Towards Artificial Forms of Intelligence, Creativity, and Surprise.

Mark W. Peters (markpeters@cse.unsw.edu.au)
Department of Artificial Intelligence, School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia.

Abstract

This paper starts out from two observations: firstly, that there are complex links between what we term intelligence and what we term creativity and, secondly, that the phenomenon of surprise has a significant role in both the genesis and evaluation of creativity, and is tightly coupled to perception. We argue that for machines to develop to the point where we attribute to them intelligence and, therefore, their own degree of creativity, they must first develop a sensibility of surprise. This, we show, is predicated upon a multi-level organisation of perception, and a method of representing the interest, or novelty, of events and actions taking place in the physical world. A sensibility of surprise further depends on an ability to recognise the novelty of actions the system itself is contemplating. We describe methods of encoding surprise in perceptual robots, and show how this enables them to focus on what is interesting in their environment – a prerequisite to the production of behaviour both creative and intelligent.

Introduction

The Gestalt psychologists, Wertheimer (1959) in particular, made a strong distinction between productive and reproductive thinking. Reproductive thinking is what humans do when presented with a problem of an easily identifiable class, such as finding the length of the hypotenuse of a right-angled triangle. It is called reproductive because it is a question of recalling and reproducing a familiar algorithm. Productive thinking is the solving of problems in a manner that is significantly new. This involves creating rather than recalling a solution; it is volitional as well as selective. Productive intelligence encompasses moments of insight (Köhler 1927) and, until such moments are experienced, an inability on the part of the subject to estimate how close they are to a solution (Metcalfe 1986a, 1986b, Metcalfe & Wiebe 1987). It is a step into unexplored problem space. In practice, the productive-reproductive distinction becomes stronger, because the reproductive form of intelligence can, as its name implies, be easily transported to other individuals or machines, whereas productive intelligence proves to be difficult to copy. It has been argued (Weisberg, 1992) that the distinction is merely questions of degree, which suggests that the greater the intelligence, the harder it is to introspect, predict, and transport, as might be expected.

Intelligence cast in productive terms is demonstrably close to creativity once we remove the domain-dependent connotations from the two concepts. What is an intelligent solution is often a creative solution, and vice versa. Koestler (1975) identified many cross-correlations between creativity in the arts and intelligence and insight in the sciences (and surprise in humour, too). This argument has since been taken up by Boden (1992) who argues that artificial intelligence is the appropriate research apparatus for the scientific study of creativity, since intelligence and creativity in their pure forms are inseparable. Far from supporting an argument against AI (e.g., Penrose 1989), the phenomenon of insight may provide useful clues about where AI research should go, if its agenda is truly to make the phenomenon of intelligence understandable to us (Nilsson 1995). Paulos (1980) also takes up Koestler's baton, drawing links between the exploratory toying with abstract structures and novel combinations of ideas that specifically characterises mathematicians at work and the intellectual play of humour, but also aptly describes intelligent approaches to novel problems.

The connections of intelligence and creativity therefore appear to extend to mental exploration (play, toying, the terra incognita of insight problem space) in which the unexpected, novel, or surprising is given high significance.

While it appears that much biological activity is homeostatic, designed to maintain equilibrium in a changing environment, there is another goal, often conflicting: the active seeking of new information, new stimuli, new situations, novelty, or surprise. Such curiosity has been informally noted as a characteristic of, not only remarkably intelligent or creative humans, and small children - even *very* small children (Eimas, *et al* 1971) - but also other species whose behaviour we particularly acknowledge as intelligent, such as apes (Köhler 1927). What could drive such a predisposition to novelty? Williams (1996) has argued that the aesthetic response is commensurate with surprise, and in proportion to the degree of change a cognitive state undergoes to accommodate new information. This kind of raw pleasure stimulus would be readily co-opted as a reinforcement function to power a curiosity drive.

