Towards a Control Theory of Attention

John G. Taylor

King's College, Dept. of Mathematics,
Strand, London WC2R 2LS, UK
john.g.taylor@kcl.ac.uk

Abstract. An engineering control approach to attention is developed here, based on the original CODAM (COrollary Discharge of Attention Movement) model. Support for the existence in the brain of the various modules thereby introduced is presented, especially those components involving an observer. The manner in which the model can be extended to executive functions involving the prefrontal cortices is then outlined, Finally the manner in which conscious experience may be supported by the architecture is described.

1 Introduction

Attention, claimed William James, is understood by everybody. But it is still unclear how it works in detail and it is still trying to be understood by attention researchers in a variety of ways. This is partly because most of the processes carried out by the brain involve attention in one way or another, but are complex in their overall use of the many different modules present in the brain. This complexity has delayed the separation of these active networks of modules into those most closely involved in attention and those which are lesser so. However considerable progress has now occurred using brain imaging and it has been shown convincingly that there are two regions of brain tissue involved in attention: those carrying activity being attended to and those doing the attending [1, 2].

The modules observed as being controlled by attention are relatively easy to understand: they function so as to have attended activity being amplified by attention and unattended activity reduced. This is a filter process, so that only the attended activity becomes activated enough to become of note for higher level processing. It is the higher level stage that is of concern in this paper. That is now thought to occur by some sort of threshold process on the attended lower-level activity. Attended activity above the threshold is thought to gain access to one of various working memory buffers in posterior sites (mainly parietal). The resulting buffered activity is then accessible to manipulation by various executive function sites in prefrontal cortex.

It is how these executive functions work that is presently becoming of great interest. Numerous studies are showing how such functions as rehearsal, memory encoding and retrieval and others depend heavily on attention control sites. This is to be expected if the executive functions themselves are under the control of attention, which enables the singling out, by the attention amplification/inhibition process, of

S. Kollias et al. (Eds.): ICANN 2006, Part II, LNCS 4132, pp. 461 – 470, 2006.

suitable manipulations of posterior buffered activity (and its related lower level components). Thus rehearsal itself could be achieved by having attention drawn to the decay below a threshold of buffered activity, thereby amplifying it, and so rescuing it from oblivion.

At the same time awareness of stimuli is only achieved if they are attended to. Given a good model of attention, is it possible to begin to understand how the model might begin to explain the most important aspects of awareness?

In this paper I present a brief review of the earlier CODAM engineering control model of attention, and consider its recent support from brain science. I then extend the model so as to be able to handle some of the executive processes involved in higher order cognitive processes. I conclude the paper with a discussion of the way that CODAM and its extensions can help begin to explain how consciousness, and especially the pre-reflective self, can be understood in CODAM terms.

2 The CODAM Engineering Control Model of Attention

Attention, as mentioned in section 1, arises from a control system in higher order cortex (parietal and prefrontal) which initially generates a signal which amplifies a specific target representation in posterior cortex, at the same time inhibiting those of distracters. We apply the language of engineering control theory to this process, so assume the existence in higher cortical sites of an inverse model for attention movement, as an IMC (inverse model controller), the signal being created by use of a bias signal from prefrontal goal sites. The resulting IMC signal amplifies (by contrast gain singling out the synapses from lower order attended stimulus representations) posterior activity in semantic memory sites (early occipital, temporal and parietal cortices). This leads to the following ballistic model of attention control:

Goal bias (PFC) → Inverse model controller IMC (Parietal lobe) → Amplified lower level representation of attended stimulus (in various modalities in posterior CX) (1)

We denote the state of the lower level representation as **x**(, t), where the unwritten internal variable denotes a set of co-ordinate positions of the component neurons in a set of lower level modules in posterior cortex. Also we take the states of the goal and IMC modules to be x(, t; goal), x(,t; IMC).

The set of equations representing the processes in equation (1) are

$$\tau dx(goal)/dt = -x(goal) + bias \tag{2a}$$

$$\tau dx(IMC)/dt = - x(IMC) + x(goal) \tag{2b}$$

$$\tau dx(\ ,t)/dt = -x(\ , t) + w{*}x((IMC) + w'{*}{*}x(IMC)I(t) \tag{2c}$$

In (2c) the single-starred quantity w*x denotes the standard convolution product $\int w(r, r')IMC(r')dr'$ and w**x(IMC)I(t) denotes the double convolution product $\int w(r, r', r'')\, x(r'; IMC)I(r'')$, where I® is the external input at r. These two terms involving the weights w and w' and single and double convolution products correspond to the additive feedback and contrast gain suggested by various researchers.

