Consciousness and Commentaries

L. Weiskrantz
University of Oxford, UK

There are several neuropsychological syndromes in which good residual function is retained in the absence of acknowledged awareness, among them blindsight, amnesia, and unilateral neglect. All of them point to the need in studying conscious awareness not only for an "on-line" demonstration of the relevant capacity, but for an independent commentary or classificatory response by the subject, whether human or animal. The parametric limits of blindsight (visual discrimination without awareness) can be measured using a "commentary key" psychophysical paradigm, and the results may possibly allow an approach to identifying neural structures involved in visual awareness.

Dans plusieurs syndromes neuropsychologiques, notamment la vision aveugle, l'amnésie et la négligence unilatérale, on observe un bon fonctionnement résiduel alors que le patient semble ne pas prendre conscience de perceptions ou des souvenirs. Tous ces syndromes indiquent que dans l'étude de la prise de conscience, il faut non seulement démontrer "on line" que le patient est conscient ou non des événements mais il faut aussi obtenir des indices indépendant sur l'état de conscience du sujet humain ou animal, par un commentaire ou par une réponse de classification. Les limites paramétriques de la vision aveugle (discrimination visuelle dans prise de conscience) peuvent être mesurées en utilisant le paradigme psychophysique du "commentaire-clé" et les résultats obtenus peuvent rendre possible une approche visant à identifier les structures nerveuses jouant un rôle dans la prise de conscience visuelle.

In recent years a surprising fact has emerged from neuropsychological studies of brain-damaged patients with cognitive disorders: in all of the syndromes, robust residual capacities remain of which the subjects themselves are unaware: they are opaque to the patient but not to the experimenter (cf. reviews by Milner & Rugg, 1992; Weiskrantz, 1986, 1991, 1996; Schacter, McAndrews, & Moscovitch, 1988). Thus, an amnesic patient disclaims any recognition or recall of recent events, and yet one can show by indirect methods such as priming or conditioning that the earlier event has been stored. The prosopagnosic patient shows no recognition of familiar faces, and yet his autonomic nervous system clearly distinguishes between familiar and unfamiliar faces. Moreover, the subject can link names to the familiar faces appropriately, and more efficiently than he can to unfamiliar faces, even though he does not recognize the familiar faces. Even in the most "human" of capacities, namely language, aphasic patients can show good preservation of the syntactical and semantic content when tested with reaction times to target words in normal vs. degraded sentences. The subject does not comprehend the sentences nor discriminate normal from degraded sentences, and his reaction time is slowed to targets in syntactically or semantically degraded sentences just as it is with normal subjects. Somewhere in the brain there still lurks a good capacity to do so. Again, subjects with unilateral neglect of the left half of visual space, a condition associated with damage to the posterior right hemisphere, can still show good evidence of processing visual events to which they do not respond explicitly. And in patients with damage to visual cortex, which causes "blindness" of the contralateral hemifield of vision, it is possible to demonstrate that they have an ability to detect, locate, and disciminate visual events in their blind field, a condition known as "blindsight". Even though the patients

Requests for reprints should be addressed to Dr. L. Weiskrantz, Department of Experimental Psychology, University of Oxford, Oxford, UK.

are blind to the events, with forced-choice guessing or other indirect methods they can discriminate certain events within certain limits (Weiskrantz, 1986, 1990, 1996; Weiskrantz, Warrington, Sanders, & Marshall, 1974)

Each of these conditions requires its own particular techniques with which to reveal the covert capacity, but they all share a dissociation of conscious awareness from a capacity. They are also all caused by known or knowable brain damage. Therefore they offer an interesting possible route to the study of brain mechanisms actually involved in conscious awareness or its disjunction. This neuropsychological approach differs, it will be noted, but also complements other approaches in this symposium, in that it deals with *changes* in awareness rather than the analysis of its properties or its electrophysiological correlates. Empirical analysis is usually helped by studying change rather than static states. Of all the syndromes (and the list above is by no means exhaustive), blindsight perhaps offers the most promising candidate for further analysis, because more is known about the physiology and anatomy of the visual system than any other brain system, and also the psychophysical methods for studying visual capacity are well established.