There is a strong relationship between surprise and perception, maybe even an equality. It seems that we can only learn from those things that are sufficiently reinforced by repetition or direct importance to us. Having perceived the more persistent of phenomena (gravity, for instance) we become unconscious of them, though via adaptive homeostasis we behave to all intents and purposes *as if* we are conscious of them - we can be observed to actively take them into account as we behave. It is normally only when phenomena depart from their norm that they spring back into our consciousness, surprise us, and become concrete percepts. Where particularly creative individuals often make their mark is in the re-seeing of the mundane, and its re-

presentation to the rest of us. They provide a service of re-acquainting us with our milieu. This has been described by Berger (1972) as different ways of seeing, necessary in similar ways for both the author and the audience of creative activity.

To reinforce the notion of selective perception of novelty at the expense of stasis in the world it is instructive to refer to simple experiments that show that our faculty of vision is near wholly dependent on variance in the input. If we are forced to fixate on a static scene our conscious perception of it dissolves in about three seconds. (Crick & Koch 1992). By extension we can argue that it is only the presence of change in the world (more accurately, change in the relation between ourselves and the world) that creates any need for perception in the first place.

Additionally, it has to be recognised that much everyday activity performed spontaneously by humans and animals has proved to be extremely difficult for robots. Either the robots' perceptions (representations) of the world are full of the wrong kind of information, or the robots are not responding to them appropriately. Yet either way the representations are to blame, for they cannot depend on a homuncular *deus ex machina* to get them out of trouble, they must be responsible for *ensuring* the correct response themselves. To summarise the arguments so far, the more intelligent the behaviour, the greater its creative content. Creativity is a linking of seeing the new and acting upon it. Far from being the preserve of the gifted few, creativity is present to a greater or lesser degree in much everyday activity. We shall now discuss the notion that surprise, rather than being a cognitive *response* to perceptual *stimuli*, actually forms the substance of perception, and that therefore, as Berger argues, it is greater perceptivity that engenders creativity, and thus intelligence.

Encoding Surprise

For all biological systems some states are more conducive to life than others, and the basic biological function is to preserve homeostasis by moving from less conducive states to more conducive ones. This applies to states both internal and external. Once homeostasis is achieved, a system need not do anything different until there is some change of state. In other words, it does not even need to keep telling itself that things are still the same. It is not surprising then, that over time individual neurons and even semi-discrete neural systems exhibit reduction in response to unvarying stimuli (Day 1972). The processes are called habituation, adaptation, or depletion, depending on the context. Importantly, such neural units effectively report onset and offset, not absolute values.

Yet it grossly oversimplifies to say that perception is 'on' in the presence of change, and 'off' otherwise, because perception is actually 'off' much more frequently than suggested. We adapt not only to no change, but also to consistency of change, or a derivative of change. Tuning out constant background phenomena such as a clock ticking is an example. That this happens should not be surprising, as it is a predictable effect of certain multi-layered neural systems we shall be describing, in which the output of one layer is the input of another. A constant signal will cause an anterior layer to habituate, and thus cease to activate the change-measuring function in a posterior layer. A system built on surprise-perception equivalence can be readily instantiated (Peters & Sowmya 1987, 1998).

To decompose the phenomenon of surprise let us agree that there are two primary components: an *expectation* and a *departure*, without either of which no surprise can be experienced. The expectation is set up on the basis of previous experience, and might loosely be thought of as a form of pattern recognition. The departure is the discrepancy between what was projected to happen and what actually did happen.

To measure the discrepancy we need memory to compare what is happening now with what has just happened. We start by providing each pixel location in a robot's visual field with a miniature processing unit (which we shall call a memory unit). This possesses a proto-memory consisting of a single value, *memory*. The input to each unit is also a single value, which we shall call *signal*. In the initial case, *signal* is just the brightness of the pixel at the memory unit's location. The value of *memory* at time t is updated from *signal*, using the equation:

$$memory_t = signal + (memory_{(t-1)} \times retention)$$

where *retention* is a constant between 0 (when previous values of *signal* have no effect on *memory*) and 1 (when all previous values of *signal* are summed and stored in *memory*). These extreme cases roughly characterise remembering nothing and remembering everything, respectively, and neither of them is very useful, but many intermediate values are. Varying *retention* adjusts the relative weight given to more recent values of *signal*, effectively determining whether *memory* can be thought of as long-term or short-term, and thereby having significant influence over the behaviour of the system.

