



2007 Special Issue

# CODAM: A neural network model of consciousness

J.G. Taylor

*Department of Mathematics, King's College, London, WC2R 2LS, UK*

## Abstract

We present a review of the CODAM neural network control model of consciousness and develop it to arrive at a functional account of consciousness. The main feature is as a speed-up and error-correcting mechanism known, in engineering control theory, to be efficient in improving the speed of response and accuracy of any control system. We use the CODAM model to generate a set of predictions as to how such speed-up is achieved, as well as relate to previous explanations by CODAM of various attention-based phenomena.

© 2007 Published by Elsevier Ltd

*Keywords:* Attention; Corollary discharge; Efference copy; Pre-reflective self; Working memory

## 1. Introduction

There has only been slow progress over the last few decades on the increasingly scientifically respectable subject of consciousness, in spite of the valiant endeavours of numerous groups and individuals. Various models of the neural bases of consciousness of differing levels of specificity have been proposed; none has gained universal acceptance. This difference of opinion has a variety of causes, among these being

- (a) The nature of consciousness itself has not been universally agreed upon;
- (b) The manner in which objective experimental results can indicate how subjective experience is created appears still a difficult bridge to cross (the old London Underground adage for passengers to ‘mind the gap’ springs to mind here);
- (c) There has been considerable debate in the past about the possible usefulness of consciousness. With no clear function agreed on, the nature of any explanation is not yet properly tied down to suitable functional ground truth.

As noted under (a), the nature of consciousness is still unclear. The epiphenomenal (dualist) approach to the topic posits that consciousness floats above the material world on a parallel dynamic track to the grinding machinery of the physical world. How these two apparently similar worlds interact is then at issue, a problem not yet resolved satisfactorily by anyone:

the gap to be joined between the two appears unbridgeable. Furthermore consciousness would appear to have no function in terms of physical activity, although it ‘feels’ from the inside as if it certainly does have effects. Making an act of will to do something does not seem to be irrelevant, yet would seem to be so according to the strict dualist thesis. Moreover the ability of a person when unconscious, such as under the influence of anaesthetic, is considerably reduced, especially in any purposive manner. This implies that consciousness plays an important role in decision making, although exactly what is unclear; that uncertainty does not seem clarified by the dualist (nor presently any other) approach.

The alternative to dualism is some form of physicalism: all is based on the physical activity of the brain. The other extreme possibility of mentalism, that all is mind, is not able to bear the weight of advances in the physical sciences. For example grand unified gauge theories of matter and the discovery of the *W* and *Z* mesons are infinitely more precise and involve energies and distances completely remote from the normal human range.

In physicalism, various levels of dependence on brain activity have been posited. These go from supervenience (a very limited form of dependence on the physical) to out-and-out physicalism (brain activity somehow describes every aspect of inner experience). We will assume this latter approach here, although a reduced form of it would still be compatible with our discussion.

There have been physiological results which support the point of view that consciousness is an epiphenomenon, especially from the work of Benjamin Libet and his colleagues

*E-mail address:* [john.g.taylor@kcl.ac.uk](mailto:john.g.taylor@kcl.ac.uk).

(Libet, Wright, Feinstein, & Pearl, 1979). This has shown, for example, that when a subject makes a self-determined movement, the ‘readiness potential’ measurable by EEG from the frontal cortex starts about 550 ms before the movement, whilst awareness of making the movement occurs some 350 ms later, at about 200 ms before the movement is made. It was concluded that consciousness could only have a veto function in such a situation, but certainly not be involved in the initiation of the movement itself.

Further analyses of brain functioning have shown that there is processing of words up to a semantic level even if there is no consciousness of a stimulus (Vogel, Luck, & Shapiro, 1998). This has been taken to indicate that consciousness is not really necessary to enable responses to be made. Furthermore the existence of dissociation between attention and consciousness has been claimed to be detected by several groups (Koch & Tsuchiya, 2006; Lamme, 2003, 2006). Thus the exact involvement in consciousness of well established brain processes, such as attention, are also presently unclear. Again dualism seems partially supported by such results, but still leaves the bigger question of any interaction between mind and matter untouched, as is also the question of the nature of the usefulness of consciousness. The same can be said for how the gap of matter to mind can be closed or equivalently of how the outside objective view can in any way be related to the inner subjective experience of an individual.

In this paper we consider the CODAM model as a possible neural network model for the brain-based creation of consciousness, using the attention copy signal as the crucial component. We will briefly describe CODAM and the attention copy signal, but then turn to a more extensive discussion of the basic question as to the purpose of consciousness. One extreme is the epiphenomenal answer: there is none (as in the dualist approach to the mind). The other extreme is that consciousness is crucial to brain processing, and without it there is considerable inefficiency and degradation of ability, especially to handle complex tasks. We will present an answer that is in between the two extremes, based on both experimental data as well as on modelling studies of how consciousness and attention could function in the brain. The paper can be seen as an extension of the CODAM model to an in-depth discussion of the functionality thereby achieved of the model, and how this relates to what would be expected for consciousness seen as part of the overall control system of attention.

We start in the next section with a brief account of what aspects of consciousness we propose to analyse for determining its functionality. This shows the bias of our approach, leaving out some aspects dear to the hearts of some, bringing in other aspects which are anathema to them. But we have to restrict our analysis due to the enormous breadth of features that consciousness possesses. In the following section we review experimental and model-based approaches supporting the notion of attention as a control system; in Section 4 we briefly present the CODAM control model for attention, which has both generality for taking account of new understanding as well as specificity of the modules proposed in the model and their functionality to fit presently known data on attention.

How such a model of attention could help create the experience of consciousness of stimuli or movement is also discussed in this section. Possible answers to the basic question of the functionality of consciousness are then considered in the light of this discussion in Section 5. Predictions and possible tests are described in the final section.