A point that emerges transparently from the phenomenon of blindsight, but is a feature of all the examples of residual capacity, is that one cannot draw any conclusions about whether a subject is or is not consciously aware of events (or of a capacity) simply by studying how good his performance is. The blindsight subject can discriminate relatively fine differences between wavelengths, or relatively fine differences between the orientations of gratings. He does so in his blind hemifield, and of course can also do it in his intact hemifield. From such evidence alone, it follows, one could not conclude whether this is done with or without awareness of colour or orientation. Similarly, from analyzing the reaction times of an aphasic patient, which demonstrate an intact syntactical and semantic capacity, one could not tell from that information alone that the subject cannot engage in conversation. That is, the "on-line" study of a capacity is not a sufficient basis for drawing any conclusions about consciousness. Instead one must go "off-line" (Tyler, 1988, 1992); one must obtain some independent evidence about the subject's state of awareness, and compare this with the subject's performance with and without awareness. In most cases we do this by asking the subject directly whether or not he is "aware" or "confident" (not the same thing, actually) of his discriminative choice. That is, we obtain a "commentary" from the subject. But, as we shall see, the "commentary" need not be verbal, nor restricted to human subjects.

This distinction—between performance with and without awareness—is sometimes discussed as a difference between "explicit" and "implicit" processing. But the point is the same. When a subject is performing in an "explicit" mode an appeal is made, not actually or necessarily spelled out as such, to an off-line commentary. For example, when the amnesic subject is asked to respond to items that are "recognized" (on which he is typically at chance), in effect he is being asked a question about whether he acknowledges that he has a specific memory. This is in contrast to a priming task (on which he typically performs well), in which no such question arises either directly or indirectly.

In clinical testing with blindsight or subjects with related syndromes (e.g. neglect, blind touch) usually the "commentary" phase arises first. It is first determined as part of the clinical screening that a subject with, say, occipital brain damage is phenomenally "blind" in the affected hemifield. Only later is it then determined (and historically this next phase took about 100 years) whether or not the subject can discriminate events in his blind field by forced-choice guessing or some indirect approach. In the course of doing so, one might conduct a block of trials with forced-choice guessing, let us say, and after each block simply ask the subject whether he was "aware" of any of the events. But in principle both the commentary and the discriminative response can be brought together and be made *after each trial*—the "commentary key paradigm" (Weiskrantz, 1986)—and in practice there is a considerable gain in doing this, as we shall see.

Before turning to specific results, some background to the topic of blindsight would be useful. It is not always appreciated that the eye sends not just one pathway to the brain—the often-studied pathway to lateral-geniculate nucleus and thence to the striate cortex (also known as V1 or Brodmann's area 17, or just "visual cortex")—but also to nine other pathways ending in different subcortical targets in the brain. Therefore when the striate cortex is removed in monkeys, it may not be surprising that primates can still carry out visual discriminations (Humphrey, 1974; Pasik & Pasik, 1982). Their

capacity is altered both qualitatively and quantitatively, the details of which we cannot go into here, but nevertheless their capacity is still quite impressive, e.g. a visual acuity of about 8 cycles/degree, an ability to discriminate orientation differences of about 8 degrees, and an almost normal ability to localize small, brief targets in space. The surprise is that human subjects with supposedly comparable lesions of striate cortex are "blind" in the corresponding part of their visual fields. As the human and monkey visual anatomy and capacities are closely similar, why should the apparent outcome be so different? It was only when humans were tested in the way in which one must of necessity test monkeys to discriminate that the gap began to narrow. That is, one cannot ask a monkey to tell one what it "sees"; one must give it a choice between alternatives or allow it to reach to a spatial location or to retrieve an object. When similar methods were used with human subjects in their "blind" fields (Pöppel, Held, & Frost, 1973; Weiskrantz, 1986; Weiskrantz et al., 1974), they too could perform at least some of the tasks that the monkeys could do with striate cortex lesions.

Striate cortex is connected to several other visual association areas, either directly or indirectly, via a rich network. But the removal or blockade of striate cortex in no way isolates this visual association complex from a retinal input. Well-known pathways exist, for example, from the midbrain (superior colliculus) to the thalamus (pulvinar) to visual association areas, which remain patent even without V1. This was demonstrated directly in electrophysiological recordings of area MT (also known as area V5) by Rodman, Gross, and Albright (1989) in the monkey. Neurones in MT continue to fire in the absence of V1, and indeed it was established that the route that allowed this to happen was via the superior colliculus. On the other hand, visual association areas may not be necessary for all visual function in "blindsight". Midbrain neurons are, after all, neurons with their own rich set of connections to more anterior regions of cortex (e.g. frontal lobes) as well as downstream (e.g. to the cerebellum). Recently it has been shown that even in hemispherectomy, when *all* cortex in one hemisphere is removed surgically (for the treatment, usually, of intractable epilepsy), there may still be demonstrable residual visual function in the "blind" hemifield (Tomaiuolo, Ptito, Marzi, Paus, & Ptito, 1997).

