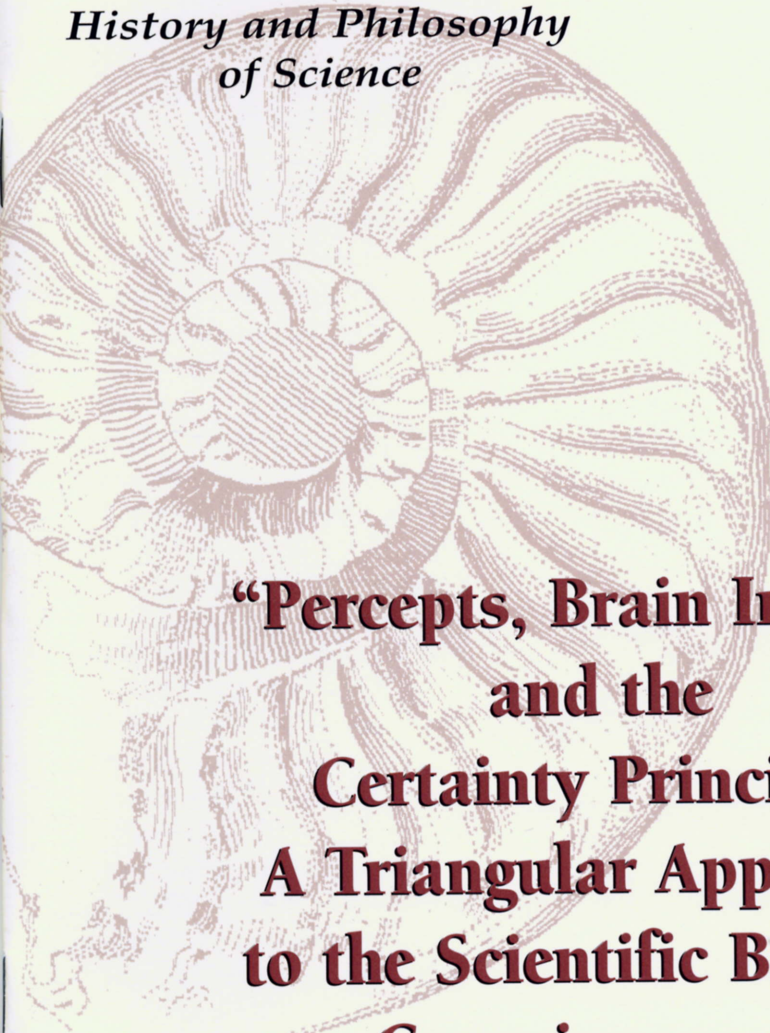
The 1999
Herbert H. Reynolds
Lectureship in the
History and Philosophy
of Science



"Percepts, Brain Imaging and the Certainty Principle: A Triangular Approach to the Scientific Basis of Consciousness"

Dr. Lawrence Weiskrantz

April 12, 1999
Baylor University

The Herbert H. Reynolds Lectureship in the History and Philosophy of Science

Percents, Brain Imaging, and the Certainty Principle: A Triangular Approach to the Scientific Basis of Consciousness

L. Weiskrantz
University of Oxford

As presented April 12, 1999, at Baylor University

The Herbert H. Reynolds Lectureship in the History and Philosophy of Science, established in 1998 by a gift to the Baylor University Department of Philosophy from the Herbert H. and Joy C. Reynolds Endowment Fund for University Excellence, brings to the Baylor University campus an internationally recognized scholar in the sciences or in the philosophy or history of science.

The recipient of the Lectureship serves in residence at Baylor for two weeks, presents a public lecture, participates as a co-professor in a philosophy of science seminar, and is available for classroom participation in courses in appropriate disciplines.

A committee comprised of faculty members from the following departments plans the annual lecturship: biology, chemistry/biochemistry, geology, history, philosophy, physics, and psychology/neuroscience.

BAYLOR
UNIVERSITY

Percepts, Brain Imaging, and the Certainty Principle: A Triangular Approach to the Scientific Basis of Consciousness

L. Weiskrantz
University of Oxford

As presented April 12, 1999, at Baylor University

It is indeed an honour to be invited to give this very first Herbert H. Reynolds Lecture. I have known and have been in awe of your Chancellor and former President for more years than I — probably either of us — would like to acknowledge. We first met more than 35 years ago when I was working at the Cambridge University, and he was a research officer of the US Aerospace Agency supporting some of our research. He and I and our respective colleagues had some notable research sessions together, but who could have imagined so long ago that I would have the pleasure of reestablishing a link with him when he became the President of this major university, which he has served with such outstanding leadership and effectiveness and, I might add, with such humanity?

This lectureship is allocated to the philosophy department, and so — as a psychologist — I wish to consider some aspects of the relationship between psychology and philosophy. The area of philosophy that most closely concerns psychology is, of course, the philosophy of **mind**. This has a long and by no means peaceful history; it is by no means peaceful today. Psychology itself, whatever else it does, must concern itself with **mind** and the **menta**l. What is the proper study of the mind and the mental in scientific terms — that is, in terms that allow for systematic, coordinated study, and prediction? And, crucially, that places it in the same universe in which all other scientific endeavours are located — the universe of matter and energy? An incontrovertible point is that we know that changes in brain function have massive as well as very subtle effects on our mental life. Without the brain, there is no mental life. We also know that mental life, in turn, is reflected in brain activity. If we grant that the brain is a physical organ subject to scientific laws, then the relationship between mind and brain could also be assumed to be lawful.

