Available online at www.sciencedirect.com

ScienceDirect

Neural Networks 20 (2007) 962–982

Neural Networks

www.elsevier.com/locate/neunet

2007 Special Issue

A computational neuroscience approach to consciousness

Edmund T. Rolls*

University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, England, United Kingdom

Abstract

Simultaneous recordings from populations of neurons in the inferior temporal visual cortex show that most of the information about which stimulus was shown is available in the number of spikes (or firing rate) of each neuron, and not from stimulus-dependent synchrony, so that it is unlikely that stimulus-dependent synchrony (or indeed oscillations) is an essential aspect of visual object perception. Neurophysiological investigations of backward masking show that the threshold for conscious visual perception may be set to be higher than the level at which small but significant information is present in neuronal firing and which allows humans to guess which stimulus was shown without conscious awareness. The adaptive value of this may be that the systems in the brain that implement the type of information processing involved in conscious thoughts are not interrupted by small signals that could be noise in sensory pathways. I then consider what computational processes are closely related to conscious processing, and describe a higher order syntactic thought (HOST) computational theory of consciousness. It is argued that the adaptive value of higher order thoughts is to solve the credit assignment problem that arises if a multistep syntactic plan needs to be corrected. It is then suggested that it feels like something to be an organism that can think about its own linguistic, and semantically-based thoughts. It is suggested that qualia, raw sensory and emotional feels, arise secondarily to having evolved such a higher order thought system, and that sensory and emotional processing feels like something because it would be unparsimonious for it to enter the planning, higher order thought, system and *not* feel like something.

Keywords: Consciousness; Higher order thought; Synchrony; Oscillations; Backward masking; Binding

1. Introduction

Neural network theories of visual object recognition in the ventral visual stream are being developed that are consistent with much of the related neurophysiology (Rolls, 2008; Rolls & Deco, 2002; Rolls & Stringer, 2006). Neural network theories of attention are also being developed (Deco & Rolls, 2005a, 2005b; Rolls, 2008; Rolls & Deco, 2002). In this paper I consider whether some of the computational processing involved in these perceptual and attentional computations is closely linked to consciousness. Then I consider what computational processing may be closely related to consciousness. The architecture of some of the processing regions discussed is shown in Fig. 1.

2. Oscillations and stimulus-dependent synchrony: Their role in information processing in the ventral visual system, and in consciousness

It has been suggested that syntax in real neuronal networks is implemented by temporal binding (see Malsburg, 1990), which would be evident as for example stimulus-dependent synchrony (Singer, 1999). According to this hypothesis, the binding between features common to an object could be implemented by neurons coding for that object firing in synchrony, whereas if the features belong to different objects, the neurons would not fire in synchrony. Crick and Koch (1990) postulated that oscillations and synchronization are necessary bases of consciousness. It is difficult to see what useful purpose oscillations per se could perform for neural information processing, apart from perhaps resetting a population of neurons to low activity so that they can restart some attractor process (see e.g. Rolls & Treves, 1998), or acting as a reference phase to allow neurons to provide some additional information by virtue of the time that they fire with respect to the reference waveform (Huxter, Burgess, & O'Keefe, 2003). Neither putative function seems to be closely related to consciousness. However, stimulus-dependent synchrony, by implementing binding, a function that has been related to attention (Treisman, 1996), might perhaps be related to consciousness. Let us consider the evidence on whether stimulus-dependent synchrony between neurons

* Tel.: +44 1865 271348; fax: +44 1865 310447.
E-mail address: Edmund.Rolls@psy.ox.ac.uk.
URL: http://www.cns.ox.ac.uk.

0893-6080/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2007.10.001

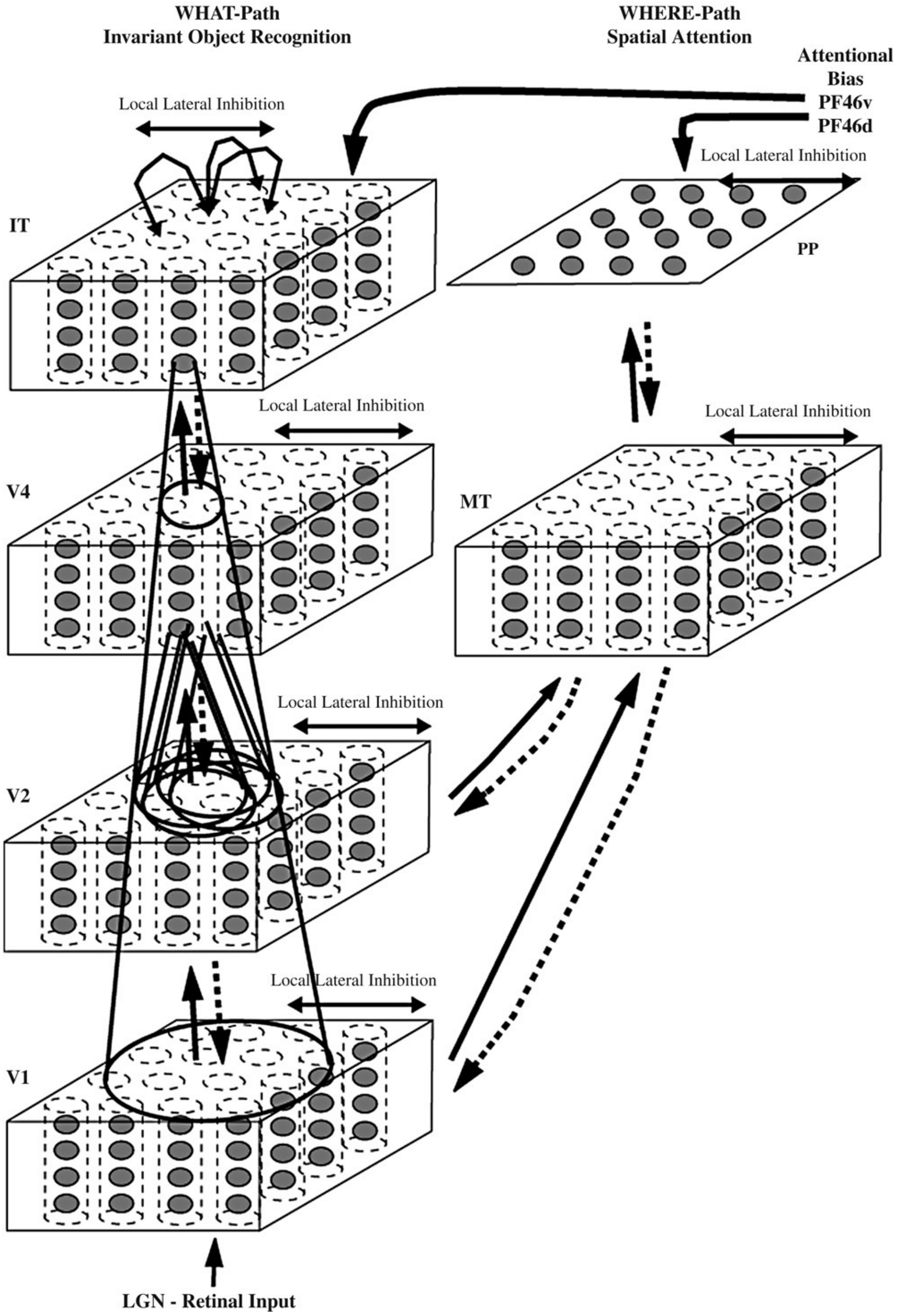

Fig. 1. Cortical architecture for hierarchical and attention-based visual perception after Deco and Rolls (2004). The system is essentially composed of five modules structured such that they resemble the two known main visual paths of the mammalian visual cortex. Information from the retino-geniculo-striate pathway enters the visual cortex through area V1 in the occipital lobe and proceeds into two processing streams. The occipital–temporal stream leads ventrally through V2–V4 and IT (inferior temporal visual cortex), and is mainly concerned with object recognition. The occipital–parietal stream leads dorsally into PP (posterior parietal complex), and is responsible for maintaining a spatial map of an object's location. The solid lines with arrows between levels show the forward connections, and the dashed lines the top-down backprojections. Short-term memory systems in the prefrontal cortex (PF46) apply top-down attentional bias to the object or spatial processing streams. (After Deco and Rolls 2004).

provides significant information related to object recognition and top-down attention in the ventral visual system.

This has been investigated by developing information theoretic methods for measuring the information present in stimulus-dependent synchrony (Franco, Rolls, Aggelopoulos, & Treves, 2004; Panzeri, Schultz, Treves, & Rolls, 1999; Rolls, 2003), and applying them to the analysis of neuronal activity in the macaque inferior temporal visual cortex during object recognition and attention. This brain region represents both features such as parts of objects and faces,

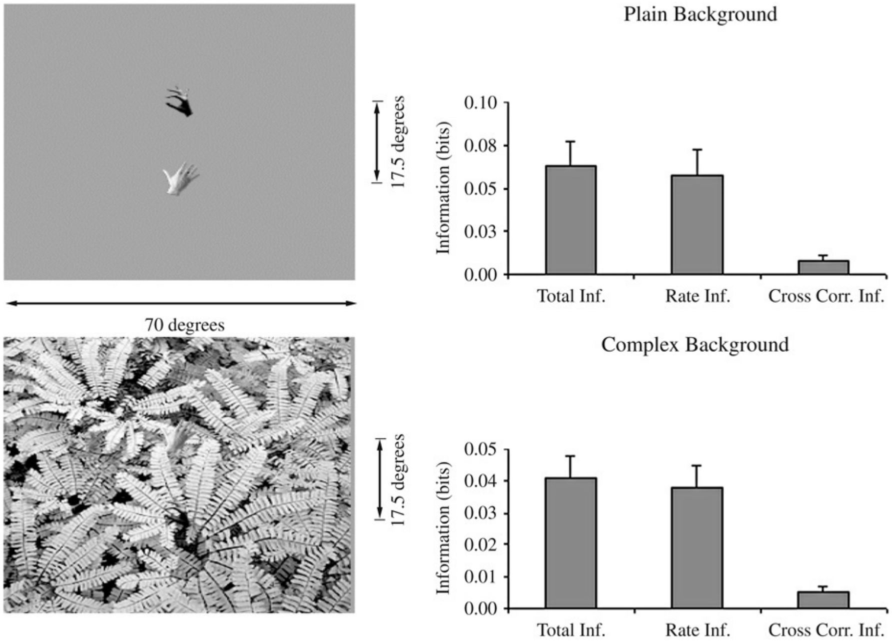

Fig. 2. Information encoding in natural scenes. Left: the objects against the plain background, and in a natural scene. Right: the information available (mean ± sem) from the firing rates (Rate Inf) or from stimulus-dependent synchrony (Cross-Corr Inf) from 30 populations of 2–4 simultaneously recorded inferior temporal cortex neurons about which stimulus had been presented in a complex natural scene. The total information (Total Inf) is that available from both the rate and the stimulus-dependent synchrony, which do not necessarily contribute independently (after Aggelopoulos et al., 2005).

and whole objects in which the features must be bound in the correct spatial relationship for the neurons to respond (Rolls, 2007d, 2008; Rolls & Deco, 2002). It has been shown that simultaneously recorded single neurons do sometimes show stimulus-dependent synchrony, but that the information available is less than 5% of that available from the spike counts (Aggelopoulos, Franco, & Rolls, 2005; Franco et al., 2004; Rolls, 2008; Rolls, Aggelopoulos, Franco, & Treves, 2004; Rolls, Franco, Aggelopoulos, & Reece, 2003).

The neurophysiological studies performed have included situations in which feature binding is likely to be needed, that is when the monkey had to choose to touch one of two simultaneously presented objects, with the stimuli presented in a complex natural background in a top-down attentional task (Aggelopoulos et al., 2005). We found that between 99% and 94% of the information was present in the firing rates of inferior temporal cortex neurons, and less that 5% in any stimulus-dependent synchrony that was present, as illustrated in Fig. 3. The implication of these results is that any stimulus-dependent synchrony that is present is not quantitatively important as measured by information theoretic analyses under natural scene conditions.

The point of the experimental design used was to test whether when the visual system is operating normally, in natural scenes and even searching for a particular object, stimulus-dependent synchrony is quantitatively important for encoding information about objects, and it was found not to be in the inferior temporal visual cortex. It will be of interest to apply the same quantitative information theoretic methods to earlier cortical visual areas, but the clear implication of the findings is that even when features must be bound together in the correct relative spatial positions to form object representations, and these must be segmented from the background, then stimulus-dependent synchrony is not quantitatively important in information encoding (Aggelopoulos et al., 2005; Rolls, 2008). Further, it was shown that there was little redundancy (less than 6%) between the information provided by the spike counts of the simultaneously recorded neurons, making spike counts an efficient population code with a high encoding capacity (Rolls, 2008).