Equation (2a) indicates how a bias signal (from lower level cortex) as in exogenous attention, an already present continued bias as in endogenous attention, or in both a form of value bias as is known to arise from orbito-frontal cortex and amygdala. The goal signal is then used in (2b) to guide the direction of the IMC signal (which may be a spatial direction or in object feature space). Finally this IMC signal is sent back to lower level cortices in either a contrast gain manner (modulating the weights arising from a particular stimulus, as determined by the goal bias, to amplify relevant inputs) or in an additive manner. Which of these two is relevant is presently controversial, so we delay that choice by taking both possibilities. That may indeed be the case.

The amplified target activity in the lower sites is then able to access a buffer working memory site in posterior cortices (temporal and parietal) which acts as an attended state estimator. The access to this buffer has been modelled in the more extended CODAM model [2, 3] as a threshold process, arising possibly from two-state neurons being sent from the down to the up-state (more specifically by two reciprocally coupled neurons almost in bifurcation, so possessing long lifetime against decay of activity). Such a process of threshold access to a buffer site corresponds to the equation

$$x(WM) = xY[x - \text{threshold}] \tag{3}$$

where Y is the step function or hard threshold function. Such a threshold process has been shown to occur by means of modelling of experiments on priming [4] as well as in detailed analysis of the temporal flow of activity in the attentional blink (AB) [5]; the activity in the buffer only arises from input activity above the threshold. Several mechanisms for this threshold process have been suggested but will not occupy us further here, in spite of their importance.

The resulting threshold model of attended state access to the buffer working memory site is different from that usual in control theory. State estimation usually involves a form of corollary discharge of the control signal so as to allow for rapid updating of the control signal if any error occurs. But the state being estimated is usually that of the whole plant being controlled. In attention it is only the attended stimulus whose internal activity representation is being estimated by its being allowed to access the relevant working memory buffer. This is a big difference from standard control theory and embodying the filtration process being carried out by attention. Indeed in modern control theory partial measurement on a state leads to the requirement of state reconstruction for the remainder of the state. This is so-called reduced-order estimation [6]. In attention control it is not the missing component that is important but that which is present as the attended component.

The access to the sensory buffer, as noted above, is aided by an efference copy of the attention movement control signal generated by the inverse attention model. The existence of an efference copy of attention was predicted as being observable by its effect on the sensory buffer signal (as represented by its P3) [3]; this has just been observed in an experiment on the Attentional Blink, where the N2 of the second target is observed to inhibit the P3 of the first when T2 is detected. [3, 4, 5].

The ballistic model of (1) is extended by addition of a copy signal – termed corollary discharge - of the attention movement control signal (from the IMC), and used to help speed up the attention movement and reduce error in attention control [2, 3]. The corollary discharge activity can be represented as

$$x(CD) = x(IMC) \tag{4}$$

The presence of this copy signal modifies the manner in which updates are made to the IMC and to the Monitor

$$\tau dx(IMC)/dt = - x(IMC) + x(goal) + w''*x(CD) \tag{5a}$$

$$\tau dx(,t)/dt = -x(, t) + w*x((IMC) + w'**x(IMC)I(t) + x(MON) \tag{5b}$$

$$x(MON) = |x(goal) – x(CD)| + |x(goal) – x(WM)| \qquad 5(c)$$

where the monitor is set up so as to take whichever is first of the error signals from the corollary discharge and the buffer activations, but then discard the first for the latter when it arrives (the first having died away in the meantime).

It is the corollary discharge of the attention control model that is beyond that of earlier models, such as of ‘biased competition’ [7]. It is important to appreciate this as acting in two different ways (as emphasized in [2, 3]):

1) As a contributor to the threshold ‘battle’ ongoing in the posterior buffer in order to gain access by the attended lower level stimulus. In [2, 3] it was conjectured that this amplification occurred by a direct feedback of a corollary discharge copy of the IMC signal to the buffer (at the same time with inhibition of any distracter activity arriving there).

2) Used as a temporally early proxy for the attention-amplified stimulus activity, being used in a monitor module to determine how close the resulting attended stimulus achieves the pre-frontally held goal.

Both of these processes were shown to be important in a simulation of the attentional blink [2]; the spatially separated and temporally detailed EEG data of [4] required especially the first of these as the interaction of the N2 of the second target T2 and the P3 of the first target T1.

The resulting CODAM model [1, 2, 8] takes the form of figure 1.