Through what may be a convenient by-product of evolutionary history, the residual function in "blindsight" is sometimes at a half-way house between total loss of visual awareness and normal vision. Sharply transient events—a flash with a sharp onset, or a rapidly moving event—in the affected hemifield can produce what subjects report as a kind of "awareness" or a "gut feeling" that something has happened, but this is said by them not to be "seeing" as such. They may even be able to locate it and even sense in what direction a moving stimulus is travelling. But for non-transient events, such as slowly moving targets, or stimuli with smooth and shallow Gaussian temporal envelopes, there is no awareness, and yet discriminative performance can be good. And for qualitative aspects of stimuli, e.g. colour, or orientation, or spatial frequency of a grating, or shape of an object, there is no awareness as such even though in all these domains there can be good residual function. It is, as it were, responding in the absence of "qualia", even qualia of which philosophers are so fond, such as colour.

This distinction—between transient and non-transient events in the blind hemifield—can be combined with the "commentary key paradigm" in a way that is potentially of some special interest. My colleagues, Drs. John Barbur and Arash Sahraie, and I have exploited it in the case of one well-studied subject, GY, who sustained striate cortex damage in his left hemisphere after a head injury when he was 8 years old (he is now 40), and has a corresponding total field defect in his right visual hemifield (except for a small area of "macular sparing", a common feature in many cases of striate cortex damage). He has a "feeling" with rapidly moving stimuli that something has moved, although he does not "see" anything as such. But outside the range of rapidly transient stimuli, either slow moving, or of weak contrast, he can still discriminate the direction of movement quite well, even though he has no experience of anything at all.

This was put on a quantitative basis by asking GY on every trial to respond (with the usual two-alternative forced-choice method of psychophysics) by pressing one key if a target (it was actually a projected red laser beam) moved horizontally, and to press a second if it moved non-horizontally. But on every trial he was also provided with two other "commentary" keys, numbers 3 and 4. In addition to responding on key 1 or 2, he was instructed that he was to press key 3 if he had any experience of the event, even a

feeling or even a faint tickle. And he was to press key 4 if he had absolutely no experience whatever. Thus, we could compare discriminative performance with his commentary performance. We varied stimulus velocity, stimulus excursion, and also contrast (by varying background luminance).

As shown in Fig. 1, it can be seen that performance could remain high relatively independently of acknowledged awareness. That is, as background luminance increased (thereby lowering stimulus contrast), the percentage of trials on which he signalled "aware" dropped sharply, without any change in performance. Similarly, with a slow velocity, performance could be good without any reported awareness. As velocity increased to approach the "transient" range, the percentage of awareness reports increased, and performance remained good (Fig. 2). Above a certain level when he was well into the transient range, of course, the subject reported awareness on all trials.

One implication of such a pattern of results is that it ought, in principle, to match the performance of the subject when he reports "awareness" with that when he reports "no awareness." For example, in Fig. 3, the results are shown for an experiment when we varied the angle between horizontal vs. non-horizontal movement. The results are plotted for "aware" reports on every trial, and for "unaware" reports on every trial, using different velocities for these two modes. As angular separation increases, not surprisingly, discriminative performance also improves. But note that performance in both the "aware" mode and the "unaware" mode show parallel functions, and actually converge on the same high value, approximately 95% correct (chance being 50%).

This implication carries another: it ought to be possible to carry out brain imaging in each of the two modes, and thus to see whether there is a pattern of activity associated with visual awareness as such. Such an experiment was recently carried out by myself, John Barbur, and Arash Sahraie (for the psychophysics), and colleagues at the FMRI imaging centre at the Institute of Psychiatry in London (Drs. S. Williams, A. Simmons, and their team). The analysis is still in progress, but a short summary appears as an addendum to a book (Weiskrantz, 1997) and a full multi-authored paper will be submitted when the analysis is complete[1]. The main result appears to

