16 Ross, B.H. and Murphy, G.L. Food for thought: cross-classification and category organization in a complex real-world domain *Cognit. Psychol.* (in press)
17 Heit, E. and Rubinstein, J. (1994) Similarity and property effects in inductive reasoning *J. Exp. Psychol. Learn. Mem. Cognit.* 20, 411–422
18 Wisniewski, E.J. (1996) Construal and similarity in conceptual combination *J. Mem. Lang.* 35, 1–20
19 Wisniewski, E.J. (1997) When concepts combine *Psych. Bull. Rev.* 4, 167–183
20 Wisniewski, E.J. and Bassok, M. (1996) On putting milk in coffee: the effect of thematic relations on similarity judgments *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, pp. 464–468, Erlbaum

The neural correlates of conscious experience: an experimental framework

Chris Frith, Richard Perry and Erik Lumer

Demonstrating that neural activity 'represents' physical properties of the world such as the orientation of a line in the receptive field of a nerve cell is a standard procedure in neuroscience. However, not all such neural activity will be associated with the mental representations that form the contents of consciousness. In some cases, such as when patients with blindsight correctly 'guess' the location of a stimulus, neural activity is associated with physical stimulation and with appropriate behaviour, but not with awareness. To identify the neural correlates of conscious experience we need to identify patterns of neural activity that are specifically associated with awareness. Experiments aimed at making such identifications require that subjects report some aspect of their conscious experience either verbally or through some pre-arranged non-verbal report while neural activity is measured. If there is some characteristic neural signature of consciousness, then this should be distinguishable from the kinds of neural activity associated with stimulation and/or behaviour in the absence of awareness. It remains to be seen whether the neural signature of consciousness relates to the location of the neural activity, the temporal properties of the neural activity or the form of the interaction between activity in different brain regions.

In the last few years there has been a dramatic increase in the willingness of neuroscientists to speculate about the biological basis of consciousness[1,2]. At the same time philosophers increasingly refer to neuropsychological data when discussing the nature of consciousness[3,4]. From both sides of the debate we are told that questions about the neural basis of consciousness can now be answered through experimentation. In this article we shall try to make as clear as possible the assumptions that underlie such experiments and indicate the areas where progress is likely to be made. We shall consider only the problem of the *association* between consciousness and neural activity. We believe that systematic exploration of the neural correlates of consciousness will increase our understanding of the nature and purpose of consciousness.

A fundamental assumption is that for every mental state (state of consciousness) there is an associated neural state; it is impossible for there to be a change of mental state without a corresponding change in neural state. Questions about the neural correlates of consciousness are essentially questions about the relationships between mental states and neural states.

Levels versus contents of consciousness

A useful distinction can be made between factors influencing the overall level of consciousness and those determining its content. The former are generally associated with arousal, which is controlled at least in part by the ascending reticular activating system of the pons. Other structures implicated in arousal include the locus coeruleus, involved in vigilance, and the intralaminar nuclei of the thalamus,

C. Frith, R. Perry and E. Lumer are at the Wellcome Department of Cognitive Neurology, Institute of Neurology, 12 Queen Square, London, UK WC1N 3BG.
tel: +44 171 833 7456
fax: +44 171 813 1445
e-mail: cfrith@fil.ion.ucl.ac.uk

1364-6613/99/$ – see front matter © 1999 Elsevier Science. All rights reserved. PII: S1364-6613(99)01281-4

Box 1. Do animals dream?

Inferring consciousness in the absence of a verbal report

While there continues to be argument about whether animals other than man are conscious and what they are conscious of, we are confident that animals (at least mammals) dream. What is the basis of this confidence? Essentially it is that we have a very reliable and specific physiological marker for dreaming.

On the basis of measures of electroencephalography (EEG) and electromyography (EMG), sleep can be divided into several clear stages. In one of these stages the EEG shows the low-voltage, fast activity characteristic of waking, but muscle tone is actively inhibited and behavioural sleep persists. Stereotypic bursts of saccadic eye movements called rapid eye movements occur, giving this state the name REM sleep. In 90–95% of cases, arousal from REM sleep yields reports of dreaming when the person is asked to report immediately. Dreams are characterized by vivid hallucinatory imagery, usually in the visual domain, and illusions of self-motion. Although thought-like mentation can occur in other stages of sleep this does not have the vivid quality of dreams. Awakening during non-REM sleep yields reports of sensation and motion in only 5–10% of cases and these reports are of considerably reduced intensity. Dreaming resembles waking consciousness in that both involve mental representations. Dreaming differs from waking consciousness in that the dreamer has no insight (except at the moment of waking) and regards the dream events as completely real. Dreaming also differs from waking consciousness in that memory of the dream fades rapidly and is lost unless rehearsed immediately upon waking. Within five minutes of the termination of a period of REM sleep, awakening yields no report of dreaming and it is estimated that there is amnesia (in the sense of a lack of episodic memory) for over 95% of dreams. What we do remember of our dreams derives from those fragments dwelt upon immediately after waking (Ref. a).

Mammals show the same stages of sleep as are observed in humans. In particular they show REM sleep characterized by a 'waking' pattern of EEG, rapid eye movements and inhibited muscle tone. Observing the twitches and growls of a dog during REM sleep, one is strongly inclined to conclude that it is having dream experiences similar to our own. While we cannot obtain a report of an animal's dream experiences, it is possible to make lesions that remove the paralysis normally associated with REM sleep (Ref. b). Cats with such lesions get up during REM sleep and appear to the observer to be acting out dreams. These results strongly suggest that, during dreams, animals are having experiences the same as those humans have. Because, in intact animals, these experiences are not expressed in overt behaviour, it would seem reasonable to conclude these experiences are occurring in a mental domain. During dreams, at least, it is likely that animals form mental representations and have conscious experiences very similar to those of humans. The same is likely to be true in the waking state. The problem is that we have, as yet, no physiological marker analogous to REM that indicates that a mental representation has been formed.

The neural correlates of the different stages of sleep have been studied extensively in mammals. Acetylcholine levels are at a minimum during slow wave sleep (SWS) and maximal during REM. In contrast, noradrenaline and serotonin levels are high during SWS and near zero during REM. Associated with these neuromodulatory changes there are shifts in the direction of flow of neural information between the neocortex and the hippocampus. During REM sleep (and during active exploration in the waking state) information flows from neocortex to hippocampus. During SWS (and quiet rest) information flows in the opposite direction. These patterns of neural activity seem to be associated with different aspects of the consolidation and updating of long-term memories (Ref. c). It has been possible to acquire this extensive body of knowledge about the neural correlates of a form of consciousness (dreaming) because there is a very reliable physiological marker (REM) for this state in humans, which can also be observed in other mammals.