If *retention* is near 1 then the value of *memory* will be large in relation to that of *signal*, due to its accumulative effect. So, if we now wish to compare *signal* to *memory* to derive a *surprise* value, we first need to renormalise *memory*. To do this we use another constant, which we derive from *retention*:

$$persistence = 1 / (1 - retention)$$

hence:

$$prediction = memory / persistence$$

which is an exponentially decaying moving average of *signal*. Then:

$$surprise = | signal - prediction |$$

The value of *surprise* represents the difference (departure) between the *signal* just detected and the *prediction* (the expectation). Thus *surprise* corresponds to the 'figural' content set against the 'ground' of *prediction* (Pribram 1991). What is figure and what is ground is clearly predicated on the context of the *surprise*.

Distinctions between ordinary and extraordinary events must be made at multiple levels. Indeed, a single event may be simultaneously novel at one level, but quite unremarkable at the next. For example, one morning a person who has been lying down suddenly rises and walks about, but is known to do this every day. Within the context of the day their activity is novel, within the context of the week, month, year, lifetime, it is not. If this is an ordinary morning rising it may get a second or two of our attention, but if this is the first rising in many months it will be paid far more attention. We need now to develop mechanisms that respond to an event *according to all its temporal contexts*.

We measure change as a conjunction of the outputs of several layers of *surprise* generators, each building upon its predecessor and effectively measuring a derivative of change, or the way change itself, in a previous layer, has been changing. Each layer tunes itself to recognise a pattern (the current pattern) in its input and reacts only when this pattern is interrupted. Its reaction is passed to the next (higher) layer, forming another input, with a pattern of its own. The sum of the parts is thus a pattern reaction system in which *only the changes between patterns are reacted to*.

If we create these layers from similar memory units they need to be connected, not to raw data such as *signal*, but to a data stream which has already had more superficial frequencies removed, and *prediction* fits this role perfectly. It is now possible to connect memory units serially, representing progressively deeper memory layers, each input connected to the *prediction* output of the preceding unit. Note that once memory units are arranged serially, there must be some relative weighting between the layers, even if this weighting is uniform and gives precedence to no layer in particular. The centroid of surprise is thus

$$x = \sum_{l=1}^{m} \frac{l_w \sum_{u=1}^{n} u_x u_{dl}}{\sum_{u=1}^{n} u_{dl}} \qquad y = \sum_{l=1}^{m} \frac{l_w \sum_{u=1}^{n} u_y u_{dl}}{\sum_{u=1}^{n} u_{dl}} \qquad (1)$$

where n is the number of memory units, each of which (u), has an x co-ordinate, a y co-ordinate, and a difference d, and where m is the number of memory layers, each of which (l) has a weight w.

Activities that we call ordinary (a highly contextual term) can be accommodated in a finite number of these perceptual layers, ranging upward from 1 in the case of inanimate, stationary objects, to some arbitrary number. Watching people working on a production line may fail to excite more than, say, 6 layers in our kind of perceptual system. Extraordinary patterns, on the other hand, are defined as those that exceed some arbitrary number of activation levels. A system built of several layers may therefore be able, based on the conjunction of reactions at different levels, to develop responses that represent activities at many different levels in the world, separating the extraordinary from the ordinary, and performing a natural bottom-up hierarchical segmentation. If interrogated, such a system would be able to rate the novelty of the current activity at several different levels.

Pattern intervals are complex multi-level phenomena. Thibadeau (1986) noted the importance of identifying the moments when particular actions start and finish. We attempt to make a system react to these special moments both adaptively, and without a priori knowledge in any form, so that future higher functions, that might have specific recognition-based tasks, may receive representations that have already conveniently carved the world at its joints.