But if we are going to pursue the putative science, how should we do so? Several candidates, with allegiance to separate university departments, rush forward to claim priority. There are some who will deny, from the outset, that our mental life can be amenable to scientific enquiry. And there are others who admit that there might be a scientific account in principle, but that it is, and will remain, beyond our limited hominoid capabilities to grasp it. There are others who might say that mental events can only be inferred from behaviour, and that it is folly to study mental phenomena, because they are unknowable directly to the scientist. Just study behaviour. Nothing funda-mental about the mental. In psychology this was at the core of the behaviourist movement, but in philosophy we also have those who tell us — eliminative materialists, for example — that we need not study anything other than the physical organ, i.e., the brain, that instantiates — or is thought to instantiate — these subjective mental events. Indeed, they go further and declare that our own understanding of our own mental life is inherently flawed — down with "folk psychology!". There are others who tell us that there are theoretical systems that can simulate anything worth simulating, and hence we only need turn to such systems — connexionism, neural network theory, artificial intelligence — to gain our understanding. More extreme still are those neuroscientists who accept the reality of the mental, but only as a way station: they assume that the intensive and persistent study of the physical device itself, in this case, the brain, will itself reveal the nature of the mind. We can, according to them, dispense with psychology and philosophy, as no doubt amusing, even colourful potential remnants, mere harmless epiphenomena in the intellectual universe. Psychology and philosophy are like the whistle on the steam locomotive — the locomotive can go even if the whistle is broken. And yet each of these departments is apt to argue, with force, that it is the steam engine. It is a wonder that university administrators are not derailed more often.

And so, I return: what is the proper study of the mind for its scientific understanding? I start with an argument that no one discipline can claim it for its own, although this is not to conclude, I shall argue, that we should have a free-for-all. First of all, psychology sets the phenomena to be studied, from this there is no escape — and it is obvious that I reject the view that the study of raw behaviour alone is sufficient. For deeper understanding we want to know how psychological phenomena come to be, how they work. Let's take the example of colour perception. We can study colour perception objectively, even though the percepts are subjective. Psychophysicists have been doing so for decades, it is not a problem. One should not confuse ontological status with epistemology. No one, however, can truly understand how we perceive colour without knowing about specific receptors in the eye with particular absorption properties, or how stimuli with different chromatic composition influence each other, without

knowing, that is, about lateral inhibition between ganglion cells, or how colours change their appearance in the dusk without knowing that we have separate rod and cone receptor mechanisms with different spectral sensitivities. And so forth. We have to get inside the organism. Retinal ganglion cells and receptors may be necessary, but they are hardly sufficient — one cannot see with a retina alone. We must go into the brain.

Well, why not carry on along these lines as a general strategy — follow the visual pathways up into the brain, record from them, study their anatomy, their biochemistry, their DNA? Everything, in fact. Of course, we can do so in principle, but it is extremely unlikely to work in practice. The nervous system is not merely complex — it is fantastically complex, with 10^{10} neurones, each with as many as thousands of connexions. Years ago Horace Barlow (1961) pointed out that if we treated each of the neurones in the brain as simply being active or inactive, a simple binary element — and of course neurones are more complicated than simple binary devices — this would generate 2 to the power of 10^{10} possible brain states. And he noted that this adds up to more states than there are particles in the Universe, according to Eddington. Even if we assigned a particle to every possible state, we could not do it.

And, indeed, there are very few examples of where we have even begun to be helped fundamentally by such a bottom-up approach. One could dissect the brain as anatomical organ as carefully as one wished and one would never conclude, I submit, that this slab of jelly-like matter is capable of controlling colour vision, speech, let alone of thought, or memory, or consciousness. We are helped not by anatomy alone, but starting with the knowledge that we do have speech and thought, and we can then try, haltingly, to relate these to brain events. Probably the way we would first discover a connexion would be to see that isolated damage to particular regions of the brain interferes with speech, another with colour, and so forth. But this is the point: *We cannot understand the brain without knowing what it is trying to understand.* Full stop.

Well, if the nervous system is so hopelessly complex, mathematical theorists have no difficulty in dealing with very large populations of elements, and in analyzing how they might work in complex interactions with each other. Is this the way into the problem — to have an abstract theory of the brain? It is one way, but, again, not a sufficient way. Indeed, it can be grossly wrong even when it achieves famous victories. We have all grown up with particular theories of this and that in psychology and in neuroscience. I can remember well the models of pattern perception that prevailed when I was a student. They were essentially systems of transforming spatial arrays into temporal patterns, as is done by television cameras. The models worked. Good predictions could be made from them, tested with people and with animals. For example, one could predict that horizontal and vertical are more discriminable than diagonals at 90 degrees to each other, because of the orientation of the scanning signal. Such theories were quietly but effectively put to rest by the discoveries of Hubel and Wiesel (1962), among others, that neurons in the visual cortex did not code by temporal sequencing, but by selective sensitivities to target features, such as lines and edges. An abstract theory can make correct predictions but still be fundamentally wrong. We still do not have an adequate theory of pattern perception, but we cannot have one that ignores the physiology. We need the r.n.s. (the real nervous system) not just the c.n.s. (the conceptual nervous system). The actual hardware of the brain constrains and shapes the properties of the software.

And so psychology, neuroscience, or theory alone will not suffice: psychology on its own because it does not get inside the skull, theory on its own because it must accommodate what is inside the brain, and neuroscience on its own because it cannot understand what it has to understand without outside guidance. To return to the example of colour perception, even when we get as far as the ganglion cells in the retina, we need a good theorist to make sense of the possible interactions of ganglion cells to see how colour contrast and colour mixing might be generated. But theory alone will not suffice, as exemplified in the early arguments between the contrasting theories of Herring and Helmholtz at the very beginnings of the story in the 19th century — both of them worked as theories. And none of them could have predicted Land phenomena, or the electrophysiological or functional brain imaging evidence of cortical systems relevant to colour constancy.