It has often been stated that when we dream we experience insanity. This is because we have no insight into the hallucinatory nature of our dreams and because we accept the improbabilities and impossibilities of dream sequences without question. However, it is the insanity associated with delirium rather than psychosis that dreams most resemble (Ref. d). While the hallucinations and delusions associated with psychosis represent a similar deviation from reality, the patient does not simply accept them, but reflects upon them and develops coherent and complex explanations to explain their occurrence. We would suggest that the form of consciousness associated with dreams is consciousness without reflection; that is, it is consciousness but not self-consciousness; experience but without reflection upon that experience. In consequence we have no insight and no episodic memory. During dreams do we discover what it is like to be a dog or a cat?

References

a Hobson, J.A. (1988) *The Dreaming Brain*, Basic Books

b Jouvet, M. (1979). What does a cat dream about? *Trends Neurosci.* 2, 15–16

c Stickgold, R. (1998) Sleep: off-line memory reprocessing *Trends Cognit. Sci.* 2, 484–492

d Hobson, J.A. (1997) Dreaming as delirium: a mental status analysis of our nightly madness *Semin. Neurol.* 17, 121–128

where lesions can lead to coma and vegetative states[5]. The level of consciousness can also be altered by exogenous substances, such as anaesthetics and psychoactive drugs[6].

A second aspect of consciousness concerns the content of subjective experience, that is, what one is conscious of. This is determined by the interaction between exogenous factors derived from our environment and endogenous factors, such as attention. The contents of consciousness include percepts of the objects around us, memories of past events and intentions concerning future actions. The contents of consciousness are associated with activity in specific cortical areas. Awareness of visual movement, for example, is associated with activity in visual area V5 (also called MT) at the parietal-occipital-temporal junction[7], while conscious recollection of events requires that the medial temporal lobe and its inputs are intact[8].

The relationship between level of arousal and contents of consciousness is complex and yet to be determined. Both very low and very high levels of arousal are generally associated with impoverished contents of consciousness[9]. Of particular interest is the dream state in which a low level of arousal (sleep) is associated with vivid sensory imagery[10]. As suggested by Llinas and Pare[11], insights into the nature of consciousness might be gained by analysis of the similarities

and differences between wakefulness and dreaming, and between these states and slow-wave sleep. Both dreaming and wakefulness are endowed with subjective experience, although dreaming is dissociated from awareness of external (and internal) sensory events in the rapid-eye-movement sleep phases during which dreams mostly occur. These two states are characterized by a similar complex pattern of activation in the thalamo–cortical system (see Box 1). Another potential route to link states of consciousness to brain function is based on a careful examination of the physiological effects of anaesthesia. For example, barbiturates greatly depress spontaneous neuronal activity; they also produce a decrease in excitability of brainstem reticular networks. Moreover, the EEG remains uniformly synchronous under barbiturate anaesthesia, in contrast with scalp potentials recorded during wakefulness[12]. Which of these physiological changes is critically associated with the altered state of arousal under general anaesthesia remains to be determined. In the rest of this paper, however, our discussion of the neural correlates of consciousness will largely be concerned with subjective awareness and the contents of consciousness. To aid this discussion we shall make frequent use of the term 'representation', which in this context refers to a mental entity that stands for something in the external world.

Mental representations

Cognitive neuroscientists persistently talk about neural representations. For example, a recent paper stated that 'the role of the prefrontal cortex in visual attention is to provide neural representations of to-be-attended information'[13]. We prefer to use more neutral terms, such as 'patterns of neural activity', for neurophysiological states. However, we suggest that it might be convenient to refer to the contents of consciousness (that is, phenomenal consciousness) as mental representations; mental entities that can stand for things in the outside world, and can usually be reported. When I remember something I have a mental representation of a past event. When I imagine something I have a mental representation of something that could occur in the outside world. When I perceive something I have a mental representation of something currently in the outside world. To say that we are conscious of something (or aware of something) is equivalent to saying that we have a mental representation of something. When we speak of the neural correlates of mental representations it is clear that we are referring to the contents of consciousness rather than to the level of consciousness.

Given that all mental activity derives from brain activity, it follows that all mental representations have corresponding neural activities. However, not all neural activities have corresponding mental representations. This is the crucial lesson taught us by phenomena such as blindsight[14]; behaviour can be guided by neural responses to visual stimuli in the absence of any awareness. In some cases, highly processed information can be used unconsciously, as evidenced in masked priming experiments[15]. A major part of the programme for studying the neural correlates of consciousness must be to investigate the difference between neural activities that are associated with awareness and those that are not.

The closely related lesson learnt from phenomena like blindsight is that goal-directed behaviour is not a reliable indicator of mental representation or subjective experience. To discover what someone is conscious of we need them to give us some form of report about their subjective experience. Such reports are qualitatively different from behaviour; reports, like consciousness, have content. They are about something. Behaviour simply occurs.

Verbal reports

How do we know about the contents of consciousness? The most direct way is by verbal report. I learn about your current mental representations from what you tell me about your perceptions, memories and intentions. You can describe the colour of an object, rate the intensity of a sensation, report whether one experience is the same or different from another. Such reports depend upon a shared communication system, such as language. Obviously there will be occasions where such reporting is not perfect (see Box 2), but, in general, this system works well. After all, outside the laboratory, such reporting is the basis of most of human communication. This communication system is constantly tuned to maintain the successful sharing of experience. I validate my understanding of your report by matching your description to my experience.

Behavioural reports

However, we do not need to use language to report our mental experiences. Gestures and movements can be made with a deliberate communicative intent. In many experiments the observer will indicate that he sees something by pressing a button. In this case a behavioural response (the button press) is a report in the sense that it means something: observer and experimenter have agreed, prior to the experiment, that a button press will mean that the stimulus has been perceived. More complex behavioural reports can be agreed. For example, in an experiment on binocular rivalry (see below) the observer can press the left button for one percept and the right button for the other. The advantage of a behavioural report is that it can be used in situations where a verbal report is not possible. A patient in the 'locked-in' state can indicate his or her wishes by minimal finger movements or eyeblinks[16]. The same procedure can be used in studies of animals. In a binocular rivalry experiment, a monkey can be trained to press the left button for one percept and the right button for the other[17]. The disadvantage of the behavioural report is that it can all too easily be confused with mere behaviour. Is that man who is violently waving his arm around swatting a wasp or hailing a taxi? Furthermore, appropriate behaviour can occur in the absence of any awareness of what is guiding the behaviour. This phenomenon is most obvious in neurological cases such as blindsight and the agnosias that follow lesions in the ventral visual pathways[18].

