Axiomatic Consciousness Theory For Visual Phenomenology In Artificial Intelligence

Igor Aleksander[1] and Helen Morton[2]

[1]Department of Electrical Engineering
ImperialCollege
London SW7 2BT, UK

[2]School of Social Sciences
Brunel University
Uxbridge, UB8 3PH, UK
(also Imperial College)

[1]i.aleksander@imperial.ac.uk [2]helen.morton@brunel.ac.uk

Abstract

A model of conscious mechanisms called Axiomatic Consciousness Theory (ACT) is used to develop a theoretical computational model of visual phenomenology . The result is an extension of concepts in AI towards phenomenal intentionality: the lack of which is a common starting point for critiques of AI . Here the argument is developed at four interacting grain levels of computational description and the associated theoretical analysis. The visual domain is highlighted due to its dominance in discussions involving inner mental states.

I. Introduction

Most of the major attacks on artificial intelligence have come from those who argue that existing effort fails to capture a 'self' with a point of view from which the system can develop its intelligent behavior in the world. That is, there is no first person representation within the system. Phenomenology is a movement (of the early part of the20th Century) which develops philosophy on the basis of reality as it is perceived rather than the way it is construed. It is therefore a study of consciousness of perceived reality. To base a computational approach on this, that is, to design systems that act intelligently in the world, mechanisms need to be provided for the system to have a first person point of view. In this paper we look at a computational theory of consciousness, the Axiomatic Consciousness Theory (ACT) (Aleksander 2005) to investigate an early approach to such a design based on visual systems.

ACT consists of an introspectively found set of five interlocking components of experienced consciousness and attempts to find underlying computational mechanisms which when taken together constitute a computational model of consciousness. A distinctive feature of this model is that it has the capacity to address the issue of phenomenology (a representation of a 'self' in a real world). We first look briefly at the reported lack of success of AI in the context of phenomenology and then apply aspects of ACT to see how this might lead to computational descriptions of visual phenomenology into AI. This contribution aims to have both the character of a review and the description of an advance on previously published material on models with phenomenological content (Aleksander and Morton, 2007). The specific advance discussed here is how a neural mechanism can provide a sensation of seemingly infinite spaces. The key to this is an interaction between the content of the current phenomenological states of a system and imaginational phenomenology of earlier experience.

II. AI and Phenomenology

As pointed out by Margaret Boden (2006), Dreyfus' (1979) attack on artificial intelligence was, loosely speaking, based on the absence of the phenomenological view of mind which is summarized briefly below.

As it is reasonable to say that one cannot have thoughts that are about nothing, internal representations in brains or

their models must be about something. That is, they need to be intentional (Brentano, 1874). The representation has to have the detail of the object itself and interactions the organism can have with it (Husserl, 1913). Another classical objection by Searle (1992) known as the *Chinese Room Argument* applies to the lack of phenomenology in symbolic systems. But what is a symbol? Generally it is a mark (or a variable in a computer) that 'stands for something'. So symbol C could stand for 'cat' and in order to approach some form of intentional representation, cognitive scientists have created semantic networks which as in 'C <has> 4P' or 'C<has> four-paws' attempts to capture the 'aboutness' or intentionality of symbol C. According do Searle this fails unless the symbol is shorthand for other mental, experiential representation. However were the mark on the piece of paper a good picture of a cat, it could act instead of the semantic network in, say, driving a robot appropriately in a real world containing a cat. In this paper we define fine-grain neural representations (that is, symbols as fine-grain marks) that are better candidates for visual phenomenology than classical computational symbols.

Twentieth century thought on phenomenology extended 'aboutness' to experiences of one's own body (Merleau-Ponty, 1945) and the existence of the body as a 'self' in some containing environment (Heidegger, 1975). Positive suggestions that approach phenomenology in a computational system were first indicated by connectionists (e.g. Rumelhart and McClelland, 1986) where distributed representations were favored over symbolic ones. In this paper we argue that, to be phenomenological, representations are necessarily distributed, but also, they need to represent the process of controlling the interaction of an organism with its environment. Harnad (1990) pointed out that distal representation of the world should be included in order to 'ground' the symbolic system within a computational model. But, generally, a phenomenological system would be expected to develop its behavior from its own, introspective representation of the world. Symbol grounding develops its behavior from the symbolic algorithms even in Harnad's grounded systems.