Results

These processes are shown pictorially in Figure 1 displaying four two-image blocks in each of which the upper image is made from *prediction* values and the lower image is made from *surprise* values. The images show the state of the system in an instant of time. It can be seen that memory layer 2 *surprise* values are activated by the movement of the subject's head rather than by the movement of his hand. This is because the hand has been waving almost continuously (evidenced by its high levels in both the *history* image and memory unit 2's *prediction* image), whereas the head has until this moment been quite still. In other words the system is no longer attracted by the frantic hand, and has just transferred its attention to the head.

To show quantifiably how our system, known as WRAITH, (Peters & Sowmya, 1996) behaves under different memory configurations we set up a scene containing two 5 cm discs with alternating black and white quadrants, attached to small motors set 12 cm apart. Figure 2 plots the focus of attention of three layers of memory units. The short-term plot is that of two parallel units named channel 1 and channel 2, the medium-term plot is that of a memory unit that receives input from the two channels, and the long-term plot is that of a third layer which receives input from the second layer. The plot is actually the horizontal x co-ordinate of the various foci, plotted vertically against time (measured in tenths of a second). The upper trace and lower trace represent approximately 31 seconds and 40 seconds, respectively. In each trace the grey bars correspond to the times that each disc was spinning. In the upper trace the first disc starts to spin after about a second and has soon attracted the attention of all three memory layers. It can be seen that the short-term memory is highly reactive. The second and third layers are slower to react. About 1.5 seconds later the second disc starts to spin. The short-term memory simply centres its attention precisely halfway between both discs because it simply looks for the centroid of raw change. The medium-term memory is momentarily attracted close to the second disc, but also soon centres its attention, meanwhile the long-term memory, though slower to react, takes a hard look at the second disc before also finally centring.

A more interesting phenomenon occurs when the second disc suddenly ceases to move. The short-term memory is now free to return to the first disc, but long-term memory, having fully adapted to the motion of the second disc, is now dramatically attracted to the sudden *cessation* of rotation, despite the fact that *all motion is on the other side of its visual field*. The same goes for the medium-term memory, though to a lesser extent. Eventually all memory is attracted