The findings (Aggelopoulos et al., 2005) are consistent with the hypothesis that feature binding is implemented by neurons that respond to features in the correct relative spatial locations (Elliffe, Rolls, & Stringer, 2002; Rolls, 2008; Rolls & Deco, 2002), and not by temporal synchrony and attention (Malsburg, 1990; Singer, 1999). In any case, the computational point is that even if stimulus-dependent synchrony was useful for grouping, it would not without much extra machinery be useful for binding the relative spatial positions of features within an object, or for that matter of the positions of objects in a scene which appears to be encoded in a different way by using receptive fields that become asymmetric in crowded scenes (Aggelopoulos & Rolls, 2005). The computational problem is that synchronization does not by itself define the spatial relations between the features being bound, so is not as a binding mechanism adequate for shape recognition. For example, temporal binding might enable features 1, 2 and 3, which might define one stimulus to be bound together and kept separate from for example another stimulus consisting of features 2, 3 and 4, but would require a further temporal binding (leading in the end potentially to a combinatorial explosion) to indicate the relative spatial positions of the 1, 2 and 3 in the 123 stimulus, so that it can be discriminated from e.g. 312 (Rolls, 2008). However, the required computation for binding can be

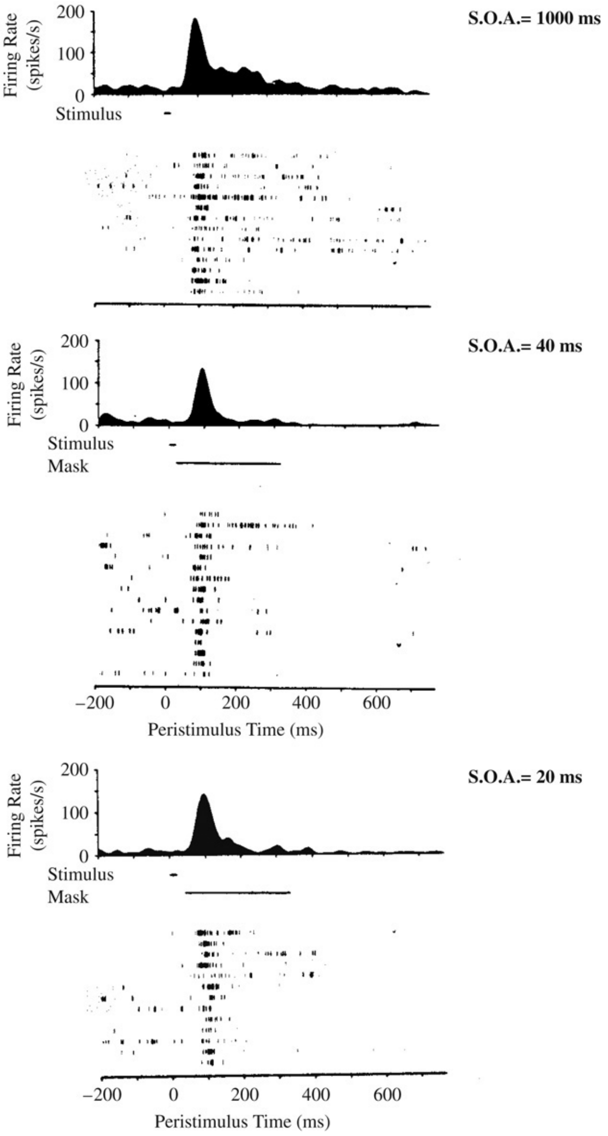

Fig. 3. Backward masking. Peristimulus rastergrams and smoothed peristimulus spike density histograms based on the responses of an inferior temporal cortex neuron in 8–16 trials to the test face alone (top raster–histogram pair), and to the test face followed by a masking stimulus (which was a face that was ineffective in activating the neuron) with S.O.As. of 40 ms and 20 ms. (S.O.A. = Stimulus Onset Asynchrony, the delay between the onset of the test stimulus and the onset of the mask stimulus.) The mask alone did not produce firing in the neuron. The target stimulus was shown for 16 ms starting at time 0. (The top trace shows the response to the target stimulus alone, in that with this 1000 ms S.O.A., the mask stimulus was delayed until well after the end of the recording period shown.) The effect of the mask is to interrupt the neuronal firing, which otherwise continues for up to several hundred ms after the short test stimulus (after Rolls & Tovee, 1994).

performed by the use of neurons that respond to combinations of features with a particular spatial arrangement (Elliffe et al., 2002; Rolls, 2008; Rolls & Deco, 2002; Rolls & Stringer, 2006).

Another type of evidence that stimulus-dependent neuronal synchrony is not likely to be crucial for information encoding, at least in the ventral visual system, is that the code about which visual stimulus has been shown can be read off from the end of the visual system in short times of 20–50 ms, and cortical neurons need fire for only this long during the identification of objects (Rolls, 2008; Rolls, Franco, Aggelopoulos, & Perez, 2006; Rolls & Tovee, 1994; Rolls, Tovee, Purcell, Stewart, & Azzopardi, 1994b; Tovee & Rolls, 1995; Tovee, Rolls, Treves, & Bellis, 1993). These are rather short time windows for the expression of multiple separate populations of synchronized neurons.

If large populations of neurons become synchronized, oscillations are likely to be evident in cortical recordings. In fact, oscillations are not an obvious property of neuronal firing in the primate temporal cortical visual areas involved in the representation of faces and objects when the system is operating normally in the awake behaving macaque (Tovee & Rolls, 1992). The fact that oscillations and neuronal synchronization are especially evident in anaesthetized cats does not impress as strong evidence that oscillations and synchronization are critical features of consciousness, for most people would hold that anaesthetized cats are *not* conscious. The fact that oscillations and synchronization are much more difficult to demonstrate in the temporal cortical visual areas of awake behaving monkeys (Aggelopoulos et al., 2005) might just mean that during evolution to primates the cortex has become better able to avoid parasitic oscillations, as a result of developing better feedforward and feedback inhibitory circuits (see Rolls & Deco, 2002; Rolls & Treves, 1998).

However, in addition there is an interesting computational argument against the utility of oscillations. The computational argument is related to the speed of information processing in cortical circuits with recurrent collateral connections. It has been shown that if attractor networks have integrate-and-fire neurons, and spontaneous activity, then memory recall into a basin of attraction can occur in approximately 1.5 time constants of the synapses, i.e. in times in the order of 15 ms (Battaglia & Treves, 1998; Panzeri, Rolls, Battaglia, & Lavis, 2001; Rolls & Treves, 1998; Simmen, Treves, & Rolls, 1996; Treves, 1993). One factor in this rapid dynamics of autoassociative networks with brain-like integrate-and-fire membrane and synaptic properties is that with some spontaneous activity, some of the neurons in the network are close to threshold already before the recall cue is applied, and hence some of the neurons are very quickly pushed by the recall cue into firing, so that information starts to be exchanged very rapidly (within 1–2 ms of brain time) through the modified synapses by the neurons in the network. The progressive exchange of information starting early on within what would otherwise be thought of as an iteration period (of perhaps 20 ms, corresponding to a neuronal firing rate of 50 spikes/s) is the mechanism accounting for rapid recall in an autoassociative neuronal network made biologically realistic in this way. However, this process relies on spontaneous random firings of different neurons, so that some will always be close to threshold when the retrieval cue is applied. If many of the neurons were firing synchronously in an oscillatory pattern, then there might be no neurons close to threshold and ready to be activated by the retrieval cue, so that the network might act much more like a discrete time network with fixed timesteps, which typically takes 8–15 iterations to settle, equivalent to perhaps 100 ms of brain time, and much too slow for cortical processing within any one area (Panzeri et al., 2001; Rolls, 2008; Rolls & Treves, 1998). The implication is that oscillations would tend to be detrimental to cortical computation, by slowing down any process using attractor dynamics. Attractor dynamics are likely to be implemented not only by the recurrent collateral connections between pyramidal neurons in a given cortical area, but also by the reciprocal feedforward and feedback connections between adjacent layers in cortical processing hierarchies (Rolls, 2008).

Another computational argument is that it is possible to account for many aspects of attention, including the non-linear interactions between top-down and bottom-up inputs, in integrate-and-fire neuronal networks that do not oscillate or show stimulus-dependent synchrony (Deco & Rolls, 2005a, 2005b, 2005c; Rolls, 2008).

The implication of these findings is that stimulus-dependent neuronal synchronization, and oscillatory activity, are unlikely to be quantitatively important in cortical processing, at least in the ventral visual stream. To the extent that we can be conscious of activity that has been processed in the ventral visual stream (made evident for example by reports of the appearance of objects), stimulus-dependent synchrony and oscillations are unlikely to be important in the neural mechanisms of consciousness.

3. A neural threshold for consciousness: The neurophysiology of backward masking

Damage to the primary (striate) visual cortex can result in blindsight, in which patients report that they do not see stimuli consciously, yet when making forced choices can discriminate some properties of the stimuli such as motion, position, some aspects of form, and even face expression (Azzopardi & Cowey, 1997; De Gelder, Vroomen, Pourtois, & Weiskrantz, 1999; Weiskrantz, 1997, 1998). In normal human subjects, backward masking of visual stimuli, in which another visual stimulus closely follows the short presentation of a test stimulus, reduces the visual perception of the test visual stimulus, and this paradigm has been widely used in psychophysics (Humphreys & Bruce, 1991). In this Section 1 consider how much information is present in neuronal firing in the part of the visual system that represents faces and objects, the inferior temporal visual cortex (Rolls, 2008; Rolls & Deco, 2002), when human subjects can discriminate face identity in forced choice testing, but cannot consciously perceive the person's face; and how much information is present when they become conscious of perceiving the stimulus. From this evidence I argue that even *within* a particular processing stream the processing may not be conscious yet can lead to behaviour; and that with higher and longer neuronal firing, events in that system become conscious. From this evidence I argue that the threshold for consciousness is normally higher than for some behavioural response. I then suggest a computational hypothesis for why this might be adaptive.

3.1. The neurophysiology of backward masking

The responses of single neurons in the macaque inferior temporal visual cortex have been investigated during backward visual masking (Rolls & Tovee, 1994; Rolls et al., 1994b). Recordings were made from neurons that were selective for faces, using distributed encoding (Rolls, 2007d, 2008; Rolls & Deco, 2002), during presentation of a test stimulus, a face, that lasted for 16 ms. The test stimulus was followed on different trials by a mask with stimulus onset asynchrony (S.O.A.) values of 20, 40, 60, 100 or 1000 ms. (The Stimulus Onset Asynchrony is the time between the onset of the test stimulus and the onset of the mask.) The duration of the pattern masking stimulus (letters of the alphabet) was 300 ms, and the neuron did not respond to the masking stimulus. Fig. 3 shows examples of the firing without a masking stimulus in the analysis period (S.O.A. = 1000 ms). Relative to the prestimulus rate, there was an increase in the firing produced with a latency of approximately 80 ms, and this firing lasted for 200–300 ms, that is for much longer than the 16 ms presentation of the target stimulus. With an S.O.A. of 20 ms, it is shown that *the effect of the mask is to interrupt the firing*, so that at this S.O.A. the neuron fired for approximately 30 ms. With longer S.O.As., the neuron fired for approximately 10 ms longer than the S.O.A., so that for example with an S.O.A. of 40 ms, the neuron fired for approximately 50 ms. Because of this, there were more spikes at the longer than the shorter S.O.As. Similar experiments, each involving effective and non-effective face stimuli for the cell, were repeated on 42 different neurons, and in all cases the temporal aspects of the masking were similar to those shown in Fig. 3 and in more detail elsewhere (Rolls & Tovee, 1994; Rolls et al., 1994b).

One important conclusion from these results is that the effect of a backward masking stimulus on cortical visual information processing is to limit the duration of neuronal responses, by interrupting neuronal firing. This persistence of cortical neuronal firing when a masking stimulus is not present is probably related to cortical recurrent collateral connections which could implement an autoassociative network with attractor and short-term memory properties (see Rolls, 2008; Rolls & Deco, 2002; Rolls & Treves, 1998), because such continuing post-stimulus neuronal firing is not observed in the lateral geniculate nucleus (K. Martin, personal communication).

3.2. Information available in inferior temporal cortex visual neurons during backward masking

To fully understand quantitatively the responses of inferior temporal cortex neurons at the threshold for visual perception, Rolls, Treves, Tovée, and Panzeri (1999) applied information theoretic methods to the analysis of the neurophysiological data with backward masking obtained by Rolls et al. (1994b) and Rolls and Tovee (1994). One advantage of this analysis is that it shows how well the neurons discriminate between the stimuli under different conditions, by taking into account not only the number of spikes (which does increase with the

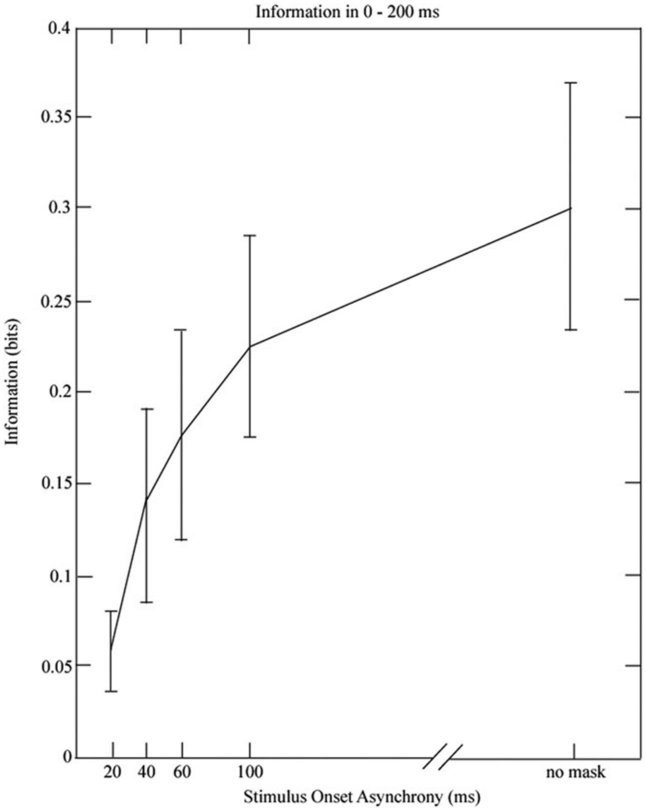

Fig. 4. Backward masking. The average (±sem) across the cells of the cumulated information available in a 200 ms period from stimulus onset from the responses of the cells as a function of the S.O.A. (After Rolls et al., 1999).