Visual input enters by the INPUT module and feeds to the object map. At the same time this input alerts exogenous goals which alert the attention movement generator IMC so as to amplify the input to the object map. The corollary discharge of the IMC signal is sent to a corollary discharge short-term buffer, which is then used either to aid the access to the WM buffer site of the object map activity, or to update the error monitor (by comparison of the corollary discharge signal with an endogenous goal) so as to boost the attention signal in the IMC so as to better achieve the access of the posterior activation in the object map of the attended stimulus so it achieves access to the WM buffer.

Numerous other features have been added to the CODAM model:

a) More detailed perception/concept processing system (GNOSYS)
b) Addition of emotional evaluation modules, especially modeled on the amygdala [9]
c) Addition of a value-learning system similar to the OFC [10]

The relation of this approach contained in equations to standard engineering control theory is summarised in table 1.

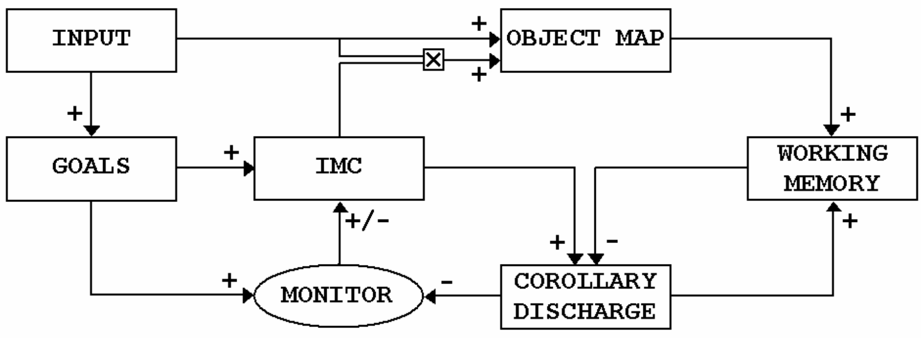

Fig. 1. The CODAM Model

Table 1. Comparison of Variables in Engineering Control Theory and Attetion

Variable	In Engineering control	In Attention
x(, t)	State of plant	State of lower level cortical activity
x(IMC)	Control signal to control plant in some manner	Control signal to move attention to a spatial position or to object features
x(goal)	Desired state of plant	Desired goal causing attention to move
x(CD)	Corollary discharge signal to be used for control speed-up	Corollary discharge to speed-up attention movement
x(WM)	Estimated state of plant (as at present time or as predictor for future use) often termed an observer	Estimated state of attended lower level activity (at present time or as predictor for future use)

We note in table 1 that the main difference between the two columns is in the entries in the lowest row, where the buffer working memory in attention control contains an estimate of only the state of the attended activity in lower level cortex; this is clearly distinguished from that for standard engineering control theory, where the estimated state in the equivalent site is that of the total plant and not just a component of it. There may in control theory be an estimate of the unobserved state of the plant only [6], but that is even more different from attention, where the estimate is only of the attended – so observed – state of lower level brain activity.

3 Executive Functions Under Attention Control

There are numerous executive functions of interest. These arise in reasoning, thinking and planning, including:

1) Storage and retrieval of memories in hippocampus (HC) and related areas;

2) Rehearsal of desired inputs in working memory;

3) Comparison of goals with new posterior activity;.

4) Transformation of buffered material into a new, goal-directed form (such as spatial rotation of an image held in the mind);

5) Inhibition of pre-potent responses [11];

6) The development of forward maps of attention in both sensory and motor modalities, so that possibly consequences of attended actions on the world can be imagined, and used in reasoning and planning;

7) Determination of the value of elements of sequences of sensory-motor states as they are being activated in forward model recurrence;

8) Learning of automatic sequences (chunks) so as to speed up the cognitive process

The rehearsal, transformation, inhibition and retrieval processes are those that can be carried out already by a CODAM model [2, 3] (with additional hippocampus for encoding & retrieval). CODAM can be used to set up a goal, such as the transformed state of the buffered image, or its preserved level of activity on the buffer, and transform what is presently on the buffer by the inverse attention controller into the desired goal state. Such transformations arise by use of the monitor in CODAM to enable the original image to be transformed or preserved under an attention feedback signal, generated by an error signal from the monitor and returning to the inverse model generating the attention movement control signal so as to modify (or preserve) attention and hence what is changed (or held) in the buffer, for later report. Longer term storage of material for much later use would proceed in the HC, under attention control. The comparison process involves yet again the monitor of CODAM. The use of forward models mentioned in (6) allows for careful planning of actions and the realization and possible valuation of the consequences. Multiple recurrence through forward models and associated inverse model controllers allow further look-ahead, and prediction of consequences of several further action steps. Automatic processing is created by sequence learning in the frontal cortex, using FCX → basal ganglia → Thalamus → FCX, as well as with Cerebellum involvement, so as to obtain the recurrent architecture needed for learning chunks (although shorter chunks are also learnt in hippocampus). Attention agents have been constructed {12], and most recently combined with reward learning [13].