More pertinently, Varela (1996) coined the term *Neurophenomenology* to initiate studies on the way inner sensations of the world might be represented in the brain. Lutz and Thomson (2003) relate this to the accuracy of introspective data and show that considering potential brain representation reduces the 'gap' between first person and third person data although this is hardly done in the context of AI.

III. ACT and AI: a résumé

This paper concerns formalizing the 'gap' between first and third person visual experience in the context of AI. While ACT was first enunciated in 2003 (Aleksander and Dunmall, 2003 – here called *Axioms*) this was done by quoting high level non-operative relations involved in five introspectively derived axioms (presence, imagination, attention, volition and emotion). In this paper we take the opportunity of presenting a new operational analysis of these relations in the context of neural aspects of AI.

Why 'Neural'? By introspection, the primary sensation for which we wish to find a computational model is one of being in a space which is an unvarying reality made up of detail. In *Axioms* we suggested that there was a 'minimal world event' of which we could be conscious probably through multiple senses. While in *Axioms* we argued that the world is made up of such minimal events, here we feel that this is wrong – it is the perceiving mechanism that provides the sensation of this fine grain world. The reason is that the minimal event of which we can become conscious is determined by the graininess of the machinery of the brain. In the brain it is a matter of neurons. In a computational system could it be the pixel of a stored image? We strenuously argue that this cannot be the case as these pixels need to take part in a dynamic process which underpins the ACT as described below. So in a computational system too it becomes a matter of a fine-grained cellular system for which a neuron model is an obvious candidate. Below, we briefly examine the axioms in *Axioms* to maintain a complete foundation for this paper and introduce a progression towards a phenomenological system..

Axiom 1: Presence. In *Axioms* it is noted that a sensation of presence may be modeled by making use of the existence of neurons that are indexed on signals from muscles as is known to occur in brains. In this paper we extend this notion to show how large visual fields can be accommodated to obtain a new formulation for visual phenomenology.

Axiom 2: Imagination. In *Axioms*, the sensation of recall of meaningful experienced scenes is modeled through getting the machinery of axiom 1 both to create the states and the transitions between states in a separate recursive neural system that can then enter perceptually meaningful inner states even in the absence of perceptual input. Here we also extend this to dealing with continuous memories of experienced spaces.

Axiom 3: Attention. In *Axioms,* attention is presented as the introspective feeling of selecting input in the process of building up a sensation (visual in the examples given). It is modeled by mimicking the actions of the superior colliculus in the brain, which causes the eye fovea to move to areas of high spatial frequency or changes with time in the visual scene. Here we assume that attention is a mechanism that controls the division between restricted

phenomenological states and transitions between such states (both in perception and imagination) that leads to fuller phenomenological sensations.

Axiom 4/5: Volition/Emotion. In *Axioms* we introduced a way of traversing the content of axiom 2 imaginative machinery that is controlled by an affective evaluation of imagined outcomes of actions. This is not pursued further in the current paper.

IV. Defining levels of visual sensation.

We now address the central aim of this paper: the apparently unlimited nature of visual phenomenology. Imagine entering a space that has not been experienced before such as, for example, the Houses of Parliament in London. It must be recognized that the overall phenomenological visual experience is made up of different elements which we delineate as follows.

Space Sensation is the sensation described from introspection as being in a substantial location (e.g. of being in the Houses of Parliament). We assert that the sensation of space is that which requires the movement of the body for its completion. Another example is 'knowing what is behind one' where one can turn around to look but then remains in one's sensation.

Frontal Sensation is the sensation that would be described as 'being in front of me'. For example, "the entrance to the Commons debating chamber is in front of me as is the functionary who is controlling access". We assert that frontal sensation is that which requires muscular movement of the head and torso without locomotion.

Eyefield Sensation is that which can be experienced without head movement. It is that which is acquired by eye movement, that is, movement of the foveal content of vision without head movement.

Foveal View is generally not in complete conscious experience (Crick and Koch, 1998) but is still present in the primary visual cortex to be brought into consciousness as an eyefield sensation deeper in the visual sysem as explained below.

V. System needs.

In the published ACT so far, it is the transition between the foveal view and the eyefield sensation that is addressed.