We add that in this paper we consider the overall information flow at a global level, not down at the level of neurons or even clusters of them. In other words we are attacking the problem of the function of consciousness from a brain-based functional point of view. We take guidance from the brain, but do not try slavishly to incorporate into our functional models the multifarious highways and byways the brain possesses. Thereby it is hoped that the models being discussed can possibly be implemented in a number of ways. This possibility has been made clear in the case of the comparison of bird and mammalian powers of reasoning, for example – in which the crow and the chimpanzee are thought to be of equivalent status – whereas recent research has indicated the very different cortical architectures on which the bird brain and the mammalian brain are based (Jarvis & the Avian Brain Nomenclature Forum, 2005): one has a six-layered cortex, the other a one-layered one.

## 2. The nature of consciousness

To start, I briefly list some of the most important aspects of consciousness, so as to ground the discussion of consciousness in those aspects that are recognised as playing important roles. These components can be listed as follows, starting with arguably the two main components into which I suggest consciousness can be divided:

- (1) The presence of content in consciousness/awareness of external stimuli. This corresponds to the smell of the rose, the taste of the glass of wine, the feel of the smooth texture of satin or skin, and many, many other sensory experiences we have in of the external world, together with the sensations we have of memories and of our imaginations or day-dreams;
- (2) The presence of an ‘owner’ of these experiences, such that the content mentioned above is experienced by ‘some one’, not by nobody. This leads to the problem of who that internal being or owner could be. This relates closely to the meaning of the pronoun ‘I’ and to the difficult question: to whom does ‘I’ refer? It does not refer to the person I see in my mirror when I shave or wash my face or comb my hair. No list of attributes of this ‘I’ could be error-free, since they could always have been set up, without my knowledge, by mischievous friends or enemies. For example they could have replaced the image of my face in my mirror by that of someone else altogether; I would be in error of thinking that face denoted ‘I’. So the ‘I’ has the property of knowing that it is indeed the error-free ‘I’: it possesses the property of “lack of error through misidentification of the first person pronoun” (Shoemaker, 1968). Thus I cannot ask you, when you tell me you are in pain, ‘are you sure that it is you who is in pain?’ You just are sure.

I will denote the owner to be the ‘inner self’. That there are these two components of consciousness – inner self and content – is controversial and has not been accepted by many in cognitive science or cognitive neuroscience (until possibly very recently). However the influential paper of Block (1995) indicated precisely that such a bipartite division is necessary and underlined the strong evidence for such a division from Western phenomenology developed over the previous century and even before (Zahavi, 1999). In the light of these developments, it is apparent that the proposed bipartite division of consciousness is now becoming more accepted. However it is absolutely necessary to put the bipartite model to experimental test, and that will be discussed later in the paper.

There are numerous further aspects of these two main components of consciousness that need to be explored so as to make it clearer what are the problems we face in explaining and modelling the associated phenomena. These cover (though not necessarily completely):

- (a) **Unity:** The world is experienced as a unified structure. I am a singlet in terms of this experience. I do not exist doubly or trebly or in any other multiple. Nor does part of ‘me’ experience one thing, and another part of ‘me’ a second. This is in spite of the neural activities in my brain by which I represent the world being split up across the vast expanses of brain and across different modalities. Such fusion could arise by attention, since it singles out only one component of the complexity of the world, although the attention system and what it acts on is also spread out across the brain. A similar spread is expected of any set of ‘I’ modules coding for the inner self. There is only one ‘I’.
- There are indications that this unity of ‘I’ might break down in the presence of degraded stimuli, where different responses, for example from the eyes, the hands or the feet can be quite different in the presence of the same degraded input (Marcel, 1994);
- (b) **Binding:** This is based on the important feature that we see the external world in terms of single entities, with all features bound together, not split up into its components such as the various features into which a visual stimulus is split by the hierarchical set of modules in the visual cortex (so this feature is related to that under (a) of unity). Binding involves the mechanisms by which different components of neural processing are fused into a single percept, such as by synchronised oscillations or by common attention amplification of the various components into which the neural representations of sensory stimuli are decomposed;
- (c) **Possibility of (and explanation of) a broad range of states of consciousness** (sleeping, dreaming, drug-induced hallucinations, ranges of mental dysfunction such as schizophrenia, Alzheimer’s disease, autism spectrum disorder, and so on). Both content and inner self are expected to be involved in different ways in these experiences;
- (d) **Transparency:** We experience objects in a fashion which seems not to involve any intermediaries that ‘clog up’ the vision we have of these stimuli. Thus the inner self, for example, does not intrude on our direct experience, nor do

the lower-level processing stages being bound together as under (b) above;

- (e) **Infinite closeness:** This is a property related to transparency: we have no distance at all between our experience of external stimuli and the stimuli themselves, as if ‘we’ are bound up intimately with ‘them’ without any gap between oneself and the experienced stimulus.

Elsewhere (Taylor, 2001) I have discussed how the properties (a)–(e) may be obtained from the neural networks (as based on standard neural networks) of the brain. Moreover there are numerous models of continued activity in various modules that could support the nature of the contents of consciousness, so providing a detailed response to point (1) above. So let us turn to the more difficult question associated with point (2) above, that of the inner self, or the ‘I’. It is exactly this component that is most difficult to construct as part of any neural model of consciousness, and presents also the greatest difficulty for experimental investigation.

### 3. A general control model for attention: CODAM

It has been suggested by many investigators since the time of Aristotle that attention is a crucial pre-requisite for awareness or consciousness. As such it appears necessary to investigate the powers that attention possesses most carefully in order to further probe inside its intimate recesses so as to tease out how consciousness can thereby be supported by attentive processing (if it so can — perhaps it cannot which would itself be a very interesting result). The studies reported in the numerous papers on attention can thus be seen as helping progress towards uncovering those parts of attention that are necessary, if not sufficient, for consciousness.