No one discipline by itself will do. We need all three disciplines, in triangular formation. Not just any triangle, but one with a particular formation and with a particular orientation. And a triangle that does not lead to rivalry or divorce — indeed it might even spark more than a touch of romance! It is an equilateral triangle, but with a leading apex and, like the point of a spear, pointing forwards. At the leading edge are the psychological phenomena we wish to understand — whether they be mundane, such as why we can see the stars at night better by not looking straight at them, or why we suffer from jet-lag, or why memories get confused with each other or become distorted, to more challenging ones — including questions of consciousness, which is the one I wish to tackle myself a little later. The two other corners of the triangle are occupied by neuroscience and by

theory. But obviously the idea behind the metaphor is that these corners are simply the origins of paths of three-way communication. It is obvious that occupants at each of the corners vary enormously not only in scope but in their level of analysis. Thus, psychology of consciousness spreads from minute details of introspection, to dreams and sleep, to a plethora of neuropsychological syndromes involving, for example, conscious and unconscious memory systems, to fine-grained analysis by signal-detection theory of subjective criteria, and so forth. I need not say anything about the complexity of neuroscience, extending from broad-band electroencephalography, to functional brain imaging, to neurochemistry, to molecular structure. But theory is less straight-forward. It includes the usual gamut which we have known for a long time in psychology, namely model building and simulation, recently rapidly expanding to neural networks and parallel distributed connectionist models, shown to be extraordinarily rich in explanatory power. It also includes developments as abstruse as quantum mechanics, as in the theoretical concerns of Roger Penrose (1989).

But it also includes philosophy. I do not think that philosophy on its own, separate from empirical enquiry and substance, involves theory construction, despite its claims for "theories of meaning" or a "theories of language". Philosophy, as it enters into the picture here, takes three rather different forms. First, it is concerned with the analysis of concepts and word-meanings, with analytic rather than with factual or synthetic truths. But philosophers help us to understand why it is that scientists sometimes get the wrong end of the conceptual stick, and thus they keep its two other triangular partners alive to linguistic and logical dangers in their descriptions of these complex events, which necessarily involve complex language. I have, for example, found John Searle's guidance through the epistemological and ontological minefield that scientists sometimes enter with complete disdain and disregard, very helpful (1998). Beyond that, philosophers are, by inclination and by profession, thinkers. They do think about psychological phenomena, admittedly often in the comfort of their armchairs, but also often with insight. For example, they think about thought, and its properties. As Searle has put it, "much of philosophy is concerned with questions that we do not know how to answer in the systematic way that is characteristic of science, and many of the results are efforts to revise questions to the point that they can become scientific questions." (*op. cit.*, pp. 157-8) For example, there are philosophers who develop their own theoretical concepts from such thoughts, such as David Rosenthal (1986, 1990, 1993) and other "higher order thought" theorists or, again, Ned Block

(1995), who distinguishes between two types of meanings of consciousness, Michael Lockwood (1996) who tries to address quantum mechanical aspects in relation to the real brain. But, finally, there is also another species of philosophy that is emerging, that is very much at home at the theory corner of our romantic triangle. Increasingly some philosophers are concerning themselves with critical facts from experiment and with addressing their implications for scientific endeavour, as they see them. Such has long been evident in the philosophy of science of physics and mathematics, but it is now extending to psychological and biological science. Thus we have Daniel Dennett (1991) commenting on a range of empirical perceptual phenomena, Patricia Churchland (1988) on a huge spectrum of neuroscience, Ned Block (1995) commenting on a theory that the frontal lobes have a particular key role for understanding consciousness, Thomas Natsoulas (1981) on split-brain and blindsight subjects, and so forth.

And so, even for the most intransigent and challenging of topics, human consciousness, there are theorists, including philosophers, patently (if not always, patiently!) interacting with neuroscientists and with psychologists, and psychologists interacting with both neuroscientists and theorists. And neuroscientists interacting with....? well, as I have suggested already, here we run into a wall, not one of brick fortunately, but more like a thorn-hedge full of prickles, also inter-mixed with a few wild roses. There are many neuroscientists, of course, who communicate between the other two apices, and earnestly share and pursue the same aims. I include, for example, Francis Crick (1994; Crick and Koch, 1990, 1995), who is, as it were, a converted neuroscientist. And Gerald Edelman (1989). And Logothetis (1989) with his fascinating work on neuronal activity correlated with reversible figures. And Ben Libet and colleagues (1979, 1993), whose methods estimate the additional time after the arrival of an electrical signal from a sensory nerve in the sensory cortex for the subject actually to become consciously aware of it (about 1/2 second). And a veritable army of scientists working with electrical potentials of the brain directly correlated with mental events. And an even larger army now doing functional brain imaging of human subjects in many centres around the world. All of them are interested in inter-relating brain events with conscious activities of subjects, cognitive as well as emotive.

But among this army of neuroscientists, as I have already indicated, there are some who hold an implicit belief that they are simply at a temporary staging post, on their way to being able to reduce consciousness, or any other mental category, to the operation of neural events. This is a

take-over bid from one corner of the triangle, its occupants tolerantly patronizing their other bed partners within the triangle for the moment, but with their eyes on the whole pad. Or, if they are going to admit another bed partner, it will be a molecular type of scientist to allow for even further reduction. Some of them have anticipated their supposed famous victory by absorbing psychology into their own nomenclature, referring to it as cognitive neuroscience, but with an ever diminishing role for the cognitive modifier.

There are two related issues here that must be addressed: to preserve the right kind of balance within the triangle, to prevent a takeover; and, secondly, to defend us against what is sometimes called "nothing but" reductionism. Let us go back again to a less exalted and more manageable topic than consciousness, to our friend, colour vision. I look at a rainbow and marvel at its beauty. The existence of the rainbow is not a matter of my own construction, because the rainbow can be photographed. Its existence needs explaining — not only in terms of what physical processes are going on in the atmosphere — of light diffraction, but in terms of why we see the colours in the first place — not only in the rainbow itself, but in a picture of a rainbow. We have good explanations of light diffraction, and also of differential absorption of different wavelengths by receptors in the eye and, as I have indicated, of processing by specialized retinal ganglion cells, and beyond that by systems in the brain. But there is no way in which the sight of the rainbow can be reduced to any of these. Our explanation may make it less mysterious, less mystical, but a rainbow is a rainbow is a rainbow. The explanation of something does not eliminate that something. It remains the object that requires explanation or for which an explanation has been advanced, and hence continues to be the leading apex of the triangle, the head of the spear. And one cannot throw a spear sideways to good effect. The psychological phenomena that remain are those that either demand or have secured explanations, but do not disappear in the process. An example is sometimes given of the explanation of living processes in terms of organic chemistry. Again, the mystery may be reduced — that is the nature of explanation — but in no way does organic chemistry abolish the properties of living matter, nor the need to study those properties in their own right.

