensional cases. In both cases we note that as W_m is increased, say by increase of cell density, the corresponding lifetime increases.

For the other extreme $\tau' > > \tau$ then τ and τ' must be interchanged in the lifetime formulae (28) and (32).

In conclusion, for the case $\tau > > \tau'$ the bubble lifetime is effectively proportional to τ, so dependent on whatever mechanism produces the bubble itself. In the opposite case the bubble lifetime is proportional to τ'. The latter quantity is expected to be an intrinsic characteristic of the single (pyramidal) neuron, so very likely constant throughout cortex. This characteristic difference between the bubble lifetimes in the two cases (in the former depending on the cortical area under consideration, in the latter constant throughout cortex) has interesting experimental consequences for cortical processing, which are explored in Taylor (1997a).

Finally the lifetime is seen to increase (logarithmically) as the lateral connection strength (so W_m in (32)) increases. Such an increase is slow, according to the formula (32), but this may be only a result of the assumption of a hard limiting threshold output function for the neurons. A smooth sigmoid response (which can still support bubble creation) can lead to a linear increase in lifetime. This may therefore explain the observed increase of lifetime as observed in Lu et al. (1992), as well as be the source of the buffer capability of working memory modules.

4.2.4. *The Buffer Stores.* A simple explanation of the recency effect, that more recent items in a short list are remembered better and faster, has been given by many groups; one of these, with a useful 'universal forgetting formula' was in Hastings and Taylor (1994). This models the short-term buffer store as a set of dedicated nodes which have decaying activity on them. A short list of items coded by these nodes has decaying activity on the store in which the strongest activity is that for the latest input. Recognition of the items presented then occurs from probe inputs which cause a temporal increase in activity until it reaches some criterial threshold for response. The form of this reactivation is shown in Figure 12. It is clear from the figure that the most recent input reaches the criterial threshold soonest, the earliest one taking the longest time. The resulting set of reaction times $RT(n, N)$ for the n'th item in a list of length N may be shown to be given by the 'forgetting' formula

$$RT(n, N) = a.ln\{b + c.\exp[d(n-N)]\} \tag{33}$$

where a, b, c, d are constants and in particular d is determined by the decay constant of the nodes. A very good fit to the experimental observations leads to a value of the lifetime of the activations on the nodes of about 1.5 s, which is in the same 'ball-park' as the lifetimes observed by Williamson et al. noted above.

We conclude that there is support for the existence of bubbles in cortex at the basis of phenomenal experience and that these can help to explain the somewhat activity-independent lifetimes observed in short-term memory tests and modeled by very simple dedicated nodes with lifetimes of about 100 times the decay constant of the single neurons themselves. This latter feature is explicable in terms of adaptation-driven bubble decay. The resulting lifetime of a bubble depends on the density of the cells in the area supporting it, with the longest

lifetime occurring for highest density. Finally we note that these bubble may be used to help explain the nature of the 'qualia'-like aspects of phenomenal experience, in particular its apparently intrinsic and non-relational characteristics Taylor (1997b).

5. THE THREE-STAGE MODEL

We now turn briefly to the third stage of the three-stage model of Figure 3. This stage receives conscious material from the second stage discussed above and uses it to achieve goals by planning and reasoning. We would expect that the regions of frontal cortex so involved would be the working memory sites of the prefrontal cortex, particularly areas 46 and 9, those observed to cause deficits in long-term planning and delayed tasks when lost due to accident or disease. There is also a clear change in personality and social responsiveness when there is loss of the mesial oribtofrontal regions. Such regions would also be implicated in the construction of higher order consciousness.

We have already given support, at the end of Section 3, to the suggested dissociability between the first two stages and that at the third, highest level assumed in the three-stage model of Figure 3. There is considerable further material for such dissociation between the stages from deficits of frontal patients, who appear to posses awareness of their deficits but not the concern that would be expected in their situations. Thus there is 'a dissociation between what the frontal patient knows or says, and how he or she behaves' (Stuss, 1991), a feature indicating dissociation between posterior knowledge and anterior action systems.

The earlier approach, in terms of the use of memory structures, termed the 'Relational Mind' model in Taylor (1973, 1991) can also be used to explain how content arises in higher consciousness. Thus there are representations of earlier experienced objects encoded in appropriate sites in frontal lobe and their re-activations by later input gives content to that input. These inputs arise from the posterior buffer sites discussed earlier, so that the coding of the frontal sites is guided by posterior activity and at the same time feedback to those sites would help their own representations. However the late onset of prefrontal myelination would prevent much use being made of such frontal sites in the first few years of life, a period when there is extensive coding of words and objects in posterior sites. Thus the prefrontal representations may only help in the development of more advanced and sophisticated concept representations than carried posteriorly.

