BEHAVIORAL AND BRAIN SCIENCES (1995) 18, 659–722
Printed in the United States of America

The contents of consciousness: A neuropsychological conjecture

Jeffrey A. Gray
Department of Psychology, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, England
Electronic mail: *spjtjag@ucl.ac.uk*

Abstract: Drawing on previous models of anxiety, intermediate memory, the positive symptoms of schizophrenia, and goal-directed behaviour, a neuropsychological hypothesis is proposed for the generation of the contents of consciousness. It is suggested that these correspond to the outputs of a comparator that, on a moment-by-moment basis, compares the current state of the organism's perceptual world with a predicted state. An outline is given of the information-processing functions of the comparator system and of the neural systems which mediate them. The hypothesis appears to be able to account for a number of key features of the contents of consciousness. However, it is argued that neither this nor any existing comparable hypothesis is yet able to explain why the brain should generate conscious experience of any kind at all.

Keywords: anxiety; comparator; consciousness; contents of consciousness; schizophrenia; septohippocampal system

Introduction

How does the brain generate conscious experience? Twenty years ago most scientists thought this a question for the philosophers; and most philosophers regarded it as a symptom of some kind (though what kind never became clear) of deep linguistic confusion. In sharp contrast to this picture, there is now a large measure of agreement among both scientists and philosophers that not only is there a real problem about consciousness, but it is a scientific problem and the time has come for the scientists to tackle it. Hardly anyone today doubts that consciousness is in some way a product of the brain, a product that is intimately connected with the brain's role in behaviour and the processing of information. Cartesian dualism – the notion that brain-stuff and mind-stuff are essentially separate, though able to communicate with each other – has virtually no contemporary followers. The main debate (Marsh 1993) now centres on just how to go about determining the way in which consciousness in fact relates to brain function.

Contemporary functionalists appear to hold the view that there are no major philosophical or theoretical issues that need to be resolved. It will be sufficient merely to gather more, and more detailed, data regarding the empirical relationships that link together environmental events, brain function, conscious experience, information processing (computation), and behaviour; the resulting data set will itself then provide a sufficient account of how consciousness fits into the overall scientific picture. Dennett (in Marsh 1993, pp. 7–8), for example, argues that one can already take what is known about the way the brain codes and recodes visual stimuli, put it together with existing data from psychophysics, and use this knowledge to predict, successfully, new visual phenomena such as illusory experiences. One can even "inject" such experiences into the brains of experimental animals, as demonstrated for example in elegant experiments (Newsome & Salzman 1993) in which monkeys responded (behaviourally) to microstimulation of a circuit encoding a particular direction of motion in the same way that they had been trained to respond to an exteroceptive stimulus having the same directional value. What more than this, Dennett asks, do we need?

> If your science can tell you under exactly what conditions a subject will hate a certain stimulus or prefer one circumstance to another, it seems to me that to say, "but that just gives you the observable physical surroundings and circumstances of phenomenology" isn't a fair representation of the facts. As far as I can see, what's left to be inaccessible is pared down to something which one has to take on faith as making a difference even though it doesn't make any detectable difference. At that point, it seems to me, science would say, "there's nothing left to explain." (Dennett in Marsh 1993, pp. 7–8)

Kinsbourne (1993, p. 43) makes the same point still more firmly: "I take as my premise a view that I call neurofunctional – that awareness is an irreducible property of the activity of functionally entrained neuronal assemblies and therefore is amenable to no further explanation."

Some, however – probably the majority at a recent symposium (Marsh 1993) – including myself, remain unconvinced that it is going to be so simple. In this view, more than just dedicated gathering of experimental data is going to be needed. What is needed, rather, is a new theory, one that will render the relations between brain events and conscious experiences, to use Nagel's (Marsh 1993, p. 4) felicitous term, "transparent" (as, say, the theory of heat renders transparent the relation between the gas flame and the boiling of the kettle), rather than a series of brute correlations between environmental inputs, brain events, and behavioural outputs, on the one hand, and conscious experiences on the other. This, after all, is the standard set in all other scientific domains, so why not here?

 0140-525X/95 $9.00+.10

What would such a theory need to explain? I have considered this issue before (Gray 1971, p. 251) and concluded that:

> We stand in need of a scientific account of how conscious experiences (in the sense which is best illustrated by the experience of primitive sensations, that is, qualia) (1) evolved and (2) confer survival value on organisms possessing them (the evolutionary questions); and of how they (3) arise out of brain events and (4) alter behaviour (the mechanistic questions).

A successful theory of consciousness would be one from which answers to these questions could be deduced from one unifying set of concepts. Consider the analogy with the dual wave and particle aspects of elementary physical particles (Marsh 1993, pp. 242–43). It is a brute fact that such a duality of aspects exists in our universe, just as it is a brute fact that there is a duality of brain events and conscious experiences. But the theory of quantum mechanics provides a single set of equations from which both the wave-like and the particle-like aspects of the behaviour of fundamental particles can be predicted and understood; whereas no such theory yet exists from which one might derive the duality of brain events and conscious experiences. Moreover, such a theory is at present unimaginable – though only in the sense that no one could have imagined relativity or quantum mechanics before *they* were invented – not because we are dealing with the unknowable or a bad language habit.

With this background in mind, we can now state the main aim of this target article, which is to propose a new and specific hypothesis concerning a limited version of question 3 distinguished above, namely: How do conscious experiences arise out of brain events? (The hypothesis is termed "new" principally in order to distinguish it from the overall theoretical model to which, as indicated below, it is an addition; it has, of course, many antecedents, as is made clear in later sections of the paper.) According to this hypothesis a specific set of brain processes, linked to a specific set of psychological (information-processing) functions, provides the basis for conscious experiences. I shall attempt to show how the nature of these processes and functions might give rise to a number of specific features of conscious experience. In this sense, the paper takes a theoretical step beyond just the gathering of data on correlations between brain events and conscious experiences. But it is a step that still falls far short of Nagel's standard of transparency, as I shall show in the final section of the paper.[1]

The new hypothesis is stated fully in section 3, below. In brief, it proposes that the contents of consciousness consist of the outputs of a comparator system (Gray 1982a; 1982b) that has the general function of predicting, on a moment-by-moment basis, the next perceived state of the world, comparing this to the actual next perceived state of the world, and determining whether the predicted and actual states match or fail to do so ("mismatch"). The "contents of consciousness" in this hypothesis refer to the subjective experiences that make up what Jackendoff (1987, pp. 3–4) calls "primary awareness," including above all the perceived world with all its various qualities, but also bodily sensations, proprioception, mental images, dreams, internal speech, hallucinations, and so on. What Jackendoff calls "reflective awareness" – including for example beliefs and self-awareness – lies outside the scope of the hypothesis. The hypothesis states, therefore, that the contents of primary awareness consist of subjective qualities whose neural and computational equivalents have been processed by the comparator system, and have been determined by that system to be familiar or novel. Because the comparator system has itself previously been described within the context of a detailed neuropsychological model of anxiety (Gray 1982a; 1982b), intermediate memory (Gray & Rawlins 1986; Rawlins 1985), the positive psychotic symptoms of schizophrenia (Gray et al. 1991a; 1991b; Schmajuk et al., in preparation) and goal directed behaviour (Gray 1994), the new hypothesis necessarily also contains proposals concerning the specific neural mechanisms that generate the contents of consciousness. The general neuropsychological model is outlined in sections 1 and 2 below. No attempt is made to summarise the data or arguments on which it is based; for these, the reader is referred to the original sources. In section 4, I consider the level of analysis occupied by the new hypothesis; and in section 5, I attempt to show how it is able to account for a range of special features of consciousness noted in previous discussions of the problem, for example, O'Keefe (1985), Marcel & Bisiach (1988) and, most recently, Marsh (1993). Finally, in section 6, I consider the limitations in principle of this general kind of hypothesis.

In proposing any hypothesis concerning the contents of primary awareness, one encounters the problem posed by the private nature of conscious experience, in contrast to the public, intersubjectively confirmable domain of scientific data (Meehl 1966). It is not the privacy of conscious experience *per se* that poses the problem, however. Scientists need have no more difficulty in principle in agreeing on observations about conscious experiences than they do in agreeing about meter readings; witness the whole of psychophysics. Furthermore, reports of such experiences can be used to provide tests of specific hypotheses concerning their nature; research on the rotation of mental images (Cooper & Shepard 1973) provides a good, if controversial (Pylyshyn 1984), example. From a scientific point of view, the problem may be put as follows (Gray 1987): One's own conscious experience is a *datum* that stands *in need of explanation*; the conscious experiences of others, however, can function only as a *hypothesis* by which to *explain* their behaviour. The reason the problem posed by consciousness seems so acute, at least to nonfunctionalists, is the following: nothing that we know so far about behaviour, physiology, the evolution of either behaviour or physiology, or the possibilities of constructing automata to carry out complex forms of behaviour is such that the hypothesis of consciousness would arise if it did not occur in addition as a datum in our own experience; nor, having arisen, does it provide a useful explanation of the phenomena observed in those domains. The hypothesis proposed here approaches the problem principally from the point of view of consciousness-as-datum; that is, it seeks to account for some of the phenomenological features of conscious experiences by postulating specific neural and psychological functions that might give rise to them. In this respect, it resembles the approach taken, on a much larger scale, by Jackendoff (1987).

Sections 1 and 2 consider a further issue: how far does the normal scientific process of building neuropsychological models of specific psychological phenomena – as we have done for anxiety (Gray 1982a; 1982b) and positive psychotic symptoms (Gray et al. 1991a; 1991b) – lead *ipso*

facto to progress in solving the general problem of consciousness? If Dennett (in Marsh 1993, pp. 7–10) is right, the successful construction of such models should lead to sufficient (and sufficiently detailed) understanding of the relationships between brain, behaviour, and conscious experience for the general problem of consciousness gradually to wither away.