With human observers we know about the discrepancy between behaviour and awareness because verbal reports are available – the observer tells us that he was guessing – but how do we know that a monkey is not guessing? One solution is to use parameters that are well outside the guessing range in humans and to demonstrate that the performance of the monkey closely resembles that of the human in terms of response times and their distribution. In this way the

Box 2. Why should I believe what you tell me about your mental states?

When verbal reports are inadequate

There is a fundamental distinction between reports (whether verbal or behavioural) and behaviour. Reports are about something, are intended to convey meaning and can be true or false. Behaviour, on the other hand, simply occurs. Behaviour is not about anything. We might wrongly interpret its significance, but ideas of truth and falsity do not apply. This objectivity makes behaviour an attractive subject for scientific study, but such study tells us nothing about consciousness. To study consciousness we depend upon reports about mental states, but what if these reports cannot be relied upon? In psychophysical studies with normal volunteers we take it for granted that their reports are accurate. However, if we study patients with obvious brain damage, or psychosis or 'psychogenic' disorders, we have to be very wary about our reliance upon reports about mental states. If an abnormal mental state is reported does this mean: (a) the patient is experiencing an abnormal mental state?; (b) the patient is not able to report certain mental states correctly?; (c) the patient is trying to deceive us?

A large and controversial industry has grown up around the use of physiological and behavioural markers of lying. In one version of 'lie detecting' the subject might be shown a series of pictures and asked if he recognizes them. For the critical item, the subject denies having seen it, but shows a large increase in skin conductance and characteristic evoked potentials in the EEG. Clearly he has guilty knowledge about this item, but is trying to make us believe he knows nothing about it (Ref. a). The same procedure has been applied to patients with obvious brain damage. For example, a patient with prosopagnosia might be unable to recognize the faces of familiar people. If, however, we measure skin conductance when the faces are presented, then we will see an enhanced response to familiar, but unrecognized, faces (Ref. b). Similar effects can be observed in patients with amnesia (Ref. c). Nevertheless, we do not conclude that these patients are lying. Rather, we use these observations as evidence that, in these cases, as in blindsight, there are two neural pathways concerned with recognition; one conscious and the other unconscious. In these cases of organic damage we conclude that the conscious pathway has been damaged while the other remains intact. Why do we not conclude that these patients are lying? Because, firstly, they have obvious brain damage, and secondly, they do not gain anything by lying. There is nothing intrinsic to the physiological measures so far developed to indicate whether or not they are lying.

Our situation becomes even more uncertain when examining patients with 'functional' blindness or 'psychogenic' amnesia (Ref. d). These are patients in whom no obvious organic damage can be found. Sometimes they seem to obtain short-term gains from their symptoms. When asked to guess they might perform worse than chance. This could be the sign of a naive liar who avoids giving what he knows to be the correct answer. Yet these patients seem distinctly different from malingerers who are faking injury in order to obtain compensation and can be caught out, for instance, by using a 'damaged' limb when they think they are unobserved. If their symptoms are indeed faked, then hysterical patients seem abnormally devoted to keeping up the pretence. The long-term disadvantages resulting from the symptom more than outweighs the short-term gains. What is the neural concomitant of the dissociations present in hysteria in which it seems as if the patient does not know that he is 'lying'? If a specific physiological marker for the presence of mental representations can be found, then our understanding of disorders like hysteria will be greatly increased.

References

a Farwell, L.A. and Donchin, E. (1991) The truth will out: interrogative polygraphy ('lie detection') with event-related brain potentials *Psychophysiology* 28, 531–547

b Bauer, R.M. (1994) Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the guilty knowledge test *Neuropsychologia* 22, 457–469

c Diamond, B.J., Mayes, A.R. and Meudell, P.R. (1996) Autonomic and recognition indices of memory in amnesic and healthy control subjects *Cortex* 32, 439–459

d Fukuda, M. *et al.* (1996) Event-related potential correlates of functional hearing loss: reduced P3 amplitude with preserved N1 and N2 components in a unilateral case *Psychiatry Clin. Neurosci.* 50, 85–87

experimenter uses the verbal report of the human in an attempt to validate the behavioural report of the monkey[17]. Another solution is to develop tasks in which the monkey learns to report the presence of a stimulus rather than making a forced choice about the nature of a stimulus[19].

Identifying the neural correlates of mental representations

In our search for the neural correlates of consciousness we aim to identify that kind of neural activity which has a formal relationship with reports about mental representations. There are two parts to this aim. The first part is to identify neural activity that is related to mental representations. We are searching for situations in which changes in neural activity predict changes in mental representation and the converse. We can predict either the direction of the change or the time at which the change occurs. For example, in a study of threshold determination, the presence (or absence) of a certain pattern of neural activity would predict the presence (or absence) of a particular mental representation. The time at which the change in neural activity occurred would also predict the time of the change of mental representation. It should be noted, however, that for this prediction it is not necessary for the changes in the two domains to occur precisely at the same time. It is merely necessary that an early change in the neural activity is associated with an early change in the mental representation. Prediction is still possible if there is systematic time difference between the changes in the two domains.

It is also necessary to show that these relationships are not simply the concomitant consequence of changes in stimulation or changes in behaviour. The most direct way of 'partialling out' these unwanted effects is to keep them constant. Hence the emphasis on paradigms in which, for example, subjective experience changes while stimulation remains constant (e.g. binocular rivalry).

The second part of our aim is to show that neural activity associated with mental representations is qualitatively

Box 3. Neural correlates and signatures of consciousness

A systematic investigation of the brain bases of consciousness is predicated on the assumption that at any moment subjective experience is associated with a corresponding pattern of neuronal activity. This basic tenet underlies current research programs on consciousness, in which the first objective is to characterize the neural correlates of consciousness (NCC) (Ref. a). Key questions that must be addressed empirically include: in which areas are such correlates present (or absent)? Are the neurons involved of a particular type? Do they fire in a particular way, either by themselves or with respect to other neurons?

An issue of current focus concerns the localization of brain activity correlated with phenomenal awareness. For example, Schacter and co-workers have proposed that there is a specific brain system, separate from the structures involved in perception and action, that mediates conscious awareness (Ref. b). Conversely, it has been suggested that certain brain regions and pathways do not contribute directly to conscious experience. Thus, Crick and Koch have suggested that activity in the primary visual cortex (V1) is not directly linked to visual awareness (Ref. c). In a variant of this 'eliminative' view, Milner and Goodale suggest that the occipito-parietal pathway subserves an on-line system for visually guided actions that operates in the absence of awareness (Ref. d). One difficulty in testing these hypotheses is that in any given area, the NCCs could in principle take many forms. Possibilities that have been considered include a modulation of firing rates, neuronal synchronization, and activity of cells in specific layers or with particular spike trains (e.g. bursting cells). When assessing whether activity in a given area reflects conscious experience, caution must therefore be taken in drawing conclusions from restricted measures of neuronal activity, such as those pertaining to mean firing rates. For instance, neuronal synchronization in V1 has been shown to reflect the functional dominance of one eye in amblyopia in cats, whereas no dominance-related modulation of firing rate was observed in this area (Ref. e); such synchronization might contribute to the formation of perceptual states.