S.O.A. (Rolls, 2003, 2005a; Rolls, Tovee, & Panzeri, 1999)), but also the variability from trial to trial in the number of spikes. Another advantage of this analysis is that it evaluates the extent to which the neurons discriminate between stimuli in bits, which can then be directly compared with evidence about discriminability obtained using different measures, such as human psychophysical performance. The analysis quantifies what can be determined about which of the set of faces was presented from a single trial of neuronal firing.

The transmitted information carried by the neuronal firing rates about the stimuli was computed with the use of techniques that have been described elsewhere (e.g. Rolls, 2008; Rolls et al., 1999; Rolls & Treves, 1998). Fig. 4 shows the average across the cells of the cumulated information available in a 200 ms period from stimulus onset from the responses of the 15 neurons as a function of the S.O.A. This emphasizes how as the S.O.A. is reduced towards 20 ms the information does reduce rapidly, but that nevertheless at an S.O.A. of 20 ms there is still considerable information about which stimulus was shown. The reduction of the information at different S.O.A.s was highly significant (one way ANOVA) at $p < 0.001$. It was notable that the information reduced much more than the number of spikes on each trial as the S.O.A. was shortened. The explanation for this is that at short S.O.A.s the neuronal responses become noisy, as shown by Rolls et al. (1999).

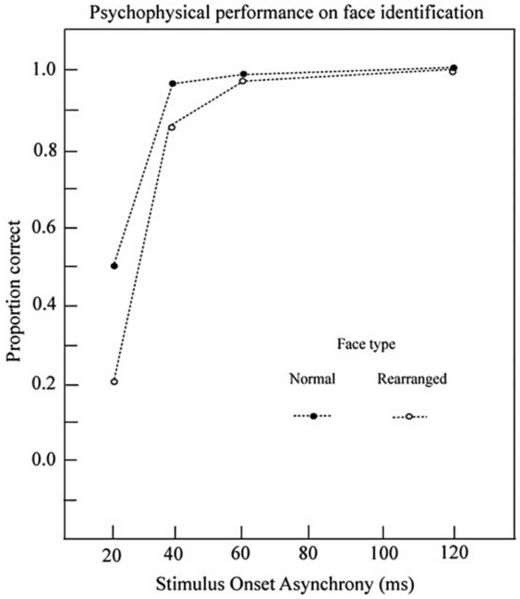

Fig. 5. Backward masking. Psychophysical performance of humans with the same stimuli as used in the neurophysiological experiments. The subjects were shown "normal" faces or faces with the parts "rearranged", and were asked to state which of 5 faces had been shown ("Identification"). The plots show the proportion correct on the task of determination of the identity of the faces for faces in the Normal or Rearranged spatial configuration of face features, as a function of Stimulus Onset Asynchrony (S.O.A). The data have been corrected for guessing. The means of the proportions correct are shown. The test stimulus was presented for 16 ms. (After Rolls et al., 1994b).

3.3. Human psychophysical performance with the same set of stimuli

Rolls et al. (1994b) performed human psychophysical experiments with the same set of stimuli and with the same apparatus used for the neurophysiological experiments so that the neuronal responses could be closely related to the identification that was possible of which face was shown. Five different faces were used as stimuli. All the faces were well-known to each of the 8 observers who participated in the experiment. In the forced choice paradigm, the observers specified whether the face was normal or rearranged (i.e. with the features jumbled), and identified whose face they thought had been presented. Even if the observers were unsure of their judgement they were instructed to respond with their best guess. The data were corrected for guessing. This correction arranged that chance performance would be shown as 0% correct on the graphs, and perfect performance as 100% correct.

The mean proportions of correct responses for the face identification task are shown in Fig. 5. (The proportion correct data was submitted to an arc sin transformation (to normalise the data) and a repeated measures ANOVA was performed. This analysis showed statistically significant effects of S.O.A. $[F(4, 28) = 61.52, p < 0.0001]$.)

Forced choice discrimination of face identity was better than chance at an S.O.A. of 20 ms. However, at this S.O.A., the subjects were not conscious of seeing the face, or of the identity of the face, and felt that their guessing about which face had been shown was not correct. The subjects did know that something had changed on the screen (and this was not just brightness, as this was constant throughout a trial). Sometimes the subjects had some conscious feeling that a part of a face (such as a mouth) had been shown. However, the subjects were not conscious of seeing a whole face, or of seeing the face of a particular person. At an S.O.A. of 40 ms, the subjects' forced choice performance of face identification was close to 100% (see Fig. 5), and at this S.O.A., the subjects became much more consciously aware of the identity of which face had been shown (Rolls et al., 1994b).

3.4. Comparison of neuronal data with the identification, and with the conscious perception, of visual stimuli

The neurophysiological data (Rolls & Tovee, 1994; Rolls et al., 1994b), and the results of the information theoretic analysis (Rolls et al., 1999), can now be compared directly with the effects of backward masking in human observers, studied in the same apparatus with the same stimuli (Rolls et al., 1994b). For the human observers, identification of which face from a set of six had been seen was 50% correct (with 0% correct corresponding to chance performance) with an S.O.A. of 20 ms, and 97% correct with an S.O.A. of 40 ms (Rolls et al., 1994b). Comparing the human performance purely with the changes in firing rate under the same stimulus conditions suggested that when it is just possible to identify which face has been seen, neurons in a given cortical area may be responding for only approximately 30 ms (Rolls & Tovee, 1994; Rolls et al., 1994b). The implication is that 30 ms is enough time for a neuron to perform sufficient computation to enable its output to be used for identification. The results based on an analysis of the information encoded in the spike trains at different S.O.A.s support this hypothesis by showing that a significant proportion of information is available in these few spikes (see Fig. 4), with on average 0.1 bits available from each neuron at an S.O.A. of 20 ms. Thus when subjects feel that they are guessing, and are not conscious of seeing whose face has been shown, macaque inferior temporal cortex neurons provide small but significant amounts of information about which face has been shown. When the S.O.A. was increased to 40 ms, the inferior temporal cortex neurons responded for approximately 50 ms, and encoded approximately 0.14 bits of information (cumulated information, for the subset of face-selective neurons tested, see Fig. 4). At this S.O.A., not only was face identification 97% correct, but the subjects were much more likely to be able to report consciously seeing a face and/or whose face had been shown. One further way in which the conscious perception of the faces was measured quantitatively was by asking subjects to rate the clarity of the faces. This was a subjective assessment and therefore reflected conscious processing, and was made using magnitude estimation. It was found that the subjective clarity of the stimuli was low at 20 ms S.O.A., was higher at 40 ms S.O.A., and was almost complete by 60 ms S.O.A (Rolls, 2003, 2005a; Rolls et al., 1994b).

It is suggested that the threshold for conscious visual perception may be set to be higher than the level at which small but significant sensory information is present so that the systems in the brain that implement the type of information processing involved in conscious thoughts are not interrupted by small signals that could be noise in sensory pathways. It is suggested below that the processing related to consciousness involves a higher order syntactic thought system used to correct first-order syntactic thoughts, and that these processes are inherently serial because of the way that the binding problems associated with the syntactic binding of symbols are treated by the brain. The argument is that is would be inefficient and would not be adaptive to interrupt this serial processing if the signal was very small and might be related to noise. Interruption of the serial processing would mean that the processing would need to start again, as when a train of thought is interrupted. The small signals that do not interrupt conscious processing but are present in sensory systems may nevertheless be useful for some implicit (non-conscious) functions, such as orienting the eyes towards the source of the input, and may be reflected in the better than chance recognition performance at short S.O.A.s even without conscious awareness.

3.5. Relation to blindsight

The quantitative analyses of neuronal activity in an area of the ventral visual system involved in face and object identification described here which show that significant neuronal processing can occur that is sufficient to support forced choice but implicit (unconscious) discrimination in the absence of conscious awareness of the identity of the face is of interest in relation to studies of blindsight (Azzopardi & Cowey, 1997; De Gelder et al., 1999; Weiskrantz, 1997, 1998). It has been argued that the results in blindsight are not due just to reduced visual processing, because some aspects of visual processing are less impaired than others (Azzopardi & Cowey, 1997; Weiskrantz, 1997, 1998, 2001). It is though suggested that some of the visual capacities that do remain in blindsight reflect processing via visual pathways that are alternatives to the V1 processing stream (Weiskrantz, 1997, 1998, 2001). If some of those pathways are normally involved in implicit processing, this may help us to give an account of why some implicit (unconscious) performance is possible in blindsight patients. Further, it has been suggested that ventral visual stream processing is especially involved in consciousness, because it is information about objects and faces that needs to enter a system to select and plan actions (Milner & Goodale, 1995; Rolls, 2008); and the planning of actions that involve the operation and correction of flexible one-time multiple step plans may be closely related to conscious processing (Rolls, 1999a, 2005b, 2008). In contrast, dorsal stream visual processing may be more closely related to executing an action on an object once the action has been selected, and the details of this action execution can take place implicitly (unconsciously) (Milner & Goodale, 1995; Rolls, 2008), perhaps because they do not require multiple step syntactic planning (Rolls, 1999a, 2005b, 2008).

One of the implications of blindsight thus seems to be that some visual pathways are more involved in implicit processing, and other pathways in explicit processing. In contrast, the results described here suggest that short and information-poor signals in a sensory system involved in conscious processing do not reach consciousness, and do not interrupt ongoing or engage conscious processing. This evidence described here thus provides interesting and direct evidence that there may be a threshold for activity in a sensory stream that must be exceeded in order to lead to consciousness, even when that activity is sufficient for some types of visual processing such as visual identification of faces at well above chance in an implicit mode. The latter implicit mode processing can be revealed by forced choice tests and by direct measurements of neuronal responses. Complementary evidence at the purely psychophysical level using backward masking has been obtained by Marcel (1983a, 1983b) and discussed by Weiskrantz (1998, 2001). Possible reasons for this relatively high threshold for consciousness are considered above in Section 3.4.

3.6. The speed of visual processing within a cortical visual area shows that top-down interactions with bottom-up processes are not essential for conscious visual perception

The results of the information analysis of backward masking (Rolls et al., 1999) emphasize that very considerable information about which stimulus was shown is available in a short epoch of, for example, 50 ms of neuronal firing. This confirms and is consistent with many further findings on the speed of processing of inferior temporal cortex neurons (Rolls, 2008; Rolls et al., 2006; Tovee & Rolls, 1995; Tovee et al., 1993), and facilitates the rapid read-out of information from the inferior temporal visual cortex. One direct implication of the 30 ms firing with the 20 ms S.O.A. is that this is sufficient time both for a cortical area to perform its computation, and for the information to be read out from a cortical area, given that psychophysical performance is 50% correct at this S.O.A. Another implication is that the recognition of visual stimuli can be performed using feedforward processing in the multistage hierarchically organised ventral visual system comprising at least V1–V2–V4-Inferior Temporal Visual Cortex, in that the typical shortest neuronal response latencies in macaque V1 are approximately 40 ms, and increase by approximately 15–17 ms per stage to produce a value of approximately 90 ms in the inferior temporal visual cortex (Dinse & Kruger, 1994; Nowak & Bullier, 1997; Rolls, 2008; Rolls & Deco, 2002). Given these timings, it would not be possible in the 20 ms S.O.A. condition for inferior temporal cortex neuronal responses to feed back to influence V1 neuronal responses to the test stimulus before the mask stimulus produced its effects on the V1 neurons. [In an example, in the 20 ms S.O.A. condition with 30 ms of firing, the V1 neurons would stop responding to the stimulus at $40+30 = 70$ ms, but would not be influenced by backprojected information from the inferior temporal cortex until $90 + (3$ stages $\times$ 15 ms per stage$) = 135$ ms. In another example for conscious processing, in the 40 ms S.O.A. condition with 50 ms of firing, the V1 neurons would stop responding to the stimulus at $40 + 50 = 90$ ms, but would not be influenced by backprojected information from the inferior temporal cortex until $90 + (3$ stages $\times$ 15 ms per stage$) = 135$ ms.] This shows that not only recognition, but also conscious awareness, of visual stimuli is possible without top-down backprojection effects from the inferior temporal visual cortex to early cortical processing areas that could interact with the processing in the early cortical areas.

The same information theoretic analyses (Rolls, 2008; Rolls et al., 2006, 1999; Tovee & Rolls, 1995; Tovee et al., 1993) show that from the earliest spikes of the anterior inferior temporal cortex neurons described here after they have started to respond (at approximately 80 ms after stimulus onset), the neuronal response is specific to the stimulus, and it is only in more posterior parts of the inferior temporal visual cortex that neurons may have an earlier short period of firing (of perhaps 20 ms) which is not selective for a particular stimulus. The neurons described by Sugase, Yamane, Ueno, and Kawano (1999) thus behaved like more posterior inferior temporal cortex neurons, not like typical anterior inferior temporal cortex neurons. This evidence thus suggests that in the anterior inferior temporal cortex, recurrent processing may help us to sharpen up representations to minimize early non-specific firing (cf. Lamme & Roelfsema, 2000).

4. To what extent is consciousness involved in the different types of processing initiated by perceptual and emotional states?