There is another meaning of reductionism in science, which is to appeal to a lower level of analysis and explanation, so that physical heat, for example, as measured by thermometers, is given another account in terms of the movement of molecules in a gas. But psychological phenomena are not like that — we cannot change an experience into a set of neurones firing, or chemicals coursing across synapses. On the contrary, for psychological phenomena we must go — if in any way — in the opposite direction. We must preserve the psychological phenomena but find brain activities that are worthy of making us understand the psychology. We must elevate neuroscience, not reduce psychology. In this sense, but not with any metaphysical implications intended, I call myself an elevationist rather than reductionist. Not only will bottom-up not work, as we have seen, but even more so bottom-down will capsize the boat.

Having got that pontification out of the way, let me return to the question of consciousness and the brain. What would the romantic triangle — of psychology, theory, and neuroscience — look like? Well, not the whole of consciousness — that is too huge a topic. Here the philosophers have helped to remind us that there are two rather different meanings of the term, indeed many, many different ones, but two major ones. The first is what has been called "creature consciousness," which means the difference between being awake or asleep, or more extremely, being awake or in coma. It is, in other words, the level of activation, and there is quite a bit of evidence — electrophysiological, anatomical, and neurochemical — about the neural processes with which this level correlates. I am not going to deal with this aspect further here. But David Rosenthal (1986) contrasts this with "state consciousness", by which is meant our experience of ongoing events. One can be fully awake and yet be unaware experientially. Indeed, a car driver had better be fully awake to drive, but he also had better suppress his conscious and continuous attention to each and every other detail of all objects in the road, otherwise he will surely crash. Much of what we do is, and must necessarily be, on "automatic pilot." Thinking is a privileged operation, best reserved for when it is needed. Here is Alfred North Whitehead on the subject:

> "It is a profoundly erroneous truism.....that we should cultivate the habit of thinking what we are doing. The precise opposite is the case. Civilisation advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle — they are strictly limited in number, they require fresh horses, and must only be made at decisive moments." (1948, p. 61).

Now, it happens that there are many examples of dramatic disturbances of state consciousness that turn up in the neurology which can serve as the provocative leading edges of the triangle. They require full

descriptions in their own right, and challenge one for explanations. Coming from neurology, we are immediately helped at one corner of the triangle, because the brain damage is either known already, or is in principle knowable. But brain damage is, of course, only a fragment of the relevant knowledge that we seek. As far as theory is concerned, we need to consider different ways in which the phenomena and the neuroscience can map onto a common scheme.

Let me give you an example, one familiar to me from my own work on vision. The eye of the higher primates, including ourselves, sends nerve fibres to a large number of different targets in the brain, about 10 of them (Cowey and Stoerig, 1995). The main one — occupying about 85% of the capacity of the nerve bundle leaving the eye — ends up in the visual cortex, so-called V1, over a route that is known as the geniculo-striate pathway. If the geniculo-striate pathway is damaged or interrupted in a monkey, it is not surprising that the animal is still able to make visual discriminations — 9 other pathways, which go to lower and older levels of the brain, remain intact. Of course, there are changes in visual capacity, it would be surprising if there were not, but the visual performance of a monkey without any striate cortex can be impressive.

Now the paradox: if the human has the same damage, he or she appears blind. Fortunately, the person (or the monkey) is not severely handicapped, because each geniculate striate pathway serves only one half of the visual field, and it is rare for both geniculostriate pathways to be damaged because they are in different cerebral hemispheres. And so a person with such damage will retain a perfectly good half field of vision, but be blind in the other half, with both eyes. It is that half we are concerned with in this research. Now, if this whole constellation of non-striate cortical pathways still remain intact, why cannot such persons — like the monkey — discriminate visual events in their blind fields? Well, it is a long story, but it has been discovered that they can, or at least some of them can. But you do not find this out by asking them to describe whether they can see a letter on a chart, or whether they are aware of the colour of a square, or to describe a shape — they will tell you that they cannot see them. Nor, of course, can we ask the monkey such questions; instead we must ask the monkey to choose between two alternatives and reward it for picking the correct one. And, if you ask the human subject to guess the identity of the alternative stimuli in the blind field, even though they cannot actually see them — a strange question — lo and behold the performance can be excellent, much to the surprise of the subjects themselves when shown the results afterwards. We called this condition blindsight (Weiskrantz et al., 1974). It is visual discrimination in the absence of visual awareness. it is a loss of state consciousness of visual events without the loss of a visual capacity, caused by loss of primary visual cortex.

A few examples: a blindsight subject, by forced-choice guessing, can discriminate different colours of light, such red vs. green, at a high level of success and yet never actually sees any colours. Or can discriminate the difference in orientation of lines, or the direction of a moving spot, or the presence or absence of visual structure, or reach for randomly located lights in the blind field, or even adjust the hands to the shape of the unseen object towards which they are reaching (for reviews, cf. Weiskrantz, 1986, 1997, 1998). All this without the experience of "seeing."

Blindsight is by no means the only example. Indeed, there is no category of cortical dysfunction in humans in which comparable losses of state consciousness cannot be found (cf. Weiskrantz, 1991, 1997). There is, for example, "deaf hearing" after auditory cortex lesions, in the tactile modality there is a comparable condition to blindsight called "numbsense (Rossetti et al., 1995). In a recent issue of Brain (Sobel et al., 1999) there appeared a report of "blind smell." One of the best known examples is in the field of memory disorders. Amnesic patients who cannot remember an event for more than a minute or two can be shown to have very good storage of new information, but they do not know it, nor can they use it in everyday life because they are not aware of such knowledge *as an experienced memory* (Warrington and Weiskrantz, 1968, 1982; Weiskrantz and Warrington, 1979). And perhaps the best-known examples of dissociations between perception and acknowledged awareness came from the evidence from "split-brain" patients so tellingly analyzed by Sperry and his colleagues (Sperry, 1974, 1977).