Cognitive Architecture: A possible architecture is a) CODAM as an attention controller (with both sensory and motor forms and containing forward models) b) Extension of CODAM by inclusion of value maps and the reward error prediction delta; c) Extension of CODAM to include a HC able to be attended to and to learn short sequences d) Further extension of CODAM by addition of cerebellum to act as an error learner for 'glueing' chunked sequences together, with further extension to addition of basal ganglia (especially SNc) so as to have the requisite automated chunks embedded in attended control of sequential progression. The goal systems in PFC are composed of basal ganglia/thalamus architecture, in addition to prefrontal cortex, as in [14], [15], and observed in [16]. This can allow both for cortico-cortico recurrence as well as cortico-basal ganglia-thalamo-cortical recurrence as a source of long-lifetime activity (as well as through possible dopaminergic modulation of prefrontal neuron activity).

4 Modelling the Cognitive Task of Rehearsal

Rehearsal is a crucial aspect of executive function. It allows the holding of various forms of activity buffered in posterior sites to have its lifetime extended for as long as some form of rehearsal continues. As such, delayed response can be made to achieve remembered goals using activity held long past its usual sell-by date. Such a sell-by date is know to occur for the posterior buffered activity by numerous experimental paradigms [17]. The central executive was introduced by Baddeley as a crucial component of his distributed theory of working memory. The modes of action of this rehearsal process were conjectured as being based in prefrontal sites. More recent brain imaging ([18] & earlier references) have shown that there is a network of parietal and prefrontal sites involved in rehearsal. Let us consider the possible mechanism for such rehearsal to occur.

One of the natural processes to use is that of setting up a goal whose purpose is to refresh the posterior buffered activity at the attended site or object if this activity drops below a certain level. Thus if the buffered posterior activity satisfies equations (3) and (5b) then when the activity drops below a threshold then the monitor is turned on and there is the driving of attention back to the appropriate place or object needing to be preserved.

There are a number of components of this overall process which are still unexplored, so lead to ambiguities. These are as follows:

1) Is the rehearsal signal directed back from the rehearsal goal site to the IMC, and thence to boost the decaying but required input activity, or is there a direct refreshing of activity on the posterior WM buffer?

2) This question leads to the further question: is there a distinction between the posterior WM buffer site and that coding for the inputs at semantic level? It is known, for spatial maps, that there is a visuo-spatial sketchpad buffering spatial representations; is this distinct from a shorter-decaying representation of space? Also is there a separate set of object representations from that of an object WMM buffer?

3) A further question is that, if there is refreshment attention directed to the WM buffer, how does this act? Is it by contrast gain on recurrent synaptic weights that are generators of the buffering property? Or does it occur by an additive bias so as to directly boost activity in the buffer WM

The difference between the answers to question 3) above – how attention is fed back to the WM buffered activity to refresh it – can be seen by analysis. For the membrane potential u of a recurrent neuron (with recurrent weight w) in the WM buffer there is the graded simple dynamic equation:

$$du/dt = -u + wf(u) \tag{6}$$

If attention feedback is by contrast gain then the effect in equation (6) is to increase the weight w by a multiplicative factor, as in the double convolution term in equation (2a). This will have one of two possible effects on the steady-state solution to 9^):

a) The bifurcating value (the non-zero steady state solution to u = wf(u)) is increased, so amplifying the final steady state value of the WM buffer activity;
b) If there is no bifurcation, then the decay lifetime of activity in (6) will increase.

But either case a) or b) above will not necessarily help. If bifurcation has occurred, so case a) applies, then there will not be any decay of WM buffer activity, so there is no need for refreshing in the first place. If the system has not bifurcated but has a long lifetime for decay, as in case b) above, an increase in the lifetime may not help boost back the original activity, but only prolong its decay. Thus it would appear that, barring a more complex picture than present in (6), it will be necessary to have some additive feedback to the WM buffer. This would then boost the threshold activity of the WM buffer, and so help prevent its loss (by keeping it above noise). Thus both a contrast gain and an additive feedback mode of action of refreshment attention would enable the WM buffer to hold onto its activity, and so allow later use of the continued activity in the WM buffer.