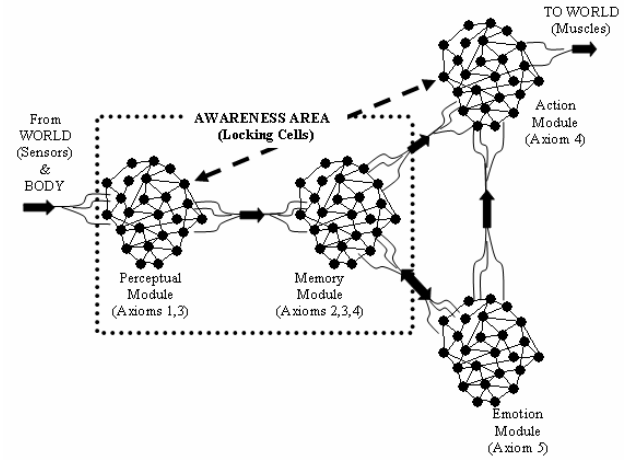

Figure 1: A 'kernel architecture (reproduced from Aleksander and Morton, 2006). This shows how the responsibility for implementing the axiomatic mechanisms is distributed across four neural areas.

This is posited to occur because the signals from the muscular actions produce signals which are used to index neurons according to the position of the foveal view within the eyefield area. In other words the eyefield sensation exists as a state in the perceptual module of a 'kernel' architecture shown in figure 1.

In other words, the dotted line between action neurology and the perceptual module controls which state variables in the latter are influenced by the content of a foveal image. The only memory requirement for the perceptual module is that it holds the positioned foveal image for long enough during a fading process for the eyefield image to be created. As the eyefield representation mirrors reality it has been called ***depictive*** in other publications (e.g. Aleksander 2005)

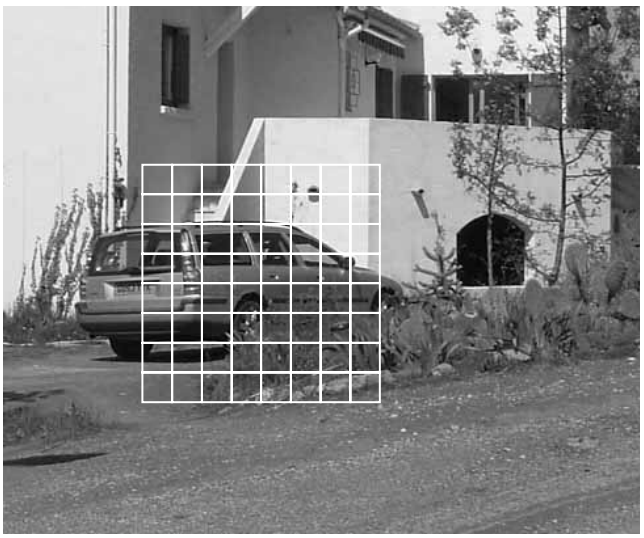

Figure 2: A 'frontal' sensation with a superimposed grid..

Now consider head movement. In previous work, it was assumed that a process similar to the above was at work. However, this must be questioned for two reasons. First, despite the vastness of available neural tissue in the brain, representing a whole history of eyefield images in a space somatotopically related to real space has a lack of economy that is unlikely and needs to be examined for alternatives. Second, this kind of representation has no clear point of view, and this too needs examining.

While the method is entirely proper for an eyefield representation (with the reference point in the centre of the eyefield being indicative of the pointing direction of the head) it does not align with introspection of the visual sensation which is of being at the centre of a space larger than that subtended by the eyefield.

The strategy we propose here is to cast the representation of physical space into state space. For example consider the image in figure 2. The whole figure is a candidate for a 2-D frontal sensation. In figure 3, this is decomposed into a set of states where each state is an eyefield sensation in turn made up of nine foveal sensations. Looking again at figure 1 we impose a division of labour. As indicated, the main role for the perceptual module (PM) is to create the eyefield representation responsible for the eyefield sensation (relationships between sensation and representation are discussed at length in Aleksander, 2005). This corresponds to a group of 3x3 foveal representations. Now, it is the task of the memory module (MM) of figure 1 to provide the memory of previous eyefield representations so as to lead to a representation that aligns with the frontal sensation.