1. The comparator model: Anxiety

As developed so far, the comparator model encompasses three interlinked levels of analysis: behavioural, neural, and cognitive (i.e., information-processing). Notice that these levels of analysis have not hitherto included an experiential component (except in a limited sense that is clarified below). Indeed, the aim of the present paper may be seen as that of extending the model to the fourth level.

The model was first developed as a theory of anxiety (Gray 1982a; 1982b). At the behavioural level, this theory postulated a "behavioural inhibition system" (BIS), as presented in Figure 1. The critical eliciting stimuli for activity in the BIS are conditioned stimuli associated with punishment, conditioned stimuli associated with the omission, or termination of reward ("frustrative nonreward"; Amsel 1962; 1992), and novel stimuli. The behaviour elicited by these stimuli (right-hand side of Fig. 1) consists in behavioural inhibition (interruption of any ongoing behaviour); an increment in the level of arousal, such that the next behaviour to occur is carried out with extra vigour and/or speed; and an increment in attention, such that more information is taken in, especially concerning novel features of the environment. Any one of the inputs to the BIS elicits all the outputs; furthermore, a range of interventions is capable of blocking all the outputs to any of the inputs, while leaving intact other input-output relationships (including some that involve inputs to or outputs from the BIS, but not both). These are some of the reasons for regarding the BIS as indeed a unified system, rather than a congeries of separate input-output relationships.

Among the interventions which specifically abolish the input-output relationships that define the BIS is the administration of drugs, such as the benzodiazepines, barbiturates, and alcohol, which reduce anxiety in human beings (Gray 1977); indeed, the study of such drugs was a major impetus to the formation of the concept of the BIS (Gray 1982a; 1982b). It is on this basis that I identified the subjective state that accompanies activity in the BIS as anxiety. This identification gains plausibility from the fact that it leads to a face-valid description of human anxiety: that is, a state in which one responds to threat (stimuli associated with punishment or nonreward) or uncertainty (novelty) with the reaction, "stop, look and listen, and get ready for action" (right-hand side of Fig. 1).

In this sense, then, the model did include an experiential component from its inception. However, let us suppose that the identification of activity in the BIS with the subjective state of anxiety is correct in every detail. Note, even so, that we are still left with no more than a brute correlation: the identification offers no account of how such activity gives rise to the specific subjective features of felt anxiety, nor of why activity in the BIS should give rise to any subjective features at all. Thus the attribution of subjective features to the model in this way does nothing to solve the central problem of consciousness, it merely evades it.

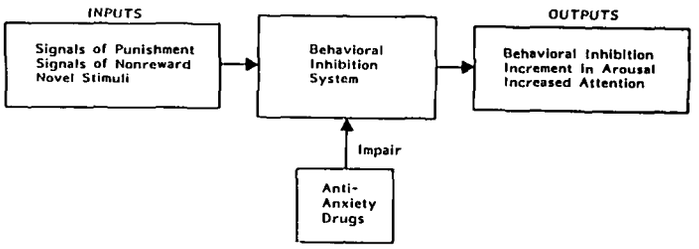

Figure 1. The behavioural inhibition system (BIS) as defined by its inputs and outputs.

The set of neurological structures which appear to discharge the functions of the BIS is illustrated in Figure 2, including notably the septohippocampal system (SHS) and its associated "Papez loop," areas of the temporal and frontal neocortex, and the ascending noradrenergic and serotonergic pathways that innervate these forebrain regions. We shall leave until later a more detailed consideration of the neural activity mediated by these structures. The proposal that the BIS, as defined behaviourally (Fig. 1), consists of such activity depends upon a variety of sources of information (Gray 1982a). For the sake of the argument pursued here, let us again suppose that this proposal is correct in every detail: would this bring us any closer to an understanding of the relationship between, on the one hand, the behavioural and neurological aspects of anxiety and, on the other, its experiential features? Clearly not: we merely have another brute correlation. To see that this is so, suppose that, instead of the set of brain structures depicted in Figure 2, another set turns out to mediate the functions of the BIS. Would this lead to any testable predictions about

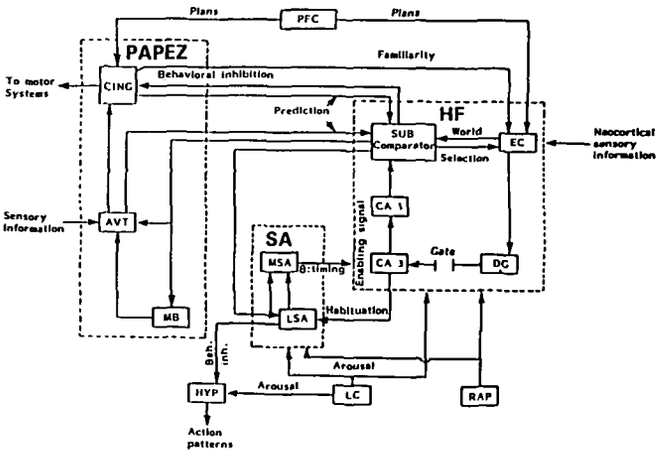

Figure 2. The septohippocampal system: the three major building blocks are shown in heavy print: HF, the hippocampal formation, made up of the entorhinal cortex, EC, the dentate gyrus, DG, 3, CA 1, and the subicular area, SUB: SA, the septal area, containing the medial and lateral septal areas, MSA and LSA; and the Papez circuit, which receives projections from and returns them to the subicular area via the mammillary bodies, MB, anteroventral thalamus, AVT, and cingulate cortex, CING. Other structures shown are the hypothalamus, HYP, the locus coeruleus, LC, the raphe nuclei, RAP, and the prefrontal cortex, PFC. Arrows show direction of projection; the projection form SUB to MSA lacks anatomical confirmation. Words in lower case show postulated functions; beh. inhib., behavioural inhibition. From Gray (1982b).

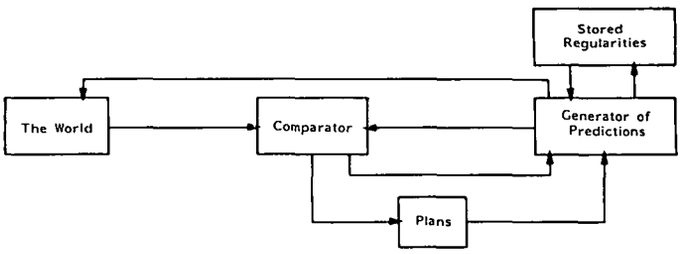

Figure 3. Information processing required for the comparator function of the septohippocampal system.

change in the subjective features of anxiety? The ability in principle to derive such predictions would take us beyond brute correlations and towards "transparency"; but the theory that links the hypotheses depicted in Figures 1 and 2 to each other and to anxiety offers no way of deriving them.

Now, the main function of the brain is to process information. Thus, faced with a neurological flow diagram, one should ask not only how the structures illustrated therein produce their relevant behavioural outputs, but also what cognitive (i.e., information processing) operations they perform in order to do so. The information-processing functions attributed in the model to the interlinked set of structures depicted in Figure 2 are themselves illustrated in Figure 3; detailed justification for the ideas contained in this figure can be found in Gray (1982a; 1982b). The key concept is that of the comparator, that is, a system which, moment by moment, predicts the next likely event and compares this prediction to the actual event. The experimental evidence upon which this concept rests is complex and taken from many sources, especially those summarised by Gray (1982a; 1982b) in relation to anxiety. Eichenbaum et al. (1994) have recently used a similar conceptual analysis in the attempt to integrate experimental evidence concerning the memorial functions of the hippocampal system.

The system depicted in Figures 2 and 3 (i) takes in information describing the current state of the (perceived) world; (ii) adds to this further information concerning the subject's current motor program; (iii) makes use of information stored in memory and describing past regularities that relate stimulus events to other stimulus events (derived from the process of classical conditioning); (iv) similarly makes use of stored information describing past regularities that relate responses to subsequent stimulus events (derived from instrumental conditioning); (v) from these sources of information predicts the next expected state of the world; (vi) compares the predicted to the actual next state of the world; (vii) decides either that there is a match or that there is a mismatch between the predicted and actual states of the world; (viii) if there is a match, proceeds to run through steps (i) to (vii) again; but (ix) brings the current motor program to a halt (i.e., operates the outputs of the BIS; Fig. 1) if there is a mismatch, or (x) if the predicted state of the world is associated with punishment or nonreward; and (xi) in that case takes in further information to resolve the difficulty that has interrupted the current motor program.

Figure 3 depicts, as it were, the software of the comparator proposed by Gray (1982a); as we have seen, the corresponding hardware is as illustrated in Figure 2. At this neural level, the core structure is the SHS, composed of the septal area, entorhinal cortex, dentate gyrus, hippocampus, and subicular area. Here we note only the following points (refer to Figure 2 for the circuitry).

First, the heart of the comparator function is attributed to the subicular area. This is postulated (1) to receive elaborated descriptions of the perceptual world from the entorhinal cortex, itself the recipient of input from all cortical sensory association areas; (2) to receive predictions from, and to initiate the generation of the next prediction in, the Papez circuit (i.e., the circuit from the subiculum to the mammillary bodies, the anteroventral thalamus, the cingulate cortex and back to the subiculum); and (3) to interface with motor programming systems (not themselves included in Fig. 2; see below) so as either to bring them to a halt or to permit them to continue.

Second, the prefrontal cortex is allotted the role of providing the comparator system with information concerning the current motor program (via its projections to the entorhinal and cingulate cortices, the latter forming part of the Papez circuit).

Third, the monoaminergic pathways which ascend from the mesencephalon to innervate the SHS (consisting of noradrenergic fibres originating in the locus coeruleus, and serotonergic fibres originating in the median raphe) are charged with alerting the whole system under conditions of threat and diverting its activities to deal with the threat; in the absence of threat, the information-processing activities of the system can be put to other, nonemotional purposes (Gray 1984).