It might be possible to build a computer which would perform the functions of perception described above, and yet we might not want to ascribe *feelings* of perception to the computer. Indeed, as described above, the operation of the circuitry can occur to allow identification (by "guessing") in the absence of conscious awareness of the face being processed. The same argument applies to emotions, some of the processing for which can occur without conscious awareness (Rolls, 2005b, 2007c, 2008). We might even build the computer with some of the main processing stages present in the brain, and implement it using neural networks which simulate the operation of the real neural networks in the brain (Rolls, 2008; Rolls & Deco, 2002), yet we might not still wish to ascribe emotional feelings to this computer. In a sense, the functions of reward and punishment in emotional behaviour are described by the types of process and their underlying brain mechanisms in structures such as the amygdala and orbitofrontal cortex as described by Rolls (2005b), but what about the subjective aspects of emotion, what about the feeling of pleasure? A similar point arises when we consider the parts of the taste, olfactory, and visual systems in which the reward value of the taste, smell and sight of food is represented. One such brain region is the orbitofrontal cortex (Rolls, 2004a, 2005b, 2006b). Although the neuronal representation in the orbitofrontal cortex is clearly related to the reward value of food, is this where the pleasantness (the subjective hedonic aspect) of the taste, smell and sight of food is represented? Again, we could (in principle at least) build a computer with neural networks to simulate each of the processing stages for the taste, smell and sight of food that are described by Rolls (2005b) (and more formally in terms of neural networks by Rolls (2008) and Rolls and Deco (2002)), and yet would probably not wish to ascribe feelings of pleasantness to the system we have simulated on the computer.

What is it about neural processing that makes it feel like something when some types of information processing are taking place? It is clearly not a general property of processing in neural networks, for there is much processing, for example that concerned with the control of our blood pressure and heart rate, of which we are not aware. Is it then that awareness arises when a certain type of information processing is being performed? If so, what type of information processing? And how do emotional feelings, and sensory events, come to feel like anything? These feels are called qualia. These are great mysteries that have puzzled philosophers for centuries. They are at the heart of the problem of consciousness, for why it should feel like something at all is the great mystery. Other aspects of consciousness, such as the fact that often when we "pay attention" to events in the world, we can process those events in some better way, that is process or access as opposed to phenomenal aspects of consciousness, may be easier to analyse (Allport, 1988; Block, 1995; Chalmers, 1996). The puzzle of qualia, that is of the phenomenal aspect of consciousness, seems to be rather different from normal investigations in science, in that there is no agreement on criteria by which to assess whether we have made progress. So, although the aim of what follows in this paper is to address the issue of consciousness, especially of qualia, what is written cannot be regarded as being establishable by the normal methods of scientific enquiry. Accordingly, I emphasize that the view on consciousness that I describe is only preliminary, and theories of consciousness are likely to develop considerably. Partly for these reasons, this theory of consciousness, at least, should not be taken to have practical implications.

5. A theory of consciousness

5.1. *Multiple routes to action*

A starting point is that many actions can be performed relatively automatically, without apparent conscious intervention. An example sometimes given is driving a car. Another example is the identification of a visual stimulus that can occur without conscious awareness as described above. Such actions could involve control of behaviour by brain systems that are old in evolutionary terms such as the basal ganglia. It is of interest that the basal ganglia (and cerebellum) do not have backprojection systems to most of the parts of the cerebral cortex from which they receive inputs (Rolls, 2005b; Rolls & Treves, 1998). In contrast, parts of the brain such as the hippocampus and amygdala, involved in functions such as episodic memory and emotion respectively, about which we can make (verbal) declarations (hence declarative memory, Squire and Zola (1996)) do have major backprojection systems to the high parts of the cerebral cortex from which they receive forward projections (Rolls, 2008; Treves & Rolls, 1994). It may be that evolutionarily newer parts of the brain, such as the language areas and parts of the prefrontal cortex, are involved in an alternative type of control of behaviour, in which actions can be planned with the use of a (language) system which allows relatively arbitrary (syntactic) manipulation of semantic entities (symbols).

The general view that there are many routes to behavioural output is supported by the evidence that there are many input systems to the basal ganglia (from almost all areas of the cerebral cortex), and that neuronal activity in each part of the striatum reflects the activity in the overlying cortical area (Rolls, 1994, 2005b). The evidence is consistent with the possibility that different cortical areas, each specialized for a different type of computation, have their outputs directed to the basal ganglia, which then select the strongest input, and map this into action (via outputs directed for example to the premotor cortex) (Rolls, 2005b). Within this scheme, the language areas would offer one of many routes to action, but a route particularly suited to planning actions, because of the syntactic manipulation of semantic entities which may make long-term planning possible. A schematic diagram of this suggestion is provided in Fig. 6.

Consistent with the hypothesis of multiple routes to action, only some of which utilise language, is the evidence that split-brain patients may not be aware of actions being performed by the "non-dominant" hemisphere (Cooney & Gazzaniga, 2003; Gazzaniga, 1988, 1995; Gazzaniga & LeDoux, 1978). Also consistent with multiple including non-verbal routes to action, patients with focal brain damage, for example to the prefrontal cortex, may perform actions, yet comment verbally that they should not be performing those actions (Hornak et al., 2003, 2004; Rolls, 1999b, 2005b; Rolls, Hornak, Wade, & McGrath, 1994a). The actions which appear to be performed implicitly, with surprise expressed later by the explicit system, include making behavioural responses to a no-longer rewarded visual stimulus in a visual discrimination reversal (Hornak et al., 2004; Rolls et al., 1994a). In both these types of patients, confabulation may occur, in that a verbal account of why the action was performed may be given, and this may not be related at all to the environmental event which actually triggered the action (Gazzaniga, 1988, 1995; Gazzaniga & LeDoux, 1978; LeDoux, 2007; Rolls, 2005b; Rolls et al., 1994a).

Also consistent with multiple including non-verbal routes to action is the evidence that in backward masking at short time delays between the stimulus and the mask, neurons in the inferior temporal visual cortex respond selectively to different faces, and humans guess which face was presented 50% better than chance, yet report having not seen the face consciously (see Section 3).

It is possible that sometimes in normal humans when actions are initiated as a result of processing in a specialized brain region such as those involved in some types of rewarded behaviour, the language system may subsequently elaborate a coherent account of why that action was performed (i.e., confabulate). This would be consistent with a general view of brain evolution in which as areas of the cortex evolve, they are laid on top of existing circuitry connecting inputs to outputs, and in which each level in this hierarchy of separate input–output pathways may control behaviour according to the specialized function it can perform (see schematic in Fig. 6). (It is of interest that mathematicians may get a hunch

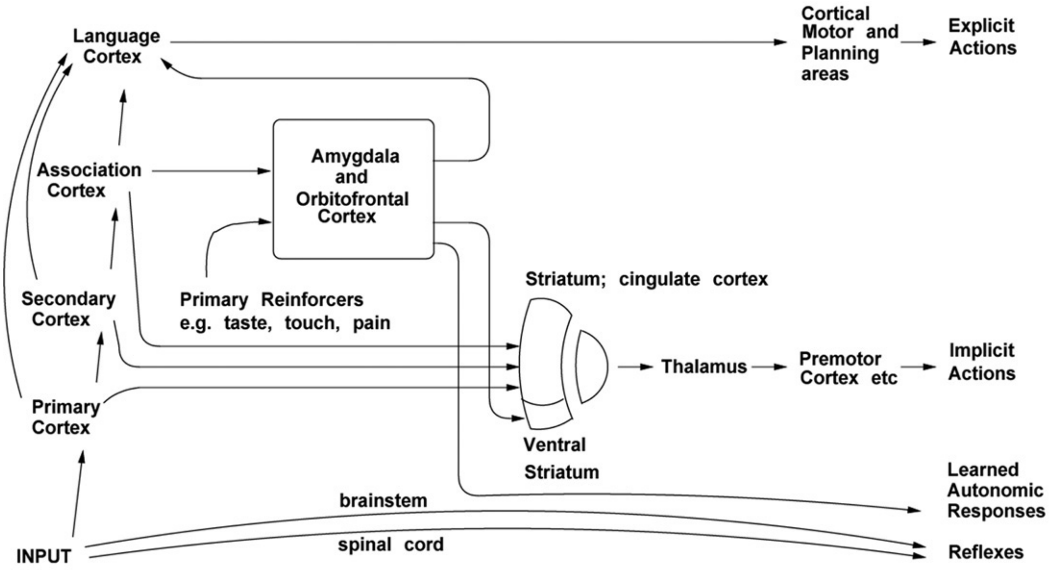

Fig. 6. Dual routes to the initiation of action in response to rewarding and punishing stimuli. The inputs from different sensory systems to brain structures such as the orbitofrontal cortex and amygdala allow these brain structures to evaluate the reward- or punishment-related value of incoming stimuli, or of remembered stimuli. The different sensory inputs enable evaluations within the orbitofrontal cortex and amygdala based mainly on the primary (unlearned) reinforcement value for taste, touch and olfactory stimuli, and on the secondary (learned) reinforcement value for visual and auditory stimuli. In the case of vision, the 'association cortex' which outputs representations of objects to the amygdala and orbitofrontal cortex is the inferior temporal visual cortex. One route for the outputs from these evaluative brain structures is via projections directly to structures such as the basal ganglia (including the striatum and ventral striatum) to enable implicit, direct behavioural responses based on the reward- or punishment-related evaluation of the stimuli to be made. The second route is via the language systems of the brain, which allow explicit decisions involving multistep syntactic planning to be implemented.

that something is correct, yet not be able to verbalise why. They may then resort to formal, more serial and language-like, theorems to prove the case, and these seem to require conscious processing. This is a further indication of a close association between linguistic processing, and consciousness. The linguistic processing need not, as in reading, involve an inner articulatory loop.)

We may next examine some of the advantages and behavioural functions that language, present as the most recently added layer to the above system, would confer. One major advantage would be the ability to plan actions through many potential stages and to evaluate the consequences of those actions without having to perform the actions. For this, the ability to form propositional statements, and to perform syntactic operations on the semantic representations of states in the world, would be important. Also important in this system would be the ability to have second-order thoughts about the type of thought that I have just described (e.g., I think that he thinks that...), as this would allow much better modelling and prediction of others' behaviour, and therefore of planning, particularly planning when it involves others.[1] This capability for higher order thoughts would also enable reflection on past events, which would also be useful in planning. In contrast, non-linguistic behaviour would be driven by learned reinforcement associations, learned rules etc., but not by flexible planning for many steps ahead involving a model of the world including others' behaviour. (For an earlier view which is close to this part of the argument see Humphrey (1980).) (The examples of behaviour from non-humans that may reflect planning may reflect much more limited and inflexible planning. For example, the dance of the honey-bee to signal to other bees the location of food may be said to reflect planning, but the symbol manipulation is not arbitrary. There are likely to be interesting examples of non-human primate behaviour, perhaps in the great apes, that reflect the evolution of an arbitrary symbol-manipulation system that could be useful for flexible planning, cf. Cheney and Seyfarth (1990). It is important to state that the language ability referred to here is not necessarily human verbal language (though this would be an example). What it is suggested is important to planning is the syntactic manipulation of symbols, and it is this syntactic manipulation of symbols which is the sense in which language is defined and used here.

5.2. A computational hypothesis of consciousness

It is next suggested that this arbitrary symbol manipulation using important aspects of language processing and used for planning but not in initiating all types of behaviour is close to what consciousness is about. In particular, consciousness may *be* the state which arises in a system that can think about (or reflect on) its own (or other peoples') thoughts, that is in a system capable of second or higher order thoughts (Carruthers, 1996; Dennett, 1991; Gennaro, 2004; Rolls, 1995, 1997a, 1997b, 1999a, 2004b, 2005b, 2007a, 2007c; Rosenthal, 1986, 1990, 1993, 2004, 2005). On this account, a mental state is non-introspectively (i.e., non-reflectively) conscious if one has a roughly simultaneous thought that one is in that mental

[1] Second-order thoughts are thoughts about thoughts. Higher order thoughts refer to second order, third order etc. thoughts about thoughts.

state. Following from this, introspective consciousness (or reflexive consciousness, or selfconsciousness) is the attentive, deliberately focused consciousness of one's mental states. It is noted that not all of the higher order thoughts need themselves be conscious (many mental states are not). However, according to the analysis, having a higher order thought about a lower order thought is necessary for the lower order thought to be conscious. A slightly weaker position than Rosenthal's (and mine) on this is that a conscious state corresponds to a first-order thought that has the *capacity* to cause a second-order thought or judgement about it (Carruthers, 1996). [Another position which is close in some respects to that of Carruthers and the present position is that of Chalmers (1996), that awareness is something that has *direct availability for behavioural control*, which amounts effectively for him in humans to saying that consciousness is what we can report (verbally) about.] This analysis is consistent with the points made above that the brain systems that are required for consciousness and language are similar. In particular, a system that can have second or higher order thoughts about its own operation, including its planning and linguistic operation, must itself be a language processor, in that it must be able to bind correctly to the symbols and syntax in the first-order system. According to this explanation, the feeling of anything is the state that is present when linguistic processing that involves second or higher order thoughts is being performed.

It might be objected that this captures some of the process aspects of consciousness, what is being performed in the relevant information processing system, but does not capture the phenomenal aspect of consciousness. I agree that there is an element of "mystery" that is invoked at this step of the argument, when I say that it feels like something for a machine with higher order thoughts to be thinking about its own first or lower order thoughts. But the return point (discussed further below) is the following: *if a human with second-order thoughts is thinking about its own first-order thoughts, surely it is very difficult for us to conceive that this would not feel like something?* (Perhaps the higher order thoughts in thinking about the first-order thoughts would need to have in doing this some sense of continuity or self, so that the first-order thoughts would be related to the same system that had thought of something else a few minutes ago. But even this continuity aspect may not be a requirement for consciousness. Humans with anterograde amnesia cannot remember what they felt a few minutes ago; yet their current state does feel like something.)