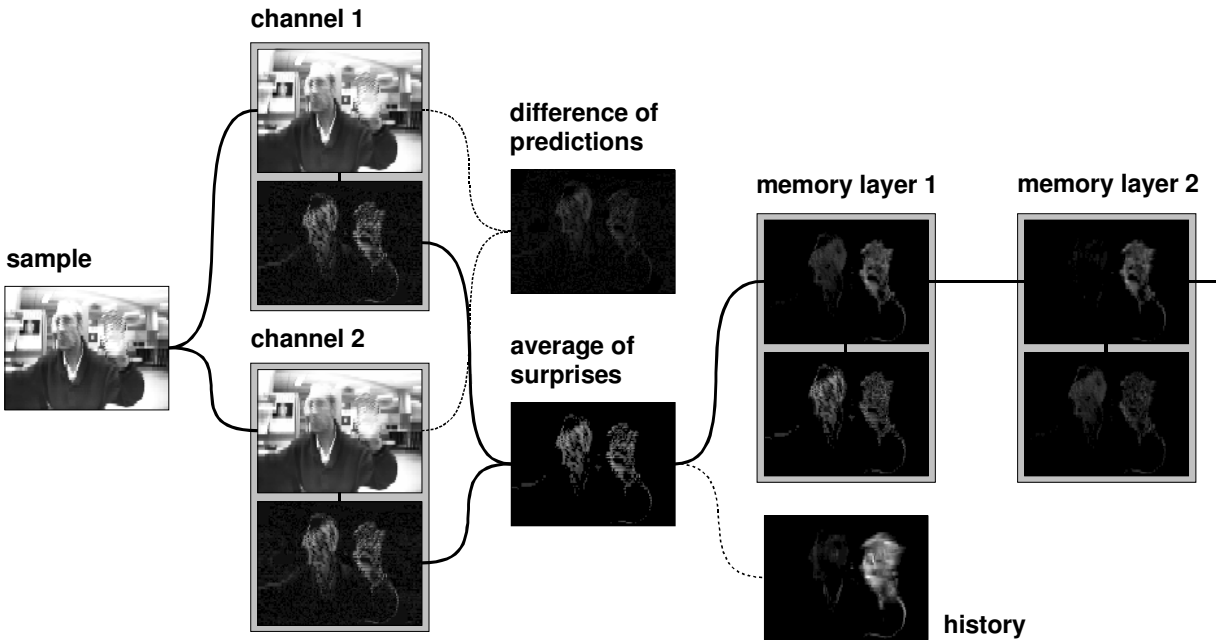

Figure 1: Pictorial representation of activity and surprise.

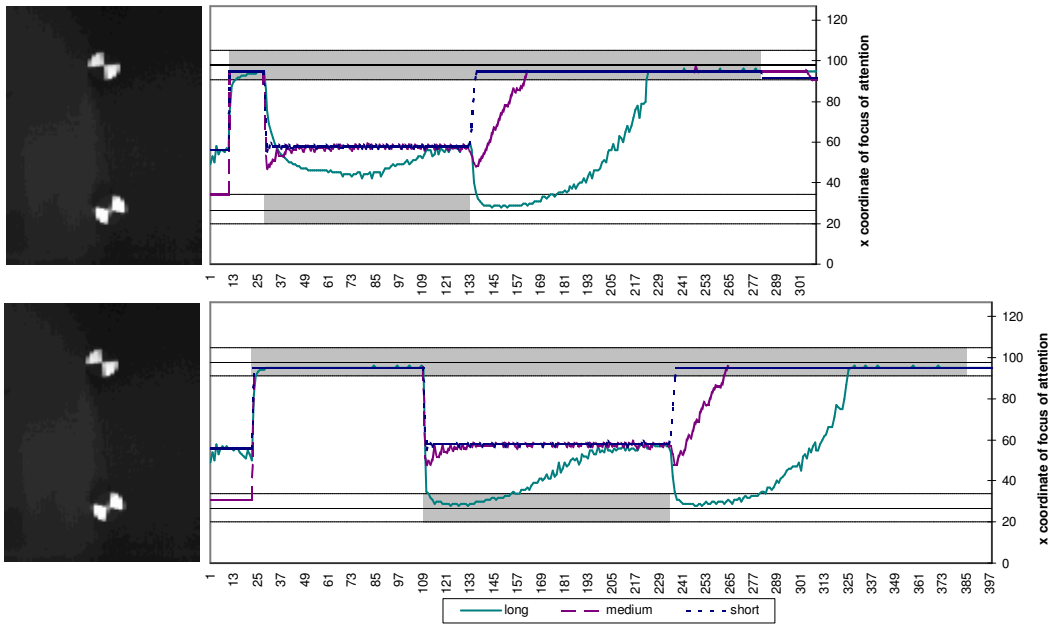

Figure 2: Graphical representation of activity and surprise.

to the sole remaining spinning disc, where attention remains once that disc too finally stops.

In the lower trace the sequence of onsets and offsets is the same, except that now the memory is given more time to adapt to the rotation of the first disc. Instead of only 1.5 seconds, it remains spinning for about 9 seconds before the second disc starts. By its reaction, it can be seen that the long-term memory has fully adapted to the constant rotation of the first disc, so it moves fully to the centre of the second disc when it starts, and remains preoccupied with it for several seconds before eventually accommodating to that motion too.

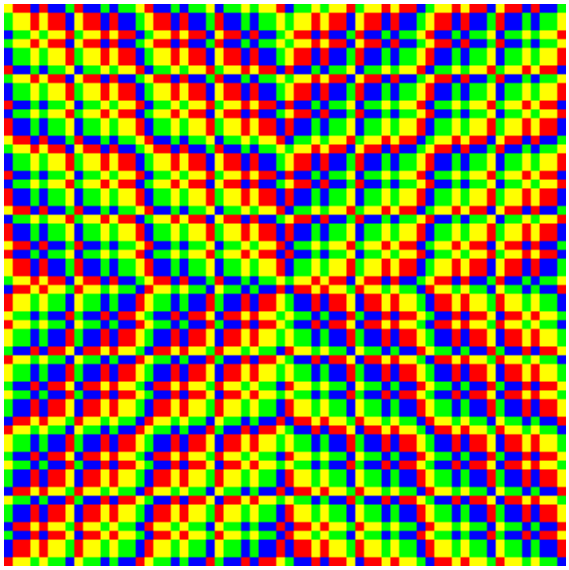

Figure 3: (Peters 1992a).