There is now considerable experimental evidence from non-invasive instruments for the involvement of the above mentioned areas in so-called 'working memory' tasks, those requiring the holding of activity over long and variable periods of time for transformation and later response. Such tasks, for example in the encoding and later retrieval of words or pictures, have shown up a network of modules involving area 46 on both left and right as well as Broca's speech area 44/45. These new modules appear to have considerably longer lifetimes than those of the posterior cortex (Goldman-Rakic, 1992) so posing a question as to the mechanism behind such long activation. We will turn in the next section to consider a possible model for such long and adaptive lifetimes.

6. A NEURAL NETWORK FOR THE THREE-STAGE MODEL.

One of the clearest differences between the frontal and posterior cortices is the existence of strong cortico-thalamic recurrence between frontal sites and the medio-dorsal thalamus. There is also known crucial involvement of the basal ganglia in motor and cognitive processing, as defects brought about by loss of the basal ganglia show in patients with Parkinson's disease. The problem we are facing is that of modeling what Fuster (Fuster, 1993) calls 'active memory', that is persistent neural activities which can be acted upon and transformed into other ones so as to achieve some goal or other. It is useful to turn to consider what form representations of motor actions take in motor and related cortices. 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 such processing 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). It has been applied to modeling delayed tasks and their deficits in frontal patients (Monchi and Taylor, 1995, 1997). There are also a further range of neural models of frontal components of working memory (Carpenter and Grossberg, 1993; Dominey and Arbib, 1992; Dominey et al., 1995: Guigon et al., 1994; Kirillov et al., 1993; Zipser et al., 1993). We will concentrate here on the ACTION network as being most closely related to the neuroanatomy of the frontal cortex.

6.1. The Action Network

The ACTION net is composed of the basal ganglia acting in a disinhibitory manner on the thalamo-cortical recurrent loops (with the presence of lateral cortico-cortical

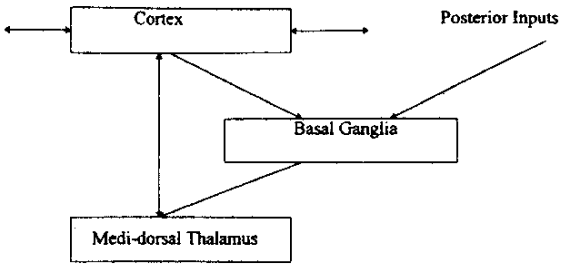

FIGURE 14. A schematic version of the ACTION network. See text for details.

connections present 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, as shown schematically in Figure 14. This generalises the two-dimensional population-vector action 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, 1997), and has been considered in some detail in Dominey et al. (1995). These connections would allow the active memories to have written on them material from posterior buffer sites, from autobiographical/ episodic memory databases related to this material and from goal or other memory buffered in hippocampus.

The process of 'intermingling' of these activities with ongoing activity so as to achieve higher level consciousness is expected to require competitive processing for a similar reason to that discussed in connection with low-level consciousness and for which the NRT was employed earlier in the paper. That such competition can be supported by the ACTION network was suggested in 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 five possible ACTION networks, corresponding to the 5 great frontal loops of Alexander and colleagues (Alexander et al., 1986) 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, as mentioned earlier, and so supporting the 'well-connected character' required for the modules supporting consciousness noted in Table 1.

From the discussions so far it is clear that the re-excitation of episodic memories from hippocampal and mesio-orbital areas needs to be of a whole host of earlier memories, which can give the consciousness 'color' to experience. That might not best be achieved by a pattern completion or attractor network (Amit, 1990), since only a single pattern would result at one time. The most suitable memory structure is an associative matrix memory for a feedforward net, in which a given input pattern will excite as many outputs as previously stored patterns with strong overlap with the input. However the use of an attractor net may in fact help explain the length of time taken in memory search, as various of the attractors are activated and then rejected by comparison to some actively held template as part of the processing of the frontal lobes' ACTION network. This has been discussed from an interesting neural basis in Amit (1995).

7. DISCUSSION

Having constructed some of the possible neural machinery to support higher consciousness, it has now to be put to work. In particular it is important to show, though in the space available only very cursorily, 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 higher consciousness, such as intentionality and introspection, are considered elsewhere (Taylor, 1997a).

A number of these processes have already been hinted at in Taylor (1995). Thus attention can be decomposed into the component sub-processes of comparison, activation/inhibition and monitoring. 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 in Taylor (1995), using action transformations on representations written on prefrontal cortex so as to make such representations be 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—whilst the latter emphasises the set of transformations and not the states.