Last, the system depicted in Figure 3 needs to be quantized in time, to allow appropriate comparisons between specific perceived states of the world and corresponding predictions, followed by initiation of the next prediction and next intake of information describing the world. This function is attributed to the hippocampal theta rhythm, giving rise to an "instant" within the model of about one-tenth of a second. (More specifically, the theta rhythm in animals such as the rat has a frequency ranging from about 6 to about 12 Hz; so an "instant" must have a duration of approximately 80–160 msec.)

In the application of this model to anxiety (Gray 1982a; 1982b), the main focus of the analysis was on steps (ix)–(xi), outlined above, and the further consequences of these steps. A detailed attempt was made to show how the combination of the three levels of analysis shown here in Figures 1–3 (henceforth collectively termed simply "the BIS") was able to account for a variety of features of human anxiety and anxiety disorders, including many highly subjective features (e.g., the compulsive ruminations of patients with obsessive-compulsive disorder). A key feature of this application to human anxiety was the distinction between two modes of operation of the BIS. The first, "checking," mode applies when the comparator repeatedly declares "match" and recursively runs through steps (i)–(vii) listed above. The second, "control," mode applies when either a predicted threat or a mismatch is detected, leading to operation of the behavioural outputs of the BIS (Fig. 1). The phobic symptoms of human anxiety were attributed to activity of the BIS in control mode; the cognitive symptoms (worry, rumination, etc.), in part to the operation of step (xi) in control mode, and in part to certain special features of activity in checking mode that become possible uniquely in human beings because of descending control over the BIS from the prefrontal cortex, conveying influences from language-based cortical systems.

Does the addition of the information-processing level of analysis, summarised in Figure 3, to the other two levels (Figs. 1 and 2) bring us any closer to a transparent theory of the linkage between brain-and-behaviour and conscious experience? Certainly, there are many passages in Gray's (1982a) detailed account of features of human anxiety which might give the unwary reader the impression that real progress in this direction has been made (see, e.g., pp. 442–44). This impression is strengthened, in particular, by the use in the description of the software of the model (Fig. 3) of terms that have a rather ambiguous status insofar as consciousness is concerned. Consider, for example, "the use of information stored in memory and describing past regularities of experience in order to predict the next state of the world" (steps iii–v), or "the comparison between this predicted state and the actual state of the world" (steps vi, vii). Are these conscious or unconscious processes? If we were to model them in a computer program (as can be done; e.g., Schmajuk et al., in preparation), we would surely suppose them to lack conscious components; and it is in this way that I intended to use these concepts. But the fact that such predictions and comparisons are also familiar to us as elements in our conscious experience allows one to slip all too readily into the assumption that, by the simple postulation of equivalence between certain aspects of neural activity (e.g., the passage of impulses around the Papez loop) and certain features of the processing of information (e.g., the making of a prediction as to the next input into the system) – itself a controversial but not a mysterious step in theory building (see Marsh 1993, pp. 273–75) – one has made a contribution to the substantive problem of consciousness. But it is precisely the fact that we may leave the status – conscious or unconscious – of such processes undefined in the construction of this kind of information-processing model that most clearly demonstrates our lack of progress in understanding consciousness. For consciousness must make a behavioural difference, otherwise it could not have been the subject of Darwinian selection and evolution (Gray 1971). If our theories make identical predictions whether the processes they postulate are treated as conscious or not, then this critical difference that is due to consciousness has clearly been left out.

2. The comparator model: Schizophrenia

More recently, Gray et al. (1991a; 1991b) have extended the comparator model and applied it to some of the bizarre cognitive features of acute schizophrenia with positive symptoms. Whereas applications of the model to anxiety were principally concerned with the consequences of disrupting the monitoring of motor programs (when threats or mismatch is detected), the extended model is more concerned with the details of the monitoring process itself, and the way in which this interacts with motor programs. The running of the latter is attributed to a system separate from the BIS, the "behavioural approach system" (BAS; Gray 1994). Like the BIS, the BAS has been submitted to three interlinked levels of behavioural, neurological, and information-processing analysis.

The input-output relations that define the BAS at the behavioural level are set out in Figure 4. In essence, this depicts a simple positive feedback system, activated by stimuli associated with reward or with the termination or omission of punishment ("relieving nonpunishment";

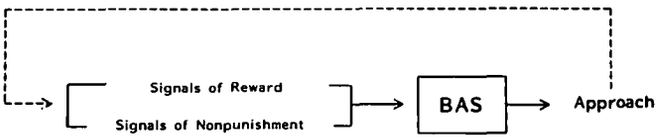

Figure 4. The behavioural approach (BAS) as defined by its inputs and outputs.

Mowrer 1960), and operating so as to increase spatiotemporal proximity to such stimuli. By adding the postulate that conditioned appetitive stimuli of this kind activate the BAS to a degree proportional in their spatiotemporal proximity to the unconditioned appetitive stimulus ("goal") with which they are associated, we have in Figure 4 a system that is in general capable of guiding the organism to the goals it needs to attain (food, water, etc.) for survival (Deutsch 1964; Gray 1975, Ch. 5; Gray 1994).

The last decade has seen rapid progress at the neurological level (Gray et al. 1991a; Groves 1983; Penney & Young 1981; Swerdlow & Koob 1987) in the construction of plausible neuropsychological models of the BAS (though this phrase has not itself been used, in the relevant literature; terms such as "motor programming system" have been preferred). The key components are the basal ganglia (the dorsal and ventral striatum, and dorsal and ventral pallidum); the dopaminergic fibres that ascend from the mesencephalon (substantia nigra and nucleus A10 in the ventral tegmental area) to innervate the basal ganglia; thalamic nuclei closely linked to the basal ganglia; and similarly neocortical areas (motor, sensorimotor, and prefrontal cortex) closely linked to the basal ganglia. These components are best seen as forming two closely interrelated subsystems, as illustrated in Figure 5 (based on Groves 1983; Penney & Young 1981; and chiefly Swerdlow & Koob 1987). The upper part of this figure shows the interrelations between nonlimbic cortex (i.e., motor, sensorimotor, and association cortices), the caudate-putamen (or dorsal-striatum), the dorsal globus pallidus, nn. ventralis anterior (VA) and ventralis lateralis (VL) of the thalamus, and the ascending dopaminergic pathway from the substantia nigra; for the sake of brevity we shall refer to this set of structures as the "caudate" motor system. Similarly, the lower part of Figure 5 shows the interrelations between the limbic cortex (i.e., prefrontal and cingulate cortices), n. accumbens (ventral striatum), the ventral globus pallidus, the dorsomedial (DM) thalamic nucleus, and the ascending dopaminergic projection from A10; for brevity, we shall call this set of structures the "accumbens" motor system. It is important to note that n. accumbens also receives projections from two major limbic structures: the subiculum (output station for the SHS) and the amygdala. The overall system and its links to the BIS are shown in Figure 6.

At the information-processing level, the key proposal (Gray et al. 1991a) is that a particular set of neurons firing at a particular time in the basal ganglia: (1) represents a step in a goal-directed motor program; and (2) is selected for this function by instrumental reinforcement mediated by the connectivity of the neurons that make up the set (Rolls & Williams 1987). Within this overall function, the role played by the caudate subsystem is that of encoding the specific content (in terms of relationships between stimuli, responses, and reinforcement) of successive steps in the program. The complementary role played by the accum-

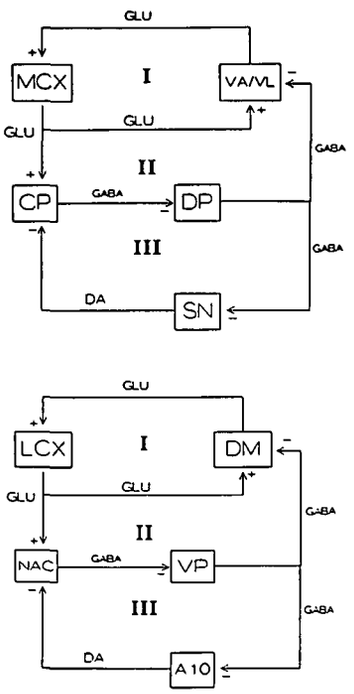

Figure 5. *Above* The caudate motor system: non-limbic cortico-striato-pallido-thalamic-midbrain circuitry. MCX: motor and sensorimotor cortex. VA/VL: ventral anterior and ventrolateral thalamic nuclei. CP: caudate-putamen (dorsal striatum). DP: dorsal pallidum. SN: Substantia nigra.
Below The accumbens motor system: limbic cortico-striato-pallido-thalamic-midbrain circuitry. LCX: limbic cortex, including prefrontal and cingulate areas. DM: dorsomedial thalamic nucleus: NAC: nucleus accumbens (ventral striatum). VP: ventral pallidum. A 10: dopaminergic nucleus A 10 in the ventral tegmental area.
GLU, CABA and DA: the neurotramsmitters, glutamate, gamma-aminobutyric acid and dopamine. + −: excitation and inhibition. I, II, III: feedback loops, the first two positive, the third negative. Based on Swerdlow and Koob (1987).

bens subsystem is that of (1) switching between steps in the motor program; and (2), in interaction with the SHS, monitoring the smooth running of the motor program in terms of progress towards the intended goal. In more detail, the specific hypotheses are as follows.

1. The caudate system, by way of its connections with sensory and motor cortices, encodes the specific sensorimotor content of each step in a motor program (e.g., for a rat, turn left at a junction in a maze).

2. The accumbens system operates in tandem with the caudate system so as to permit switching from one step to the next in a program. In addition, there are outputs to exploratory behavioural systems (via the superior colliculus and mesencephalic locomotor region) that are activated in response to novel stimuli.

3. Both the establishment of the sequence of steps that makes up a given motor program and the subsequent orderly (i.e., goal-directed) running of the program are guided by the projection to n. accumbens from the amygdala; this projection conveys information concerning cue-reinforcement associations (Rolls & Williams 1987).

