COMMENTARY RESPONSES AND CONSCIOUS AWARENESS IN HUMANS: THE IMPLICATIONS FOR AWARENESS IN NON-HUMAN ANIMALS

L Weiskrantz

Department of Experimental Psychology, University of Oxford,
South Parks Road, Oxford OX1 3UD, UK

Abstract *Animal Welfare 2001, **10:*** S41-46

There are several examples from human neuropsychology of the intact preservation of processes and capacities in the absence of conscious awareness by the patients. These include loss of visual awareness (blindsight), episodic memory (amnesic syndrome), attention (unilateral neglect) and language (aphasia). The implication of these and other clinical phenomena is that descriptions of ongoing behaviour are necessary but quite insufficient for making inferences about conscious awareness, because even quite 'high level' behaviour can be run off in the absence of awareness. A commentary, or independent off-line response, is a prerequisite for determining whether the subject is consciously aware. Whether or not the commentary allows an inference about awareness in animals rests ultimately on an argument from analogy, just as is the case when we make judgements about fellow humans. But when parallel disjunctions between on-line behaviour and off-line classifications are found for both human and infra-human subjects, as is demonstrable for blindsight and amnesia, not only do they bolster inferences about common neural mechanisms, but they strengthen inferences for analogous processing and hence for conscious experience.

Keywords: *amnesia, animal consciousness, animal welfare, awareness, commentary responses, neuropsychology*

There are many examples from human clinical neuropsychology that demonstrate the preservation of intact behavioural capacities in the absence of acknowledged conscious awareness (see Weiskrantz [1997] for review). For example, lesions of the visual cortex produce 'blindness' in corresponding parts of the visual field, and yet visual discriminations can be shown to be possible even though the patient does not 'see'. Similarly, it has been demonstrated that patients with the amnesic syndrome, caused by lesions to medial temporal lobe regions, can readily acquire and store new information even though they are severely disabled because they do not treat the stored items as 'memories'. Functionally, the patients are detached from their intact memory system. Aphasic patients can be shown to respond to semantic and syntactic features of language by using reaction time measures, even though such patients cannot discriminate or be guided by such features in their daily lives. Patients with unilateral neglect, an attentional disorder associated with right posterior brain lesions, can nevertheless show sensitivity to stimuli of which they claim no existence. Even at the level of spinal cord mechanisms, paraplegic patients still respond to noxious stimuli with appropriate reflexes even

though they do not feel any pain below the level of the spinal cut. To quote Charles Sherrington (1957 p 257):

> A needle-prick causes invariably the drawing up of the limb ... It is those stimuli which are most fitted to excite pain which are prepotent ... The nervous arcs of pain-nerves, broadly speaking, dominate the spinal centres ... where pain is, of course, non-existent.

The flexible and adaptive features of the frog's spinal cord isolated from the brain are impressive, so much so that William James (1890) concluded that the isolated spinal cord of the frog demonstrates 'conscious intelligence'. But all of these examples from neuropsychology, and the evidence from paraplegia, demonstrate one salient point: one cannot make any valid judgment from the contemporary ongoing stream of behaviour as to whether a subject is consciously aware of the ongoing events and the capacities they demonstrate. Of course, there are lots of examples in everyday life where this is true – we are not normally aware of oxygen levels that control breathing rates, or the intricate perceptual skills involved in balancing, or in avoiding bumping into passers-by on a crowded pavement, etc. Indeed, the philosopher Alfred North Whitehead (1948 p 61) has gone so far as to applaud the value of non-thinking behaviour:

> It is a profoundly erroneous truism ... that we should cultivate the habit of thinking of 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.

But the examples from neuropsychology are more informative, because normal subjects do have an awareness in just those domains where the patients do not. If a human paraplegic cannot feel the painful stimuli, yet shows a clear withdrawal response, it follows equally we cannot tell from on-going behavioural withdrawal or other elicited responses alone that an intact frog, cat or monkey feels pain. On-line descriptions are insufficient, although they are necessary.