Emotions are 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

consciousness so as to cause modification of on-going behavior. This may be crucial in order to change schemata (Frijda and Moffat, 1993) as well as having an important effect in motivation and intelligence. Such an approach has been used in the neural models of delayed tasks and the Wisconsin card sorting task in Monchi and Taylor (1995, 1997) where there are three coupled ACTION networks with switching betweeen them according to valuation (concerns) by a net modeling the amygdala.

The above models of neural networks for consciousness have not only been sketched very briefly but also there has been omission of models for the limbic system, in particular the hippocampus and related areas. There is presently considerable interest in such areas and they clearly play an important role in determining the contents of consciousness. However this is something that the interested reader will have to look elsewhere (see, for example, Taylor, 1996c, 1998).

REFERENCES

Alavi, F. and Taylor, J. G. (1992). A simulation of the gated thalamocortical model (pp 929–932) In I. Aleksander and J. G. Taylor (Eds), *Artificial Neural Network 2*, North Holland: Amsterdam.

Alavi, F. and Taylor, J. G. (1993). Mathematical analysis of a global network for Attention, In J.G. Taylor (Ed.), *Mathematical Approaches to Neural Networks*, Elsevier.

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

Alavi, F. and Taylor, J. G. (1996). A basis for long range inhibition across cortex, In J. Sirosh, R. Miikulainen and Y. Choe (Eds.), *Lateral Interactions in the Cortex: Structure and Function*, Hpertext Book ISBN 0-9647060-8 (sited at http://www.cs.utexas.edu/users/nn/lateral-interaction-book/cover.html.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organisation of functionally segregated circuits linking basal ganglia and cprtex. *Annu Rev Neuroscience, 9,* 357–381.

Amari, S.-I. (1977). Dynamical study of formation of cortical maps. *Biological Cybernetics, 27,* 77–87.

Amit, D. (1990). *A Model of the Brain*, Cambridge: Cambridge University Press.

Amit, D. (1995). Psychology, neurobiology and modeling: the science of hebbian reverberations, In F. Ventriglia (Ed.), ch11 *Neural Modeling and Neural Networks*, Oxford: Pergamon Press.

Baars, B. J., (1988). *A Cognitive Theory of Consciousness*, Cambridge University Press.

Baars, B. J., & Newman, J. B. (1993). A neural attentional model for access to consciousness: a global workspace perspective. *Concepts in Neuroscience, 4,* 255–290.

Baars, and B. J. and Newman, J. B. (1994). A neurobiological interpretation of global workspace theory, In A. Revonsuo and M. Kamppinenpp (Eds), *Consciousness in Philosophy and Cognitive Neuroscience*, (211–226) Hillsdale NJ: Lawrence Erlbaum.

Baddeley, A. D. and Hitch, G. (1974). Working memory, In G. A. Bower (Ed.), *The Psychology of learning and motivation, 8,* 47–89, Academic Press, New York.

Baddeley, A. D. (1986). *Working Memory*, Oxford: Oxford University Press.

Ben-Yishai, R., Bar-Or, R.L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Nat Acad Sci (USA), 92,* 3844–3848.

Beurle, R. L. (1956). Properties of a mass of cells capable of regenerating pulses. *Trans. Roy. Soc. London B, 240,* 55–94.

Bisiach E. (1988) The (haunted) brain and consciousness, In A. J. Marcel and E. Bisiach (eds.) *Consciousness in Contemporary Science*, New York: Oxford University Press, pp 101–120.

Cavanagh, and Mather, (1990). Motion: The long and short of it, *Spatial Vision,* 4, 103–129.

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.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E.E. (1996). Temporal dynamics of brain activation during a working memory task. *Nature, 386,* 604–608.

Connors, B. W., Gutnick, M. J., & Prince, D. A. (1982). Electrophysiological properties of neocortical neurons in vitro. *Journal of Neurophysiology, 48,* 1302–1320.

Crane, H. D., & Piantanida, T. P. (1983). On seeing Reddish Green and Yellowish Blue. *Science, 221,* 1078–1080.

Dawson, M.R.W. (1991). The how and why of what went where in apparent motion: modeling solutions to the motion correspondence problem. *Psychological Review, 98,* 569–603.

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

Dominey, P., Arbib, M., & Joseph, J.-P. (1995). A Model of corticostriatal plasticity for learning oculomotor associations and sequences. *J Cognitive Neuroscience, 7,* 311–336.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. C., & Suarez, H. H. (1995). Recurrent excitation in neocortical circuits. *Science, 269,* 981–985.