In terms of the romantic triangle, the historical discovery and the analysis of these phenomena lay wholly at the psychology apex. They were counter-intuitive and surprising, and still continue to surprise. Of course, there were many controls one had to carry out to ensure that phenomena such as blindsight were genuine, but I think we can be satisfied on that score now (cf. Weiskrantz, 1998 for a recent review). How can one try to understand blindsight phenomena in terms of brain events at the neuroscience level, and how does theory help? We are just at the beginning of this journey, but certain features are emerging.

First, the neuroscience was helped inadvertently by the mis-understanding by a philosopher of the psychological aspects of a case of blindsight. It happened that Daniel Dennett and I were at a symposium together in which I played a video tape of an informal demonstration of a blindsight subject mimicking with his arm movements the path of a moving spot of light that was projected onto the wall by a hand-held laser pointer. The subject (G.Y.) could mimic the movement with impressive ease if the spot of light moved very quickly (15° per second or faster). When the spot moved very slowly, he simply failed to respond. Dennett later described this demonstration in his book on consciousness (1991) as follows:

> D.B [in fact, it was G.Y, another subject] one of the subjects studied by Weiskrantz, has a right hemianopia and shows the classic blindsight capacity to guess above chance when cued. For instance, if a light is slowly moved across his scotoma horizontally or vertically and he is prompted to guess 'vertical' or 'horizontal,' he does so extremely well, while denying all consciousness of the motion. However, if the light is moved more swiftly, it becomes self-cueing. D.B. [G.Y.] can volunteer without prompting a highly accurate report of the motion, and even mimic the motion with a hand gesture as soon as it occurs. And when asked, D.B. [G.Y.] insists that of course he consciously experiences the motion — how else would he be able to report it? (1991, p. 332).

Dennett's misunderstanding — understandable given the briefness of this demonstration, which had not been published — was this: In point of fact, we never did the "classic blindsight" experiment with slowly moving stimuli (nor, for that matter, did G.Y. "insist" that he was conscious when the spot moved quickly — the tape records his hesitant comment that he "knew" that something had moved and where, but he "did not see anything"). The tape simply shows that the subject failed to respond at all with slowly moving stimuli. In fact, it would have been technically very difficult to do the experiment. To generate a truly smoothly moving projected spot of light with a hand-held laser pointer is virtually impossible — there are too many tremors and jiggles of the hand, magnified by the projection, and we know that such transients are important in blindsight (see below). But the experiment seemed interesting to do, as a proper controlled study, and so my colleagues and I designed and built the equipment to move a laser beam over a wide range of velocities and directions. The subject (G.Y.) was instructed to indicate on one of two keys in which of two possible directions he judged or guessed the light had moved. But, importantly, we also had to obtain evidence about the other side of the coin — whether the subject actually reported any experience of the slowly moving stimuli. For this purpose we introduced two more keys, "commentary keys"; on every trial he was required to indicate whether he had any experience whatever of the visual event, even the vaguest of a "gut feeling", by pressing the "yes" key, or had absolutely none whatever, by pressing the "no" key. And so we could plot trial-by-trial performance in relation to reported awareness.

The result was that Dennett would have been right about the experiment we had not done, had we done it (Weiskrantz et al., 1995). Even when the subject reported no awareness whatever of the slowly moving stimuli, he could still discriminate the direction of movement at high levels of success, i.e., he demonstrated "classical blindsight". Now, that in turn led us to realize that we could compare these two states of consciousness in a functional brain imaging study — i.e., complete unawareness with good performance versus "unseeing" feeling of awareness with good, matched levels of performance. (It turns out that this kind of "unseeing awareness" can be eliminated by reducing the transient components of visual stimuli — for example, by making their onset gradual rather than abrupt, making stimuli move slowly rather than fast, and so forth, cf. Weiskrantz et al., in press). Moreover, we could compare the brain images of both of those states with real seeing in the intact half-field of a blindsight subject. We did the experiment, and found an interesting difference in the brain activity when the subject was aware and when there was no reported awareness. I do not want to go into the fine details here, which have been published (Sahraie et al., 1997). Briefly, there was a shift from activation of dorsolateral cortical structures, especially in the frontal lobes, in the aware state, to structures lying in the midbrain or ventral and medial cortex in the unaware state. One of those structures in the latter case was the superior colliculus in the midbrain, which is the second largest recipient of fibres from the eye to the brain, and which remains intact when the larger striate pathway is damaged.

But what about theory? Well, it is necessary first to consider another neuroscience experiment that illuminates theoretical aspects. We know that the visual cortex and its closely neighbouring association areas at the back of the brain are <u>necessary</u> for conscious visual sight. But are they <u>sufficient</u>? Can one see by visual cortex alone? An approach to the answer came from work at National Institutes of Health by Nakamara and Mishkin (1980, 1986), and from even earlier work by Roger Sperry and coworkers at the California Institute of Technology (Sperry et al., 1960). If all the visual cortices are left intact, but all non-visual cortex is removed in the monkey, the monkey does not respond to visual events — even though the neurones in the visual cortex continue to respond to them. One interpretation of this result is that one evidently needs structures beyond visual cortex to <u>see</u>. The visual cortex is necessary but not sufficient. The differentially active structures in our brain images of the visually aware mode turned out to lie far away from the visual cortex, in dorsolateral regions of the frontal lobes and elsewhere. We can get some idea of the routes over which the information might flow from visual cortex to these frontal loci because quite a bit of neuroanatomy is available. In fact, there are strong inter-connexions between posterior visual cortex (V1, etc.) and other regions of the brain, and in particular a variety of routes that originate in the visual cortex eventually converge heavily in a region of the brain far removed from it, in area 46 of the frontal lobes (Young, 1992, 1993), precisely that area that was selectively activated in the "aware" mode of visual discrimination in blindsight (Sahraie et al., 1997).