4. The septohippocampal system is responsible for checking whether the outcome of a particular motor step matches the expected outcome; this information is transmitted to n. accumbens by the projection from the subiculum. For "match" a topographically localised signal is sent to the accumbens, permitting switching to the next step in the motor program. For "mismatch" a generalised subicular input to accumbens interrupts all accumbens function relating to motor steps, enabling instead the activation of exploratory behaviour outputs.

5. The activities of the caudate, accumbens, and septohippocampal systems are coordinated and kept in step with one another by the prefrontal cortex, acting by way of its interconnections, respectively, with (a) the cortical components of the caudate system, (b) n. accumbens, dorsomedial thalamus, and amygdala and (c) the entorhinal and cingulate cortices.

6. Timing is coordinated between the septohippocampal monitoring system and the basal ganglia motor programming system; given the assumption that time is quantized in the SHS by the theta rhythm (Gray 1982a), corresponding to an "instant" of about a tenth of a second, this must also be the duration of a step in the motor program.

The structure of this part of the overall model is obviously similar to the part outlined in section 1. There is therefore no need to repeat points made already concerning the light, if any, that the model is able in and of itself to throw upon the relationship between brain-and-behaviour and consciousness. As we have seen, the fundamental questions raised by this relationship remain opaque to such models. However, applying the model to the cognitive abnormalities of acute schizophrenia throws this opacity into particularly stark relief (Gray 1993).

The theory of the neuropsychology of schizophrenia that my colleagues and I have developed (Gray et al. 1991a; 1991b; Schmajuk et al., in preparation) is intended to span the complete range of explanation from a malfunction in the brain to the psychological symptoms of the condition (Fig. 7). It integrates four levels of description: (1) a structural abnormality in the brain (specifically, in the limbic forebrain, affecting the hippocampal formation, amygdala, and temporal and frontal neocortex) gives rise to (2) a functional neurochemical abnormality in the brain (specifically, relative hyperactivity of transmission in the ascending mesolimbic dopaminergic pathway); this in turn (3) disrupts a cognitive process (specifically, the integration of past regularities of experience with current stimulus recognition, learning and action), and so produces (4) the positive symptoms characteristic of acute schizophrenic psychosis (Fig. 8). Notice that, if the explicandum (step 4) in this chain were susceptible of definition in ordinary biological terms (e.g., a failure in thermoregulation), this would be a familiar type of integrative neuroscientific explanation and would pose no theoretical or philosophical problems.

Unlike thermoregularion, however, most positive (Crow 1980) symptoms of schizophrenia (e.g., auditory hallucinations, delusional beliefs, enhanced sensory awareness, difficulties in the focussing of attention, etc.; see Freedman, 1974, for a distillation of autobiographical accounts) are necessarily linked to conscious experience. So the question arises: how do we get from the first, clearly neuroscientific steps (1 and 2) in this chain of explanation to the fourth, which apparently belongs to a different universe of discourse? The answer lies in step 3. This utilises the same deliberate ambiguity that we noted at the end of the previous section: a "weakening of the influence of stored memories of regularities of previous input on current

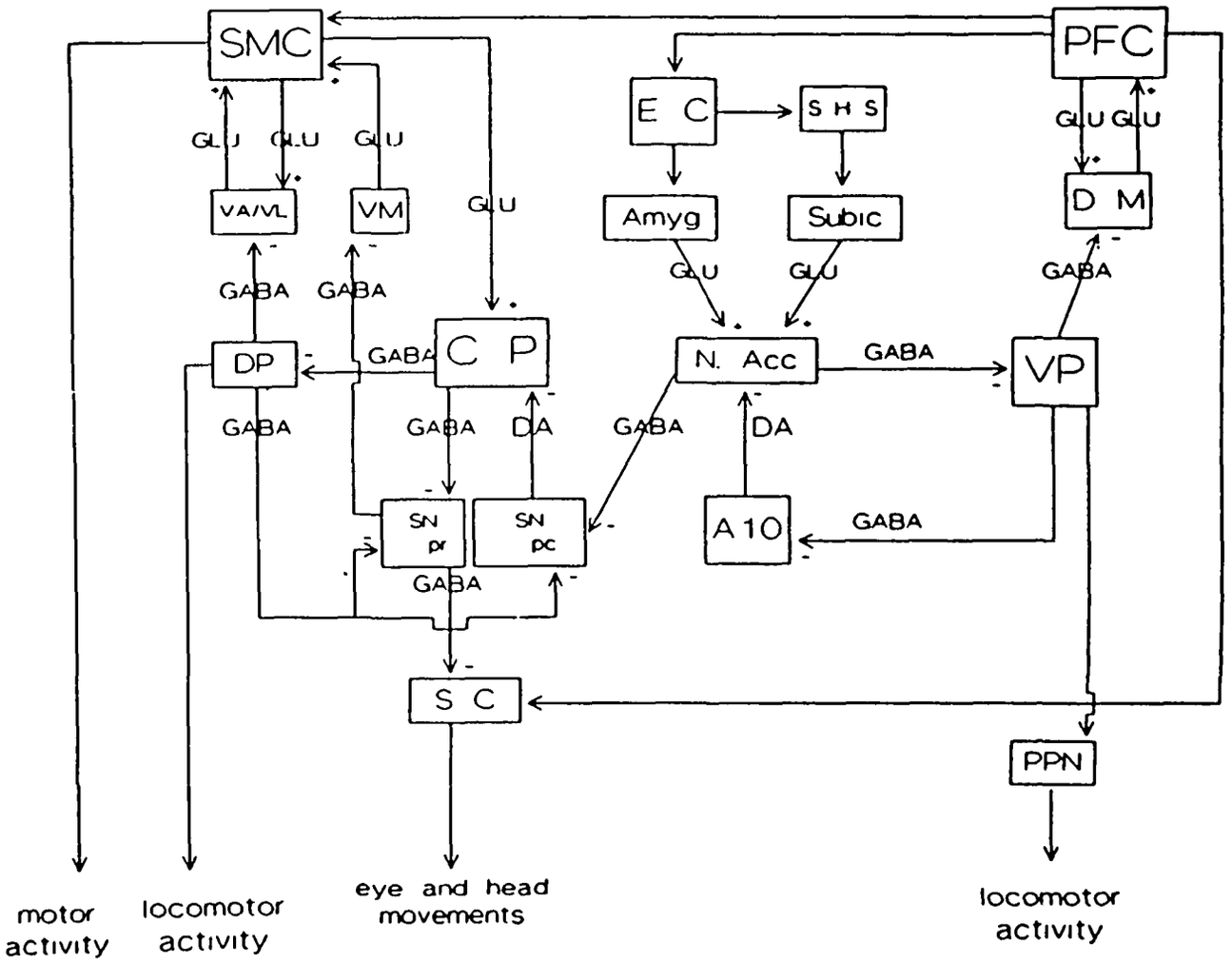

Figure 6. The basal ganglia and their connections with the limbic system. Structures: SMC = sensorimotor cortex; PFC = prefrontal cortex; EC = entorhinal cortex; SHS = Septohippocampal system; Subic = subicular area; Amyg = amygdala; VA/VL = nucleus (N.) ventralis anterior and ventralis lateralis thalami; VM = N. ventralis medialis thalami; DM = dorsalis medialis thalami; DP = dorsal pallidum; VP = ventral pallidum; CP = caudate-putamen; N. Acc = N. accumbens; SNpr = substantia nigra, pars reticulata; SNpc = substantia nigra, pars compacta; A 10 = A 10 in ventral tegmental area; SC = superior colliculus; PPN = pedunculopantine nucleus. Transmitters: GLU = glutamate; DA = dopamine; GABA = gamma-aminobutyric acid. From Gray et al. (1991a).

perception," proposed by Hemsley (1987) to underlie the positive symptoms of schizophrenia, describes a process that can be treated either as conscious or not; and this ambiguity is deliberately left unresolved (Gray 1993). This failure to disambiguate is not due to oversight or insufficient analysis; it is a general, and probably (given current understanding) inevitable, feature of all similar contemporary attempts to apply neuroscientific concepts to the explanation of conscious phenomena.

Nor are such theories rendered useless by their inherent ambiguity. On the contrary, it has, for example, been possible to derive from our model of schizophrenia and the earlier hypotheses (Solomon et al. 1981; Weiner et al. 1981) of which this model is an elaboration, a variety of detailed predictions, and to test them successfully in experiments with both animal and human subjects, utilising both behavioural and neuroscientific data. Much of the evidence in support of the theory derives from two key behavioural paradigms, latent inhibition and the Kamin blocking effect, both first studied in experimental animals and also demonstrable with human subjects. Here, I shall consider only latent inhibition (Lubow 1973; 1989). This is an extremely simple phenomenon. If a potential conditioned stimulus (CS) (e.g., a tone or white noise) is presented to the subject a number of times (typically, 20–40) without consequence, this preexposure retards subsequent learning when the stimulus does have consequence (e.g., it now predicts the occurrence of a second stimulus: footshock if the subject is a rat, an increment on a counter for human subjects). In Hemsley's (1987) terms, a past regularity (CS with no consequence) adversely affects the current learning of a new association. If, therefore, one were unable normally to integrate current learning with such past regularities, latent inhibition should be blocked, that is, learning with a preexposed CS should resemble learning to a novel CS.