Attention has already been recognised through many experiments to act as a filter processing system. In order to clarify how this might be achieved in general we consider the two sorts of attention that are now appreciated as different in character: top-down (endogenous) and bottom-up (exogenous). The former of these uses guidance from relatively long-term goals (very likely held in prefrontal cortex) set up when a specific psychological task is being carried out by a subject (such as the attentional blink or a GO/NOGO task, etc). The goal bias is thought to bias attention to move to the next site for attention focus in posterior cortices, so comes under the heading of the influential ‘biased competition’ model (Desimone & Duncan, 1995).

On the other hand exogenous or bottom-up attention requires some form of ‘break-through’ of a lower level stimulus representation which carries great salience (such as a fire alarm in the building you are sitting in as you read this paper). The nature of salience has been studied experimentally by numbers of experimental groups as well as encompassing models built to incorporate how salience can be used to guide the direction of the focus of attention (Koch & Ullman, 1985; Walther & Koch, 2006). It is also known that there is considerable overlap of brain sites involved in bottom-up attention control as compared to top-down control, although the exact degree of such overlap is still under discussion (Kincade, Abrams, Astafiev, Shulman,



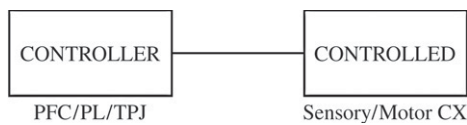


Fig. 1. The two-component nature of attention. The two-component model of attention is shown in the figure as being composed of the controlled part (sensory and motor cortex) and the controlling part (prefrontal cortex, PFC, parietal lobe, PL and the temporo-parietal junction TPJ).

& Corbetta, 2005). It is possible that the bottom-up salient goals are also sent to the prefrontal cortex, similarly to the endogenous goals, as shown by observation of early prefrontal activation of visual stimuli (Foxe & Simpson, 2002).

These experimental results and theoretical approaches can be summarised as involving top-down bias as goals, which will have been set up in prefrontal cortices either as endogenous signals entered as task rules, say, from experimental instruction or as exogenous biases from lower cortices from salient inputs. This bias is transmitted to an attention movement signal generator (inverse model controller or IMC) which then sends a new attention signal to lower level cortical stimulus activity; this can be summarised as a two-stage model, in which the higher level control system generators (goals and IMC) send attention signals to lower level cortical representations (Corbetta & Shulman, 2002; Corbetta et al., 2005; Kanwisher & Wojciulik, 2000). A simplified version of this is shown in Fig. 1. In the figure, the controller component is composed of the goals module acting as a bias to send out an attention signal from the IMC to feedback to the input modules acting as those which are controlled by attention.

There are already various models of attention which have been studied in the recent past, ranging from those of a descriptive form, such as the already-mentioned influential 'biased competition' model of attention (Desimone & Duncan, 1995) to the more detailed neural-network based models involving large-scale simulations, such as those of Deco and Rolls (2005) or of Mozer and Sitton (1998). However these and other neural models of attention have not had a clear overarching functional model guiding their construction. If we consider the recent results on attention of brain imaging experiments (Corbetta & Shulman, 2002; Corbetta et al., 2005; Kanwisher & Wojciulik, 2000) then we find that the language of engineering control theory (see for example Phillips and Harbor (2000)) could be applied to help understand the complex-looking network of modules observed to be involved in attention effects. Already the 2-stage model of attention as a control generator system together with a controlled system described in the previous paragraph indicates the simplest 'ballistic control' model of this sort. A more general engineering control approach will be employed in this paper, encompassing and surpassing the simple ballistic control model. This will allow us to develop a more detailed neural modelling framework to help understand the nature of networks involved in higher order cognitive processes, even leading to suggestions for the creation of consciousness. More specifically it will allow us to suggest some very specific functions for consciousness which have detailed experimental implications.

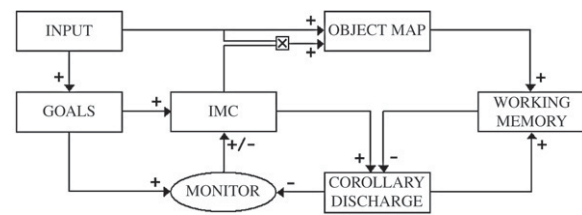


Fig. 2. The CODAM model architecture. The figure shows the modules of the CODAM model of attention control, based on engineering control mechanisms. Visual input, for example, enters at the INPUT module and is sent, through a hierarchy of visual processing modules, to activate the object map module, OBJECT MAP. At the same time in the exogenous case it rapidly accesses the GOAL module, so causing bias to be sent to the inverse model controller denoted IMC in the figure (the generator of the signal to move the focus of attention). This sends a modulatory feedback signal to the object map, of multiplicative or additive form, to amplify the requisite target activity entering the object map. As this happens the corollary discharge of the signal from the IMC is sent to the MONITOR module, acting as a buffer for the corollary discharge signal. This can then be used both to support the target activity from the object map accessing its sensory buffer, the WORKING MEMORY module, and to be compared with the requisite goal from the GOAL module. The resulting error signal from the monitor module is then used to enhance the IMC attention movement signal and so help speed up access as well as reduce the activities of possible distracters.

These will be explored later, after we have briefly reviewed the CODAM model.

The engineering control approach to attention was developed in the Corollary Discharge of Attention Movement (CODAM) model in Taylor (2000, 2003) (see also Taylor (2002a, 2002b), Taylor and Fragopanagos (2005)) and used in Taylor and Rogers (2002) to simulate the Posner benefit effect in vision. It was further developed in the CODAM model application to the attentional blink in Fragopanagos, Kockelkoren, and Taylor (2005), and more recently in numerous applications of CODAM to working memory tasks (Korsten, Fragopanagos, Hartley, Taylor, & Taylor, 2006) as well as to help understand results observed by brain imaging of paradigms involving emotion and cognition in interaction (Taylor & Fragopanagos, 2005). Here I will use these various applications, and their associated models, to provide a unified description of the observed effects and to lay a framework for further extensions into cognition: to reasoning, thinking and planning and ultimately to consciousness.