Now let us go back to our neuropsychological syndromes. How do we know when a subject is visually aware or not? In principle, simple: we ask him or her. In other words, we need not only the evidence of a visual discrimination but an independent acknowledgement or "commentary" by the subject that it was or was not experienced. We can put this under operational control, as described above, by using "commentary keys" to be pressed after every trial.

But in practice the matter is not always so simple. We cannot always depend upon a verbal acknowledgement even when the subject is presumably aware — we cannot do so with human infants, with aphasic patients, or with animals. And the problem can also be acute with some blindsight subjects, who have no verbal problem as such, but find it strange to be asked to respond to stimuli which they insist that they cannot see. We are all so strongly tuned into the belief that we can only discriminate visual events if we are visually aware of them that some subjects simply refuse to believe that they can discriminate correctly the "unseen" visual stimuli in their blind visual fields by forced-choice guessing, and some subjects refuse outright to do so. Fortunately there are indirect methods that get around the problem. We were greatly helped by the discovery by my colleague, Prof. John Barbur and others, that the pupil of the eye is a very sensitive monitor of visual processing (Barbur et al., 1986, 1994b, 1987a,b). The surprising fact is that the pupil constricts not only to increases in light levels, but to a whole host of visual properties even when there are absolutely no changes in light level. Among other dimensions, it responds to spatial structure, to colour, to movement. One can measure a person's visual acuity and spectral sensitivity by careful measurement of the pupil's constriction. And so one can also do this in the blind visual field of the blindsight subject, or in the visual field of an animal, or a human infant, in the absence of verbal responses. We have used pupillometry in just this way with human blindsight subjects and also with monkeys whose visual cortex was removed from one hemisphere. In both, the results show that there is a narrowly tuned visual channel that can be precisely specified, with a peak sensitivity at about 1 cycle/degree. There is also very close agreement between the profile of the visual channel as demonstrated by pupillometry and that found by using forced-choice guessing by the human subject (Barbur et al., 1994a; Weiskrantz et al., 1998) in the affected part of the visual field. In some respects pupillometry is actually well ahead of psychophysical methods, and can reveal visual blindsight capacities that would be extremely difficult to discover by more conventional methods.

Let us return to the main theme: As I mentioned, the evidence from experimental neurology suggests that visual processes themselves require an additional step in the nervous system for them to be said to be "seen", to give rise to visual awareness. And, similarly, the very operation of testing for awareness itself requires an additional step beyond the operation of the keys involved in the visual discrimination — blindsight demonstrates that the discrimination can be excellent even in the absence of acknowledged visual awareness. There is never any way of knowing whether a subject is aware without such stepping sideways and getting such an independent assessment. In contemporary jargon, one must go "off-line". Staying "on-line" is blind to its own status. *Awareness always requires a parallel and separate response to the discrimination itself for it to be identified and acknowledged as such by an observer and, I would suggest, by the subject himself or herself.* The general operational point is that awareness of sensory events is a two-component, a two-stage affair. This makes it attractive to consider philosophical theories of consciousness that also postulate a two-stage operation. The class of theories that come to mind are the so-called higher order thought theories of consciousness. Rosenthal (1993), for example, has suggested that a thought cannot have state consciousness without there being a higher order thought about the thought. I believe it is interesting and helpful to have a philosophical theory onto which both the psychology and the neuroscience can be mapped, and vice versa, at least in outline.

In a slightly different and more general context I am tempted to be bold and postulate a quite strong hypothesis, namely that the commentary itself actually *endows* one with the experience of awareness. That is to say that it is the very achieving of the ability to make a commentary of any particular event that gives rise to awareness, and it is what we mean by

being conscious. To borrow a metaphor: the medium is the message — the achieving of the commentary is the awareness. It is what we mean by being aware, for an event to be experienced as such. The commentary need not be uttered; what matters is the state of "comment-ability". Indeed, I would be even bolder and say that for the measurement of consciousness we are in exactly the converse situation that Heisenberg postulated for the measurement of physical phenomena. Heisenberg's *uncertainty* principle states that the very measurement of a physical event changes that event itself — the most infuriating example I have in my everyday life is when I try to measure the pressure in my bicycle tire: the gauge invariably releases a spurt of air when it is attached and detached so that the reading is bound to be wrong! For consciousness, my position is that the very act of being able to comment on a sensory event is what actually endows it with awareness. It is of interest that Benjamin Libet's (1979, 1993) fascinating research indicates that when a human subject's hand is touched, it not until about a half second after the arrival of electrical potential in the sensory cortex in the brain that a human subject indicates conscious awareness of the touch. I think it may well be the arrival of the off-line commentary signal that is the critical event, which takes time to reach those non-sensory regions of the brain that are involved in comment-ability. Consciousness, on this view, arises from the inverse of the Heisenberg principle. It is the ability to make the commentary — not necessarily verbally — in any particular instance that endows awareness and certainty. Consciousness is a Heisenberg certainty principle. I recognize that Heisenberg's uncertainty principle applies at the level of particles, but I use the inversion of Heisenberg as a convenient metaphor to make my point.

There is a weaker form of the hypothesis, namely that awareness arises first, and that it enables a commentary to be made. I know of no certain way at this stage of testing the difference between the weak and strong form of the status of the for commentary, that is between the enablement and the endowment positions, although I prefer the strong form. The advantage of this approach, while it admittedly but tolerantly suffers from the pitfalls of folk-psychological discourse, is that it is directly linked to the very methodology and the operations themselves that allow the limits of awareness to be defined, and it maps onto both the two-stage neurology and philosophical positions.