A number of experimental manipulations have been shown to influence latent inhibition in ways predicted by the theory (references in Gray et al. 1991a unless others given). Thus, the indirect dopamine agonist and psychotomimetic, amphetamine, blocks latent inhibition in both rats and human volunteers, showing in both cases an inverse dependence upon dose (N. S. Gray et al. 1992a); this effect is reversed by dopamine receptor antagonists with anti-psychotic action. Given on their own, such anti-

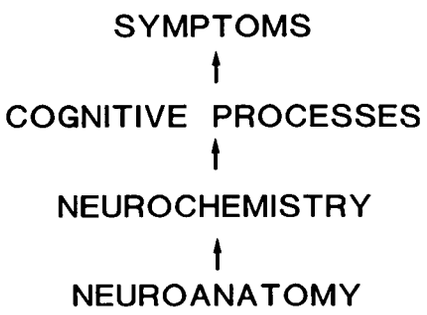

Figure 7. An integrative theory (Gray et al. 1991a) of positive schizophrenic symptoms (top), seen as arising from a structural abnormality in the brain (bottom), which gives rise to a functional neurochemical abnormality, and hence to an abnormality in cognitive processing.

psychotic drugs increase latent inhibition in both animal (Dunn et al. 1993) and human (Williams et al. 1994) subjects. The Gray et al. (1991a; 1991b) model attributes the effect of amphetamine on latent inhibition to the release of dopamine specifically in n. accumbens (Fig. 8); intense dopaminergic activity in this structure is thought to interrupt motor programming and substitute exploratory behaviour (Kelly et al. 1975). This prediction is supported by several observations. First, latent inhibition is also abolished by nicotine (Joseph et al. 1993), a drug which causes dopamine release in n. accumbens but not in the caudate-putamen (Brazell et al. 1990). Second, a CS paired with footshock elicits conditioned dopamine release in n. accumbens, and this response is blocked if the CS is first preexposed, paralleling behavioural latent inhibition (Young et al. 1993). Third, amphetamine injected directly into n. accumbens, but not into the caudate-putamen, has been reported to block latent inhibition by Solomon & Staton (1982), although Killcross & Robbins (1993) failed to replicate this finding. Fourth, excitotoxic lesions of the shell of n. accumbens also block latent inhibition (Tai et al., in press). Fifth, destruction of the dopaminergic afferents to the n. accumbens (by local injection of the catecholamine-specific neurotoxin, 6-hydroxydopamine) has the same effect as systemic administration of dopamine-receptor antagonists, that is, this lesion potentiates latent inhibition (Peters et al., in preparation).

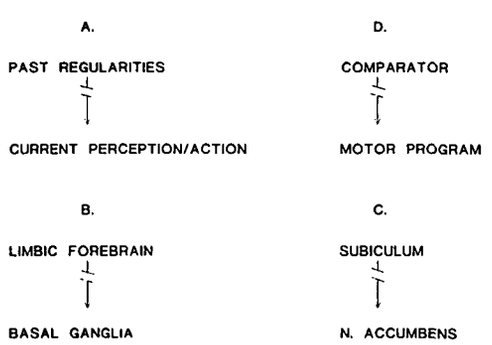

Figure 8. A schematic summary of the theory of schizophrenia proposed by Gray et al. (1991a). (A) The abnormality of cognitive processing consists of a failure to integrate past regularities of experience with the current control of perception and action. (B) This reflects a dysfunctional connection between the limbic forebrain and the basal ganglia. (C) The specific pathway carrying the dysfunctional connection is from the subiculum (in the limbic forebrain) to the nucleus accumbens (in the basal ganglia). (D) The computing functions thus disrupted are the passage of information from a comparator system, utilizing stored traces of past regularities (limbic forebrain), to a motor programming system (located in the basal ganglia) controlling perception and action.

The Gray et al. (1991a) model also predicts that latent inhibition should be affected by manipulation of the input from the hippocampal formation, via the subiculum, to n. accumbens. This prediction too is supported by several lines of observation. First, latent inhibition is abolished after large hippocampal-system lesions and after destruction of the serotonergic innervation of the hippocampus (Cassaday et al. 1993a); and it loses its normal dependence upon context after more selective hippocampal lesions (Good & Honey 1993). Second, section of the pathway connecting the subiculum to n. accumbens also blocks latent inhibition, and this effect is reversed by administering the dopamine-receptor antagonist, haloperidol, consistent with the hypothesis that sectioning the subiculo-accumbens projection affects behaviour by functionally increasing dopaminergic transmission in n. accumbens (Tarrasch et al. 1992). Third, the abolition of latent inhibition seen after large hippocampal lesions is also reversed by systemic haloperidol (Schmajuk & Christiansen, in preparation). Fourth, Cassaday et al. (1993b) were able to block latent inhibition by systemic administration of certain serotonergic agonists but not others; and Boulenguez et al. (1994) have demonstrated, after intrasubicular injection of these compounds, that those which block latent inhibition, but not those without this effect, give rise to dopamine release in n. accumbens. Fifth, work in J. N. P. Rawlins's laboratory (Yee et al. 1995) has shown that lesions in the retrohippocampal region (entorhinal cortex plus subicular area) also block latent inhibition and that this effect too is reversed by systemic haloperidol.[2]

With few exceptions (e.g., Killcross & Robbins 1993), then, these observations are consistent with the hypothesis that latent inhibition is disrupted by enhanced dopaminergic transmission in n. accumbens, whether this is primary or secondary to structural damage in the hippocampal formation. Furthermore, the limited human data suggest that in Man the same processes underlie latent inhibitions as in the rat, although of course the specific details of the experimental procedures differ widely in the two cases. Finally, and critically, as initially reported by Baruch et al. (1988a) and replicated by N. S. Gray et al. (in press), latent inhibition is absent in the early stages of an acute schizophrenic episode. Medication with antipsychotic drugs normalises latent inhibition over a period of several weeks, the time it also normally takes for these drugs to reduce psychotic symptoms. In unmedicated patients normalisation of latent inhibition is more drawn out, taking over a year (N. S. Gray et al., in press). What is not yet clear is the exact relationship between disrupted latent inhibition and the occurrence of positive psychotic symptoms. Whereas in the Baruch et al. (1988a) study this relationship was close, N. S. Gray et al. (1992b) selected chronic schizophrenics in whom, despite neuroleptic treatment, the positive-symptom score remained as high as in Baruch's acute patients; nonetheless, the acute but not the chronic patients showed loss of latent inhibition. Thus, loss of latent inhibition appears to be a more sensitive index of changes in dopaminergic transmission (as indicated also by

the ability of oral amphetamine to disrupt latent inhibition in normal volunteers; N. S. Gray et al. 1992a) than positive psychotic symptoms. This is an issue that requires further research.

It is important to note that acute schizophrenics in the preexposed condition of the latent inhibition paradigm learn the association *faster* than do preexposed normal controls. This eliminates the possibility that the aberrant cognitive performance of the patients is an artefact arising from interference from some aspect of their illness (hearing voices in the head, insufficient motivation, unwillingness to cooperate, etc.). Patient groups other than schizophrenics have not so far shown comparable changes, although data are limited to agoraphobics (Baruch 1988) and children with attention deficit disorder (Lubow & Josman 1994). However, normal individuals with high scores on questionnaire measures of schizotypy have reduced latent inhibition compared to individuals low on schizotypy (Baruch et al. 1988b; Lipp & Vaitl 1992; Lubow et al. 1992). Thus, latent inhibition is affected by both state (increased dopaminergic transmission due to drug administration in normal subjects, or during the acute psychotic period in patients) and trait (schizotypy) factors, as indeed is the occurrence of schizophrenic symptoms.

A possible inference from these experiments is that the conscious experience of the amphetamine-treated rat shares important features with that of amphetamine-treated human beings and schizophrenic patients (Gray 1993). The form of the theory these experiments have been designed to test, combined with the nature of the schizophrenic symptoms that it attempts to explain, strongly imply that the effect of amphetamine in blocking latent inhibition in human subjects should be understood as reflecting a change in conscious experience (taking the form, roughly speaking, that the drug enhances the salience in consciousness of the preexposed CS; see the descriptions, summarised by Freedman (1974), of feelings of "enhanced sensory awareness" in schizophrenics' autobiographical accounts of their illness). But similar effects require similar explanations, especially if they are encompassed by the same theory based upon the same evidence. It should therefore follow that, if the blockade of latent inhibition by amphetamine in human subjects reflects a change in conscious experience, so it does in the rat.

Since it is still disputed whether animals other than Man have conscious experiences of any kind, the construction of a body of data and theory from which such inferences may be drawn, even if they remain tentative, is no mean achievement. Nonetheless, a point made in connection with the similar discussion of the application of the model to anxiety (sect. 1) remains valid: to the extent that models of this kind rely on processes whose status – conscious or otherwise – is left ambiguous, they necessarily fail to contribute to a resolution of the substantive problem of consciousness, namely, what difference is created by some neural processes achieving consciousness while others do not?

3. A new hypothesis

Although, as we have seen, applying the comparator model to both anxiety and schizophrenia has had implications for the general study of consciousness, this was not the primary purpose for which the model was developed. In this section, in contrast, I propose a new hypothesis, based upon the comparator model, which aims explicitly to provide an account of a number of specific features of conscious experience in general. Given the extensive previous elaboration of the comparator model, outlined above and in earlier publications (especially Gray 1982a and Gray et al. 1991a), the hypothesis is (to a first approximation) simply stated: *the contents of consciousness consist of the outputs of the subicular comparator*. This hypothesis is not simply an extension of the way the model applies to anxiety or schizophrenia, nor does it seek support in the same data as did these earlier applications. Rather, it represents a new insight (remembering, of course, that insights can be false as well as true): a range of well-known features of conscious experience appear to fall out naturally from the hypothesis. In the next section we consider what these features are. First, however, the hypothesis needs to be stated in more detail.

It will be recalled (sect. 1) that the subicular area is the region, within the model, in which predicted states of the world (arriving from the Papez loop) are compared to actual (perceived) states of the world (arriving from the entorhinal cortex, part of the temporal lobe). Each such comparison (occurring on the order of once every hundred milliseconds) results in a decision of "match" or "mismatch," the two types of decision having different consequences for the outputs distributed from the subicular area to other parts of the brain. From an informational point of view, the contents of the neural messages that are submitted to the process of comparison are each equivalent to a multi-modal and highly elaborated perceptual description (e.g., "a scented red rose, in a glass vase in the corner of the room, rustling in a draft near an open door"). But these equivalences are not, as it were, written in the neural code carrying the information; they derive, rather, from the relevant sensory inputs, motor programs, potential behavioural outputs, and links to associative memory stores that jointly give rise to the neural activity that reaches the subicular area from the neocortex and from the Papez loop.