In contrast with accounts of consciousness based on particular brain systems, integrative theories of brain function propose that subjective experience emerges as the result of interactions among widely distributed brain regions that mediate perception, attention, memory and action (Ref. f). Accordingly, the NCCs should be expressed in terms of patterns of correlation among separate and distant brain areas.

In exploring these possibilities, it will be important to guard against spurious correlates of conscious experience – neuronal activity that, although correlated with a particular subjective state, does not account for its emergence. To discover bona fide neural correlates of conscious experience, brain activity must be measured under various experimental conditions and with a range of measurement techniques, in the search for patterns of activity that consistently correlate with conscious experience across multiple conditions. We refer to these patterns as neural signatures of consciousness (NSC). The identification of such robust neural signatures will be essential for the formulation of hypotheses regarding the neural mechanisms that give rise to consciousness, and for the design of experiments aimed at interfering with such mechanisms.

Measuring neural activity

Direct measures of neural activity can be made at the level of the single cell, but this work is largely restricted to animals. Changes in electrical activity occurring at the scalp (EEG and MEG) can be recorded in humans. These measures reflect post-synaptic potentials in large populations of cells (Ref. g). Indirect measures of neural activity at the level of large populations of cells can be made in humans by measuring blood flow (PET) or blood oxygenation level (fMRI). These indirect measures largely reflect pre-synaptic activity (Ref. h). Functional MRI and PET are good measures for localizing activity, while EEG and MEG are good for revealing the precise timing of activity. All these techniques can also be used to examine the interactions between distant brain areas using measures of effective connectivity (Ref. i) and structural equation modelling (Ref. j).

Our emphasis on the need for verbal report to identify the contents of consciousness implies that the search for the neural correlates of consciousness will largely be restricted to humans. At present, studies of neural activity in the human brain are restricted to techniques that measure activity in large populations of cells. However, if the neural signature of consciousness relates to activity in certain cortical layers or to the precise timing of spike trains, then consideration of activity at the level of the single cell will be critical. It will therefore be very important to show how activity in large populations of neurons relates to activity at the single-cell level. This is likely to be achieved through computational modelling (Ref. k) and also through application of fMRI to the monkey brain.

References

a Crick, F. and Koch, C. (1998) Consciousness and neuroscience *Cereb. Cortex* 8, 97–107

b Schacter, D.L., McAndrews, M.P. and Moscovitch, M. (1988) Access to consciousness: dissociations between implicit and explicit knowledge in neuropsychological syndromes, in *Thought Without Language* (Weiskrantz, L, ed.) pp. 242–278, Oxford University Press

c Crick, F. and Koch, C. (1995) Are we aware of neural activity in primary visual cortex? *Nature* 373, 121–123

d Milner, A.D. and Goodale, M.A. (1995) *The Visual Brain in Action*, Oxford University Press

e Roelfsema, P.R. *et al.* (1994) Reduced synchronization in the visual cortex of cats with strabismic amplyopia *Eur. J. Neurosci.* 6, 1645–1655

f Kinsbourne, M. (1988) Integrated field theory of consciousness, in *Consciousness in Contemporary Science* (Marcel, A.J. and Bisiach, E., eds), pp. 239–256, Oxford University Press

g Lopes da Silva, F. (1991) Neuronal mechanisms underlying brainwaves: from neural membranes to networks *Electroencephalogr. Clin. Neurophysiol.* 90, 81–93

h Jueptner, M. and Weiller, C. (1995) Review: does measurement of regional cerebral blood flow reflect synaptic activity? – implications for PET and fMRI *NeuroImage* 2, 148–156

i Gerstein, G.L., Bedenbaugh, P. and Aertsen, A. (1989) Neuronal assemblies *IEEE Trans. Biomed. Eng.* 36, 4–14

j McIntosh, A.R. and Gonzalez-Lima, F. (1991) Structural equation modelling of functional neural pathways mapped with 2-deoxyglucose: effects of acoustic startle habituation on the auditory system *Brain Res.* 546, 295–302

k Lumer, E.D., Edelman, G.M. and Tononi, G. (1997) Neural dynamics in a model of the thalamocortical system: II. The role of neural synchrony tested through perturbations of spike timing *Cereb. Cortex* 7, 228–236

different from other kinds of neural activity. To achieve this aim it is necessary to identify neural correlates of changing stimulation and changing behaviour that occur in the absence of changes in mental representations. Comparison of these different kinds of neural activity will demonstrate that there is a certain pattern or type of neural activity that is uniquely associated with mental representations: the neural 'signature' of consciousness (see Box 3).

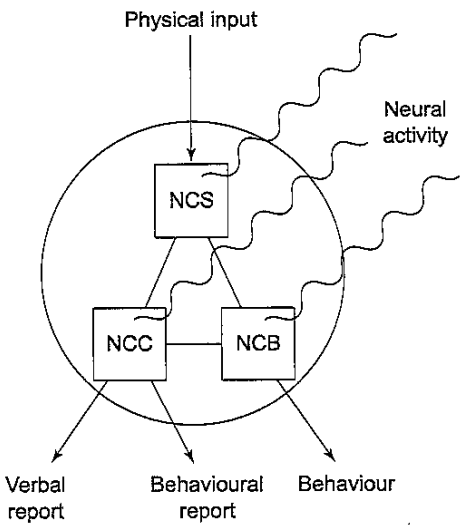

Fig. 1. Schematic illustration of the three neural correlates that must be contrasted in order to specify the neural correlates of consciousness. NCS: neural correlates of sensory stimulation. NCB: neural correlates of behaviour. NCC: neural correlates of consciousness. The three different classes of activity need not be spatially segregated in discrete brain regions.

We suggest that it will be useful to distinguish three kinds of neural activity: (1) neural activity associated with mental representations; (2) neural activity associated with changes in a sensory stimulus, in the absence of changes in mental representations; and (3) neural activity associated with behaviour in the absence of mental representations (see Fig. 1). Each of these three kinds of activity can be identified using specific experimental paradigms.