The spear with our romantic triangle at its leading point has only just started traversing its universe, and has a very long way to go. But at least I think the example I have given suggests how all of the partners in the triangle can and must talk to each other, to make their own contributions and support each other, without causing untoward jealously or divorce or take-over. For this particular example, it was the demonstration that set the problem, by showing that visual awareness and discrimination can be dissociated. That required the use of an "off-line" commentary response, separated from the ongoing discriminative responses. As blindsight actually emanated from animal work, the circle has recently been completed by the ingenious demonstration that monkeys lacking V1 likewise have blindsight: the animals have excellent sensitive and accurate responses to visual events in their affected visual fields, but by using a separate "off-line" mode Cowey and Stoerig (1995, 1997) demonstrated that the animals classify those very same visual events as being "non-lights," exactly as a human blindsight subject would do. Note that it was a philosopher commenting on the psychology that led to the application of a neuroscientific measurement, using brain imaging. And that neuroscientific and behavioural evidence suggests that the visual cortex may be necessary but not sufficient for visual awareness. And that a two-stage interpretation from neuroscience and psychology can be related to and enriched by philosophical and theoretical accounts that themselves posit a two-stage process.

Let me briefly pull the strings together. My title was "Percepts, Brain Imaging, and the Certainty Principle." I hope the allusions will be clear. Percepts or images or directed attention or memories or meanings, and all other psychological categories, lie at the apex of the spearhead and can only be specified in full by the methods of psychology; brain imaging is at the neuroscientific apex aligned with the enormous variety of other rapidly developing methods of neuroscience, and the certainty principle is at the theoretical cum philosophical apex, together a range of other rich theoretical developments. Each of the corners covers 120 degrees — all of them, I venture to say, honourable degrees. And with honourable collaboration, hopefully completion over 360 degrees can be achieved.

References

Barbur, J.L. and Forsyth, P.M. (1986). Can the pupil response be used as a measure of the visual input associated with the geniculo-striate pathway? *Clin. Vis. Sci.* **1**, 107-111.

Barbur, J.L., Harlow, J.A., Weiskrantz, L. (1994a). Spatial and temporal response properties of residual vision in a case of hemianopia. *Phil. Trans. Roy. Soc.* B, **343**, 157-166.

Barbur, J.L., Harlow, J.A., Sahraie, A., Stoerig, P., and Weiskrantz, L. (1994b). Responses to chromatic stimuli in the absence of V1: pupillometric and psychophysical studies. In: Vision science and its applications. *Optical Soc. Amer. Technical Digest*, **2**, 312-315.

Barbur, J.L. and Thomson, W.D. (1987). Pupil response as an objective measure of visual acuity. *Ophthal. Physiol. Opt.*, **7**, 425-429.

Barbur J.L., Thomson W.D., and Forsyth P.M. (1987). A new system for the simultaneous measurement of pupil size and two-dimensional eye movements. *Clin. Vis. Sci.*, **2**, 131-142.

Barbur, J.L., Sahraie, A., Simmons, A., Weiskrantz, L., and Williams, S.C.R. (1998). Residual processing of chromatic signals in the absence of a geniculostriate projection. *Vis. Res.*, **38**, 3447-3453. Barlow, H.B. (1961). The coding of sensory messages. In *Current problems in animal behaviour*, (ed. W.H. Thorpe and O.L. Zangwill), pp. 331-360. Cambridge Univ. Press, Cambridge.

Block, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.*, **18**, 227-247.

Churchland, P.S. (1988). *Neurophilosophy: Toward a unified science of mind and brain.* MIT Press, Cambridge.

Cowey, A. and Stoerig, P. (1991a). The neurobiology of blindsight. *Trends Neurosci.*, **29**, 65-80.

Cowey, A. and Stoerig, P. (1995). Blindsight in monkeys. *Nature Lond*, 1995, **373**, 247-249.

Cowey, A. and Stoerig, P. (1997). Visual detection in monkeys with blindsight. *Neuropsychologia*, **35**, 929-939.

Crick, F. (1994). *The Astonishing Hypothesis. The scientific search for the soul.* Charles Scribner's Sons, New York.

Crick, F. and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars Neurosci.*, **2**, 263-275.

Crick, F. and Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature Lond*, **375**, 121-123.

Dennett, D.C. (1991). *Consciousness Explained.* Penguin Press, London.

Edelman, G.M. (1989). *The remembered present: A biological theory of consciousness.* Basic Books, New York.

Frederici, A.D. (1982). Syntactic and semantic processes in aphasic deficits: the availability of prepositions. *Brain and Language*, **15**, 245-258.

Hubel, D.H. and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, **160**, 106-154.

Libet, B. (1993). The neural time factor in conscious and unconscious events. In *Experimental and theoretical studies of consciousness* (Ciba Foundation Symposium). Wiley, New York.

Libet, B., Wright, W.W., Jr., Feinstein, B., and Pearl, D.K. (1979). Subjective referral of the timing for a conscious sensory experience. *Brain*, **102**, 193-224.

Logothetis, N. and Schall, J. (1989). Neuronal correlates of subjective visual perception. *Science*, **245**, 761-763.

Lockwood, M. (1996). *Mind, brain, and the quantum*. Blackwells, Oxford.

Nakamura, R.K. and Mishkin, M. (1980). Blindness in monkeys following non-visual cortical lesions. *Brain Res*, **188**, 572-577.

Nakamura, R.K. and Mishkin, M. (1986). Chronic blindness following lesions of nonvisual cortex in the monkey. *Exp Brain Res.*, **62**, 173-184.

Natsoulas, T. (1981). Basic problems of consciousness. *J. personality and social psychol.*, **41**, 132-178.

Penrose, R. (1989). *The emperor's new mind*. Oxford Univ. Press, Oxford.

Rosenthal, D. (1986). Two concepts of consciousness. *Phil. Studies*, **49**, 329-359.