Exactly how the process of comparison is made is unknown; but one can envisage sets of neurons wired up in such a way that they constitute "and" or "or" gates, etc., and which automatically ensure that the result of the prediction initiated around the Papez loop one "instant" (one hundred milliseconds) ago enters the appropriate set of gates in the same instant as the next current perceived state of the world arriving from the entorhinal cortex. Once the match/mismatch decision is made in the subicular comparator, this decision is sent on to other structures. A match decision is passed on as a topographically discrete message to the nucleus accumbens, where it confirms the successful outcome of the last step in the motor program and permits the input to the accumbens from the amygdala to trigger the initiation of the next step in the program. A mismatch decision is passed on in two ways: as a message to the cingulate cortex, where it brings the current motor program to a halt; and as a generalised message to the nucleus accumbens, which triggers exploratory behaviour. In either case (match or mismatch), a message is transmitted back to memory stores in the temporal lobe, indicating either that a stored associative regularity has been confirmed or that memory stores need to be updated (Gray 1982a, p. 283; Gray & Rawlins 1986).

From one point of view, the outputs of this comparison process are simple binary decisions: match or mismatch. Clearly, however, the contents of consciousness contain

much more than this. Thus, the proposal is that the informational equivalences (or some portion of them) that correspond to neural activity in the subicular area (and/or in the associated circuits and regions from which that region receives, and to which it sends, neural messages) are jointly instantiated into conscious experience and that they are marked as "familiar/expected" or "novel/unexpected," depending upon the outcome of the comparison process. Now, the bracketed phrases in the previous sentence indicate several theoretical alternatives that present themselves at this juncture.

The first alternative is that conscious experience is tightly linked to activity in a highly restricted region of the brain corresponding to the point at which the binary match/mismatch decision is made (*ex hypothesi*, the subicular area). As pointed out by B. Libet (personal communication, June 17, 1993), this proposal should predict that "destruction of the subicular area should abolish the contents of consciousness." There is no evidence to support this prediction.[3] It is not clear, however, how fatal Libet's objection is. The subicular area, and even more the whole hippocampal system, has a complex spatial organisation within the human brain and, when it suffers damage, this is in consequence usually only partial. It is difficult to destroy these regions totally even in experimental animals. Supposing that we nonetheless succeed in doing so, we would still not know how to detect altered consciousness (as distinct from altered learning, memory, etc.) in animals – precisely because we lack a transparent theory of the relationship between consciousness and brain-and-behaviour.

A second alternative would involve neural feedback from the subicular area to those regions from which the activity feeding into its match/mismatch decision has been derived, such feedback occurring at the time that each binary decision is made. However, if feedback of this kind is necessary for the entry into consciousness of activity in the regions concerned, then this proposal would appear to make the same prediction (apparently false but, as noted above, neither adequately tested nor probably as yet testable) with regard to the effects of subicular damage as the first. There is a further possible objection to this alternative. If *all* the informational equivalences to which subicular activity corresponds, and the inputs that give rise to them, are jointly instantiated into consciousness, then conscious experience should contain more than is found there. Recall that the subicular comparison process requires inputs from all sensory modalities, from motor programming systems, from descriptions of potential behavioural outputs, and from associative memory stores. But, as argued persuasively and in detail by Jackendoff (1987), one is typically conscious only of the perceptual components of this conglomeration, not, for example, of the processes of motor programming or memory guided processes of selection or semantic interpretation. Thus, one would need to suppose that only some of the systems linked to the subicular comparison process receive what one might call "consciousness-instigating" feedback – a qualification that would appear to weaken this second alternative.

Nonetheless, it is the second alternative that I adopt. I am prompted to do so by a further feature of consciousness, pointed out by Jackendoff (1987, p. 51), the fact that "experience is sharply differentiated by modality. There is no mistaking visual awareness, for example, for auditory awareness, or either of these for tactile awareness." It would be consistent with this feature of consciousness that at least part of the neural activity that gives rise to conscious experience should remain closely tied to the different perceptual systems themselves.

However, the "disunity of awareness," as Jackendoff (1987) calls this phenomenon, must be considered together with its complement: the fact that the various contents of consciousness, though each of irreducibly different modalities, are nonetheless somehow combined into a single whole. As Searle (1993) puts it: "we never just have, for example, a pain in the elbow, a feeling of warmth, or an experience of seeing something red; we have them all occurring simultaneously as part of one unified conscious experience."

This *unity* of awareness has traditionally been one of the arguments used to support the notion that conscious experience is linked to activity in one, privileged part of the brain. The prototypical hypothesis of this kind is that of Descartes, who linked mental and spiritual life to the pineal gland. Dennett and Kinsbourne (1992; Dennett 1991) accordingly refer (disparagingly) to such a privileged locus in the brain as the "Cartesian theatre." As part of his argument against theories that invoke a Cartesian theatre (such as the one proposed here), Kinsbourne (1993, p. 45) states that "there is no area in the brain that receives inputs from all sensory sources." This statement is incorrect, however. Just such a sensory convergence occurs onto the region of the temporal cortex that projects into the hippocampal formation and in the hippocampus itself; and single-unit recording from neurons in the latter structure demonstrate responses to a wide variety of multimodal environmental features (for a review, see Eichenbaum et al. 1994; Gray 1982a). (A second part of Kinsbourne's argument against the Cartesian theatre is also incorrect: "Which representations, then, contribute to awareness? None are specially designated for this purpose. Any one is capable of doing so, should it become entrained with the dominant neuronal action pattern in the cortex" [1993, p. 46]. However, as noted above, and argued in detail by Jackendoff (1987), it is very largely *perceptual* material that enters consciousness, to the exclusion of other potential sources of material, even when these dominate behaviour, as in the case when one is executing a complex and fast-moving motor skill.)

How, then, can one accommodate both Jackendoff's point about the disunity of awareness and the equally striking phenomenal experience of unity in awareness (O'Keefe 1985, p. 90; Searle 1993)? An account capable of this reconciliation would attribute unity to activity in the hippocampal system (as discussed in more detail below), upon which all perceptual systems converge, and disunity to additional, linked activity in the perceptual systems themselves. Thus, the hypothesis can now be restated as follows: *the contents of consciousness consist of activity in the subicular comparator, together with feedback from the comparator to those sets of neurons in perceptual systems which have just provided input to the comparator in respect of the current process of comparison.*

Adoption of this second alternative of the "comparator" hypothesis gives rise to two further theoretical options. Should we regard "feedback" from the comparator to perceptual systems: (1) as merely flagging the activity present in the latter as "expected/familiar" or "unexpected/novel"; or (2) as contributing in a more nuanced

manner to the description of the perceived world that finally enters consciousness?

An advantage of choosing option (2) is that the argument pursued here then joins hands with those that have been advanced, with much success, by such "constructivist" theorists as Neisser (1976) and Jackendoff (1987). Indeed, the present approach in general shares much with Neisser's (1976) concept of "the perceptual cycle." As applied to the visual case, he states this concept as follows.

> The cognitive structures crucial for vision are the anticipatory schemata that prepare the perceiver to accept certain kinds of information rather than others and thus control the activity of looking. Because we can see only what we know how to look for, it is these schemata (together with the information actually available) that determine what will be perceived. (Neisser 1976, p. 20)

Neisser's contrast of schemata with "information actually available" corresponds, within the present approach, to that between predictions and descriptions of the current perceived state of the world, the former resulting from processing in the subicular comparator circuits, the latter from processing in cortical perceptual systems.

Jackendoff (1987) takes this type of analysis further, concentrating in particular on the question: what is the level at which informational structures enter consciousness? His closely argued answer to this question takes the form of an "intermediate-level" theory of consciousness. Roughly speaking, this theory holds that the contents of consciousness reflect informational structures that are derived from a combination (within each modality of perception) of bottom-up and top-down processing. As Jackendoff shows, normally one is not (and perhaps one never is) aware of sensation unaffected by conceptual interpretation, nor of pure conceptual structure, but only of an admixture of the two that optimizes the fit between them. In line with the intermediate-level theory (Jackendoff 1987, p. 298), "feedback" from the subicular comparator to cortical perceptual systems will be understood here to mean that this feedback (together with inputs originating at sensory surfaces) is actively used in the construction of an optimal fit between bottom-up and top-down processing. In such a constructional process, the flagging of some parts of the current informational structures as "novel" and others as "familiar" would be expected to play a critical role in indicating the currently needed adjustments between predicted and actual perceived states of the world.

Up to this point, Jackendoff's analysis (with which the present argument now converges) has been applied to modality-specific processing. He then goes on to ask (1987, p. 300):

> How is it that entities detected in multiple modalities can be experienced as unified? For instance, when I look at something and handle it at the same time, how can I experience it as the *same* object, if my awareness is disunified into visual and haptic modalities? The answer comes from the character of processing. The visual and haptic representations that support awareness of the object are each in registration with 3D model [in Marr's 1982 sense] and conceptual structures that encode the shape, identity and category of the object. If it so happens that they are in registration with the *same* 3D model and conceptual structure, then the two modalities will be understood and experienced as simultaneous manifestations of the same object.

Within the present theory, the conceptual structure that performs this unifying function is part of the information that circulates in the subicular comparator system, where it aids the making of predictions that are then fed back to modality-specific perceptual systems.

I conclude this section by considering how the new hypothesis relates to the known anatomy and physiology of the hippocampal system, and to previous hypotheses concerning the functions of this system.