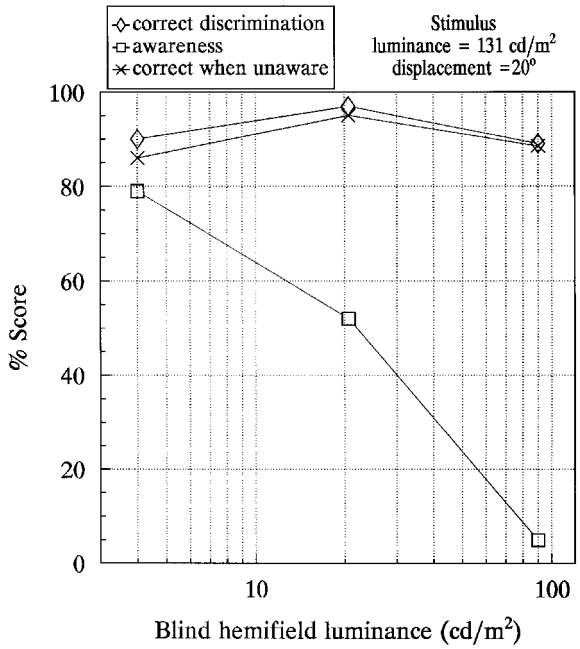

FIG. 1. Discrimination of horizontal vs. vertical movement as a function of stimulus contrast. The subject had to indicate (by guessing, if necessary) whether the presented stimulus was moving horizontally or vertically by pressing the appropriate response key. He also had two "commentary" keys to indicate "aware" or "unaware" on every trial. "Awareness" refers to percentage of trials on which the subject pressed the aware key. "Correct when unaware" refers to performance during those trials when the subject pressed the unaware key. The luminance of the test stimulus was held constant at 131 cd/m^2. The background luminance in the blind field was changed systematically, thus altering the contrast of the stimulus. Stimulus speed was 15°/sec, and displacement was 20°. Note the relative stability of the high level of performance independent of stimulus contrast, despite the steep decline in percentage of awareness responses with decreasing contrast (increasing luminance). (Reprinted from Weiskrantz et al., 1995, with permission. Copyright National Academy of Sciences, USA.)

be a difference between cortical and subcortical foci in the two modes, but many of the details remain to be established.

It was, of course, necessary to repeat the psychophysical determinations for the purposes of brain imaging, because the actual physical environment of the imager imposes certain constraints. This was done. But we also followed up other psychophysical questions. For example, our original "commentary key paradigm" procedure used only two responses—"aware" or "unaware". It might be that this is too crude a distinction—the "unaware" mode may contain some "smidgen" of awareness. And so we also used a six-point scale of awareness. The results were essentially the same: when GY reported zero awareness, this was so whether he used a binary

[1] Results and analysis now published; cf. Sahraie, Weiskrantz, Barbur, Simmons, Williams, and Brammer, 1997.

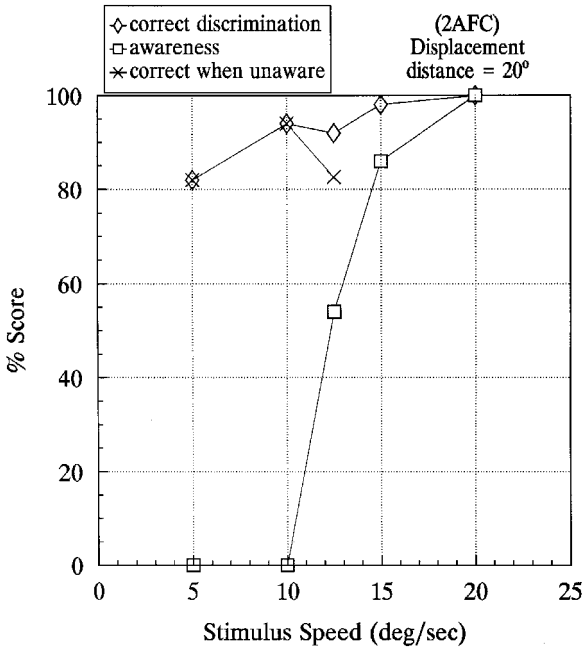

FIG. 2. Awareness and discrimination performance for a horizontal vs. a vertical movement, as a function of stimulus speed, with fixed displacement of 20°. Explanation of key as in Fig. 1. Note high levels of discriminative performance at speeds at which the subject reported no awareness. (From Weiskrantz et al., 1995, with permission. Copyright National Academy of Sciences, USA.)