What would be sufficient for drawing a valid judgment that a subject is aware of ongoing events? Let us leave sufficiency for the moment: it is a difficult criterion to meet, as we shall see. But can we say what else is necessary? Typically, with human subjects, we base a final judgment about awareness on a separate and additional off-line commentary, ie we ask the subject what he just saw or remembered, or the subject volunteers such a commentary. I would argue that such an off-line parallel response is an absolute prerequisite for forming a judgment about another person's conscious awareness of the relevant ongoing events. The commentary response, of course, need not be verbal – it can be a response key, or body language, even a smile. In blindsight and other similar syndromes, there is a serious disjunction between the discrimination of visual events and the subject's commentary. The subject can discriminate visual events, but reports that he or she was not aware of seeing anything. It is easy to put this relationship under operational control experimentally.

For example, we have provided the blindsight subject with two response keys to indicate whether a spot in the blind visual field was moving in a horizontal or a non-horizontal direction, and two further response keys to indicate whether he had any awareness of the event. The instructions were quite conservative: only press the 'no' awareness key if there is absolutely no 'feeling', 'inkling', 'knowing' or any other awareness of the movement. With appropriate parameters, high levels of correct discrimination performance were reliably obtained even though they subject consistently pressed the 'no' key (Weiskrantz *et al* 1995). Similarly, an amnesic subject can be required to guess the correct answer to a question that requires

knowledge based on acquired storage for an item, and yet consistently respond 'no' on a formal recognition test for that item (vs genuinely new items).

The reason that a commentary is not sufficient is because it can become rendered automatic under conditions of redundancy and repetition. A nice example, with literary licence for exaggeration, is what I have termed the BWCS – the British Weather Conversation Syndrome. The British, as is well known, have the remarkable capacity to discuss the weather with comments delivered at ballistic rates of delivery, in something like a state of complete and unknowing unconsciousness, much like sleep-walking, except that even the sleep-walker can sometimes recollect what he or she has done, unlike the total amnesia for a standard reflex comment like, 'nice day'. And so commentaries, in a literal sense, can become incorporated in a highly automatic routine.

But if an off-line response is necessary but not always sufficient, how do we judge its validity? When a person says he was conscious of the passing motor car, or of the blackness of the total eclipse, why do we believe him? We do so because normally that corresponds to our own experience, and so in the end the proof is inferential, it is based on an analogy. But this in itself is no great handicap – it enables us to negotiate with our fellows perfectly adequately most of the time.

Such considerations are important when we extend the same conditions of necessity and sufficiency to animals. We start with the same assumption as with humans – namely, that we cannot form a valid judgment of an animal's awareness based strictly on the ongoing stream of behaviour. We need an off-line response, something that tells us the status of ongoing behaviour. But two problems arise when we switch to non-human animals. Firstly, we are less confident about validating the off-line response based on our own experiences, and sometimes not confident at all. *What is it Like to be a Bat?* Thomas Nagel challenged famously (Nagel 1980). An assumption has to be made relating their classifications to our own, such that, when an animal is trained to detect a brief light by making a response to it, for example, its off-line classification of light vs no-light would parallel our own. The second point is that, just as with human subjects, the off-line commentary is typically redundant. We do not ask a human subject who detects a light whether he was aware of it. 'Of course I am,' will be the answer. We could soldier on, and ask whether a light is different from a non-light, and get the same answer, uttered with some exasperation. And similarly with animals, asking a dog whether a light is different from a non-light, as a separate question, will not yield any great breakthroughs in understanding.

The crux comes, as it does with humans, when there is a serious disjunction between the on-line behaviour and the off-line commentaries. This is why I think that human neuropsychology offers some useful entrées. Here is where we might be able to make some close comparisons and relate them to other features of the known anatomy and structure. Are there parallel disjunctions in non-human animals? Two types of on-line/off-line disjunctions seen in human patients have been followed up in animal experiments. The first relates to the amnesic syndrome, in which excellent learned discriminations of objects can be demonstrated by animals with lesions meant to simulate those involved in the amnesic syndrome, but the animals do not show recognition of those very same objects when tested independently on familiar vs unfamiliar objects (Gaffan 1974). That is, they can learn that object A is associated with food reward, and object B is not, but in a formal recognition test when they are rewarded for responding only to the objects they have seen before, they do not treat A or B as different from an object C that they have never seen. There is obvious and clear evidence of storage without recognition, just as in amnesia.