Rosenthal, D. (1990). A theory of consciousness. Report No. 40 Research Group on Mind and Brain. *Perspectives in Theoretical Psychology and the Philosophy of Mind*. Univ. of Bielefeld, Bielefeld.

Rosenthal, D. (1993). Thinking that one thinks. In *Consciousness. Psychology and philosophical essays*, (ed. M. Davies and G.W. Humphreys), pp. 198-223. Blackwell, Oxford.

Rossetti, Y., Rode, G., and Boisson, D. (1995). Implicit processing of somaesthetic information: a dissociation between where and how? *NeuroReport*, **6**, 506-510

Sahraie, A., Weiskrantz, L., Barbur, J.L., Simmons, A., Williams, S.C.R., and Brammer, M.L. (1997). Pattern of neuronal activity associated with conscious and unconscious processing of visual signals. *Proc. Natl. Acad. Sci. USA*, **94**, 9406-9411.

Searle, J. (1998). Mind, language, and society. Basic Books, New York.

Sobel, N., Prabkakaran, V., Hartley, C.A., Desmond, J.E., Glover, G.H., Sullivan, E.V., and Gabrieli, J.D.E. (1999). Blind smell: brain activation induced by an undetected air-borne chemical. *Brain*, **122**, 209-217.

Sperry, R.W. (1974). Lateral specialization in the surgically separated hemispheres. In *The neurosciences: Third study program*, (ed. F.O. Schmitt and F.G. Worden). M.I.T. Press, Cambridge.

Sperry, R.W. (1977). Forebrain commissurotomy and conscious awareness. *J. Medicine and Philosophy*, **2**. 101-126.

Sperry, R.W., Myers, R.E., and Schrier, A.M. (1960). Perceptual capacity in the isolated visual cortex in the cat. *Q. J. exp. Psychol.*, **12**, 65-71.

Stoerig, P. and Cowey, A. (1992). Wavelength sensitivity in blindsight. *Brain*, (1992), **115**, 425-444.

Tyler, L.K. (1988). Spoken language comprehension in a fluent aphasic patient. *Cogn. Neuropsych.*, **5**, 375-400.

Warrington, E.K. and Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature*, 217, 972-974.

Warrington, E.K. and Weiskrantz, L. (1982). Amnesia: A disconnection syndrome? *Neuropsychologia*, **20**, 233-248, 1982.

Weiskrantz, L. (1986). *Blindsight. A case study and implications*. Oxford University Press, Oxford.

Weiskrantz L. (1990). Outlooks for blindsight: explicit methodologies for implicit processes. The Ferrier Lecture. *Proc. Roy. Soc.B* , London. **B239**, 247-278.

Weiskrantz, L. (1991). Disconnected awareness for detecting, processing, and remembering in neurological patients. *J. Roy. Soc. Med.*, **84**, 466-470.

Weiskrantz, L. (1996). Blindsight revisited. *Curr. opinion in neurobiol.*, **6**, 215-220.

Weiskrantz, L. (1997). *Consciousness lost and found. A neuropsychological exploration*. Oxford Univ. Press, Oxford.

Weiskrantz, L (1998). Paperback edition, with updating, of *Blindsight. A case study and implications* (1986). Oxford Univ. Press, Oxford.

Weiskrantz, L., Barbur, J.L., and Sahraie, A. (1995). Parameters affecting conscious versus unconscious visual discrimination without V1. *Proc. Natl. Acad. Sci.* USA, **92**, 6122-6126.

Weiskrantz, L., Cowey, A., and Barbur, J.L. (1999). Differential pupillary constriction and awareness in the absence of striate cortex. *Brain*, **122**, 1533-1538.

Weiskrantz, L., Cowey, A., and Le Mare, C. (1998). Learning from the pupil: a spatial visual channel in the absence of V1 in monkey and human. *Brain*, **121**, 1065-1072.

Weiskrantz, L., Harlow, A., and Barbur, J.L. (1991). Factors affecting visual sensitivity in a hemianopic subject. *Brain*, **114**, 2269-2282.

Weiskrantz, L., Warrington, E.K., Sanders, M.D., and Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain*, **97**, 709-728.

Weiskrantz, L. and Warrington, E.K. (1979). Conditioning in amnesic patients. *Neuropsychologia*, **17**, 187-194.

Young, M.P. (1992). Objective analysis of the topological organization of the primate cortical visual system. *Nature*, **358**, 152-155.

Young, M.P. (1993). The organization of neural systems in the primate cerebral cortex. *Proc. R. Soc. Lond B*, **252**, 13-18.

Dr. Lawrence Weiskrantz

first recipient of the Herbert H. Reynolds Lectureship in the History and Philosophy of Science

author and emeritus professor of psychology, Oxford University

Dr. Weiskrantz, author and emeritus professor of psychology at Oxford University, served as professor of psychology and director of the Department of Experimental Psychology at Oxford from 1967 to 1993.

Prior to the Oxford posts, he held an appointment at Cambridge University for 11 years. Dr. Weiskrantz earned a B.Sc. degree from Oxford and a Ph.D. from Harvard University. He is a Fellow of the Royal Society of London and a member of the National Academy of Sciences.

Dr. Weiskrantz is the recipient of various awards, including the Craik Prize from Cambridge, the Hughlings Jackson medal from the Royal Society of Medicine and the WIlliam James Fellowship of the American Psychological Society. He delivered the Royal Society's Ferrier Lecture in 1989.

He is the author of *Blindsight: a case study and implications* (Oxford 1986) and *Consciousness Lost and Found: a Neuropsychological Exploration* (Oxford 1997). He is editor of and contributor to *Analysis of Behavioural Change* (Harper & Row 1968), *Neuropsychology of cognitive function* (Royal Society 1982), *Animal Intelligence* (Oxford 1985), *Thought Without Language* (Oxford 1988), *Attention: Selection, Awareness and Control: A Tribute to Donald Broadbent* (Clarendon 1995) and *The Prefontal Cortex: Executive and Cognitive Function* (Oxford 1998).