Fig. 2 is a schematic diagram of the CODAM architecture. The input enters the system through the module labelled 'visual cortex', and is passed to the 'objects' module (where high level visual representations have been stored after suitable hierarchical learning). Attention acts by a bias arising from the 'goals' module to guide the 'attention controller' module to send out a signal changing the focus of attention by altering the sites of modulation of the input to the visual cortex and object modules (or to a suitable spatial map that can be included in CODAM). The activation in the goals module can arise either top-down (from rules set up as part of an experiment, for example) or bottom-up (by fast activity going to prefrontal cortex as observed by various EEG and fMRI studies mentioned earlier). These modules: the goals, attention controller and object/visual cortex modules, form a very simple form of

ballistic controller in an engineering control framework; they fill out the simple controller/controlled dichotomy in Fig. 1.

The additional modules in Fig. 2 beyond the simpler controller of Fig. 1 include a ‘monitor’ module that computes the difference between the desired goal state and the estimated state of the system. This latter state is not that usual in control applications, since in the latter the estimated state usually consists of that of the whole system being controlled. For example in our case that would be the visual cortex activity as an estimate of the external world. But that is already present as well as being complex in general. Instead the attended state estimate is a strongly cut-down version of this complex state, the cutting down being achieved by attention acting as a filter. This state, the reduced ‘attended state of the world’, is proposed as being created in the buffer module in Fig. 2 denoted by ‘Forward’. This module acts as both a predictor of the next attended state of the world as well as its estimate. We note that there are connections to this module from two sources: (1) from the input stimulus at both visual cortex feature level and object representation level; (2) from the IMC.

The first of these inputs is to provide content to the attended state, for future use in report to higher level modules for the various cognitive processes we are considering. The second of these is well known in engineering control theory as the ‘corollary discharge’ or ‘efference copy’ of the control signal generated by the IMC. This term was originally introduced in Von Holst and Mittelstaedt (1950) to describe a copy of the oculo-motor control signal; the process was also considered in Sperry (1950). It allows there to be a rapid modification of the control signal if it is in error as determined by the monitor module. It also helps speed up the amplification/inhibition process in the lower level sites being applied to the relevant representations there. Thus the attention copy is an important component of the CODAM model of Fig. 2, using as it does good lessons from engineering control as to how to improve the overall control processing efficiency.

CODAM has been applied to the particularly important and subtle paradigm, the attentional blink, in Fragopanagos et al. (2005). This paradigm consists of subjects processing a rapid serial visual presentation of stimuli such as white digits on a black background at a rate of about 10 Hz. The subject’s task is to detect a target (denoted T1), such as a white letter, and then to report a second white letter target (denoted T2) that might appear in any of the next eight to ten slots in the presentation schedule. It has been found by numerous experiments (Vogel et al., 1998) that there is considerable reduction of accuracy if T2 follows T1 with about a 300 ms gap: this is termed the ‘attentional blink’ (AB), when attention is processing T1 to the exclusion of T2.

The explanation of the AB in terms of CODAM is that the process of T1 attaining the working memory site for report uses inhibitory processes to protect its accuracy, so that another input T2, acting as a distracter, is thereby inhibited from also entering the working memory site until this processing is completed. The inhibition can have a number of sources, such as from the working memory activation of T1 to the beginning activation on the same module of T2. We proposed in Fragopanagos

et al. (2005) that this inhibition was both of feedback from the T1 activation onto the early T2 activation on the working memory module, as well as inhibition from T2 to T1 brought about by the corollary discharge copy signal of the attention movement to focus on T2. Other sources could be associated with inhibition from T1 onto other sites of the attention copy signal.

Recent experimental data (Sergent, Baillet, & Dehaene, 2005) have provided initial evidence for such an inhibitory process occurring. The authors studied the AB and divided the subjects’ responses into those where awareness of T2 occurred and those where it did not. In the former there was a distinct reduction of the P3 signal for T1 (where P3 is taken to be an indication of the working memory activation for report of that stimulus). There was also a reduction of the N2 signal for T2, so indicating inhibition from the relevant stimulus representations of T1 to T2, so supporting the inhibitory thesis.

In conclusion there is initial evidence in support of the CODAM model structure, and in particular the presence and use of the attention copy signal for improving processing in the face of distracters.

#### 4. Creating consciousness through CODAM

We will now consider how we might implement an ‘I’, possessing an owner immune to the error of self-identification, in the CODAM model of attention control in the brain. The CODAM model gives a detailed model of attention control, containing more complex and efficient mechanisms than purely that of ballistic control discussed earlier in Section 2. CODAM thereby has the ability to explain a considerable amount of brain-based data on attention (Korsten et al., 2006), as described earlier in this paper and elsewhere (Fragopanagos et al., 2005; Taylor, 2003, 2005). The model also has the ability to enable activity of an ownership character to be located in the model (Taylor, 2002a, 2002b). I will follow here the recent discussion in Taylor (2007), but now modified to take account of the role of attention leading to the creation of an ‘owner’.

The neural basis for the owner has been suggested in CODAM (Taylor, 2000, 2002a, 2002b, 2003) as being the corollary discharge signal of the attention movement control signal being buffered for a short time on the forward module of Fig. 2. Such an approach thus comes under the heading of ‘attention copy models of consciousness’. The attention copy signal is taken here as the basis of the experience of ownership of the about-to-be experienced content of consciousness. It is this sequential process – first ownership, then content that is owned – which is taken as the basis of the two separate components of consciousness. Each component would be lost without the other: no owner implies ‘no-one’ to experience the content (which therefore loses its attribute of being ‘content’), and no content would imply absence of the external world (although the owner could still experience itself, as in the controversial experience of pure consciousness). We should add that in a paradigm like the attentional blink, in which awareness can be manipulated by specific parameter change (the T1–T2

time delay) loss of the P3, in the blink, has associated loss of the N2 (Vogel et al., 1998).