As already noted, it is consistent with the hypothesis that (largely via the entorhinal cortex in the temporal lobe) the hippocampal formation both (1) receives inputs from all modalities, after extensive prior processing at all levels of the cortical sensory systems, and (2) projects back to these systems (for reviews, see Eichenbaum et al. 1994; Gray 1982a; O'Keefe & Nadel 1978). The hippocampal formation also receives afferents from the prefrontal and cingulate cortices that are capable of delivering requisite information concerning ongoing motor programs; afferents from the amygdala and from ascending monoamine systems able to deliver information concerning motivational and reinforcing events; afferents from the temporal lobe able to convey information from memory stores; and, in virtually all these cases, there are relatively direct routes by which the hippocampal formation is able to feed back to the structures that project to it (Gray 1982a). A key feature of the anatomical connections of the hippocampal formation, and the subicular region in particular, is its location on Papez's (1937) circuit: thus, efferents from the subiculum descend to the mammillary bodies and thalamus and then re-ascend to the cingulate cortex and back to the subiculum. Finally, afferents from the prefrontal cortex constitute a likely route by which, in Man, cortical language systems can influence hippocampal processing (Gray 1982a, Chs. 13 and 14).

Also consistent with the hypothesis is the fact that neurons in the hippocampus respond to a wide diversity of multimodal environmental features (for reviews, see Eichenbaum et al. 1994; Gray 1982a; O'Keefe & Nadel 1978). It is particularly striking that these firing repertoires are flexible and rapidly disposable: that is to say, individual neurons pick up on whatever regularities the experimenter chooses to put into the animal's environment, do so quickly, and can respond to one feature in one environment but to a quite different one in a second environment (for review, see Gray 1982a). These features are what one might expect of a structure with a close relationship to conscious experience. Also consistent with the hypothesis is the capacity of hippocampal neurons to respond differentially depending upon the familiarity or novelty of environmental stimuli or combinations of environmental stimuli (Gray 1982a; O'Keefe & Nadel 1978). The important role played by novelty and match/mismatch decisions in conscious experience has been stressed by many writers (Baars 1988, p. 181; Mandler 1984; for a recent experimental example, see Johnston et al. 1990).

As pointed out by Baars (1988), any analysis of novelty/familiarity requires the concept of context: "all understandable novelty exists within a relatively stable context that is not novel." A number of theories of hippocampal function have allotted to it just such a role in the analysis of context (Hirsh 1974; Winocur 1981). In O'Keefe & Nadel's (1978) theory of hippocampal function in animals, the relevant context is specifically spatial (although, in applying their theory to Man, these authors propose the more general function of "cognitive mapping"). The key role played by spatial frameworks in conscious experience is at

once apparent to introspection. The experimental evidence for a hippocampal role in the analysis of context – a role that is not confined to spatial contexts even in animals – is abundant (for reviews, see Eichenbaum et al. 1994; Gray 1982a). In general, this evidence suggests that the hippocampus is necessary for normal associations to be formed between focal stimulus events and the context in which they occur (for a recent example, see Good & Honey 1993).

In the light of Jackendoff's (1987) intermediate-level theory of consciousness, a hippocampal function of this kind achieves particular significance. As noted above, his analysis shows that the perceptual contents of consciousness are constructs derived from an interaction between bottom-up sensory processing, on the one hand, and top-down conceptual processing, on the other, neither of which enters consciousness on its own. Both Jackendoff and Baars (1988, p. 227) give as an example of this kind of interaction the well-known tip-of-the-tongue phenomenon, in which the context primes a well-defined gap which, however, lacks clearly defined conscious properties until it is filled by a word recognised as being "correct." Putting this kind of analysis together with the neuropsychological argument pursued here, we may propose that Jackendoff's level of conceptual processing is identical to the contextual analysis that has been attributed to the hippocampal system (Hirsh 1974; O'Keefe & Nadel 1978; Winocur 1981); that the specification of context (and therefore, conceptual structure) forms part of the prediction of the next perceived state of the world computed by the Papez loop for entry into the subicular comparator; that this is brought together with current multimodal perceptual information in the comparator; and that, finally, the outputs of this comparison are fed back to the perceptual systems in the manner discussed above.

4. The level of analysis occupied by the new hypothesis

Note that the only events initially seen as constituting the various processes executed by the comparator system are trains of neural impulses, synaptic transmissions, and the like. The informational language that may also be used to describe these events does not necessarily imply that there are additional events beyond the neural ones; it simply constitutes an alternative, and heuristically useful, way of talking about the same neural events. (For a discussion of this issue, see Searle 1993 and Marsh 1993, pp. 162–64.) An example from a different domain may reinforce this point.

The genetic code, made up of sequences of the four DNA bases, may be treated as having both syntax (including, e.g., segmentation into triplets of bases) and semantics (the aminoacids that are specified by particular triplets, and the peptides, etc., specified by higher-order combinations of amino acids). No one supposes, however, that strings of DNA bases consist of more than one level of physical reality – the syntax and semantics do not constitute levels *additional* to the physicochemical level. But this is not to say that talk of syntax and semantics in this case is *just* a way of talking. On the contrary: making use of the syntax and semantics to "read" the genetic code delivers rich insights into biological reality that purely chemical analysis cannot provide. There are several features of this example which appear to have parallels in the action of the brain.

First, the applicability of *both* the physicochemical *and* the syntactic levels of discourse to the genetic code is possible because the laws of physics and chemistry can be satisfied while still leaving open different combinatorial options (Polanyi & Prosch 1975). Thus, each correct syntactic structure is instantiated by its own specific physicochemical process, but the occurrence of that rather than another process is not fully determined only by physicochemical laws. Pylyshyn (1980; 1984) has stressed the relevance of this principle to the case of the relationship of brain function to computational processes. As he puts it: "the formal syntactic structure of particular occurrences (tokens) of symbolic expressions corresponds to real physical differences in the system, differences that affect the relevant features of the system's behavior" (Pylyshyn 1984, p. 74; and see also pp. 62–69).

Second, the presence in the genetic code of a semantic level is due to *selection by consequences*. Thus, the reason that the construction of particular amino acids and higher-order biological entities are specified by particular strings of DNA lies in the Darwinian survival value that has been conferred on the ancestors of the relevant organisms by the possession of just those entities. In the same way, it appears likely, the reason the brain generates particular sequences of neural (computational) processes lies in the ontogenetic survival value (successful adaptation to the environment) conferred on the organism by those sequences; and the semantics of such processes can be regarded as lying in the types of adaptation to which they give rise. (This principle is, of course, a major positive legacy from the behaviourist tradition in psychology.)

Third, although there is only one level of physical reality represented by strings of DNA bases, no full account can be given of the particular strings that exist by appeal only to the laws of physics and chemistry. Such an account needs also to appeal to biological laws, in particular those of Darwinian survival and Mendelian genetics. Both for this reason and because, as noted above, the laws of physics and chemistry leave open options for combining DNA bases, there is a real, but un-mysterious, sense in which genetics cannot be reduced to chemistry (Polanyi & Prosch 1975). The parallel case in respect of brain function can be stated particularly clearly by using the example of language.

> Anything that is going on in the brain to produce syntax or semantics is part of neurophysiology; there are nerve cells doing all the things that nerve cells have to do. Then, however, one must ask: why are the nerve cells doing those things and not others? The answer to that question is *not* in terms of neurophysiology, it is in terms of constraints that are required for communication between individuals, because that's what speech is all about. The constraints on speech between individuals include a level of syntax, because without syntax you don't have the informational combinatorial capacities that you need for language, and a level of semantics, because without that you don't have shared referents. Neither the semantic nor the syntactic properties that are necessary are properties of neurophysiological events; they are properties of the communication system. (Gray in Marsh 1993, p. 163; and, for a more general discussion, see Pylyshyn 1984, pp. 62–69)

Up until this point, then, nothing has been said that requires the postulation of other than physicochemical events taking place in and between nerve cells; even though a full account of why these events take the specific form they do requires biological laws as well as those of physics and chemistry. The hypothesis proposed here, however,

goes one stage further to consider the level of conscious experience. As noted in section 3, according to this hypothesis, the successive contents of consciousness consist in the successive results of the match/mismatch process that takes place once per 100-millisecond instant in the subiculum, followed by feedback to the perceptual systems whose outputs have contributed to that process. It is only at this point in our theory construction that we postulate a second set of events – conscious events – as occurring besides, and in some as yet unknown way linked to the neural events that constitute the subicular comparison process. Note that, as argued above, it appears possible in principle to give a perfectly good materialist account of brain events, whether considered under a physicochemical or a syntactic (computational) or semantic description (Pylyshyn 1980; 1984), without recourse to the *hypothesis* of consciousness. Thus, the main reason for bringing consciousness into our account is the same as Hilary's reason for climbing Mount Everest: because it is there (as a *datum*).

In a penetrating discussion of the appropriate levels at which one should describe brain events, Pylyshyn (1980; 1984) introduces a further useful concept: that of "cognitive penetrability." This term is used to distinguish between "phenomena that can be explained functionally and those that must be explained by appeal to semantically interpreted representations"; the former are said to belong to the brain's "functional architecture", whereas the latter are "cognitive processes" *sensu stricto* (Pylyshyn 1984, p. 130).

The hallmark of "cognitively penetrable" – commonly also described as "intentional" (Searle 1983) – processes is that "the relation between environmental events and behavior can be radically, yet systematically, varied by a wide range of conditions that need have no more in common than that they provide certain information" (Pylyshyn 1980, p. 120). Another way of putting this distinction is to contrast a representational system for which the *semantically interpreted content* of representations plays a causal role in the unfolding of events, with one for which such representational content does not play a causal role. Clearly, while it seems legitimate to interpret portion *X* of the genetic code as representing the particular protein, *A*, for which it codes, and while it is also legitimate to suppose that the value to organisms of possessing protein *A* has played a causal role in determining the existence in the genome of portion *X* of the genetic code, it makes no sense at all to suppose that a semantically interpreted representation by *X* of *A* plays a causal role in the production by the cell of the protein. It would seem, therefore, that the existence of cognitively penetrable processes (which can hardly be in doubt, at least for our own species; Pylyshyn 1984) requires a level of analysis that goes beyond those considered earlier in this section.