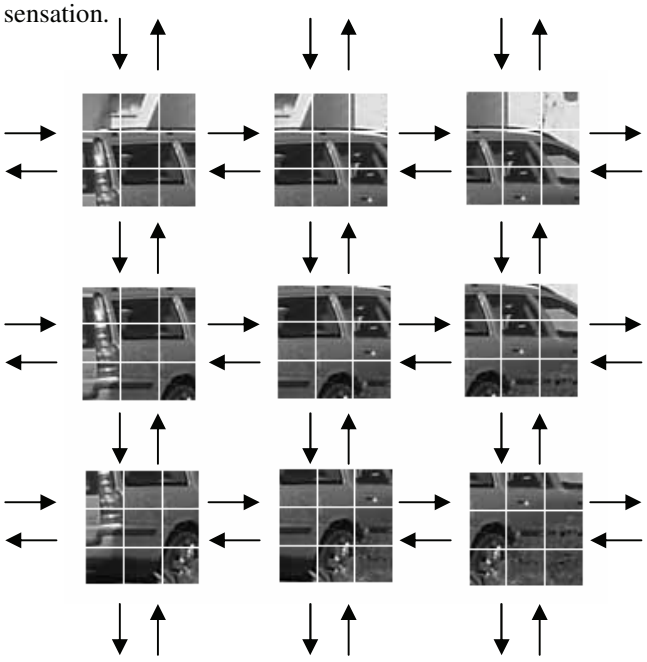

Figure 3: State space decomposition

VI. Frontal sensation memory.

The task for the memory module is to create a state space akin to figure 3 or some subset of it (the system may not explore the entire space and hence learn it). To begin a process of reverse engineering, let the current eyefield representation in PM be E_1 . Let the action (head movement) that takes place alongside this representation be A_1 . This leads to a new state E_2 and action A_2 And so on. We posit that such (E_j , A_j) pairs act input to MM. Here we evoke an 'iconic transfer' (Aleksander and Morton, 1995). Briefly, in an iconic transfer it is assumed that the input and state have the same dimensions and that the learned state is a reproduction of the input. Hence where the general training of a neural state machine is

$$(i_{m,} s_k)_t \rightarrow (s_j)_{t+1} \quad . \qquad (1)$$

$(i_{m,} s_k)_t$ being the current input/state pair and $(s_j)_{t+1}$ being the next state that is learned for the current pair, for *iconic transfer* learning this becomes

$$(i_{m,} s_k)_t \rightarrow (i_m)_{t+1} \qquad (2)$$

This implies that starting with a specific state s_k the input i_m is *recognised* through evoking a state representation of itself in full detail. Returning to the situation for MM, during the running of the process, (2) becomes:

$$((E_j , A_j), (E_k , A_k))_t \rightarrow (E_j , A_j)_{t+1} , \quad (3)$$

where the E symbols relate to eyefields and the A symbols to action symbols. Equation 3 indicates that the net anticipates an (eyefield, action) pair at $t+1$ given the previous stored pair and the currently experienced pair.

It is further commonplace to make use of the generalization properties of this type of digital neural state machine (also in Aleksander and Morton 1995). That is, given a training sequence of [input *(T)*/ output *(Z)*] pairs for a single layer net $T_1 \rightarrow Z_1$, $T_2 \rightarrow Z_2$, …… $T_n \rightarrow Z_n$, then given some unknown input T_u the response will be Z_r, where $T_r \rightarrow Z_r$ is the training step for which T_r is the pattern closest in Hamming distance to T_u . If the 'closensess' is in contention (as there may be several candidates with different outputs at the closest Hamming distance) then one of the contenders is selected at random (the details of this are beyond the scope of this paper).

Returning to the system and its exploration of the visual world through head movement , we note that (3) indicates that MM will build up a state structure such as in figure 3, understanding that the arrows represent the *As* and the 3x3 images are the *Es.* It has to be understood further that formulation in (3) is in terms of a *Moore* automaton. That is, in figure 3, each image and arrow pair is a state. We note that only up/down transitions are shown. There could be actions in many more directions.

Several effects now can be noted.