A taxonomy of experimental paradigms for studying consciousness

In Tables 1 and 2 we have made an attempt to systematize the experimental paradigms relevant to the three hypothetical types of neural activity. We have associated these with three kinds of psychological processes; those concerned with the present (perception and imagery), those concerned with the past (memory) and those concerned with the future (intentions and actions). These tables provide a basis of a programme for the development of experimental studies of the neural correlates of consciousness. Because studies of neurological cases and other disorders associated with abnormal mental states continue to play such a major role in the identification of the neural correlates of consciousness, we have included a table specifically for such studies (Table 2). We discuss some of the paradigms listed in these tables below.

Perception and attention

Many authors have suggested that the study of visual awareness is likely to provide a good starting point for investigating the neural correlates of consciousness. This notion is based on the wealth of data available on the neuroanatomy and physiology of the visual system in monkeys. Under the often implicit assumption of strong homology between the visual systems of human and monkeys, this information can be brought to bear in the design and interpretation of experimental studies. Whether we can study the neural correlates of consciousness in the monkey directly is less clear. Much important work has been reported in which the verbal reports of a human observer were correlated with activity in the brain of an anaesthetised monkey exposed to the same stimulus[20]. This approach might well identify activity that is necessary for consciousness, but it seems unlikely to identify activity that is sufficient. However, even studies of awake, behaving monkeys are problematic. We believe that monkeys do have mental representations, but it is very difficult to prove this supposition. We can train monkeys to report their perceptions behaviourally, but we cannot validate these behavioural reports with verbal ones. This situation would change dramatically if reliable neural signatures of mental representations could be found in humans (see Box 3). The observation of the same patterns of neural activity in monkeys would provide a validation of behavioural reports.

Meanwhile in the last decade, there has been a surge of experimentation on the neural basis of conscious vision in humans. Two experimental paradigms have emerged in which visually guided performance takes place in the absence of conscious perception of the stimulus.

Table 1. Experimental paradigms for studying the neural correlates of consciousness in normal states

	Perception	Memory	Action
Subjective experience changes, stimulation and/or behaviour remains constant	Neural correlates of binocular rivalry[23]	Neural correlates of episodic recall[43]	Neural correlates of the awareness of intention[30]
Stimulation changes, subjective experience remains constant	Neural correlates of changes in stimulation without awareness[41]	Neural correlates of unrecognised old items[43]	Neural correlates of stimuli eliciting action without awareness[15]
Behaviour changes, subjective experience remains constant	Neural correlates of correct guessing without awareness[42]	Neural correlates of implicit learning[44]	Neural correlates of implicit motor behaviour[45]

In each category one example is given of an experimental paradigm that has been or could be used to identify neural correlates of consciousness, sensory stimulation or behaviour.

Table 2. Experimental paradigms for studying the neural correlates of consciousness in abnormal states

	Perception	Memory	Action
Subjective experience changes, stimulation and/or behaviour remains constant	Neural correlates of hallucinations[24]	Neural correlates of confabulation[40]	Neural correlates of abnormal intentions (delusions of control)[49]
Stimulation changes, subjective experience remains constant	Neural correlates of stimulation of the blind field in blindsight[14]	Neural correlates of unrecognised items in amnesia[47]	Neural correlates of stimuli eliciting unintended actions[50]
Behaviour changes, subjective experience remains constant	Neural correlates of correct reaching in form agnosia[46]	Implicit learning in amnesia[48]	Neural correlates of unintended actions[33]

In each category one example is given of an experimental paradigm that has been or could be used to identify neural correlates of consciousness, sensory stimulation or behaviour.

Instances of the first category are given by the phenomenology of patients with lesions in the occipito-temporal pathway. Injury to the earliest stage of this pathway, in the striate cortex (V1), can lead to blindsight, in which subjects report being unaware of stimuli in the contralateral visual field but produce goal-directed behaviour that is contingent on these sensory cues[14]. Another example of a dissociation between conscious perception and behaviour is provided by a patient, D.F., who has form agnosia and is unable to report the orientation of a mail slot but can nevertheless post a letter in it. These observations have led Milner and Goodale to propose the existence of a rapid response system that is mainly unconscious, and depends on processing in the dorsal pathway, and of an object recognition system that is conscious and results from ventral stream processing[18]. This formulation implies what is perhaps the simplest possible neural signature for consciousness: activity in some brain regions is associated with mental representations, while activity in other regions is not. However, we think this is unlikely to be a general principle for the specification of the neural signature of consciousness (see Box 3).

In normal subjects, appropriate behaviour in the absence of conscious perception can be observed in masking experiments. The introduction of a mask a few tens to hundreds of milliseconds after presentation of a stimulus (such as a face) can abolish the subjective perception of the stimulus; yet, autonomic and motor responses to aspects of the stimulus may still be elicited[21]. In this paradigm the aim is to show that the presence of a certain kind of neural activity predicts autonomic and motor responses. Using masking it has been shown, for example, that activity in the amygdala changes with changes in facial expression of which the subject is unaware[22] (see Fig. 2). By studying effects close to the masking threshold it should be possible to identify a different kind of activity that predicts conscious perception of the stimulus. In particular, it should be possible to identify some pattern of neural activity that predicts whether or not the masked object will be consciously perceived.

One of the best available experimental paradigms with which to study the neural correlates of subjective visual perception is provided by the bistable perception that arises when the physical stimulus readily allows two alternative perceptual interpretations. Perceptual instability arises, for example, when dissimilar images are shown to each eye, the instability in this case is called binocular rivalry. Because the changes in perception occur despite a constant visual input, neural responses associated with perception can be distinguished from those entirely dependent on stimulus characteristics. In this example the aim is to show that the timing of the change in perception correlates with the timing of certain changes in neural activity. Recent studies with awake monkeys trained to report their percepts during rivalry have revealed that, whereas the firing of most neurons in the primary visual cortex correlates with the stimulus and not the percept, activity of neurons at higher stages of the visual pathway, such as the inferotemporal cortex, mainly reflect the percept[17]. This result has now been confirmed in human volunteers using functional MRI (Ref. 23). However, because activity throughout the brain could be monitored using this technique it was additionally shown that changes in parietal and frontal activity were also correlated with changes in perception.

In patients with blindsight, changes in stimulation or behaviour can occur in the absence of changes in perception (as discussed above). In contrast, there are various hallucinatory experiences in which perception occurs in the absence of sensation. For example, patients with schizophrenia frequently hear voices speaking to them or about them. If such patients can be persuaded to indicate when these experiences occur then the associated neural activity can be identified. One such study suggests that activity associated with hallucinations occurs in sensory association areas, but not in primary sensory areas[24]. Similar investigations could be applied to patients with phantom limbs[25].