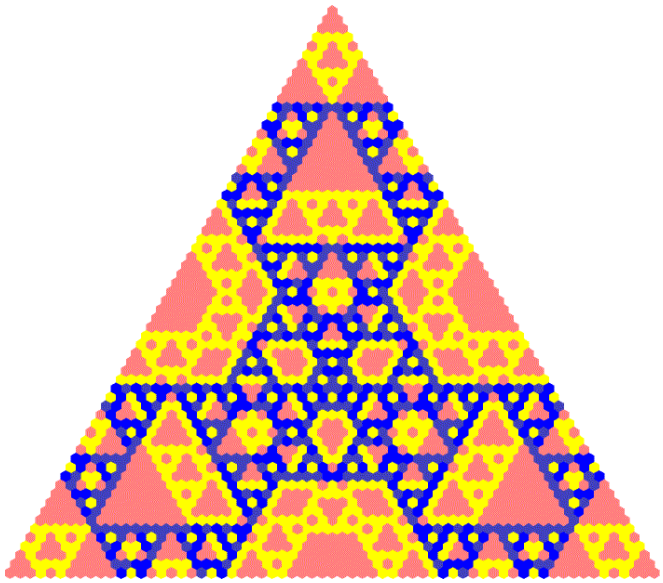

Figure 4: (Peters 1992b).

WRAITH's behaviour in response to any particular stimulus is predicated on its internal memory state. Consequently when the second disc starts turning in the upper trace, attention is diverted towards it, but not far enough to actually foveate it. This is because the onset of motion follows too soon after the onset of the first disc, which is still exerting some hold on WRAITH's attention. In the lower trace, WRAITH has had more time to accommodate to the activity of the first disc, and is therefore much more drawn to the second disc when it starts.

Creative Applications of Surprise

Figure 3 and Figure 4 show examples of algorithmically generated images. Such images can be generated *ad infinitum* using machines, but the generative power is unchecked, because machines do not yet have any form of aesthetic appreciation or evaluation. Artists such as Sims (1992) have developed sophisticated methods allowing the generative role to devolve to the machine, while retaining the editorial or quality assurance role themselves. However, given the ability to encode surprise, it should now be possible for machines to develop preliminary methods to support the faculty of aesthetic evaluation:

Method for Aesthetic Evaluation

We suggest that such methods will consist of five steps:

Generate the Material. As the ability to generate material is not the subject of this paper, we simply give the examples in Figure 3 and Figure 4. This step might also consist of simply identifying objects, either 'found' or produced by other agencies. Examples may be camera input, image databases.

Identify the Pattern (Theme). Pattern needs to be identified at multiple levels, as already demonstrated. Levels to include, ultimately, would cover not just the machines own output, but the output of others too.

Build the Expectation. The theme, as predicted, must encapsulate the extrapolation of patterns at all levels.

Identify the Departure. This might be done before production of the machine's own work.

Censor the Dull. Censorship on the basis of insufficient departure.

The implementation of such a system must be both circumspect and conscientious – and therefore highly difficult. However, the concept of surprise provides the means to do it, and chained levels of surprise detectors are a proven method for increasing the sophistication of a machine's response. We argue that they are, if not sufficient, then at least necessary.