As a point of clarification, I note that according to this theory, a language processing system (let alone a working memory, LeDoux, 2007) is not *sufficient* for consciousness. What defines a conscious system according to this analysis is the ability to have higher order thoughts, and a first-order language processor (that might be perfectly competent at language) would not be conscious, in that it could not think about its own or others' thoughts. One can perfectly well conceive of a system that obeyed the rules of language (which is the aim of much connectionist modelling), and implemented a first-order linguistic system, that would not be conscious. [Possible examples of language processing that might be performed non-consciously include computer programs implementing aspects of language, or ritualized human conversations, e.g., about the weather. These might require syntax and correctly grounded semantics, and yet be performed non-consciously. A more complex example, illustrating that syntax could be used, might be "If A does X, then B will probably do Y, and then C would be able to do Z." A first-order language system could process this statement. Moreover, the first-order language system could apply the rule usefully in the world, provided that the symbols in the language system (A, B, X, Y etc.) are grounded (have meaning) in the world.]

A second clarification is that the plan would have to be a unique string of steps, in much the same way as a sentence can be a unique and one-off (or one-time) string of words. The point here is that it is helpful to be able to think about particular one-off plans, and to correct them; and that this type of operation is very different from the slow learning of fixed rules by trial and error, or the application of fixed rules by a supervisory part of a computer program.

5.3. Adaptive value

It is suggested that part of the evolutionary *adaptive significance* of this type of higher order thought is that it enables correction of errors made in first-order linguistic or in non-linguistic processing. Indeed, the ability to reflect on previous events is extremely important for learning from them, including setting up new long-term semantic structures. It is shown elsewhere that the hippocampus may be a system for such "declarative" recall of recent memories (Rolls, 2008). Its close relation to "conscious" processing in humans (Squire and Zola (1996) have classified it as a declarative memory system) may be simply that it enables the recall of recent memories, which can then be reflected upon in conscious, higher order, processing (Rolls, 2008; Rolls & Kesner, 2006). Another part of the adaptive value of a higher order thought system may be that by thinking about its own thoughts in a given situation, it may be able to better understand the thoughts of another individual in a similar situation, and therefore predict that individual's behaviour better (cf. Barlow, 1997; Humphrey, 1980, 1986).

In line with the argument on the adaptive value of higher order thoughts and thus consciousness given above, that they are useful for correcting lower order thoughts, I now suggest that correction using higher order thoughts of lower order thoughts would have adaptive value primarily if the lower order thoughts are sufficiently complex to benefit from correction in this way. The nature of the complexity is specific: that it should involve syntactic manipulation of symbols, probably with several steps in the chain, and that the chain of steps should be a one-off (or in American, "one-time", meaning used once) set of steps, as in a sentence or in a particular plan used just once, rather than a set of well-learned rules. The first or lower order thoughts might involve a linked chain of "if ... then" statements that would be involved in planning, an example of which has been given above. It is partly because complex lower order thoughts such as these which involve syntax and language would benefit from correction by higher order thoughts, that I suggest that there is a close link between this reflective consciousness and language. The *computational hypothesis* is that by thinking about lower order thoughts, the higher order thoughts can discover what may be weak links in the chain of reasoning at the lower order level, and having detected the weak link, might alter the plan, to see if this gives better success. In our example above, if it transpired that C could not do Z, how might the plan have failed? Instead of having to go through endless random changes to the plan to see if by trial and error some combination does happen to produce results, what I am suggesting is that by thinking about the previous plan, one might for example using knowledge of the situation and the probabilities that operate in it, guess that the step where the plan failed was that B did not in fact do Y. So by thinking about the plan (the first or lower order thought), one might correct the original plan, in such a way that the weak link in that chain, that "B will probably do Y", is circumvented.

I draw a parallel with neural networks: there is a "*credit assignment*" problem in such multistep syntactic plans, in that if the whole plan fails, how does the system assign credit or blame to particular steps of the plan? (In multilayer neural networks, the credit assignment problem is that if errors are being specified at the output layer, the problem arises about how to propagate back the error to earlier, hidden, layers of the network to assign credit or blame to individual synaptic connections; see Rolls and Deco (2002), Rolls (2008) and Rumelhart, Hinton, and Williams (1986).) The suggestion is that this is the function of higher order thoughts and is why systems with higher order thoughts evolved. The suggestion I then make is that if a system were doing this type of processing (thinking about its own thoughts), it would then be very plausible that it should feel like something to be doing this. I even suggest to the reader that it is not plausible to suggest that it would not feel like anything to a system if it were doing this.

5.4. Symbol grounding

A further point in the argument should be emphasized for clarity. The system that is having syntactic thoughts about its own syntactic thoughts (higher order syntactic thoughts or HOSTs) would have to have its symbols grounded in the real world for it to feel like something to be having higher order thoughts. The intention of this clarification is to exclude systems such as a computer running a program when there is in addition some sort of control or even overseeing program checking the operation of the first program. We would want to say that in such a situation it would feel like something to be running the higher level control program only if the first-order program was symbolically performing operations on the world and receiving input about the results of those operations, and if the higher order system understood what the first-order system was trying to do in the world. The issue of symbol grounding is considered further by Rolls (2005b). The symbols (or symbolic representations) are symbols in the sense that they can take part in syntactic processing. The symbolic representations are grounded in the world in that they refer to events in the world. The symbolic representations must have a great deal of information about what is referred to in the world, including the quality and intensity of sensory events, emotional states, etc. The need for this is that the reasoning in the symbolic system must be about stimuli, events, and states, and remembered stimuli, events and states, and for the reasoning to be correct, all the information that can affect the reasoning must be represented in the symbolic system, including for example just how light or strong the touch was, etc. Indeed, it is pointed out in *Emotion Explained* (Rolls, 2005b) that it is no accident that the shape of the multidimensional phenomenal (sensory etc.) space does map so clearly onto the space defined by neuronal activity in sensory systems, for if this were not the case, reasoning about the state of affairs in the world would not map onto the world, and would not be useful. Good examples of this close correspondence are found in the taste system, in which subjective space maps simply onto the multidimensional space represented by neuronal firing in primate cortical taste areas. In particular, if a three-dimensional space reflecting the distances between the representations of different tastes provided by macaque neurons in the cortical taste areas is constructed, then the distances between the subjective ratings by humans of different tastes is very similar (Kadohisa, Rolls, & Verhagen, 2005; Smith-Swintosky, Plata-Salaman, & Scott, 1991; Yaxley, Rolls, & Sienkiewicz, 1990). Similarly, the changes in human subjective ratings of the pleasantness of the taste, smell and sight of food parallel very closely the responses of neurons in the macaque orbitofrontal cortex (see *Emotion Explained*).

The representations in the first-order linguistic processor that the HOSTs process include beliefs (for example "Food is available", or at least representations of this), and the HOST system would then have available to it the concept of a thought (so that it could represent "I believe [or there is a belief] that food is available"). However, as argued by Rolls (1999a, 2005b), representations of sensory processes and emotional states must be processed by the first-order linguistic system, and HOSTs may be about these representations of sensory processes and emotional states capable of taking part in the syntactic operations of the first-order linguistic processor. Such sensory and emotional information may reach the first-order linguistic system from many parts of the brain, including those such as the orbitofrontal cortex and amygdala implicated in emotional states (see Fig. 6 and *Emotion Explained*, Fig. 10.3). When the sensory information is about the identity of the taste, the inputs to the first-order linguistic system must come from the primary taste cortex, in that the identity of taste, independently of its pleasantness (in that the representation is independent of hunger) must come from the primary taste cortex. In contrast, when the information that reaches the first-order linguistic system is about the pleasantness of taste, it must come from the secondary taste cortex, in that there the representation of taste depends on hunger.

5.5. Qualia

This analysis does not yet give an account for sensory qualia ("raw sensory feels", for example why "red" feels red), for emotional qualia (e.g., why a rewarding touch produces an emotional feeling of pleasure), or for motivational qualia (e.g., why food deprivation makes us *feel* hungry). The view I suggest on such qualia is as follows. Information processing in and from our sensory systems (e.g., the sight of the colour red) may be relevant to planning actions using language and the conscious processing thereby implied. Given that these inputs must be represented in the system that plans, we may ask whether it is more likely that we would be conscious of them or that we would not. I suggest that it would be a very special-purpose system that would allow such sensory inputs, and emotional and motivational states, to be part of (linguistically-based) planning, and yet remain unconscious. It seems to be much more parsimonious to hold that we would be conscious of such sensory, emotional and motivational qualia because they would be being used (or are available to be used) in this type of (linguistically-based) higher order thought processing, and this is what I propose.

The explanation for perceptual, emotional and motivational subjective feelings or qualia that this discussion has led towards is thus that they should be felt as conscious because they enter into a specialized linguistic symbol-manipulation system that is part of a higher order thought system that is capable of reflecting on and correcting its lower order thoughts involved for example in the flexible planning of actions. It would require a very special machine to enable this higher order linguistically-based thought processing, which is conscious by its nature, to occur without the sensory, emotional and motivational states (which must be taken into account by the higher order thought system) becoming felt qualia. The qualia are thus accounted for by the evolution of the linguistic system that can reflect on and correct its own lower order processes, and thus has adaptive value.

This account implies that it may be especially animals with a higher order belief and thought system and with linguistic symbol manipulation that have qualia. It may be that much non-human animal behaviour, provided that it does not require flexible linguistic planning and correction by reflection, could take place according to reinforcement guidance (using e.g., stimulus-reinforcement association learning in the amygdala and orbitofrontal cortex (Rolls, 2004a, 2005b, 2008)), and rule following (implemented e.g., using habit or stimulus-response learning in the basal ganglia (Rolls, 2005b). Such behaviours might appear very similar to human behaviour performed in similar circumstances, but would not imply qualia. It would be primarily by virtue of a system for reflecting on flexible, linguistic, planning behaviour that humans (and animals with demonstrable syntactic manipulation of symbols, and the ability to think about these linguistic processes) would be different from other animals, and would have evolved qualia.

It is of interest to comment on how the evolution of a system for flexible planning might affect emotions. Consider grief which may occur when a reward is terminated and no immediate action is possible (see Rolls, 1990, 2005b). It may be adaptive by leading to a cessation of the formerly rewarded behaviour and thus facilitating the possible identification of other positive reinforcers in the environment. In humans, grief may be particularly potent because it becomes represented in a system which can plan ahead, and understand the enduring implications of the loss. (Thinking about or verbally discussing emotional states may also in these circumstances help, because this can lead towards the identification of new or alternative reinforcers, and of the realization that for example negative consequences may not be as bad as feared.)

5.6. Pathways

In order for processing in a part of our brain to be able to reach consciousness, appropriate pathways must be present. Certain constraints arise here. For example, in the sensory pathways, the nature of the representation may change as it passes through a hierarchy of processing levels, and in order to be conscious of the information in the form in which it is represented in early processing stages, the early processing stages must have access to the part of the brain necessary for consciousness (see Fig. 6). An example is provided by processing in the taste system. In the primate primary taste cortex, neurons respond to taste independently of hunger, yet in the secondary taste cortex, food-related taste neurons (e.g., responding to sweet taste) only respond to food if hunger is present, and gradually stop responding to that taste during feeding to satiety (Rolls, 2005b, 2006a). Now the quality of the tastant (sweet, salt etc.) and its intensity are not affected by hunger, but the pleasantness of its taste is decreased to zero (neutral) (or even becomes unpleasant) after we have eaten it to satiety (Rolls, 2005b). The implication of this is that for quality and intensity information about taste, we must be conscious of what is represented in the primary taste cortex (or perhaps in another area connected to it which bypasses the secondary taste cortex), and not of what is represented in the secondary taste cortex. In contrast, for the pleasantness of a taste, consciousness of this could not reflect what is represented in the primary taste cortex, but instead what is represented in the secondary taste cortex (or in an area beyond it).

The same argument arises for reward in general, and therefore for emotion, which in primates is not represented early on in processing in the sensory pathways (nor in or before the inferior temporal cortex for vision), but in the areas to which these object analysis systems project, such as the orbitofrontal cortex, where the reward value of visual stimuli is reflected in the responses of neurons to visual stimuli (Rolls, 2005b, 2006a). It is also of interest that reward signals (e.g., the taste of food when we are hungry) are associated with subjective feelings of pleasure (Rolls, 2005b, 2006a). I suggest that this correspondence arises because pleasure is the subjective state that represents in the conscious system a signal that is positively reinforcing (rewarding), and that inconsistent behaviour would result if the representations did not correspond to a signal for positive reinforcement in both the conscious and the non-conscious processing systems.

Do these arguments mean that the conscious sensation of e.g., taste quality (i.e., identity and intensity) is represented or occurs in the primary taste cortex, and of the pleasantness of taste in the secondary taste cortex, and that activity in these areas is sufficient for conscious sensations (qualia) to occur? I do not suggest this at all. Instead the arguments I have put forward above suggest that we are only conscious of representations when we have high-order thoughts about them. The implication then is that pathways must connect from each of the brain areas in which information is represented about which we can be conscious, to the system that has the higher order thoughts, which as I have argued above, requires language. Thus, in the example given, there must be connections to the language areas from the primary taste cortex, which need not be direct, but which must bypass the secondary taste cortex, in which the information is represented differently (Rolls, 2005b). There must also be pathways from the secondary taste cortex, not necessarily direct, to the language areas so that we can have higher order thoughts about the pleasantness of the representation in the secondary taste cortex. There would also need to be pathways from the hippocampus, implicated in the recall of declarative memories, back to the language areas of the cerebral cortex (at least via the cortical areas which receive backprojections from the amygdala, orbitofrontal cortex, and hippocampus, see Fig. 6, which would in turn need connections to the language areas).