Edelman, G. (1989). *The Remembered Present,* New York: Basic Books.

Ermentrout, G. B. and Cowan, J. D. (1978). *Studies in mathematics*, The Math Assoc of America, 15, 67–117.

Frijda, N.H. and Moffat, D. (1993). A model of emotions and emotion communication, In *Proc ROMAN '93: 2nd IEEE International Workshop on Robot and Human Communication*, pp. 29–34.

Fuster, J.M. (1993). Frontal lobes. *Current Biology, 3,* 160–165.

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

Goldman-Rakic, P., (1992). *Working Memory and the Mind*, Scientific American, pp. 73–79.

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

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

Harth, E. (1995). The sketchpad model. *Consciousness and Cognition, 4,* 346–368.

Hastings, S. and Taylor, J.G. (1994). Modeling the articulatory loop, P. Morasso and M. Marinaro (Eds), *Proc Int Conf Art Neural Networks (ICANN'94)*, pp. 1452–1455.

He, S., Cavenagh, P., & Intrilligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383,* 334–337.

Houk, J. C. and 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. and Cervantes, F. (Eds.), *From Neural Networks to Artificial Intelligence*, Springer, pp. 452–470.

Kilmer, (1996). Global inhibition for selecting modes of attention. *Neural Networks, 9*(4), 567–573.

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

Kolb, F.C., & Braun, J. (1995). Blindsight in normal observers. *Nature, 377,* 336–338.

Kopecz, K., & Schoner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields. *Biol. Cybern., 73,* 49–60.

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.

Leopold, and Logothetis, (1996). *Nature,* **379,** 549–553.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: a study of picture naming. *Psychological Review, 98,* 122–142.

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Phil. Quarterly, 64,* 354–361.

Logothetis, and Schall, (1989). *Nature,* **245,** 761–763.

Lu, Z.-L., Williamson, S. L., & Kaufman, L. (1992). Human auditory primary and association cortex have different lifetimes for activation traces. *Brain Research, 572,* 236–241.

Marcel, A. (1980). Conscious and preconscious recognition on polysemous words: Locating selective effects of prior verbal contexts, In Nickerson, R. S. ed. *Attention and performance VIII,* Hillsdale NJ: Erlbaum Assoc.

May, P., Himmelbach, M. and Taylor, J. G. (1997). (in preparation).

McCormick, D. (1986). Membrane properties and neurotransmitter actions, In: Sheperd (ed.) *The Synaptic Organization of the Brain,* 3rd edition, Oxford: Oxford University Press.

May, P., Tiitinen, H., Winkler, I., Naatanen, R. and Taylor, J. G. (1995). A neural network model of mismatch negativity and the underlying memory trace, Soc for Neuroscience Abstracts, san Diego.

Metzinger, T. (1995). The Problem of Consciousness, In (ed.) T. Metzinger, *Conscious Experience,* (pp. 3–40) Paderborn: Schoningh.

Monchi, O. and Taylor, J. G. (1995). A model of the prefrontal loop that includes the basal ganglia in solving the recency task, *Proc WCNN95,* Washington, INNS Press.

Monchi, O. and Taylor, J. G. (1997). A model of coupled frontal working memories for delayed response tasks and the wisconsin card sorting task, In (Ed.) P. P. Wang, (pp. 34–37) in *Int Joint Conf in Information Sciences, 2nd ICCIN'97,* Amsterdam: Elsevier.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Morton, J. (1979). Word recognition, In J. Morton and J. Marshall (eds.) *Psycholinguistics: Series 2. Structures and processes,* London: Elek, pp. 107–156.

Paulesu, E., Frith, C., & Frakowiak, R. S. J. (1993). The neural correlates of the verbal components of working memory. *Nature, 362,* 342–345.

Petersen, R. S. and Taylor, J. G. (1996a). Reorganisation of somatosensory cortex after tactile training, pp. 82–88 in *Proc Conf on Neural Information Processing NIPS '96,* New York: Kaufmann.

Petersen, R. and Taylor, J. G. (1996b). Unmasking in somatosensory cortex, *Neural Information Processing* (in press).

Pritchard, R. M. (1963). Stabilized images on the retina. *Scientific American, 204,* 72–78.

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.

Schwindt, P. C., Spain, W. J., Foehring, R. C., Stafstrom, C. E., Chubb, M. C., & Crill, W. E. (1988). Multiple potassium conductances and their functions in neurons from cat sensorimotor cortex in vitro. *J Neurophysiol, 59,* 424–449.