1. The more the system explores the frontal image, the more complete becomes the representation of the frontal image in MM. Total completion is only asymptotically achieved.
2. Say that the system (for some reason) stops exploring . Ths means that the action module is in some inactive state we call Φ. Then (3), due to generalization could become any of these:

$$((E_j\,,\,\Phi),(E_k\,,A_k))_{\,t} \rightarrow (E_j\,,A_j)_{\,t+1}$$
$$\rightarrow (E_j, A_a)_{\,t+1}$$
$$\rightarrow (E_j, A_b)_{\,t+1}$$
$$\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots$$

   where the multiplicity of mappings occurs due to the fact that many actions may have been learned for one frontal state. According to the generalization rule any one of these might be entered if the perception persits. *In other words, experiencing a particular eyefield state leads MM to explore other eyefields that resulted from head movement*
3. By similar reasoning, if neither the action nor the eyefield from PM constitute a meaningful pattern into MM (eyes closed, say), MM is can become free to 'move' around the frontal field. This is due to the fact that through the Eyefield/Action pairs the MM 'knows' what head movement is involved in the changes of eyefield patches. Indeed, appropriate movement machinery can be driven by the action part of the states in MM to execute an *imagined* movement.

Therefore, to summarise, frontal phenomenological representation consists of a state-space structure of states that are localized foveal views connected by links that correspond in magnitude and direction to head movement.are related. Here we can define a parameter that will be useful later: let $\mathbf{F_X}$ be the frontal state structure associated with the head being in a normal position with respect to the body and with the body in some position **x**. We can, indeed, associate normal eyefield and foveal positions $\mathbf{E_X}$ $\mathbf{G_X}$ from which eye movement and head movement can be measured with the body in position **x**. We return to this after considering a toy simulation.

VII. Simulation.

While major simulations of all four levels of phenomenology integrated into a single system are work in progress, here we present a toy simulation that illustrates the way that a scene viewed by a moving sensor can be represented in an MM-like mechanism.

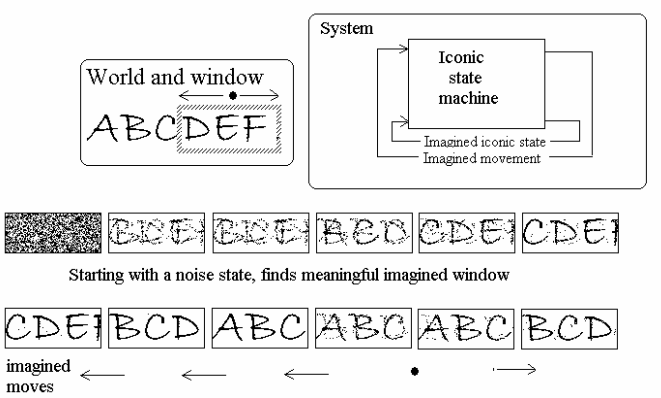

Fig. 1. In this iconic neural state machine, both the visual state and the organism movement are imagined. The imagined moves are generated at random as no move sequence has been learned. It is noted that when the system tries to imagine an impossible move (e.g. left from the leftmost position) the system simply stands still.

In figure 4 the letters are the foveal area while a group of three letters is an eyefield view. Then the frontal view is shown as 'world'. The eyefield view can be moved by one letter space at a time. There is no perceptual module here so the iconic state machine was trained by forcing states on the input of the machine as if they had come from a perceptual module. During training, the eyefield window was moved across the world from centre to limit and back to centre, repeated for the other limit. Then the system was started in an arbitrary (noise) state from which it found a meaningful eyefield state before 'exploring' (through the dynamics of 'attraction') its learned world which includes movement in that world.

VIII. Space sensation

Earlier we defined the space sensation as that which requires locomotion. For example, say that a frontal representation is obtained for the organism facing North. We use the notion of a normal head position and refer to the frontal representation garnered in this position $\mathbf{F_N}$. Now say that the organism turns its body through 180 degreees and faces South. Here it builds a frontal representation, $\mathbf{F_S}$. The question that now has to be faced is how might $\mathbf{F_N}$ and $\mathbf{F_S}$ be related in the MM? Clearly, the action signals that cause the locomotion can be distinguished from those that cause head movement. So during the buildup of ***depictive*** states in the MM an action input becomes associated with a change from a state in $\mathbf{F_N}$ to one in $\mathbf{F_S}$. Many such links may be created between these two sets of states, so that the knowledge in the MM corresponds to thinking 'if I face North this is what I will see, but if I turn to South I shall see this other scene'. Note, that due to the freedom with which the MM can run

through its state structure, the organism can 'imagine' what is at the back of its head while it is also perceiving what is in front of it.