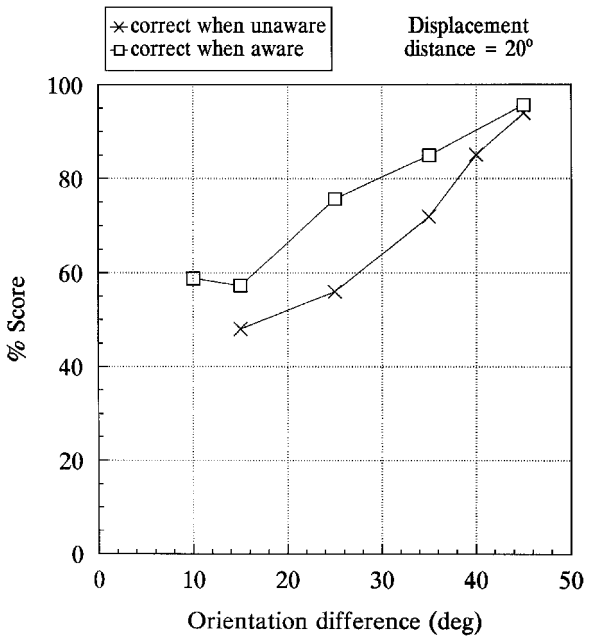

FIG. 3. Discrimination of horizontal from nonhorizontal orientation of movement, as a function of angular difference. Displacement was 20°. Key as in Fig. 1. The "correct when unaware" curve was obtained with a speed of 10°/sec, and the "correct when aware" curve with a speed of 20°/sec. The two curves converge on the same high level of performance as the horizontal-nonhorizontal difference increases. (Based on Weiskrantz et al., 1995, with permission.)

or six-point scale, and the discriminative performance was also unchanged. We also compared "awareness" with "confidence" ratings and found that they did not produce identical results. "Awareness" seemed to be a more direct and useful measure. But we were also able to show that there could be discriminative performance well above chance even with "zero" confidence. Finally, when we gave two extra keys, such that he reported both awareness and confidence levels, discriminative performance declined. It may not be surprising that you can only ask a subject to do so much in any trial!

This is but one paradigm in which "aware" vs. "unaware" modes of performance can be seen within the "blind" field. There are others that are equally interesting, e.g. the discrimination between stimuli of long wavelengths vs. achromatic targets, with varying luminance. Performance can remain well above chance in the "blind" hemifield independently of any luminance value of the achromatic comparison (Cowey, personal communication). A similar phenomenon emerges in experiments in which the cues and targets in an attention paradigm are reduced to the "unaware" level (Kentridge & Heywood, personal communication). Alternative approaches to the same question of brain mechanisms of aware vs. unaware modes are thus potentially available for brain imaging. It should also be noted that the "aware" mode had already been imaged in isolation in a PET study a few years ago with the same subject (GY), but this was before the "unaware" vs. "aware" modes became uncovered in this subject (Barbur, Watson, Frackowiak, & Zeki, 1993).

Finally, one may return the question to its historical origins. The phenomenon of blindsight emerged originally, as noted, from animal experiments in which it was found that striate cortex removal does not abolish visual discriminations. In the contralateral hemifield the animals can still carry out a range of visual discriminations at a high level, e.g. detect and locate brief visual stimuli and discriminate between gratings of different orientation. But does the animal show "blindsight"? That is, does it treat a visual stimulus as "visual"? Does it do so without "awareness"? The question was addressed experimentally in an ingenious study by Cowey and Stoerig (1995). They first confirmed, as had already been known, that monkeys with complete

unilateral striate cortex removal could respond excellently to and locate the position of small and brief lights presented to their affected hemifield. They then proceeded to a second stage of the experiment. The animals were trained, in their intact hemifields, to discriminate between "lights" and "blanks", presented in a random order. If a small light target was presented, the animals were rewarded for pressing the target. If a blank was presented, the animals were trained to press a separate panel. Now the question arises: what does the animals do when a probe light is presented in the *affected* hemifield? The answer is very clear: the animals press the "blank" panel. This is so, not withstanding the evidence in the same animals that they can detect and locate light stimuli. But when asked to classify such a stimulus, the animals treat the event as a "blank", a nonvisual event. They behave like human blindsight subjects.

The "commentary key" approach as a general solution to the question of "animal consciousness" is not, it should be stressed, complete in itself. In the human subject we start with a primitive assumption that awareness exists in the normal state, based on an argument from analogy with our own experience (our self-commentaries, if you will). The same primitive assumption must be made for the animal—there is no sleight of hand or detour around this issue. And, as with the human, we use an argument from analogy with ourselves in a similar situation, but this has no associated litmus paper test yielding a certain diagnosis, especially when the animal and the human are dissimilar in their biology. But when the human's and the animal's "off-line" responses *change* in the same way, as we have seen is the case for striate cortex lesions, the argument from analogy is strengthened, especially when we can appeal to virtually identical anatomical visual systems in humans and monkeys. There are also other approaches to the question of animal awareness, which are beyond the limits of the present discussion (Dickinson, 1988; Weiskrantz, 1997).