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

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc Nat Acad Sci, 94,* 3408–3413.

Smith, E. E. and Jonides, J. (1995). Working memory in Humans: Neurophysiological Evidence, ch 66, pp. 1090–1020 in *The Cognitive Neurosciences.*

Somers, D. C., Nelson, S. B., & Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J Neuroscience, 15,* 5448–5465.

Steriade, M., Gloor, P., Llinas, R. R., Lopes da Silva, F. H., & Mesulam, M. M. (1990). Basic mechanisms of cerebral rythmic activities. *Electroenc. and Clin. Neurophysiol, 76,* 481–508.

Stuss, D. T. (1991). Disturbance of self-awareness after frontal system damage, in *Awareness of Deficit After Brain Injury,* (eds.) G. P. Prigatano and D. L. Schachter, New York: Oxford University Press, pp. 63–83..

Takeuchi, A., & Amari, S.-I. (1979). Formation of topographic maps and columnar microstructures in nerve fields. *Biol Cybernetics, 35,* 63–72.

Taylor, J. G. (1973). A model of thinking neural networks, Seminar, Institute for Cybernetics, Univ of Tübinjen.

Taylor, J. G. (1991). Can neural networks ever be made to think?. *Neural Network World, 1,* 4–12.

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

Taylor, J. G. (1993). A global gating model of attention and consciousness, In *Neurodynamics and Psychology,* ed. Oaksford, M. and Brown, G., Academic Press, New York.

Taylor, J. G. (1994). Goals, drives and consciousness. *Neural Networks, 7,* 1181–1190.

Taylor, J. G. (1995). Modules for the mind of PSYCHE, *Proc World Congress on Neural Networks,* Washington 1995, Hillsdale NJ: L. Erlbaum/INNS Press.

Taylor, J. G. (1996). Breakthrough to awareness. *Biol Cybernetics, 75,* 59–72.

Taylor, J. G. (1996). A global competition for consciousness? *Neurcomputing, 11,* 271–296.

Taylor, J. G. (1996c). Modelling what it is like to be, In *Proc Arizona Conf on Scientific Basis of Consciousness,* (ed.), Hammeroff, S., Cambridge MA: MIT Press.

Taylor, J. G. (1997a). Neural 'bubble' dynamics in two dimensions I: Foundations; ibid II: Applications to vision, KCL/IME Preprints.

Taylor, J. G. (1997b). Cortical Activity and the Explanatory Gap, King's College preprint.

Taylor, J. G. (1998). *The Race for Consciousness,* Camrbidge MA: MIT Press.

Taylor, J. G., Jaencke, L., Shah, N. J., Noesselt, T., Schmitz, N., Himmelbach, T. and Mueller-Gaertner, H.-W. (1997). A Three-Stage Model of Awareness: Formulation and Intial Experimental Support, KFA Preprint, 1997 and Soc for Neurosci. Abstrs.

Taylor, N. and Taylor, J.G. (1997). (in preparation).

Taylor, J. G. and 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.

Tootell, R. B. H., Reppas, J. B., Dale, A. M., Look, R. B., Sereno, M. I., Malach, R., Brady, T. J., & Rosen, B. R. (1996). Visual motion aftereffect in human cortical area MT revealed by functiona magnetic resonance imaging. *Nature, 375,* 139–141.

Tootell, R. B. H. (1996b). Functional MRI maps of human visual cortical activity and consciousness, abstract 104, Consciousness Research Abstracts, J Consciousness Studies, 1996.

Ullman, S. (1979). *The Interpretation of Visual Motion,* Cambridge Mass: MIT Press.

Uusitalo, M. and Williamson, S. J. (1996). Memory Lifetimes of Human Visual Cortex, Human Brain Mapping Conference Abstract S565.

Vanni, S., Revonsuo, A., Saarinen, J., & Hari, R. (1996). Visual awareness of objects correlates with activity of right occipital cortex. *NeuroReport, 8,* 183–186.

Villa, A. (1988). Influence de L'Ecorce Cerebrale sur L'Activite Spontanee et Evoque du Thalamus auditif du Chat, These, Univ de Lausanne.

Weiskrantz, (1986). *Blindsight,* Oxford: Clarendon Press.

Weiskrantz, L., Barbur, J. L., & Sahrie, A. (1995). Parameters affecting conscious versus unconscious visual discrimination with damage to the visual cortex (V1). *Proc Nat Acad Sci (USA), 92,* 6122–6126.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head—Direction cell ensemble: a theory. *J. Neuroscience, 16,* 2112–2126.

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