The owner activity is taken to be signalled by some aspects of the higher cortical level N2 activity in the 180–250 ms post stimulus period. There are various inhibitions (of distracters) and excitations (of the target) that this signal produces to speed up the target activity reaching its buffer, as contained in the distribution of the N2 about the brain. The activity of the attention copy signal is thus activating modules coding at a high information level, so would be very likely too high to produce any experience of content (which arises only from the correlated lower level activity carrying feature information of higher level object concepts). This fits with the notion of the experience of the owner as being ‘content-free’.

There are several questions which then need to be considered. An important one is about ‘I’ in its role in episodic memory: each such memory carries with it the imprimatur of ‘I’ as a label indicating that an episodic memory is one of an event at which ‘I’ was present and it is seen through ‘my eyes’. Each of the components of such an episodic memory is composed of a pair of sequentially encoded items: firstly the ‘I’ signal, and secondly the signal of the content, containing both context and main items. The hippocampus is thought to be able to support such short memory chains, especially since this will be no longer than about a second, and will likely be shorter. Such episodic memory could be equated with the ‘episodic working memory’ of Baddeley (2000), although need not be identical with that system.

It has been suggested elsewhere (Taylor, 2007) that the ‘I’ content of episodic memory could arise from the encoding in hippocampus of the early corollary discharge signal associated with a given stimulus, with the content being coded some hundred or so milliseconds later. This would allow readout of the memory only provided that the ‘I’ component was read out first into the cortical sites from whence it came, so recreating the experience of ownership. It thus provides a clear-cut distinction between episodic and non-episodic memory. The ‘I’ thus re-activated would carry with it the ancillary components from limbic and other high-level sites which can give close relationship to the personal characteristics of the owner (which flow through into the reflective self). The manner in which this occurs needs more careful study than can be given here.

## 5. The possible functions of consciousness

The CODAM model posits that consciousness has been created by the efference copy or corollary discharge signal for the movement of the focus of attention. This copy signal is used, as usual in engineering control models, to speed up access to the state estimator. In our case this state estimator, as the working memory site, was proposed as the site creating the basic activity for the awareness of content. Correlated activity in lower level cortices is also needed to create a veridical experience of external stimuli at a certain level of detail (produced by one of the various mechanisms of binding discussed by many others). There is an associated reduction of errors in the attention focus process by early monitoring using the copy signal.

The first of these functions – of creating the amplified working memory and lower-level correlated activity – is proposed as being achieved by use of the corollary discharge signal to directly amplify/inhibit target/distracter activity reaching the working memory (report) site in parietal and lobes. The second – of error correction – is proposed to use the error signal from the monitor to amplify the attention control signal itself and hence boost the various components (including the attention copy signal) to ensure better target amplification. At the same time there is achieved better inhibition of distracter activity trying to attain report on the working memory site. These two functions – speeding access, and reducing distracter errors – have good survival value.

The above interpretations of CODAM dynamics are thus that an ownership component of consciousness is created as an intrinsic part of the speeding-up/error correcting processes: this ownership component is contained in the attention copy signal itself. From this point of view the content of consciousness, as that generated by attention applied to lower level neural activity, is complex, but not created efficiently enough to allow basic survival. The speed-up process brought about by the owner, through the use of the attention copy signal as specified above, is necessary to avoid death and destruction by fast-moving predators.

These, in a nutshell, are the functions now able to be recognised for consciousness in the CODAM model — speed-up and error correction. It might be charged that the model is so tailored that it has these two functional components built into it. However it is to be noticed that these components are only part of the more general gallery of tricks in the array of the magicians of control theory; we have done no more than borrow standard structures from such theory and apply them to attention control phenomena. It is to be expected that nature would have done the same in order to steal a march on any non-CODAM based attention systems. In the process we have been able to see how the expanded attention control system could hold in its now more ample bosom, beyond that of ballistic control, the important owner happily enfolded.

We now turn to consider the question as to how ‘content’ might seep into the ownership signal. This latter signal arises, it has been suggested, as part of the N2 complex throughout the brain. It was noted in the previous section that N2 has been observed as occurring in various parts of the brain — parietal, temporal and frontal lobes, in the hippocampus, in the limbic system, etc. There is a large amount of information in these signals, spread across a variety of modalities and functionalities. Thus memory is involved (from hippocampus), emotions (from limbic areas), rules, goals and decision (from frontal cortex), and so on. Is all (or any) of this information relevant to the owner, and if so how is it employed in the early processing at the time of creation of the N2 signal? And especially how much of the N2 signal is relevant to the ownership signal?

This question is of importance in attempting to predict the nature of any ‘inner experience’ that may be associated with the ownership signal. It was noted in Section 2 that the



ownership component of consciousness was basically content-free. Here we seem to be bringing in more content, so destroying this content-free character. How can we reconcile these two apparently opposing aspects?

Ways to achieve this reconciliation are through the fact that the representations with content activated as part of the N2 would be either at too high a level of representation to be able to contribute directly to awareness of content (so be experienced as content-free), or if they did have content their contribution would be only of a background form. They would then only arise at the end of the brief ownership period proper. The latter fits in well with the bias put into interpretations of the cessation of the phenomenon of ‘pure consciousness’ (Forman, 1999; Taylor, 2002a, 2002b).