Our implementation has chosen to deal with particular forms of pattern: patterns of movement as represented by levels of change in visual sensors. A vast body of pattern recognition research already exists in several domains, comprising visual, verbal, acoustic and numeric data. What is now required, we suggest, is that pattern recognition be co-opted as a means to another end. Incorporation of multiple pattern recognition methods (arranged both in parallel and in series) within an architecture whose main purpose is to detect, not so much pattern itself, but *departure* from identified *themes* will be essential to machines whose behaviour we will accept as intelligent. For, without such sensibilities, machine intelligence will continue to fail in con-

spicuous ways, revealing 'inhuman' perseveration with minutiae, an inability to recognise the subtexts in palimpsestic messages, an unresponsiveness to nuance, or change in the texture of discourse, and a general lack of curiosity.

Discussion

Hofstadter (1997) recently observed:

> In every intellectual field that I had encountered, ranging from mathematics to music to art to poetry, I had the sense that the moment that patterns were perceived at one level, this immediately established a higher level of abstraction, opening the door to the perception of totally unanticipated types of patterns.

It is precisely this identification and accommodation of pattern, and the consequent search for departure within pattern, and then pattern within the departure, that gives scope for the phenomenon of surprise and opportunity for creativity.

If the surprise function is applied to a machine's own output or that of other agencies, and if the measure of change accommodated within the machine is coupled with a reinforcement function, it is likely to have quite broad effects. It should provide a machine with a form of curiosity, *both* homeostatic control *and* deliberate departure from homeostasis, creative behaviour, and constantly improvement of opportunities to learn, by specifically seeking out parts of the world where change is dense and multi-level.

References

Berger, J. (1972). *Ways of seeing*. London: British Broadcasting Corporation.

Boden, M. (1992). *The creative mind*. London: Abacus.

Crick, F., & Koch, C. (1992). The problem of consciousness. *Scientific American, September,* 110-117.

Day, R. H. (1972). *Human perception* Sydney: John Wiley.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. 1971. Speech perception in Infants. *Science, 171*, 303-306.

Hofstadter, D. R. (1997). *Le ton beau de Marot*. London: Basic Books.

Koestler, A. (1975). *The act of creation*. London: Picador.

Köhler, W. (1927). *The mentality of apes*. London: Routlege Kegan and Paul.

Metcalfe, J. (1986a). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 2, 288:294.

Metcalfe, J. (1986b). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 4, 623-634.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory and Cognition, 15,* 3, 238-246.

Nilsson, N. J. (1995). Eye on the Prize. *AI Magazine, 16,* 2, 9-17.

Paulos, J. A. (1980). *Mathematics and humour*. Chicago: University of Chicago Press.

Penrose, R. (1989). *The emperor's new mind: concerning computers, minds, and the laws of physics*. London: Vintage.

Peters, M. W. (1992a). *Fantagma V*. Collection of the artist.

Peters, M. W. (1992b). *Triangle (seed 1-1-1 base 4 generation 7).* Collection of the artist.

Peters, M. W., & Sowmya, A. (1996). Active vision and adaptive learning. *Proceedings of Intelligent Robots and Computer Vision XV* (pp. 413-424). Boston: SPIE Volume 2904.

Peters, M. W., & Sowmya, A. (1997). WRAITH: ringing the changes in a changing world. *Proceedings of the Fourth Conference of the Australasian Cognitive Science Society.* Newcastle, Australia.

Peters, M. W., & Sowmya, A. (1998). Autonomous multi-domain attention control. *Pacific Rim International Conference on Articifial Intelligence '98*. Singapore.

Pribram, K. H. (1991). *Brain and perception*. New Jersey: Lawrence Erlbaum Associates.

Sims, K. (1992). Interactive evolution of equations for procedural models. *Proceedings of the Third International Symposium on Electronic Art*, (p. 128). Sydney: Australian Network for Art and Technology.

Thibadeau, R. (1986). Artificial perception of actions. *Cognitive Science, 10*, 117-149.

Weisberg, R. W. (1992). Metacognition and insight during problem solving: comment on Metcalfe. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 2, 426-431.

Wertheimer, M. (1959). *Productive thinking*. New York: Harper and Brothers.

Williams, M.-A., (1996). Aesthetics and the explication of surprise. *Languages of design, 3,* 145-157.