An even more striking and challenging example comes from blindsight research. Recall the human paradigm: a human subject with visual cortex ('V1' or 'striate cortex') damage can detect

and discriminate visual stimuli in the blind field, but reports no awareness of them on the commentary keys (Weiskrantz 1998). Is there an animal parallel? The matter has been investigated in experiments (Cowey & Stoerig 1995, 1997; Stoerig & Cowey 1997) in which the primary visual cortex was removed from one hemisphere in monkeys, simulating the clinical situation with human patients that results in cortical blindness in the contralateral visual hemifield. I should make it clear here that neither the human patients nor the experimental monkeys appear disturbed in their everyday behaviour because they have a good half-field of vision which, with intact eye and head movements, is perfectly adequate for everyday purposes.

The Cowey-Stoerig experiments proceeded as follows: the animals were first trained to fixate and then, whilst maintaining fixation, to press a briefly presented dim light projected into their affected hemifields, being rewarded with a peanut for doing so. The position of the light was changed randomly. Such hemianopic monkeys can be trained to do this very well, with virtually perfect performance. In the second part of the study, the animals were next trained on a new, separate task, using their intact, unaffected hemifields. In this task, lights were presented randomly on half of the trials, and 'blanks' on the other half. The animals were rewarded differentially for pressing the lights when they appeared, but for pressing a separate response panel for those trials when 'blanks' appeared. This, of course, is a simple task for the monkey. Now, the crucial question is: how would the animal respond when lights were presented from time to time, as probes, in their 'blind' hemifields? The answer was clear: the animals treated the lights in their blind fields as 'blanks', even though they had just demonstrated their ability to detect them. In other words, just as the human blindsight subject does, they treat visual events in their blind fields as being non-visual. Cowey and Stoerig made the lights in the blind field bright and moving or flickering, and obtained the same robust results. No-one has yet succeeded in training the monkey in a strictly parallel way to the human situation, where we combine the discrimination and the commentary keys within the same trial, but the principle in the two-stage procedure is the same.

The results still do not tell us whether the animal is completely unaware of the light, just that it treats it as a non-light. More experiments would be needed to see whether it treats it as different from a non-visual stimulus, like a click (see Humphrey [1974]). But the point is that one uses the off-line behaviour to tell us the animal's classification of what it is responding to on-line, and hence ultimately to its experience, if and when we are willing to extrapolate from our own experience in the same situation. The very fact of the parallellity of the disjunctions strengthens the inferences about the character of the on-line and off-line subjects. I think it is not difficult to infer that the animal with visual cortex is aware of visual events in a way that it is certainly not aware of in the absence of such cortex, and that therefore visual awareness is a relevant dimension on which to relate the animal to ourselves. The fact of the parallel disjunction in people and animals between performance and conscious experience strengthens the inference about the off-line response in relation to experience and our willingness to accept the existence of this domain in the animal. It also further strengthens inferences that one might draw from brain imaging experiments in relation to experience. There is now functional magnetic resonance imaging evidence from a human blindsight subject contrasting brain activity when he is visually aware as opposed to visually unaware of stimuli in his blind hemifields (Sahraie *et al* 1997; Barbur *et al* 1998). A further basis could be emerging for using neuroscientific evidence as criteria for analogous states in animals.

The much more challenging questions, at least from the animal welfare point of view, concern whether an animal experiences pain. I know of no experimental demonstrations of disjunctions involving aversive stimuli, but by extension from the evidence from human paraplegics it would seem a reasonable surmise that a spinal mammal would still be able to

discriminate a touch below the level of the cut, if only by seeing its own spinal reflex, but would still not, in an independent discrimination, classify this as a punishing stimulus. But what about the spinal frog, a focus of interest to William James (1890) and others, who claimed that the fact that the spinal reflexes were adaptive and flexible meant that it exhibited conscious intelligence? I have my doubts about the spinal frog, and do not advocate anyone doing that experiment, but note that the crux of the demonstration would depend upon first being able to demonstrate that the animal could classify punishing from non-punishing events. Only then would the dysfunction be informative, because otherwise it could be claimed that the classification task itself requires an independent validation, as with a preference task or reinforcement of an operant response.