The latter experience of pure consciousness is brought about by long periods of meditation to develop the ability to attend only to the attention circuitry itself. Thus posterior cortical sites are inhibited, as observed by brain imaging as well as from subject report. Prefrontal sites are observed active, as is expected if a goal has been set up by meditation training to inhibit all sensory input. Thus the interpretation of this experience could well be coloured by the ancillary regions activated by the N2 (limbic, mid-temporal episodic, and other related areas). On the other hand there is not usually direct awareness of codes in these sites, so fitting the first suggested mechanism for their absence from ownership awareness: they are at too high a level of coding to enter awareness as containing content. However these further activations could well colour entry and exit experiences from pure consciousness as described in many accounts; such colour is thus explicable in the above manner.

In the case of either explanation the ownership activity is clearly the forerunner to that of content, which is speeded up by the use of this attention copy signal in the ways proposed earlier and used in the simulation of the attentional blink (Fragopanagos et al., 2005). The resulting function of this attention copy signal therefore is to be identified as that associated with corollary discharge signals in standard engineering control models. Consciousness, as based on the ownership signal acting as the ‘I’, now has clear functions with a control basis.

It should be noted that the extension of the ownership component of consciousness to pure consciousness, in the manner suggested through CODAM (by means of the attention copy signal to loop back on itself), would become counter-productive. As noted by the ancient Buddhist monk Hakuin, the practitioners of meditation into pure consciousness were just ‘sitting like bumps on a log, nodding off’ (Stevens, 1999). In other words the meditative state takes one out of contact with the world of objects and agents into a world of nothing at all. That one’s attention control circuitry can be modified to such an effect that such a rare state continues over minutes or hours is remarkable. But the remarkable state is not one in which the subject is necessarily more efficient in dealing with their surroundings. It has been claimed that there is better sensitivity to external stimuli in this state; there may, however, be a decrease in levels of creativity and flexibility.

We need to return to more brain-based aspects. In particular we need to discuss the question of the relation of the N2 to ownership in terms of breakthrough into attention by salient stimuli. At a level of V4 or thereabouts we can expect that colour, for example, will act as a highly salient feature, causing breakthrough into attention control and awareness of a colour singleton (Hickey, McDonald, & Theeuwes, 2006). In the paradigm of those authors, various stimuli were presented for a short display to a subject, who had to detect a singleton of a particular sort (associated with the stimulus shape) whilst in the presence of a distracting colour singleton. An N2pc (calculated as the difference between the contra-hemispheric and ipsi-hemispheric N2 activations) was detected for the subjects. When the colour distracter and the target were presented in opposite hemispheres, the related N2pc was initially processing the salient colour distracter but then moved to the opposite side to process the target.

These results can best be understood in terms of the N2pc indicating the movement of the focus of attention from the colour singleton to the desired shape singleton. This could have been brought about by initial amplification by attention of the most salient colour singleton. A monitor then assesses the result of the initial attention-amplified putative target, with comparison being made with the required constraint, which in the case of Hickey et al. (2006) was that the target be the shape singleton. An error signal from the monitor is then sent to change the position of the focus of attention to the next salient position, which is indeed the required shape singleton.

Thus we can be certain that the N2pc is tagging where the early focus of attention is being directed. It is not so clear for the total N2 signal, which may involve more than purely attention feedback, but also involves more immediate (non-attention-based) feedback from one or more higher cortical level. As pointed out in Jolicouer, Sessa, Dell’Acqua, and Robitaille (2006) when the attentional blink is at its highest there is no observable N2pc. There may also be no N2 itself, as observed in Vogel et al. (1998), although that may depend on the paradigm used. So presently it is uncertain how much of the N2 is involved in attention control and generation of the ownership activity. All we can conclude is that there is at least some component of attention, showing up as the N2pc, which is functioning as an indicator of attention amplification/inhibition.

This interpretation is supported by the MEG results of Hopf et al. (2000), who observed an initial N2pc activation in the superior parietal lobe at about 180–200 ms after stimulus onset in a GO/NOGO paradigm, which was then followed by temporal lobe activity at 220–240 ms. The first of these activities could be ascribed, following CODAM, to the action of the IMC attention movement signal generator, followed by the activation, by means of a corollary discharge, of the temporal lobe site. This latter can be regarded as ‘preparing the ground’ for the more detailed stimulus representation activity arising from lower level visual cortex activity amplified by attention feedback. This is precisely the CODAM picture described earlier for the neural activity of the owner.

An interesting and important results reported in Jolicouer et al. (2006) was that even when there was no N2pc, in

the attentional blink, there was still a non-zero level of T2 detection. This was taken to indicate that there was possibly some attention carried by the total N2, symmetrical on either hemisphere; there thus may be some further component of the N2 involved in the ownership experience.

A further relevant result from Jolicouer et al. (2006) is the presence of a central executive, as seen by interference caused by other tasks in other modalities on the visual AB task. The presence of such an executive, also suggested by the work of others, allows the better understanding of the process of unity of the self: there is a multi-model competition which is run on the IMC, say in parietal lobe, so as to produce a unified self, although a number of tasks in different modalities may be carried out in the process. This aspect needs considerable further investigation before it is fully clear as to what is occurring.

Finally we turn back to the results of Libet et al. (1979) mentioned at the beginning: awareness of making a movement occurs some 300 msec after the movement potential has begun. But this is only similar to the time elapsed from a visual stimulus input to awareness of the stimulus. In other words awareness in both cases follows stimulus input, either as a visual or a sensori-motor input. Thus the relevant question for the motor action should be: what caused that sensori-motor input in the first case? But that must have arisen from some earlier goal set up prefrontally, so is not a mystery after all.