The studies of perception reviewed above demonstrate that stimuli can be highly processed and yet not enter awareness. Attention might be the critical mechanism by which preprocessed stimuli are selected for awareness. In consequence, studies of selective attention will continue to be of major interest in the investigation of the neural correlates of consciousness. In particular, we need to find out whether there is a characteristic difference in the neural activity associated with attended and unattended stimuli. However, the precise relationship between attention and

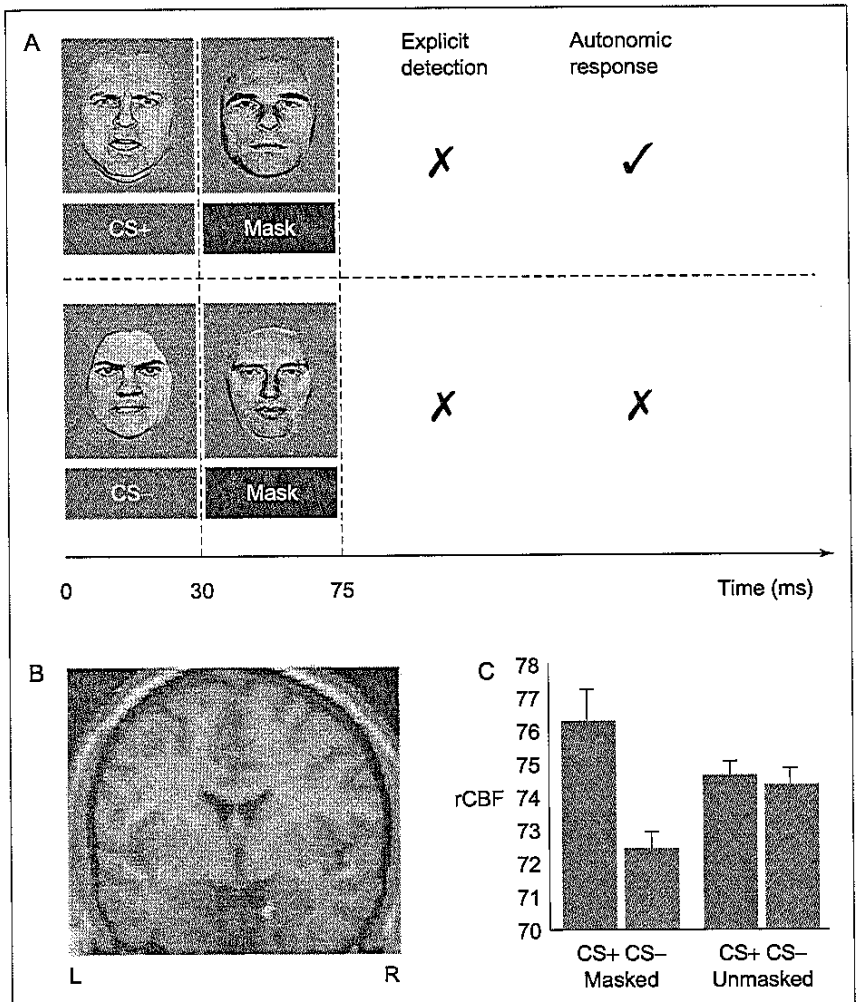

Fig. 2. Actvity in the amygdala elicited by aversive stimuli of which the subject is not aware (A) Right amygdala response to masked conditioned faces. Prior to PET scanning, healthy subjects were repeatedly shown an angry face (the CS+) paired with a 100dB noise. Another angry face (the CS–) was also shown repeatedly but never paired with noise. During scanning, the CS+ and CS– faces were presented for 30 ms, and immediately followed by a 45 ms masking stimulus (a neutral face). This masking procedure prevented subjects from explicitly reporting any awareness of the CS+ or CS– faces. Despite being 'unseen', masked CS+ presentations elicited greater autonomic responses (changes in skin conductance) than masked presentations of the CS– face. A mediobasal region of the right amygdaloid complex responded more to the masked CS+ than to the masked CS– face, shown on a coronal MRI section **(B)**. The mean adjusted regional cerebral blood flow (rCBF) for the different experimental conditions is also shown **(C)**. Although the right amygdala responded differentially to the masked conditioned faces, this differential activity was not seen when the same stimuli were shown unmasked and explicitly reported via a button press (unmasked). Modified from Ref. 22.

awareness remains unclear. It is not necessarily the case that unattended stimuli do not reach awareness. The extent to which unattended stimuli are processed depends on the perceptual 'load' imposed by the task requiring attention[26]. Study of the neural activity associated with unattended stimuli has advantages over behavioural studies because the degree to which the unattended stimuli are processed can be studied without alerting subjects to their presence[27]. Lack of any detectable physiological response to unattended stimuli would be good evidence that they did not reach awareness.

Disorders of attention can result from lesions to the inferior parietal and frontal cortex. Such lesions can lead to unilateral neglect, in which the patient is unable to attend to objects in the visual field contralateral to the lesion. Neglect differs from occipitally related blindsight in that extensive visual processing of neglected stimuli is preserved[28]. One problem in interpreting the pathological basis of this disorder, though, is that the extent and nature of the lesions (e.g. grey- versus white-matter damage) are often poorly specified. Thus, it is not clear whether right inferior parietal cortex, where lesions associated with unilateral neglect are often located, has a special role in the direction of attention, or whether damage to underlying white matter is critical owing to the disconnection of visual and motor regions[29].

Action and intention

The relationship of neural activity in the motor system to the awareness of intentions has also been studied extensively. Libet[30] has shown that brain activity that precedes a voluntary movement can be detected (in the EEG) well in advance of reports of awareness of the intention to move. This pioneering study illustrates the advantages of studying the motor system, as correlations can be explored between the times at which mental and physical events occur. Haggard and Eimer[31] present data suggesting that the time of awareness of initiating a movement correlates with the time at which the late, lateralized component of the readiness potential begins (probably reflecting the time at which the exact movement is specified), but not with the beginning of the early phase of the readiness potential. Studies that combine the high spatial resolution of fMRI and the high temporal resolution of EEG should be able to locate the brain regions associated with these various potentials.

A series of experiments from Jeannerod's group has demonstrated dissociations between behaviour and awareness in normal volunteers making rapid grasping movements. Corrections to the trajectory of a movement made in response to target movements occur several hundred milliseconds in advance of reported awareness of target movement[32] (see Fig. 3). In this respect, the distinction we have made in Tables 1 and 2 between perception and action is somewhat artificial. As Milner and Goodale have pointed out, perception and action are intimately linked[18]. In most cases the purpose of perception is to permit action. Studies of patients with blindsight or form agnosia should be particularly informative in revealing the neural activity associated with reaching and grasping in the absence of awareness of the object of the action. It is likely that the same system is involved as operates in the rapid, 'unconscious' error correction observed in normal volunteers.