5.7. Consciousness and causality

One question that has been discussed is whether there is a causal role for consciousness (e.g., Armstrong & Malcolm, 1984). The position to which the above arguments lead is that indeed conscious processing does have a causal role in the elicitation of behaviour, but only under the set of circumstances when higher order thoughts play a role in correcting or influencing lower order thoughts. The sense in which the consciousness is causal is then it is suggested, that the higher order thought is causally involved in correcting the lower order thought; and that it is a property of the higher order thought system that it feels like something when it is operating. As we have seen, some behavioural responses can be elicited when there is not this type of reflective control of lower order processing, nor indeed any contribution of language (see further Rolls (2003, 2005a) for relations between implicit and explicit processing). There are many brain processing routes to output regions, and only one of these involves conscious, verbally represented processing which can later be recalled (see Fig. 6).

This account of consciousness also leads to a suggestion about the processing that underlies the feeling of free will. Free will would in this scheme involve the use of language to check many moves ahead on a number of possible series of actions and their outcomes, and then with this information to make a choice from the likely outcomes of different possible series of actions. (If in contrast choices were made only on the basis of the reinforcement value of immediately available stimuli, without the arbitrary syntactic symbol manipulation made possible by language, then the choice strategy would be much more limited, and we might not want to use the term free will, as all the consequences of those actions would not have been computed.) It is suggested that when this type of reflective, conscious, information processing is occurring and leading to action, the system performing this processing and producing the action would have to believe that it could cause the action, for otherwise inconsistencies would arise, and the system might no longer try to initiate action. This belief held by the system may partly underlie the feeling of free will. At other times, when other brain modules are initiating actions (in the implicit systems), the conscious processor (the explicit system) may confabulate and believe that it caused the action, or at least give an account (possibly wrong) of why the action was initiated. The fact that the conscious processor may have the belief even in these circumstances that it initiated the action may arise as a property of it being inconsistent for a system which can take overall control using conscious verbal processing to believe that it was overridden by another system. This may be the reason why confabulation occurs.

In the operation of such a free will system, the uncertainties introduced by the limited information possible about the likely outcomes of series of actions, and the inability to use optimal algorithms when combining conditional probabilities, would be much more important factors than whether the brain operates deterministically or not. (The operation of brain machinery must be relatively deterministic, for it has evolved to provide reliable outputs for given inputs.)

I suggest that these concepts may help us to understand what is happening in experiments of the type described by Libet and many others in which consciousness appears to follow with a measurable latency the time when a decision was taken. This is what I predict, if the decision is being made by an implicit perhaps reward/emotion or habit-related process, for then the conscious processor confabulates an account of or commentary on the decision, so that inevitably the conscious account follows the decision. On the other hand, I predict that if the rational (multistep, reasoning) route is involved in taking the decision, as it might be during planning, or a multistep task such as mental arithmetic, then the conscious report of when the decision was taken, and behavioural or other objective evidence on when the decision was taken, would correspond much more. Under those circumstances, the brain processing taking the decision would be closely related to consciousness, and it would not be a case of just confabulating or reporting on a decision taken by an implicit processor. It would be of interest to test this hypothesis in a version of Libet's task (Libet, 2002) in which reasoning was required. The concept that the rational, conscious, processor is only in some tasks involved in taking decisions is extended further in the section on dual routes to action below.

5.8. Consciousness, a computational system for higher order syntactic manipulation of symbols, and a commentary or reporting functionality

I now consider some clarifications of the present proposal, and how it deals with some issues that arise when considering theories of the phenomenal aspects of consciousness.

First, the present proposal has as its foundation the type of computation that is being performed, and suggests that it is a property of a higher order syntactic thought (HOST) system used for correcting multistep plans with its representations grounded in the world that it would feel like something for a system to be doing this type of processing. To do this type of processing, the system would have to be able to recall previous multistep plans, and would require syntax to keep the symbols in each step of the plan separate. In a sense, the system would have to be able to recall and take into consideration its earlier multistep plans, and in this sense *report* to itself, on those earlier plans. Some approaches to consciousness take the ability to report on or make a *commentary* on events as being an important marker for consciousness (Weiskrantz, 1997), and the computational approach I propose suggests why there should be a close relation between consciousness and the ability to report or provide a commentary, for the ability to report is involved in using higher order syntactic thoughts to correct a multistep plan.

Second, the implication of the present approach is that the type of linguistic processing or reporting need not be verbal, using natural language, for what is required to correct the plan is the ability to manipulate symbols syntactically, and this could be implemented in a much simpler type of mentalese or syntactic system (Fodor, 1994; Jackendoff, 2002; Rolls, 2004b) than verbal language or natural language which implies a universal grammar.

Third, this approach to consciousness suggests that the information must be being processed in a system capable of implementing HOSTs for the information to be conscious, and in this sense is more specific than global workspace hypotheses (Baars, 1988; Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Dehaene & Naccache, 2001). Indeed, the present approach suggests that a workspace could be sufficiently global to enable even the complex processing involved in driving a car to be performed, and yet the processing might be performed unconsciously, unless HOST (supervisory, monitory, correcting) processing was involved.

Fourth, the present approach suggests that it just is a property of HOST computational processing with the representations grounded in the world that it feels like something. There is to some extent an element of mystery about why it feels like something, why it is phenomenal, but the explanatory gap does not seem so large when one holds that the system is recalling, reporting on, reflecting on, and reorganising information about itself in the world in order to prepare new or revised plans. In terms of the physicalist debate (see for a review Davies, 2007), an important aspect of my proposal is that it is a *necessary* property of this type of (HOST) computational processing that it feels like something (the philosophical description is that this is an absolute metaphysical necessity), and given this view, then it is up to one to decide whether this view is consistent with one's particular view of physicalism or not (Rolls, 2007c). Similarly, the possibility of a zombie is inconsistent with the present hypothesis, which proposes that it is by virtue of performing processing in a specialized system that can perform higher order syntactic processing with the representations grounded in the world that phenomenal consciousness is necessarily present.

An implication of these points is that my theory of consciousness is a computational theory. It argues that it is a property of a certain type of computational processing that it feels like something. In this sense, although the theory spans many levels from the neuronal to the computational, it is unlikely that any particular neuronal phenomena such as oscillations are necessary for consciousness, unless such computational processes happen to rely on some particular neuronal properties not involved in other neural computations but necessary for higher order syntactic computations. It is these computations and the system that implements them that this computational theory argues are necessary for consciousness.

These are my initial thoughts on why we have consciousness, and are conscious of sensory, emotional and motivational qualia, as well as qualia associated with first-order linguistic thoughts. However, as stated above, one does not feel that there are straightforward criteria in this philosophical field of enquiry for knowing whether the suggested theory is correct; so it is likely that theories of consciousness will continue to undergo rapid development; and current theories should not be taken to have practical implications.

6. Dual routes to Action

According to the present formulation, there are two types of route to action performed in relation to reward or punishment in humans (see also Rolls, 2003, 2005b). Examples of such actions include emotional and motivational behaviour.

The first route is via the brain systems that have been present in non-human primates such as monkeys, and to some extent in other mammals, for millions of years. These systems include the amygdala and, particularly well-developed in primates, the orbitofrontal cortex. These systems control behaviour in relation to previous associations of stimuli with reinforcement. The computation which controls the action thus involves assessment of the reinforcement-related value of a stimulus. This assessment may be based on a number of different factors. One is the previous reinforcement history, which involves stimulus-reinforcement association learning using the amygdala, and its rapid updating especially in primates using the orbitofrontal cortex. This stimulus-reinforcement association learning may involve quite specific information about a stimulus, for example of the energy associated with each type of food, by the process of conditioned appetite and satiety (Booth, 1985). A second is the current motivational state, for example whether hunger is present, whether other needs are satisfied, etc. A third factor which affects the computed reward value of the stimulus is whether that reward has been received recently. If it has been received recently but in small quantity, this may increase the reward value of the stimulus. This is known as incentive motivation or the "salted peanut" phenomenon. The adaptive value of such a process is that this positive feedback of reward value in the early stages of working for a particular reward tends to lock the organism onto behaviour being performed for that reward. This means that animals that are for example almost equally hungry and thirsty will show hysteresis in their choice of action, rather than continually switching from eating to drinking and back with each mouthful of water or food. This introduction of hysteresis into the reward evaluation system makes action selection a much more efficient process in a natural environment, for constantly switching between different types of behaviour would be very costly if all the different rewards were not available in the same place at the same time. (For example, walking half a mile between a site where water was available and a site where food was available after every mouthful would be very inefficient.) The amygdala is one structure that may be involved in this increase in the reward value of stimuli early on in a series of presentations, in that lesions of the amygdala (in rats) abolish the expression of this reward incrementing process which is normally evident in the increasing rate of working for a food reward early on in a meal (Rolls, 2005b). A fourth factor is the computed absolute value of the reward or punishment expected or being obtained from a stimulus, e.g., the sweetness of the stimulus (set by evolution so that sweet stimuli will tend to be rewarding, because they are generally associated with energy sources), or the pleasantness of touch (set by evolution to be pleasant according to the extent to which it brings animals of the opposite sex together, and depending on the investment in time that the partner is willing to put into making the touch pleasurable, a sign which indicates the commitment and value for the partner of the relationship). After the reward value of the stimulus has been assessed in these ways, behaviour is then initiated based on approach towards or withdrawal from the stimulus. A critical aspect of the behaviour produced by this type of system is that it is aimed directly towards obtaining a sensed or expected reward, by virtue of connections to brain systems such as the basal ganglia and cingulate cortex (Rolls, 2007b) which are concerned with the initiation of actions (see Fig. 6). The expectation may of course involve behaviour to obtain stimuli associated with reward, which might even be present in a chain.

Now part of the way in which the behaviour is controlled with this first route is according to the reward value of the outcome. At the same time, the animal may only work for the reward if the cost is not too high. Indeed, in the field of behavioural ecology, animals are often thought of as performing optimally on some cost-benefit curve (see e.g., Krebs & Kacelnik, 1991). This does not at all mean that the animal thinks about the rewards, and performs a cost-benefit analysis using a lot of thoughts about the costs, other rewards available and their costs, etc. Instead, it should be taken to mean that in evolution, the system has evolved in such a way that the way in which the reward varies with the different energy densities or amounts of food and the delay before it is received, can be used as part of the input to a mechanism which has also been built to track the costs of obtaining the food (e.g., energy loss in obtaining it, risk of predation, etc.), and to then select given many such types of reward and the associated cost, the current behaviour that provides the most "net reward". Part of the value of having the computation expressed in this reward-minus-cost form is that there is then a suitable "currency", or net reward value, to enable the animal to select the behaviour with currently the most net reward gain (or minimal aversive outcome).

The second route in humans involves a computation with many "if ... then" statements, to implement a plan to obtain a reward. In this case, the reward may actually be *deferred* as part of the plan, which might involve working first to obtain one reward, and only then to work for a second more highly valued reward, if this was thought to be overall an optimal strategy in terms of resource usage (e.g., time). In this case, syntax is required, because the many symbols (e.g., names of people) that are part of the plan must be correctly linked or bound. Such linking might be of the form: "if A does this, then B is likely to do this, and this will cause C to do this.". The requirement of syntax for this type of planning implies that an output to language systems in the brain is required for this type of planning (see Fig. 2). This the explicit language system in humans may allow working for deferred rewards by enabling use of a one-off, individual, plan appropriate for each situation. Another building block for such planning operations in the brain may be the type of short-term memory in which the prefrontal cortex is involved. This short-term memory may be for example in non-human primates of where in space a response has just been made. A development of this type of short-term response memory system in humans to enable multiple short-term memories to be held in place correctly, preferably with the temporal order of the different items in the short-term memory coded correctly, may be another building block for the multiple step "if ... then" type of computation in order to form a multiple step plan. Such short-term memories are implemented in the (dorsolateral and inferior convexity) prefrontal cortex of non-human primates and humans (Goldman-Rakic, 1996; Petrides, 1996; Rolls, 2008), and may be part of the reason why prefrontal cortex damage impairs planning (Shallice & Burgess, 1996).

Of these two routes (see Fig. 6), it is the second which I have suggested above is related to consciousness. The hypothesis is that consciousness is the state which arises by virtue of having the ability to think about one's own thoughts, which has the adaptive value of enabling one to correct long multistep syntactic plans. This latter system is thus the one in which explicit, declarative, processing occurs. Processing in this system is frequently associated with reason and rationality, in that many of the consequences of possible actions can be taken into account. The actual computation of how rewarding a particular stimulus or situation is or will be probably still depends on activity in the orbitofrontal and amygdala, as the reward value of stimuli is computed and represented in these regions, and in that it is found that verbalised expressions of the reward (or punishment) value of stimuli are dampened by damage to these systems. (For example, damage to the orbitofrontal cortex renders painful input still identifiable as pain, but without the strong affective, "unpleasant", reaction to it.) This language system which enables long-term planning may be contrasted with the first system in which behaviour is directed at obtaining the stimulus (including the remembered stimulus) which is currently most rewarding, as computed by brain structures that include the orbitofrontal cortex and amygdala. There are outputs from this system, perhaps those directed at the basal ganglia, which do not pass through the language system, and behaviour produced in this way is described as implicit, and verbal declarations cannot be made directly about the reasons for the choice made. When verbal declarations are made about decisions made in this first system, those verbal declarations may be confabulations, reasonable explanations or fabrications, of reasons why the choice was made. These reasonable explanations would be generated to be consistent with the sense of continuity and self that is a characteristic of reasoning in the language system.