It might be claimed that there is still something unsatisfactory about the answer: for vision we are given the input, for motor actions we decide what action to take. Thus there is still the problem as to the delay in awareness of the decision about the action relative to its commencement. Awareness/consciousness still seems not to have a real function here. However we can respond that in both the sensory and sensori-motor cases consciousness is driven by outside activity impinging onto the attention control circuitry, biased by goals of a variety of forms. In the sensori-motor case the bias was noted as the goal already set up by the experimenter. No other goal was present; no veto was asked for and was not present in the biasing goal. There could be different dynamics (as in the GO/NOGO situation) where there is important early activity associated with the frontal N2. This situation needs to be investigated in the context of the functionality of consciousness as part of attention control circuitry; that is yet to be done. But both cases – visual awareness and awareness of motor response – can be explained consistently in the CODAM picture: awareness arises as part of the overall attention dynamics. It achieves the task: ‘make the movement’ or ‘see the stimulus’ respectively. Awareness is involved as a crucial functional component as earlier described.

## 6. Predictions and tests

We have presented a brief overview of the CODAM model for the creation of consciousness by brain activity. The development has been guided by analysis of the nature of consciousness (as discussed in Section 2) and the development of an engineering control approach to attention in terms

of CODAM (as considered in Section 3). This emphasised the use of modern engineering control concepts as enabling increased efficiency to be incorporated into the simpler form of the ballistic control model employed in many discussions of attention. In the following section it was shown how the two components of consciousness – owner and content – could arise from the internal dynamics in the CODAM architecture. This led to extraction of a clear functionality for consciousness presented in Section 5, as a speeding-up and error correcting process in the movement of attention to a given stimulus. This speed-up is achieved by employing the corollary discharge or efference copy signal from the movement of attention signal to generate a predictive attended state of the world. This state can then be used to generate an error signal and so correct the attention movement, as well as speed it up if there were a large number of distracters present causing possible errors: these could be inhibited. This process was used effectively in a recent simulation of the attentional blink (Fragopanagos et al., 2005). Further details of the functionality of consciousness were discussed from this point of view in that section.

There are a number of tests and predictions that arise from the CODAM model and the paradigms to which it has been applied:

- (1) Firstly there are a number of questions or tests (with associated predictions in terms of CODAM-type models which could be developed) arising from considerations of comparative and developmental neuro-anatomy:
  - (a) Did the attention control system evolve from the simplest ballistic control to add further modules (error monitor, working memory, corollary discharge allowing early state prediction) simultaneously or did certain of these (such as the working memory site) evolve first as added to the ballistic control and then the corollary discharge and its use in early state prediction evolve later?
  - (b) In more detail could one follow up to higher-level animals possessing structures granting even more speed to the attention control response, leading to ourselves and some other high-level animals possessing, and using as in CODAM or a similar model, an attention copy signal?
  - (c) How would the attention control system appear to evolve in birds, where one would expect a similar proliferation of modules as one goes from birds with low-level cognitive powers finally up to the corvids with their renowned capabilities about equal to those of chimpanzees (Emery & Clayton, 2004; Hurley & Nudds, 2006).
  - (d) On the developmental side, does development follow evolution or is there a different order for the further CODAM modules to come on stream in the growing child?
- (2) Secondly there is the important question: what other possible paradigms could the attention copy signal show itself? It is important thereby to track this signal down in all its possible guises so as to put CODAM on a sure foundation.

The complete list of such paradigms includes:

- (a) The attentional blink, as already discussed;



- (b) Paradigms involving the N2pc, following the comments at the end of the previous section on the results of Hopf et al. (2000) and such as that of Hickey et al. (2006), in which salience is moved from one stimulus to another by means of competition between bottom-up and top-down requirements (exogenous versus endogenous goals);
- (c) Any other possibilities, such as in subliminal processing and the various paradigms considered in the adjoining paper (Taylor & Fragopanagos, 2007).
- (3) Thirdly how much dissociation can there be between the attention copy signal and the P3 of report/access to the working memory site? This is another important question since there are both neuro-physiological as well as experiential effects to explore. The former are well understood in terms of the various imaging techniques now available; the latter are also understood in terms of straight report, but here we are exploring the hinterland between conscious experience and that possibly at subliminal level (as might arise from priming by the attention copy signal) and even lower at purely unconscious level (where there has been no attention copy signal either). In particular what are the experiences of those subjects in paradigms where there is an N2 but no P3, or a P3 but no N2 (if that were possible)? Do they have a normal conscious experience or is there something missing?
- (4) Fourthly how can the various deficits in awareness due to brain damage or disease be explained through the attention copy model of CODAM? This is an important question: it has been discussed at length in Taylor (2006); due to length constraints the reader will be referred to that discussion for an initial treatment. Suffice it to say that there is considerable light that might be able to be cast in the manner in which experience is thereby modified;
- (5) Fifthly how far can the attention copy model go in helping understand the changes in experience under drug-induced states and other altered states of awareness? We have already considered the extreme state of pure consciousness as an altered state of consciousness, and given an explanation for it in terms of continued activation of the attention copy signal to preserve itself so that it inhibits all possible sensory input. The manner in which drugs act can be in a variety of ways: distortion of the input stimulus representations, of the feedback control signals (as might happen in hallucinations, where illusory stimuli are activated by feedback), of both, of goal states, and so on. Again this has been discussed in more detail in Taylor (2006), to which the interested reader is referred, although this is still only a preliminary investigation.
- (6) Sixthly there are numerous questions that can be asked about the detailed dynamical flow of activity in CODAM. Some of these have already been raised in association with the simulation of the AB through CODAM (Fragopanagos et al., 2005) and a discussion given in Taylor (2007). In any case there are more general questions as to the flow of activity, for example in comparison between exogenous and endogenous attention paradigms, for example. The data from (Fuxe & Simpson, 2002) were noted earlier as being

very relevant to this exploration, although more recent data using TMS (Chambers, Stokes, & Mattingley, 2004) has begun to uncover the complexity of the flow of information through the parietal lobe control components.

It is clear from the above lengthy list of questions and predictions that there are many open points still to be clarified, leading to possibly considerable changes to CODAM. This is for the future to decide.