The message brings us full circle not only in terms of empirical evidence, but also conceptually. From the ongoing "on-line" performance in a visual discrimination task, one cannot draw any conclusions about the meaning of the stimulus for the subject—animal or human—either in terms of awareness or whether it is "visual" experientially. (Conversely, of course, one cannot determine from on-line performance that an animal is *not* conscious, even if the creature is rather unlike ourselves.) The performance can be excellent in the "blind" field. To find out one must go "off-line." A separate commentary, or a separate classificatory response, is needed. And when one combines the "off-line" with the "on-line" within the same subject, and within the same brain imager, a possible route to the understanding of the brain mechanisms involved in awareness becomes available.

If and when such mechanisms become better understood, this is not tantamount to awareness being "reduced" to their level, and hence disappearing as a phenomenon, as is sometimes advanced by some advocates, for example, of "strong AI". Awareness remains the target of what it is that we are trying to understand. I prefer to think that we should continue to seek explanations that are fully adequate and do justice to the phenomena we wish to understand, "elevationism", if you will, and not "reductionism".

REFERENCES

Barbur, J.L., Watson, J.D.G., Frackowiak, R.S.J., & Zeki., S. (1993). Conscious visual perception without V1. *Brain*, *116*, 1293–1302.

Cowey, A., & Stoerig, P. (1995). Blindsight in monkeys. *Nature*, *373*, 247–249.

Dickinson, A. (1988). Intentionality in animal conditioning. In L. Weiskrantz (Ed.), *Thought without language* (pp. 305–325). Oxford: Oxford University Press.

Humphrey, N.K. (1974). Vision in a monkey without striate cortex: A case study. *Perception*, *3*, 241–255.

Milner, A.D., & Rugg, M.D. (1992). *Neuropsychology of consciousness*. London: Academic Press.

Pasik, P., & Pasik, T. (1982). Visual functions in monkeys after total removal of visual cerebral cortex. *Contributions to Sensory Physiology*, *7*, 147–200.

Pöppel, E., Held, R., & Frost, D. (1973). Residual visual function after brain wounds involving the central visual pathways in man. *Nature*, *243*, 295–296.

Rodman, H.T., Gross, C.G., & Albright, T.D. (1989). Afferent basis of visual response properties in area MT of the macaque. I. Effects of striate cortex removal. *Journal of Neuroscience*, *9*, 2033–2050.

Sahraie, A., Weiskrantz, L., Barbur, J.L., Simmons, A., Williams, S.C.R., & Brammer, M.L. (1997). Pattern of neuronal activity associated with conscious and unconscious processing of visual signals. *Proceedings of the National Academy of Sciences USA*, *94*, 9406–9411.

Schacter, D.L., McAndrews, M.P., & Moscovitch, M. (1988). Access to consciousness: Dissociations

between implicit and explicit knowledge in neuropsychological syndromes. In L. Weiskrantz (Ed.), *Thought without language* (pp. 242–278). Oxford: Oxford University Press.

Tomaiuolo, F., Ptito, M., Marzi, C.A., Paus, T., & Ptito, A. (1997). Blindsight in hemispherectomised patients as revealed by spatial summation across the vertical meridian. *Brain*, *120*, 795–803.

Tyler, L.K. (1988). Spoken language comprehension in a fluent aphasic patient. *Cognitive Neuropsychology*, *5*, 375–400.

Tyler, L.K. (1992). The distinction between implicit and explicit language function: Evidence from aphasia. In A.D. Milner & M.D. Rugg (Eds.), *The neuropsychology of consciouness* (pp. 159–179). London: Academic Press.

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

Weiskrantz, L. (1990). Outlooks for blindsight: Explicit methodologies for implicit processes. The Ferrier Lecture. *Proceedings of the Royal Society of London*, *B239*, 247–278.

Weiskrantz, L. (1991). Disconnected awareness for detecting, processing, and remembering in neurological patients. The Hughlings Jackson Lecture. *Journal of the Royal Society of Medicine*, *84*, 466–470.

Weiskrantz, L. (1996). Blindsight revisited. *Current Opinion in Neurobiology*, *6*, 215–220.

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

Weiskrantz, L., Barbur, J.L., & Sahraie, A. (1995). Parameters affecting conscious versus unconscious visual discrimination without V1. *Proceedings of the National Academy of Sciences, USA*, *92*, 6122–6126.

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