The question then arises of how decisions are made in animals such as humans that have both the implicit, direct reward-based, and the explicit, rational, planning systems (see Fig. 6) (Rolls, 2008). One particular situation in which the first, implicit, system may be especially important is when rapid reactions to stimuli with reward or punishment value must be made, for then the direct connections from structures such as the orbitofrontal cortex to the basal ganglia may allow rapid actions (Rolls, 2005b). Another is when there may be too many factors to be taken into account easily by the explicit, rational, planning, system, when the implicit system may be used to guide action. In contrast, when the implicit system continually makes errors, it would then be beneficial for the organism to switch from automatic, direct, action based on obtaining what the orbitofrontal cortex system decodes as being the most positively reinforcing choice currently available, to the explicit conscious control system which can evaluate with its long-term planning algorithms what action should be performed next. Indeed, it would be adaptive for the explicit system to regularly be assessing performance by the more automatic system, and to switch itself in to control behaviour quite frequently, as otherwise the adaptive value of having the explicit system would be less than optimal.

There may also be a flow of influence from the explicit, verbal system to the implicit system, in that the explicit system may decide on a plan of action or strategy, and exert an influence on the implicit system which will alter the reinforcement evaluations made by and the signals produced by the implicit system (Rolls, 2005b).

It may be expected that there is often a conflict between these systems, in that the first, implicit, system is able to guide behaviour particularly to obtain the greatest immediate reinforcement, whereas the explicit system can potentially enable immediate rewards to be deferred, and longer term, multistep, plans to be formed. This type of conflict will occur in animals with a syntactic planning ability, that is in humans and any other animals that have the ability to process a series of "if . . . then" stages of planning. This is a property of the human language system, and the extent to which it is a property of non-human primates is not yet fully clear. In any case, such a conflict may be an important aspect of the operation of at least the human mind, because it is so essential for humans to correctly decide, at every moment, whether to invest in a relationship or a group that may offer long-term benefits, or whether to directly pursue immediate benefits (Rolls, 2005b, 2008).

The thrust of the argument (Rolls, 2005b, 2008) thus is that much complex animal including human behaviour can take place using the implicit, non-conscious, route to action. We should be very careful not to postulate intentional states (i.e., states with intentions, beliefs and desires) unless the evidence for them is strong, and it seems to me that a flexible, one-off, linguistic processing system that can handle propositions is needed for intentional states. What the explicit, linguistic, system does allow is exactly this flexible, one-off, multistep planning ahead type of computation, which allows us to defer immediate rewards based on such a plan.

This discussion of dual routes to action has been with respect to the behaviour produced. There is of course in addition a third output of brain regions such as the orbitofrontal cortex and amygdala involved in emotion, that is directed to producing autonomic and endocrine responses (see Fig. 6). Although it has been argued by Rolls (2005b) that the autonomic system is not normally in a circuit through which behavioural responses are produced (i.e., against the James–Lange and related somatic theories), there may be some influence from effects produced through the endocrine system (and possibly the autonomic system, through which some endocrine responses are controlled) on behaviour, or on the dual systems just discussed that control behaviour.

7. Comparisons with other approaches to consciousness

The theory described here suggests that it feels like something to be an organism or machine that can think about its own (linguistic-, and semantically-based) thoughts. It is suggested that qualia, raw sensory and emotional subjective feelings, arise secondary to having evolved such a higher order thought system, and that sensory and emotional processing feels like something because it would be unparsimonious for it to enter the planning, higher order thought, system and *not* feel like something. The adaptive value of having sensory and emotional feelings, or qualia, is thus suggested to be that such inputs are important to the long-term planning, explicit, processing system. Raw sensory feels, and subjective states associated with emotional and motivational states, may not necessarily arise first in evolution. Some issues that arise in relation to this theory are discussed by Rolls (2000, 2004b, 2005b); reasons why the ventral visual system is more closely related to explicit than implicit processing (because reasoning about objects may be important) are considered by Rolls (2003) and by Rolls and Deco (2002); and reasons why explicit, conscious, processing may have a higher threshold in sensory processing than implicit processing are considered by Rolls (2003, 2005a, 2005b).

I now compare this approach to consciousness with those that place emphasis on working memory (LeDoux, 2007). LeDoux (1996), in line with Johnson-Laird (1988) and Baars (1988), emphasizes the role of working memory in consciousness, where he views working memory as a limited-capacity serial processor that creates and manipulates symbolic representations (p. 280). He thus holds that much emotional processing is unconscious, and that when it becomes conscious it is because emotional information is entered into a working memory system. However, LeDoux (1996) concedes that consciousness, especially its phenomenal or subjective nature, is not completely explained by the computational processes that underlie working memory (p. 281).

LeDoux (2007) notes that the term working memory can refer to a number of different processes. In attentional systems, a short-term memory is needed to hold on-line the subject of the attention, for example the position in space at which an object must be identified. There is much evidence that this short-term memory is implemented in the prefrontal cortex by an attractor network implemented by associatively modifiable recurrent collateral connections between cortical pyramidal cells, which keep the population active during the attentional task. This short-term memory then biases posterior perceptual and memory networks in the temporal and parietal lobes in a biased competition process (Deco & Rolls, 2005a, 2005b; Miller & Cohen, 2001; Rolls, 2008; Rolls & Deco, 2002). The operation of this type of short-term memory acting using biased competition to implement top-down attention does not appear to be central to consciousness, for as LeDoux (2007) agrees, prefrontal cortex lesions that have major effects on attention do not impair subjective feelings of consciousness. The same evidence suggests that attention itself is not a fundamental computational process that is necessary for consciousness, as the neural networks that implement short-term memory and operate to produce biased competition with non-linear effects do not appear to be closely related to consciousness (Deco & Rolls, 2005b; Rolls, 2008), though of course if attention is directed towards particular perceptual events, this will increase the gain of the perceptual processing (Deco & Rolls, 2005a, 2005b; Rolls, 2008), making the attended phenomena stronger.

Another process ascribed to working memory is that items can be manipulated in working memory, for example placed into a different order. This process implies at the computational level some type of syntactic processing, for each item (or symbol) could occur in any position relative to the others, and each item might occur more than once. To keep the items separate yet manipulable into any relation to each other, just having each item represented by the firing of a different set of neurons is insufficient, for this provides no information about the order or more generally the relations between the items being manipulated (Rolls, 2008; Rolls & Deco, 2002). In this sense, some form of syntax, that is a way to relate to each other the firing of the different populations of neurons each representing an item, is required. If we go this far (and LeDoux (1996) p. 280 does appear to), then we see that this aspect of working memory is very close to the concept I propose of syntactic thought in my HOST theory. My particular approach though makes it clear what function is to be performed (syntactic operations), whereas the term working memory can be used to refer to many different types of processing, and is in this sense less well-defined computationally. My approach of course argues that it is thoughts about the first-order thoughts that may be very closely linked to consciousness. In our simple case, the higher order thought might be "Do I have the items now in the correct reversed order? Should the X come before or after the Y?". To perform this syntactic manipulation, I argue that there is a special syntactic processor, perhaps in cortex near Broca's area, that performs the manipulations on the items, and that the dorsolateral prefrontal cortex itself provides the short-term store that holds the items on which the syntactic processor operates (Rolls, 2008). In this scenario, dorsolateral prefrontal cortex damage would affect the number of items that could be manipulated, but not consciousness or the ability to manipulate the items syntactically and to monitor and comment on the result to check that it is correct.

A property often attributed to consciousness is that it is *unitary*. LeDoux (2007) might relate this to the limitations of a working memory system. The current theory would account for this by the limited syntactic capability of neuronal networks in the brain, which render it difficult to implement more than a few syntactic bindings of symbols simultaneously (McLeod, Plunkett, & Rolls, 1998; Rolls, 2008). This limitation makes it difficult to run several "streams of consciousness" simultaneously. In addition, given that a linguistic system can control behavioural output, several parallel streams might produce maladaptive behaviour (apparent as e.g., indecision), and might be selected against. The close relation between, and the limited capacity of, both the stream of consciousness, and auditory-verbal short-term memory, may arise because both require implementation of the capacity for syntax in neural networks. My suggestion is that it is the difficulty the brain has in implementing the syntax required for manipulating items in working memory, and therefore for multiple step planning, and for then correcting these plans, that provides a close link between working memory concepts and my theory of higher order syntactic processing. The theory I describe makes it clear what the underlying computational problem is (how syntactic operations are performed in the system, and how they are corrected), and argues that when there are thoughts about the system, i.e. higher order syntactic thoughts (HOSTs), and the system is reflecting on its first-order thoughts (cf. Weiskrantz, 1997), then it is a property of the system that it feels conscious. As I argued above, first-order linguistic thoughts, which presumably involve working memory (which must be clearly defined for the purposes of this discussion), need not necessarily be conscious.

The theory is also different from some other theories of consciousness (Carruthers, 1996; Gennaro, 2004; Rosenthal, 2004, 2005) in that it provides an account of the evolutionary, adaptive, value of a higher order thought system in helping to solve a credit assignment problem that arises in a multistep syntactic plan, links this type of processing to consciousness, and therefore emphasizes a role for syntactic processing in consciousness. The type of syntactic processing need not be at the natural language level (which implies a universal grammar), but could be at the level of mentalese or simpler, as it involves primarily the syntactic manipulation of symbols (Fodor, 1994; Rolls, 2004b, 2005b).

The current theory holds that it is higher order *syntactic* thoughts (HOSTs) that are closely associated with consciousness, and this may differ from Rosenthal's higher order thoughts (HOTs) theory (Rosenthal, 1986, 1990, 1993, 2004, 2005), in the emphasis in the current theory on language. Language in the current theory is defined by syntactic manipulation of symbols, and does not necessarily imply verbal or "natural" language. The reason that strong emphasis is placed on language is that it is as a result of having a multistep flexible "on the fly" reasoning procedure that errors which cannot be easily corrected by reward or punishment received at the end of the reasoning, need "thoughts about thoughts", that is some type of supervisory and monitoring process, to detect where errors in the reasoning have occurred. This suggestion on the adaptive value in evolution of such a higher order linguistic thought process for multistep planning ahead, and correcting such plans, may also be different from earlier work. Put another way, this point is that *credit assignment* when reward or punishment is received is straightforward in a one-layer network (in which the reinforcement can be used directly to correct nodes in error, or responses); but is very difficult in a multistep linguistic process executed once "on the fly". Very complex mappings in a multilayer network can be learned if hundreds of learning trials are provided. But once these complex mappings are learned, their success or failure in a new situation on a given trial cannot be evaluated and corrected by the network. Indeed, the complex mappings achieved by such networks (e.g., backpropagation nets) mean that after training they operate according to fixed rules, and are often quite impenetrable and inflexible (Rolls & Deco, 2002). In contrast, to correct a multistep, single occasion, linguistically-based plan or procedure, recall of the steps just made in the reasoning or planning, and perhaps related episodic material, needs to occur, so that the link in the chain which is most likely to be in error can be identified. This may be part of the reason why there is a close relation between declarative memory systems, which can explicitly recall memories, and consciousness.

Some computer programs may have supervisory processes. Should these count as higher order linguistic thought processes? My current response to this is that they should not, to the extent that they operate with fixed rules to correct the operation of a system which does not itself involve linguistic thoughts about symbols grounded semantically in the external world. If on the other hand it were possible to implement on a computer such a higher order linguistic thought supervisory correction process to correct first-order one-off linguistic thoughts with symbols grounded in the real world, then this process would *prima facie* be conscious. If it were possible in a thought experiment to reproduce the neural connectivity and operation of a human brain on a computer, then *prima facie* it would also have the attributes of consciousness. It might continue to have those attributes for as long as power was applied to the system.

Another possible difference from other theories of consciousness is that raw sensory feels are suggested to arise as a consequence of having a system that can think about its own thoughts. Raw sensory feels, and subjective states associated with emotional and motivational states, may not necessarily arise first in evolution.

Finally, I provide a short specification of what might have to be implemented in a neural network to implement conscious processing. First, a linguistic system, not necessarily verbal, but implementing syntax between symbols grounded in the environment would be needed (e.g. a mentalese language system). Then a higher order thought system also implementing syntax and able to think about the representations in the first-order language system, and able to correct the reasoning in the first-order linguistic system in a flexible manner, would be needed. So my view is that consciousness can be implemented in neural networks of the artificial and biological type, but that the neural networks would have to implement the type of higher order linguistic processing described in this paper.

References

Aggelopoulos, N. C., Franco, L., & Rolls, E. T. (2005). Object perception in natural scenes: Encoding by inferior temporal cortex simultaneously recorded neurons. *Journal of Neurophysiology*, *93*, 1342–1357.

Aggelopoulos, N. C., & Rolls, E. T. (2005). Natural scene perception: Inferior temporal cortex neurons encode the positions of different objects in the scene. *European Journal of Neuroscience*, *22*, 2903–2916.

Allport, A. (1988). What concept of consciousness? In A. J. Marcel, & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 159–182). Oxford: Oxford University Press.

Armstrong, D. M., & Malcolm, N. (1984). *Consciousness and causality*. Oxford: Blackwell.

Azzopardi, P., & Cowey, A. (1997). Is blindsight like normal, near-threshold vision? *Proceedings of the National Academy of Sciences USA*, *94*, 14190–14194.