### Acknowledgements

The author would like to thank the EC Cognitive Systems Unit for grants under the GNOSYS and MATHESIS projects, as well as EPSRC and BBSRC funding councils of the UK for further grants under which this work has been performed in an attempt to obtain a broad view of brain processing and functionality. He would also like to thank his colleagues Neill Taylor, Matthew Hartley, Nikos Fragopanagos and Nienke Korsten for numerous stimulating discussions on brain processing.

### References

- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioural Brain Science*, 18, 227–247.
- Chambers, C. D., Stokes, M. G., & Mattingley, J. B. (2004). Modality-specific control of strategic spatial attention in parietal cortex. *Neuron*, 44, 925–930.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews, Neuroscience*, 3, 201–215.
- Corbetta, M., Tansy, A. P., Stanley, C. M., Astafiev, S. V., Snyder, A. Z., & Shulman, G. L. (2005). A functional MRI study of preparatory signals for spatial location and objects. *Neuropsychologia*, 43, 2041–2056.
- Deco, G., & Rolls, E. T. (2005). Attention, short-term memory, and action selection: A unifying theory. *Progress in Neurobiology*, 76, 236–256.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.
- Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science*, 306, 1903–1907.
- Forman, R. K. C. (1999). *The problem of pure consciousness*. Oxford: Oxford University Press.
- Fuxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. *Experimental Brain Research*, 142, 139–150.
- Fragopanagos, N., Kockelkoren, S., & Taylor, J. G. (2005). A neurodynamic model of the attentional blink. *Cognitive Brain Research*, 24, 568–586.
- Hickey, C., McDonald, J. J., & Theeuwes, J. J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of Cognitive Neuroscience*, 18(4), 604–613.
- Hopf, J. -M., et al. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex*, 10(1), 231–241.
- Hurley, S., & Nudds, M. (2006). *Rationality in animals?* Oxford: Oxford University Press.
- Jarvis, E. D., & the Avian Brain Nomenclature Forum, (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews, Neuroscience*, 6, 151–159.
- Jolicouer, P., Sessa, P., Dell'Acqua, R., & Robitaille, N. (2006). On the control of visual attention: Evidence from human electrophysiology. *Psychological Research*, 70, 414–424.
- Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience*, 1, 91–100.
- Kincade, L. M., Abrams, R. A., Astafiev, S. V., Shulman, G. L., & Corbetta, M. (2005). An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. *Journal of Neuroscience*, 25(18), 4593–4604.

- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention — towards the underlying neural circuitry. *Human Neurobiology*, 4, 214–227.
- Koch, C., & Tsuchiya, N. (2006). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences*, 11(1), 16–22.
- Korsten, N., Fragopanagos, N., Hartley, M., Taylor, N., & Taylor, J. G. (2006). Attention as a controller. *Neural Networks*, 19(9), 1408–1421.
- Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7, 12–18.
- Lamme, V. A. F. (2006). Towards a true neural stance to consciousness. *Trends in Cognitive Sciences*, 10(11), 494–501.
- Libet, B., Wright, E. W., Feinstein, B., & Pearl, D. (1979). Subjective referral of the timing for a conscious sensory experience. *Brain*, 102, 193–224.
- Marcel, A. J. (1994). In C. Peacocke (Ed.), *What is relevant to the unity of consciousness? In objectivity simulation and the unity of consciousness*. Oxford: Oxford University Press.
- Mozer, M. C., & Sitton, M. (1998). Computational modelling of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 341–393). New York: Taylor & Francis.
- Phillips, C. L., & Harbor, R. D. (2000). *Feedback control systems* (4th ed.). New Jersey, USA: Prentice Hall.
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, 8, 1391–1400.
- Shoemaker, (1968). Self-reference and self-awareness. *Journal of Philosophy*, 65, 556–570.
- Sperry, R. (1950). Neural basis of the opto-kinetic response produced by visual inversion. *Journal of Comparative Psychological Physiology*, 43, 482–489.
- Stevens, J. (1999). *Zen masters*. London: Kodansha International.
- Taylor, J. G. (2000). Attentional movement: The control basis for consciousness. *Society for Neuroscience Abstracts*, 26, 2231#839.3.
- Taylor, J. G. (2001). The importance of the parietal lobes for consciousness. *Consciousness and Cognition*, 10, 379–417.
- Taylor, J. G. (2002a). Paying attention to consciousness. *Trends in Cognitive Sciences*, 6(5), 206–210.
- Taylor, J. G. (2002b). From matter to mind. *Journal of Consciousness Studies*, 6, 3–22.
- Taylor, J. G. (2003). Paying attention to consciousness. *Progress in Neurobiology*, 71, 305–335.
- Taylor, J. G. (2005). From matter to consciousness: Towards a final solution? *Physics of Life Reviews*, 2, 1–44.
- Taylor, J. G. (2006). *The mind: A user's manual*. Chichester: Wiley & Sons.
- Taylor, J. G. (2007). On the neurodynamics of the creation of consciousness. *Cognitive Neurodynamics*, 1, 97–118.
- Taylor, J. G., & Rogers, M. (2002). A control model of the movement of attention. *Neural Networks*, 15, 309–326.
- Taylor, J. G., & Fragopanagos, N. (2005). The interaction of attention and emotion. *Neural Networks*, 18(4), 353–369.
- Taylor, J. G., & Fragopanagos, N. (2007). Resolving some confusions over attention and consciousness. *Neural Networks*, 20(9), 993–1003.
- Vogel, E. K., Luck, S. J., & Shapiro, K. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology*, 241, 1656–1674.
- Von Holst, E., & Mittelstaedt, H. (1950). *Das Reafferenzprinzip Naturwissenschaft*, 27, 464–476.
- Walther, D., & Koch, C. (2006). Modelling attentional salience of proto-objects. *Neural Networks*, 19(9), 1395–1407.
- Zahavi, D. (1999). *Self-awareness and alterity*. Evanston, IL: North-Western University Press.