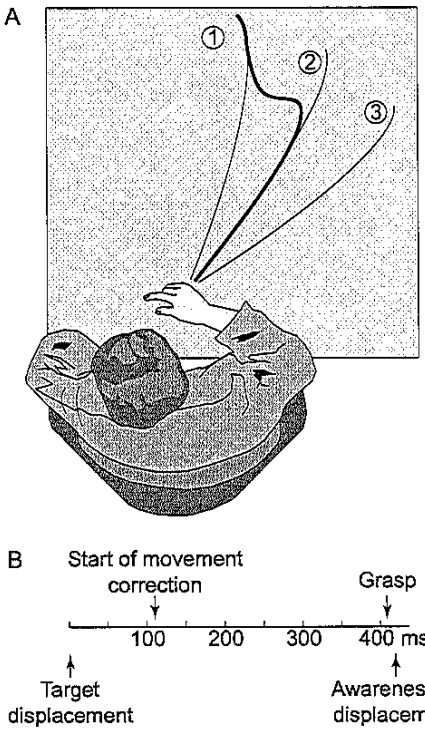

Fig. 3. Dissociation of behaviour and awareness in the actions of normal subjects. (A) When one of three rods (1, 2 and 3) was illuminated the subject moved his right hand and grasped the target rod with finger and thumb as quickly as possible. The paths followed by the subject's forefinger are indicated by the thin lines. On 20% of trials, as soon as the subject started to move, the target indicator light switched from the centre rod to one of the outer rods (perturbed trials). This switch elicited a smooth and rapid movement correction shown by the bold line. In addition to grasping the target rod, subjects gave a vocal response (saying 'tah') as soon as they perceived that the target was illuminated. On the perturbed trials they gave a second vocal response to indicate when they were aware of the target switch. **(B)** The mean times at which the various events occurred for the perturbed trials. The first signs of movement correction occurred approximately 100 ms after the target switch. However, the vocal response indicating awareness of the target switch did not occur until 300 ms later. The authors of the original paper reported that: 'on some occasions, the dissociation between motor and vocal responses was such that subjects, to their own surprise, were already lifting the target rod when they emitted the vocal response'. (Modified from data reported in Ref. 32, Figs 2 and 3 and Table 2.)

Other examples of motor behaviour in the absence of awareness are provided by patients with an 'anarchic' hand or with utilization behaviour in whom goal-directed behaviour can be observed in the absence of awareness of an intention to act[33].

In contrast, mental representation of action in the absence of behaviour is present during motor imagery, or in patients who are able to 'move' a phantom limb. During motor imagery neural activity is observed in most of the motor system excluding primary motor cortex[34].

Memory and learning

Studies of memory have close parallels with studies of perception and action. Patients with amnesia, like those with blindsight, can show behaviour influenced by past experience of which they are not conscious[35]. Normal volunteers can also show such behaviour via priming or implicit learning[36]. Studies of the neural correlates of these phenomena are beginning to appear[37]. Mental representations of past events are involved in episodic memory and explicit learning. The neural activity associated with such mental representations can be identified by contrast with that associated with the various forms of implicit memory[38]. It is also possible to study neural activity associated with false memories (analogous to hallucinations) that can be elicited in normal people in certain experimental paradigms[39] and can be observed in patients who confabulate[40].

Conclusions

It is clear from this brief summary that there are already a number of studies in the literature that address questions concerning the neural correlates of conscious experience. The purpose of this article has been to develop a framework in which such studies can be linked to one another, and to make explicit the fundamental significance of such studies for our understanding of the neural correlates of conscious experience. As the amount of evidence on this topic increases we shall be in a much better position to identify the characteristic neural signatures of consciousness (if there are any) and to address more difficult questions concerning the function of consciousness and the mechanisms by which mental representations emerge from neural activity.

Acknowledgements

This article is written on behalf of, and based on a series of discussions in, the Consciousness Club, an informal group of scientists and clinicians from the Wellcome Department of Cognitive Neurology and the Institute of Cognitive Neuroscience, University College London. Its members include Sarah-Jayne Blakemore, Raymond Dolan, Paul Fletcher, Richard Frackowiak, Karl Friston, Chris Frith, Andreas Kleinschmidt, Erik Lumer, Emiliano Macaluso, Ludovica Marini, Dave McGonigle, John Morris, Richard Perry, Chiara Portas, Geraint Rees, Mick Rugg and Bob Turner. We are grateful to Mike Martin for his comments on philosophical aspects. Preparation of this article was supported by the Wellcome Trust.

References

1 Crick, F. (1994) *The Astonishing Hypothesis*, Macmillan
2 Edelman, G.M. (1989) *The Remembered Present*, Basic Books
3 Dennett, D. (1991) *Consciousness Explained*, Penguin Books
4 Churchland, P.S. (1993) *Neurophilosophy: Towards a Unified Science of the Mind–Brain*, MIT Press

Outstanding questions

- Does the pattern of neural activity associated with conscious perception have some common feature across different sensory modalities?
- Can measurement of neural activity be used to infer the existence of mental representations in the absence of a verbal (or a non-verbal) report?
- Does the apparent unity of conscious experience require the convergence of disparate streams of neural activity into a few 'convergence zones' within the brain, or might the neural correlates of consciousness be found in multiple, independent brain areas?
- Is there evidence, at the neural level, for an intimate relationship between phenomenal consciousness and language?
- What is the minimum set of brain regions that is sufficient for the existence of mental representations?