To summarize my argument: descriptions of ongoing behaviour are necessary but quite insufficient for making inferences about conscious awareness, because of demonstrations that even quite 'high level' behaviour can be run off in the absence of awareness. A commentary, or independent off-line response, is a prerequisite. Whether or not the commentary allows an inference about awareness or not rests ultimately, just as it does when we make judgments about fellow humans, on an argument from analogy. But when parallel dysfunctions are found, not only do they bolster inferences about neural mechanisms, but they also make it easier to assume that the analogy implies that the off-line response reveals the nature of the processing involved in the on-line behaviour and the accompanying experience.

Analyses of complex sequential behaviour might break down into comparable distinctions between on-line and off-line, because the interpretation of the behaviour at one moment can be illuminated by what follows. Something of this kind is involved, perhaps, in what Dickinson (1988) calls an 'intentional' account of animal behaviour, by which one makes inferences about 'beliefs and desires' (Dickinson 1988 p 307):

> If an intentional account of behaviour is to be anything more than a redescription of behaviour, the animal must, when faced with a change in conditioning schedule, reinforcer type, or drive state, adjust its actions to maintain their goal-directedness in a way that confirms to content of the behaviours and desires supported by the new state of affair.

The situation in which Dickinson applied this approach was to ask whether rats could adjust their behaviour appropriately in an instrumental situation when the reward was weakened after training. He studied the conditions under which the rats could or could not adjust. If the reward is degraded after the animal has learned the task, will it proceed automatically in reeling off a fixed 'habit', or will it change its response appropriately to indicate that it is less interested in working for the less desirable end product? Dickinson found, among other results, that in some situations – but not in all – rats would demonstrate their altered anticipations if they were not over-trained such that the response pattern became automatic. In other words, the subsequent behaviour pattern is 'off-line' with regard to the initial 'on-line' pattern, prior to the change in reward strength, and can be taken as a commentary on the animals' initial expectations.

The evidence rests, as it must do, on formal demonstrations, but I think we use similar ideas in the states we attribute to animals in everyday life. My late pet dog used to display obvious intense signs of fear of wasps, no doubt because of some past unfortunate episode. I was sitting in my drawing room one summer day when I noticed her cowering in the distant corner, as far away as possible from a wasp buzzing by a window. Was she simply demonstrating automatic withdrawal, just as the paraplegic withdraws a limb without experiencing the noxious stimulus? I finally persuaded the wasp to rejoin the great outdoors by opening the window. Immediately afterwards the dog emerged from her corner, tail-wagging, came over to me and actively licked my hand. G C Grindley used to say that one could teach an animal to say 'please' but not 'thank

you'. I think he was wrong, and that my dog's off-line response was exactly that: 'I was aware of that damned wasp and I am so grateful for what you did'. It is the off-line commentary responses of both our human and animal partners that so illuminate and enrich our lives.

References

Barbur J L, Sahraie A, Simmons A, Weiskrantz L and Williams S C R 1998 Processing of chromatic signals in the absence of a geniculostriate projection. *Vision Research 38:* 3447-3453

Cowey A and Stoerig P 1995 Blindsight in monkeys. *Nature 373:* 247-249

Cowey A and Stoerig P 1997 Visual detection in monkeys with blindsight. *Neuropsychologia 35:* 929-1997

Dickinson A 1988 Intentionality in animal conditioning. In: Weiskrantz L (ed) *Thought without Language* pp 305-325. Oxford University Press: Oxford, UK

Gaffan D 1974 Recognition impaired and association intact in the memory of monkeys after transaction of the fornix. *Journal of Comparative and Physiological Psychology 86:* 1100-1109

Humphrey N K 1974 Vision in a monkey without striate cortex: a case study. *Brain, Perception 3:* 241-255

James W 1890 *Principles of Psychology.* Macmillan: London, UK

Nagel T 1980 What is it like to be a bat? *Philosophical Review 83:* 435-451

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. *Proceedings of the National Academy of Sciences, USA 94:* 9406-9411

Sherrington C C 1957 Spinal cord. *Encyclopaedia Britannica 21:* 227

Stoerig P and Cowey A 1997 Blindsight in man and monkey. *Brain 110:* 77-92

Weiskrantz L 1997 *Consciousness Lost and Found. A Neuropsychological Exploration.* Oxford University Press: Oxford, UK

Weiskrantz L 1998 edition *Blindsight: a Case Study and Implications.* Oxford University Press: Oxford, UK

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

Whitehead A N 1948 *Introduction to Mathematics.* Oxford University Press: Oxford, UK