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Barlow, H. B. (1997). Single neurons, communal goals, and consciousness. In M. Ito, Y. Miyashita, & E. T. Rolls (Eds.), *Cognition, computation, and consciousness* (pp. 121–136). Oxford: Oxford University Press.

Battaglia, F. P., & Treves, A. (1998). Stable and rapid recurrent processing in realistic auto-associative memories. *Neural Computation*, *10*, 431–450.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*, 227–247.

Booth, D. A. (1985). Food-conditioned eating preferences and aversions with interoceptive elements: Learned appetites and satieties. *Annals of the New York Academy of Sciences*, *443*, 22–37.

Carruthers, P. (1996). *Language, thought and consciousness*. Cambridge.: Cambridge University Press.

Chalmers, D. J. (1996). *The conscious mind*. Oxford: Oxford University Press.

Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world*. Chicago: University of Chicago Press.

Cooney, J. W., & Gazzaniga, M. S. (2003). Neurological disorders and the structure of human consciousness. *Trends in Cognitive Sciences*, *7*, 161–165.

Crick, F. H. C., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, *2*, 263–275.

Davies, M. K. (2007). Consciousness and explanation. In L. Weiskrantz, & M. K. Davies (Eds.), *Frontiers of consciousness*. Oxford: Oxford University Press.

De Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex. *Neuroreport*, *10*, 3759–3763.

Deco, G., & Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, *44*, 621–644.

Deco, G., & Rolls, E. T. (2005a). Attention, short-term memory, and action selection: A unifying theory. *Progress in Neurobiology*, *76*, 236–256.

Deco, G., & Rolls, E. T. (2005b). Neurodynamics of biased competition and co-operation for attention: A model with spiking neurons. *Journal of Neurophysiology*, *94*, 295–313.

Deco, G., & Rolls, E. T. (2005c). Synaptic and spiking dynamics underlying reward reversal in orbitofrontal cortex. *Cerebral Cortex*, *15*, 15–30.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*, 204–211.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*, 1–37.

Dennett, D. C. (1991). *Consciousness explained*. London: Penguin.

Dinse, H. R., & Kruger, K. (1994). The timing of processing along the visual pathway in the cat. *Neuroreport*, *5*, 893–897.

Elliffe, M. C. M., Rolls, E. T., & Stringer, S. M. (2002). Invariant recognition of feature combinations in the visual system. *Biological Cybernetics*, *86*, 59–71.

Fodor, J. A. (1994). *The elm and the expert: Mentalese and its semantics*. Cambridge, MA: MIT Press.

Franco, L., Rolls, E. T., Aggelopoulos, N. C., & Treves, A. (2004). The use of decoding to analyze the contribution to the information of the correlations between the firing of simultaneously recorded neurons. *Experimental Brain Research*, *155*, 370–384.

Gazzaniga, M. S. (1988). Brain modularity: Towards a philosophy of conscious experience. In A. J. Marcel, & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 218–238). Oxford: Oxford University Press.

Gazzaniga, M. S. (1995). Consciousness and the cerebral hemispheres. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1392–1400). Cambridge, MA: MIT Press.

Gazzaniga, M. S., & LeDoux, J. (1978). *The integrated mind*. New York: Plenum.

Gennaro, R. J. (Ed.) (2004). *Higher order theories of consciousness*. Amsterdam: John Benjamins.

Goldman-Rakic, P. S. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society B*, *351*, 1445–1453.

Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R., et al. (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain*, *126*, 1691–1712.

Hornak, J., O'Doherty, J., Bramham, J., Rolls, E. T., Morris, R. G., Bullock, P. R., et al. (2004). Reward-related reversal learning after surgical excisions in orbitofrontal and dorsolateral prefrontal cortex in humans. *Journal of Cognitive Neuroscience*, *16*, 463–478.

Humphrey, N. K. (1980). Nature's psychologists. In B. D. Josephson, & V. S. Ramachandran (Eds.), *Consciousness and the physical world* (pp. 57–80). Oxford: Pergamon.

Humphrey, N. K. (1986). *The inner eye*. London: Faber.

Humphreys, G. W., & Bruce, V. (1991). *Visual cognition*. Hove, East Sussex: Erlbaum.

Huxter, J., Burgess, N., & O'Keefe, J. (2003). Independent rate and temporal coding in hippocampal pyramidal cells. *Nature*, *425*, 828–832.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

Johnson-Laird, P. N. (1988). *The computer and the mind: An introduction to cognitive science*. Cambridge, MA: Harvard University Press.

Kadohisa, M., Rolls, E. T., & Verhagen, J. V. (2005). Neuronal representations of stimuli in the mouth: The primate insular taste cortex, orbitofrontal cortex, and amygdala. *Chemical Senses*, *30*, 401–419.

Krebs, J. R., & Kacelnik, A. (1991). Decision making. In J. R. Krebs, & N. B. Davies (Eds.), *Behavioural ecology* (pp. 105–136). Oxford: Blackwell.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience*, *23*, 571–579.

LeDoux, J. (2007). Thinking our way through feelings: A working memory account of conscious emotional experiences. In L. Weiskrantz, & M. Davies (Eds.), *Frontiers of consciousness*. Oxford: Oxford University Press.

LeDoux, J. E. (1996). *The emotional brain*. New York: Simon and Schuster.

Libet, B. (2002). The timing of mental events: Libet's experimental findings and their implications. *Consciousness and Cognition*, *11*, 291–299. Discussion 304–233.

Malsburg, C. v. d. (1990). A neural architecture for the representation of scenes. In J. L. McGaugh, N. M. Weinberger, & G. Lynch (Eds.), *Brain organization and memory: Cells, systems and circuits* (pp. 356–372). New York: Oxford University Press.

Marcel, A. J. (1983a). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, *15*, 238–300.

Marcel, A. J. (1983b). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, *15*, 197–237.

McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes*. Oxford: Oxford University Press.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In K. S. Rockland, J. H. Kaas, & A. Peters (Eds.), *Extrastriate visual cortex in primates* (pp. 205–241). New York: Plenum Press.

Panzeri, S., Rolls, E. T., Battaglia, F., & Lavis, R. (2001). Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons. *Network: Computation in Neural Systems*, *12*, 423–440.

Panzeri, S., Schultz, S. R., Treves, A., & Rolls, E. T. (1999). Correlations and the encoding of information in the nervous system. *Proceedings of the Royal Society of London B*, *266*, 1001–1012.

Petrides, M. (1996). Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Philosophical Transactions of the Royal Society B*, *351*, 1455–1462.

Rolls, E. T. (1990). A theory of emotion, and its application to understanding the neural basis of emotion. *Cognition and Emotion*, *4*, 161–190.

Rolls, E. T. (1994). Neurophysiology and cognitive functions of the striatum. *Revue Neurologique (Paris)*, *150*, 648–660.

Rolls, E. T. (1995). A theory of emotion and consciousness, and its application to understanding the neural basis of emotion. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1091–1106). Cambridge, MA: MIT Press.

Rolls, E. T. (1997a). Brain mechanisms of vision, memory, and consciousness. In M. Ito, Y. Miyashita, & E. T. Rolls (Eds.), *Cognition, computation, and consciousness* (pp. 81–120). Oxford: Oxford University Press.

Rolls, E. T. (1997b). *Consciousness in neural networks? Neural Networks*, *10*, 1227–1240.

Rolls, E. T. (1999a). *The brain and emotion*. Oxford: Oxford University Press.

Rolls, E. T. (1999b). The functions of the orbitofrontal cortex. *Neurocase*, *5*, 301–312.

Rolls, E. T. (2000). Précis of the brain and emotion. *Behavioral and Brain Sciences*, *23*, 177–233.

Rolls, E. T. (2003). Consciousness absent and present: A neurophysiological exploration. *Progress in Brain Research*, *144*, 95–106.

Rolls, E. T. (2004a). The functions of the orbitofrontal cortex. *Brain and Cognition*, *55*, 11–29.

Rolls, E. T. (2004b). A higher order syntactic thought (HOST) theory of consciousness. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness: An anthology* (pp. 137–172). Amsterdam: John Benjamins.

Rolls, E. T. (2005a). Consciousness absent or present: A neurophysiological exploration of masking. In H. Ogmen, & B. G. Breitmeyer (Eds.), *The first half second: The microgenesis and temporal dynamics of unconscious and conscious visual processes* (pp. 89–108). Cambridge, MA: MIT Press, chapter 106.

Rolls, E. T. (2005b). *Emotion explained*. Oxford: Oxford University Press.

Rolls, E. T. (2006a). Brain mechanisms underlying flavour and appetite. *Philosophical Transactions of the Royal Society London B*, *361*, 1123–1136.

Rolls, E. T. (2006b). The neurophysiology and functions of the orbitofrontal cortex. In D. H. Zald, & S. L. Rauch (Eds.), *The orbitofrontal cortex* (pp. 95–124). Oxford: Oxford University Press.

Rolls, E. T. (2007a). The affective neuroscience of consciousness: Higher order linguistic thoughts, dual routes to emotion and action, and consciousness.

In P. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *Cambridge handbook of consciousness* (pp. 831–859). Cambridge: Cambridge University Press.

Rolls, E. T. (2007b). The anterior and midcingulate cortices and reward. In B. A. Vogt (Ed.), *Cingulate neurobiology & disease*. Oxford: Oxford University Press.

Rolls, E. T. (2007c). Emotion, Higher Order Syntactic Thoughts, and Consciousness. In L. Weiskrantz, & M. K. Davies (Eds.), *Frontiers of consciousness*. Oxford: Oxford University Press.

Rolls, E. T. (2007d). The representation of information about faces in the temporal and frontal lobes. *Neuropsychologia*, *45*, 125–143.

Rolls, E. T. (2008). *Memory, attention, and decision-making: A unifying computational neuroscience approach*. Oxford: Oxford University Press.

Rolls, E. T., Aggelopoulos, N. C., Franco, L., & Treves, A. (2004). Information encoding in the inferior temporal cortex: Contributions of the firing rates and correlations between the firing of neurons. *Biological Cybernetics*, *90*, 19–32.

Rolls, E. T., & Deco, G. (2002). *Computational neuroscience of vision*. Oxford: Oxford University Press.

Rolls, E. T., Franco, L., Aggelopoulos, N. C., & Perez, J. M. (2006). Information in the first spike, the order of spikes, and the number of spikes provided by neurons in the inferior temporal visual cortex. *Vision Research*, *46*, 4193–4205.

Rolls, E. T., Franco, L., Aggelopoulos, N. C., & Reece, S. (2003). An information theoretic approach to the contributions of the firing rates and correlations between the firing of neurons. *Journal of Neurophysiology*, *89*, 2810–2822.

Rolls, E. T., Hornak, J., Wade, D., & McGrath, J. (1994a). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery and Psychiatry*, *57*, 1518–1524.

Rolls, E. T., & Kesner, R. P. (2006). A computational theory of hippocampal function, and empirical tests of the theory. *Progress in Neurobiology*, *79*, 1–48.

Rolls, E. T., & Stringer, S. M. (2006). Invariant visual object recognition: A model, with lighting invariance. *Journal of Physiology - Paris*, *100*, 43–62.

Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London B*, *257*, 9–15.

Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience*, *11*, 335–346.

Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., & Azzopardi, P. (1994b). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Experimental Brain Research*, *101*, 473–484.

Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function*. Oxford: Oxford University Press.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, *49*, 329–359.

Rosenthal, D. M. (1990). A theory of consciousness. In *ZIF*. Bielefeld, Germany: Zentrum für Interdisziplinaire Forschung.

Rosenthal, D. M. (1993). Thinking that one thinks. In M. Davies, & G. W. Humphreys (Eds.), *Consciousness* (pp. 197–223). Oxford: Blackwell.

Rosenthal, D. M. (2004). Varieties of higher-order theory. In R. J. Gennaro (Ed.), *Higher order theories of consciousness*. Amsterdam: John Benjamins.

Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford: Oxford University Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & T. P. R. Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Shallice, T., & Burgess, P. (1996). The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society B*, *351*, 1405–1411.

Simmen, M. W., Treves, A., & Rolls, E. T. (1996). Pattern retrieval in threshold-linear associative nets. *Network*, *7*, 109–122.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, *24*, 49–65.

Smith-Swintosky, V. L., Plata-Salaman, C. R., & Scott, T. R. (1991). Gustatory neural encoding in the monkey cortex: Stimulus quality. *Journal of Neurophysiology*, *66*, 1156–1165.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences USA*, *93*, 13515–13522.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869–873.

Tovee, M. J., & Rolls, E. T. (1992). The functional nature of neuronal oscillations. *Trends in Neurosciences*, *15*, 387.

Tovee, M. J., & Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Visual Cognition*, *2*, 35–58.

Tovee, M. J., Rolls, E. T., Treves, A., & Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology*, *70*, 640–654.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.

Treves, A. (1993). Mean-field analysis of neuronal spike dynamics. *Network*, *4*, 259–284.

Treves, A., & Rolls, E. T. (1994). A computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*, 374–391.

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

Weiskrantz, L. (1998). *Blindsight. A case study and implications*. Oxford: Oxford University Press.

Weiskrantz, L. (2001). Blindsight — putting beta (β) on the back burner. In B. De Gelder, E. De Haan, & C. Heywood (Eds.), *Out of mind: Varieties of unconscious processes* (pp. 20–31). Oxford: Oxford University Press.

Yaxley, S., Rolls, E. T., & Sienkiewicz, Z. J. (1990). Gustatory responses of single neurons in the insula of the macaque monkey. *Journal of Neurophysiology*, *63*, 689–700.