5 Purpura, K.P. and Schiff, N.D. (1997) The thalamic intralaminar nuclei: a role in visual awareness *The Neuroscientist* 3, 314–321
6 Kihlstrom, J.F. *et al.* (1990) Implicit and explicit memory following surgical anaesthesia *Psychol. Sci.* 1, 303–306
7 Barbur, J.L. *et al.* (1993) Conscious visual perception without V1. *Brain* 116, 1293–1302
8 Schacter, D.L. (1989) On the relation between memory and consciousness: dissociable interactions and conscious experience, in *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving* (Roediger, H.L. and Craik, F.I.M., eds), pp. 355–389, Erlbaum
9 Easterbrook, J.A. (1959) The effect of emotion on cue utilisation and the organisation of behaviour *Psychol. Rev.* 66, 183–201
10 Hobson, J.A. (1988) *The Dreaming Brain*, Basic Books
11 Llinas, R.R. and Pare, D. (1991) On dreaming and wakefulness *Neuroscience* 44, 521–535
12 Contreras, D. *et al.* (1997) Spatiotemporal patterns of spindle oscillations in cortex and thalamus *J. Neurosci.* 17, 1179–1196
13 Miller, E.K. and Desimone, R. The prefrontal and visual cortices: the neural basis of top-down control of visual attention, in *Attention and Performance* (Vol. XVIII) (Monsell, S. and Driver, J., eds), MIT Press (in press)
14 Weiskrantz, L. (1986) *Blindsight: A Case Study and Implications*, Oxford University Press
15 Dehaene, S. *et al.* (1998) Imaging unconscious semantic priming *Nature* 395, 597–600
16 Gallo, U.E. and Fontanarosa, P.B. (1989) Locked-in syndrome: report of a case *Am. J. Emerg. Med.* 6, 581–583
17 Leopold, D. and Logothetis, N. (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry *Nature* 379, 549–553
18 Milner, A.D. (1998) Streams and consciousness: visual awareness and the brain *Trends Cognit. Sci.* 2, 25–30
19 Cowey, A. and Stoerig, P. (1995) Blindsight in monkeys *Nature* 373, 247–249
20 Zeki, S. (1993) The physiology of the colour pathways, in *A Vision of the Brain*, pp. 256–263, Blackwell Science
21 Ohman, A. and Soares, J.J.F. (1994) 'Unconscious anxiety': phobic responses to masked stimuli *J. Abnorm. Psychol.* 103, 231–240
22 Morris, J., Ohman, A. and Dolan, R.J. (1998) Conscious and unconscious emotional learning in the human amygdala *Nature* 393, 467–470
23 Lumer, E.D., Friston, K.J. and Rees, G. (1998) Neural correlates of perceptual rivalry in the human brain *Science* 280, 1930–1934
24 Silbersweig, D.A. *et al.* (1995) A functional neuroanatomy of hallucinations in schizophrenia *Nature* 378, 176–179
25 Hari, R. *et al.* (1998) Three hands: fragmentation of bodily awareness *Neurosci. Lett.* 240, 13113–13114
26 Lavie, N. (1995) Perceptual load as a necessary condition for selective attention *J. Exp. Psychol. Hum. Percept. Perform.* 21, 451–468
27 Rees, G., Frith, C.D. and Lavie, N. (1997) Modulating irrelevant motion perception by varying attentional load in an unrelated task *Science* 278, 1616–1619
28 Driver, J. and Mattingley, J.B. (1998) Parietal neglect and visual awareness *Nat. Neurosci.* 1, 17–22
29 Gaffan, D. and Hornak, J. (1997) Visual neglect in the monkey: representation and disconnection *Brain* 120, 1647–1657
30 Libet, B. (1993) The neural time factor in conscious and unconscious events, in *Experimental and Theoretical Studies of Consciousness* (Ciba Foundation Symposium 174), pp. 123–137, John Wiley & Sons
31 Haggard, P. and Eimer, M. Brain events related to conscious representations of action *Exp. Brain Res.* (in press)
32 Castiello, U., Paulignan, Y. and Jeannerod, M. (1991) Temporal dissociation of motor responses and subjective awareness *Brain* 114, 2639–2655
33 Marchetti, C. and Della Salla, S. (1998) Disentangling the alien and anarchic hand *Cognit. Neuropsychiatry* 3, 191–208
34 Stephan, K.M. *et al.* (1995) Functional anatomy of the mental representation of upper extremity movements in healthy subjects *J. Neurophysiol.* 73, 373–386
35 Schacter, D.L. (1992) Consciousness and awareness in memory and amnesia: critical issues, in *The Neuropsychology of Consciousness* (Milner, A.D. and Rugg, M.D., eds), pp. 180–200, Academic Press
36 Kunst-Wilson, W.R. and Zajonc, R.B. (1980) Affective discrimination of stimuli that cannot be recognised *Science* 207, 557–558
37 Buckner, R.L. and Koutstaal, W. (1998) Functional neuroimaging studies of encoding, priming, and explicit memory retrieval *Proc. Natl. Acad. Sci. U. S. A.* 95, 891–898
38 Clark, R.E. and Squire, L.R. (1998) Classical conditioning and brain systems: the role of awareness *Science* 280, 77–81
39 Schacter, D.L. *et al.* (1996) Neuroanatomical correlates of veridical and illusory recognition memory: evidence from positron emission tomography *Neuron* 17, 267–274
40 Burgess, P.W. and Shallice, T. (1996) Confabulation and the control of recollection *Memory* 4, 1–53
41 He, S., Cavanagh, P. and Intriligator, J. (1996) Attentional resolution and the locus of visual awareness *Nature* 383, 334–337
42 Pierce, C.S. and Jastrow, J. (1885) On small differences in sensation *Memoirs of the National Academy of Sciences* 3, 75–83
43 Rugg, M.D. *et al.* (1998) Dissociation of the neural correlates of implicit and explicit memory *Nature* 392, 595–598
44 Hazeltine, E., Grafton, S.T. and Ivry, R. (1997) Attention and stimulus characteristics determine the locus of motor-sequence encoding: a PET study *Brain* 120, 123–140
45 Rossetti, Y. Implicit short-lived motor representations of space in brain damaged and healthy subjects *Conscious. Cognit.* (in press)
46 Milner, A.D. and Goodale, M.A. (1995) *The Visual Brain in Action*, Oxford University Press
47 Diamond, B.J., Mayes, A.R. and Meudell, P.R. (1996) Autonomic and recognition indices of memory in amnesic and healthy control subjects *Cortex* 32, 439–459
48 Knowlton, B.J., Ramus, S.J. and Squire, L.R. (1992) Intact artificial-grammar learning in amnesia: dissociation of abstract knowledge and memory for specific instances *Psychol. Sci.* 3, 172–179
49 Spence, S.A. *et al.* (1997) A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control) *Brain* 120, 1997–2011
50 Lhermitte, F. (1983) 'Utilization behaviour' and its relation to lesions of the frontal lobes *Brain* 106, 237–255

Coming soon to *Trends in Cognitive Sciences*

- What can neuropsychology tell us about category learning?, by B. Knowlton
- Mathematical modeling can help: Repsonse from R.M. Nosofski and S.R. Zaki
- Human oscillatory gamma activity and object representaion, by C. Tallon-Baudry and O. Bertrand
- Catastrophic forgetting in connectionist networks, by R.M. French
- Syntactic priming in language production, by M.J. Pickering and H.P. Branigan
- Perceptual-mnemonic functions of the perirhinal cortex, by E.A. Murray and T.J. Bussey