In summary, spaces visited by the organism are represented phenomenologiclly as Frontal sensations, but are related by transitions that represent the locomotion that causes transition learning in the MM.

IX. Concluding commentary

What has been achieved? From introspective considerations we have distinguished between four visually phenomenological sensations that require individual mechanisms that need to be integrated to provide a complete visual phenomenology. These are the *foveal* image, the *eyefield* image created by eye movement, the *frontal* image created by head and body movement while standing still and the *space* image created by locomotion. It has been shown that this representation includes the actions that need to be taken and concurs with notions of phenomenology sketched in earlier parts of the paper. As a lack of phenomenology is the criticism often leveled at artificial intelligence we submit that the Axiomatic Consciousness Theory concepts used here take AI forward towards phenomenology.

What has not been achieved? Except for a toy simulation, what is reported in this paper is purely a feasibility study for a fully structured system that demonstrates the potential mechanisms discussed here.

What kind of science is this? The process of reverse engineering of cognitive functions takes inspiration from mechanisms that are thought to be at work in the brain and tries to discover some theoretical operating principles that underlie these. In this paper the expression of such theoretical principles has been that of probabilistic neural state machines. The task for the researcher is then to build systems based on such principles to confirm the plausibility of the posited mechanisms. We note the specific concern with mechanism distinguishes this work from traditional AI where the accent is more on external behaviour.

Phenomenology and cognition. Finally, it is important to note that a vast corpus of literature exists that concerns explanations of phenomenology of space perception in the brain rather than its computational models. This is exemplified by older publications such as Paillard, 1991 or more recent papers such as Barrett et al. 2000 which, indeed, make use of the properties of neural networks. Further work is needed along the lines indicated in this paper and the observations of cognitive psychology should be examined in this context.

References

Aleksander, I.: 2005, *The World in my Mind, my Mind in the World: Key Mechanisms of Consciousness In Humans, Animals and Machines.* Exeter: Imprint Academic

Aleksander, I. and Dunmall, B.: 2003, Axioms and Tests for the Presence of Minimal Consciousness in Agents. *Journal of Consciousness Studies*,**10**, 4-5, pp. 7-19

Aleksander, I and Morton, H. B.: 1995, *Introduction to Neural Computing 2nd Ed,* London, ITCP.

Barrett, D., Kelly, S. and Kwan, H.: 2000, Phenomenology, Dynamical Neural Networks and Brain Function, *Philosophical Physiology*, **13**, 213 – 226.

Aleksander,I. and Morton H. B.: 2007, Depictive Architectures for Synthetic Phenomenology, In Chella and Manzotti, (Eds): *Artificial Consciousness*. Imprint Academic, pp.30-45

Boden, M.: 2006, *Mind as Machine,* Oxford University Press (p 1394)

Crick, F. and Koch, C.: 1998 Consciousness and Neuroscience. *Cereb Cortex*. 8**(2):**97-107

Brentano, F.: 1874, *Psychology from an Emptirical Standpoint*, Trans: Rancurello et al.Routledge (1995) Orig in German .

Harnad, S.: 1999, The Symbol Grounding System *Physica* **D 42**: 335-346

Heidegger, M.: 1975, *The Basic Problems of Phenomenology,* Trans Hofstadter, Indiana University Press, Orig in German.

Lutz, A. and Thompson E.: 2003, Neurophenomenology: integrating subjective experience and brain dynamics in the neuroscience of consciousness. *Journal of Consciousness Studie*s, **1**0, No. 9–10, pp. 31–52

Merleau-Ponty, M.: 1945, *Phenomenology of Perception*, Trans Smith, Routledge (1996), Orig in French.

Paillard, J. (ed): 1991, *Brain and Space* Oxford: Oxford University Press.

Rumelhart, D. and McCelland, J.: 1986, *Parallel Distributed Processing,* Cambridge,Mass. MIT Press.

Searle, J. 1992,: *The Rediscovery of the Mind*, Cambridge, Mass. MIT Press

Varela, F.J. 1996, 'Neurophenomenology: A methodological remedy to the hard problem', *Journal of Consciousness Studie*s, **3** (4), pp. 330–50.