5 Discussion

We discussed in section 2 the CODAM mode of attention, based on engineering control. We briefly reviewed the CODAM model, and then considered some details of the comparison between attention and standard engineering controls. An important distinction was that the estimated state of the plant (in engineering control terms usually called the observer) was replaced in attention control by the estimated state of the attended activity in lower cortical sites. This difference is crucial, since on this attention-filtered estimate is based the higher-level processing of activity going under the terms of thinking reasoning and planning. Moreover the initial stage of creating these working memory activations involves a process of taking a threshold on incoming attended activity form lower level sites, and this is regarded as a crucial component of the creation of consciousness, A further crucial component is the presence of a 'pre-signal' – the corollary discharge – suggested in CODAM as the basis of the experience of the pre-reflective self.. In section 3 there was a development of the manner in which executive control might be achieved in terms of this attention control architecture.

In section 4 the details of how rehearsal might occur was suggested in terms of an earlier monitoring model [8], together with a refreshment attention rehearsal process. Various alternatives for the way this refreshment attention could function were considered. It is necessary to wait for further data, updating that form [18], [19], [20] in order to be able to distinguish between these various possibilities. In particular the brain site where the refreshment attention signal is created, as well as the site where such refreshment actually occurs, need to be determined. The analysis of solutions of equation (6) showed a variety of mechanisms could lead to quite different dynamical and steady state activity; these could be part of the clue to how these processes occur in specific sites, so allowing regression techniques to be extended to such refreshment processing.

There is much more to be done, especially by the implementation of language, for developing such high-level cognitive processing, beyond the simple outlines of cognitive processing discussed here.

Acknowledgement

The author would like to acknowledge the support of the European Union through the FP6 IST GNOSYS project (FP6-003835) of the Cognitive Systems Initiative, and more recently of the FP6 IST MATHESIS project.

References

[1] Taylor JG (2000) Attentional Movement: the control basis for Consciousness. Soc Neurosci Abstr 26:2231 #839.3

[2] Taylor JG (2003) Paying Attention to Consciousness. Progress in Neurobiology 71:305-335

[3] Fragopanagos N, Kockelkoren S & Taylor JG (2005) Modelling the Attentional Blink Cogn Brain Research (in press)

[4] Taylor JG (1999) Race for Consciousness. Cambridge MA: MIT Press

[5] Sergent C, Baillet S, and Dehaene S (2005). Timing of the brain events underlying access to consciousness during the attentional blink.. Nat Neurosci, September 2005.

[6] Phillips CL & Harbour RD (2000) feedback Control Systems. New Jersey USA: Prentice Hall

[7] Desimone & Duncan (1995) Neural mechanisms of selective visual attention Ann. Rev. Neurosci., 18:193-222

[8] Taylor JG (2005) From Matter to Consciousness: Towards a Final Solution? Physics of Life Reviews 2:1-44

[9] Taylor JG & Fragopanagos N (2005) The interaction of attention and emotion. Neural Networks 18(4) (in press)

[10] Barto A (1995) Adaptive Critics and the basal ganglia. In: Models of Information Processing n the Basal Ganglia. JC Houk, J Davis & DC Beiser (ditors). Cambridge MA: MIT Press.

[11] Houde O & Tzourio-Mazayer N (2003) Neural foundations of logical and mathematical cognition. Nat Rev Neuroscience 4:507-514

[12] Kasderidis S & Taylor JG (2004) Attentional Agents and Robot Control. International Journal of Knowledge-based & Intelligent Systems 8:69-89

[13] Kasderidis S & Taylor JG (2005) Combining Attention and Value Maps. Proc ICANN05 to appear)

[14] Taylor N & Taylor JG (2000) Analysis of Recurrent Cortico-Basal-Ganglia-Thalamic Loops for Working Memory. Biological Cybernetics 82_415-432

[15] Monchi O, Taylor JG & Dagher A (2000) A neural model of working memory processes in normal subjects, Parkinson's disease and schizophrenia for fMRI design and predictions. Neural Networks 13(8-9):963-973

[16] Monchi O, Petrides M, Doyon J, Postuma RB, Worsley K & Dagher A (2004) Neural Bases of Set Shifting Deficits in Parkinson's Disease. Journal of Neuroscience 24:702-710

[17] Baddeley A (1986) Working Memory. Oxford: Oxford University Press

[18] Lepstein J, Griffin IC, Devlin JT & Nobre AC (2005) Directing spatial attention in mental representations: Interactions between attention orienting and working-memory load. NeuroImage 26:733-743

[19] Yoon JH, Curtis CE & D'Esposito MD (2006) Differential effects if distraction during working memory on delay-period activity in the prefrontal cortex and the visual association cortex. NeuroImage 29:1117-1126

[20] Xu Y & Chun MM (2006) Dissociable neural mechanisms supporting visual shot-term memory for objects. Nature 440:91-95