The question therefore arises: is this further level identical to the level at which conscious events enter the system? Pylyshyn's (1984, p. 265) answer to this question is explicit: the distinction between functional architecture and cognitively penetrable processes "cuts across the conscious-unconscious distinction." However, he offers little support for this assertion.

Contrary to Pylyshyn's view, it would seem parsimonious to suppose that, beyond the level of neural processes (capable of analysis in neurophysiological, syntactic, and semantic terms as indicated above, but nonetheless consisting of only one set of events), there is (at most) only one other level that need be taken into account. We know that the level of conscious events exists (while leaving open the question whether these will eventually turn out to be identical to the neural events with which they are associated, as supposed by mind-brain identity theorists; Borst 1970; Gray 1971). Thus, on grounds of parsimony we should, so long as we can, suppose that the level of intentionality (cognitive penetrability) is the same as that of conscious experience.

This assumption is aided, in the case of the argument pursued here, by the fact that according to the new hypothesis the contents of consciousness have exactly the kinds of property required by an intentional analysis: they consist of multimodal perceptual descriptions derived by comparison between predicted and actual perceived states of the world. Thus, they are constructs whose relationship to states in the real environment is indirect, and hence capable of referential slippage, although, of course, such constructs must in general bear a reasonably close relation to states in the real environment, since they derive from a process of selection by consequences in the manner outlined above. Indeed, one way of considering the new hypothesis is as providing an explanation for the existence of intentionality (a suggestion made also by Emrich 1992). If that (tentative) claim were to turn out to be correct, the theory developed here could be seen *both* as constituting a possible account of some of the features of consciousness-as-datum, *and* as proposing a hypothesis about the nature of conscious events able to explain certain features of behaviour (namely, those grouped by Pylyshyn under the rubric of "cognitive penetrability").

5. Testing the hypothesis against features of conscious experience

This section considers a number of features of conscious experience which appear to find a natural explanation in the light of the new hypothesis and some other features which do not fit the hypothesis so well.

5.1. Conscious experience is closely linked to current action plans

This feature of consciousness flows naturally from the model: the current motor program forms part of the data used to construct the next prediction; and the output of the match/mismatch decision is fed back to the motor program either to permit continuation or to bring it to a halt (Gray et al. 1991a; Gray 1994). [See also Jeannerod: "The Representing Brain" *BBS* 17(2) 1994.]

5.2. One is conscious of the outputs of motor programs, not of the program itself

It is a fact of common experience that we are not normally aware of either the planning or the execution of movements as such, but only of the end points (which may include kinaesthetic and proprioceptive feedback making up the "feel" of the movement) that constitute the successive subgoals of these movements (cf. Jackendoff 1987, p. 45; Lashley 1956; O'Keefe 1985, p. 69). Consider, for example, running towards and kicking a football; or articulating and speaking a sentence. As Velmans (in Marsh 1993, p. 121) points out, we are aware even of what we are saying only after we have said it; and the sub-vocal speech that consti-

tutes much of the process we call "thinking" consists (for me, at least) of the hearing of words – that is, *outputs*, clearly, of a linguistic motor program – in my head. Since the postulated comparator is designed precisely to compare actual with expected outputs of motor programs, it follows naturally from the hypothesis that such outputs constitute the contents of consciousness.

5.3. Consciousness occurs too late to affect the outcomes of the processes to which it is apparently linked

Velmans (1991) has reviewed a range of human information processes that are normally accompanied by conscious awareness (the analysis and selection of stimuli, learning and memory, and the production of voluntary responses, including those requiring planning and creativity). He has marshalled evidence that, in each of these cases, the relevant conscious events follow the information processes to which they are related. Libet's (1985; 1993; Libet et al. 1991) well-known experiments on the delays between (i) neural events in certain brain regions and (ii) the transitions to conscious awareness of either sensation or volition related to these neural events give rise to the same conclusion. Such a conclusion is to be expected if a monitoring system of the kind proposed here forms the basis of the contents of consciousness. However, there is a discrepancy between the duration of the delay suggested by Libet's experiments (about 500 msec) and the delay that would be expected from the comparator model: given that time is quantised into units about 100 msec long, the average delay should be about 50 msec. A possible account of such extra-long durations would postulate an additional delay prior to the initiation of the comparison process, linked perhaps to the unusual nature of Libet's experimental stimuli.

5.4. What about pain?

The arguments of sects. 5.2 and 5.3 treat consciousness as a monitoring process (see also Weiskrantz 1988). But Humphrey (in Marsh 1993, p. 166) has objected, "what has a stab of pain got to do with monitoring?" A possible answer to this objection is that a system which is concerned with predicting what should happen next, and keeping check that motor programs are proceeding "according to plan," is one that must also have an interrupt mechanism for times when things are *not* going according to plan. Pain can perhaps be construed as just such an interrupt mechanism – and a particularly powerful one. It shares with other features of consciousness that it occurs too late actually to affect responses to the noxious stimulus that gives rise to it (the hand is withdrawn from the flame before the pain is felt). Thus the information it provides must, like other forms of mismatch, gain such functional utility as it possesses from effects on *future* repetitions of the same motor program. In this sense, therefore, pain *is* a form of monitoring: it is telling you that the motor program you have just exercised is faulty and needs modification.

5.5. Conscious events occur on a time-scale that is about two orders of magnitude slower than that of the neural events that underlie them

The expected duration of a content of consciousness on the comparator model is that of a quantised time unit, that is, about 100 milliseconds. This time unit arose from considerations of the neuronal circuitry proposed to mediate the comparator function, and especially the duration of a wave of the hippocampal theta rhythm (Gray 1982a; 1982b); but it is about the right order of magnitude for the duration of conscious events (for a summary, see Baars 1988, pp. 96–97). However, the comparator model also supposes that time is indeed *quantised* into these units, in order to permit comparison between corresponding neuronal descriptions of actual and expected events. Conscious experience, in contrast, does not appear to come in separate chunks in this way.

5.6. Relative to real time, apparent time is "smeared" in conscious experience

The discussion engendered by Libet's (1985; 1993) experiments has clarified the difference between the real-time sequencing of events and the sequencing of events in experienced time (Dennett & Kinsbourne 1992; Dennett 1991; Nagel, in Marsh 1993, pp. 144–45). The hypothesis proposed here implies that what enters consciousness will reflect the prediction made by the comparator circuitry as to the next likely perceived event, and whether that prediction is sufficiently well verified or not by the actual perceived event. It follows, therefore, from the hypothesis that experienced time will be based upon these predictions and their contents, and not directly upon the sequencing of events in real time. It is also clear that the precision with which experienced events can be dated with respect to their relative time of occurrence is much coarser (Nagel in Marsh 1993, pp. 144–45) than the precision that might be expected from the time scale of occurrence of the neuronal events that underlie consciousness. The quantisation of time into 100-millisecond instants by the comparator model implies just this result: that events occurring within a single instant have no definite time sequence with respect to each other; and that any apparent time sequence within such an instant is determined by the informational content of the inputs to the comparator which determine the contents of consciousness, rather than by the objective timing of events occurring at sensory surfaces.

5.7. Conscious experience has a spatio-temporal unity that is quite unlike the patterns of neuronal events that underlie it

There are two distinct, though probably related issues raised by this contrast. One – the so-called "binding problem" – lies at the level of neuronal events. These are distributed widely in both time (on the neuronal millisecond scale) and brain location: what holds together the set that makes up a particular content of consciousness? The other issue concerns the perceptual features that drive the firing of subsets of neurons in particular brain locations. These features each only reflect aspects (e.g., in the visual domain, colour, or motion; Zeki 1978) of the fully experienced percept: what holds the separate perceptual features together? A possible answer to both questions is suggested by the comparator model, namely, that the temporally and spatially distributed neuronal events that code the separate perceptual features contributing to an overall perceptual experience are themselves nonconscious; that they combine to cause the entry of a particular prediction into the

comparator circuits; that this prediction includes the specification of spatial and temporal coordinates referred to the outside world; and that it is this prediction which constitutes the basis of the eventual conscious experience. If this line of argument is correct, the binding problem becomes more tractable, since the relationship between neuronal events and conscious experience is determined by the conjoint inputs into just one brain region, that is, the subicular area. The outputs from this region, redistributed (see sect. 3) to the sites of origin of its own inputs, then activate the detailed features that constitute the contents of consciousness.

5.8. Conscious experience consists of a series of constructions based upon a model of some aspect of the world rather than direct reflections of inputs from the world (Oatley 1988)

Dennett & Kinsbourne (1992; Dennett 1991; Kinsbourne 1993) have recently discussed a number of perceptual phenomena which provide dramatic evidence for the essentially constructive nature of consciousness. This feature of conscious experience is clearly built into the comparator model, since each content of consciousness depends upon the predicted state of the world that enters the comparator (see sect. 3). Matching is conducted upon a limited set of attributes, and only these can enter the contents of consciousness. Furthermore, "match" decisions, clearly, can be incorrect, in terms of what is actually happening in the world, even if sensory systems have correctly described the perceptual world; but it is these incorrect "match" decisions that will constitute the contents of consciousness.

5.9. There are "multiple drafts" of conscious experience

Dennett & Kinsbourne have marshalled data (e.g., from Geldard & Sherrick's 1972 "cutaneous rabbit" experiment) indicating that initial perceptual input may give rise to alternative conscious experiences depending upon later perceptual input. In principle, phenomena of this kind are compatible with the comparator model, since it could be argued that the potential multiple drafts compete with each other nonconsciously until the victor enters the comparator circuits (i.e., provides a prediction as to the next expected state of the world), thereby constituting a basis for eventual conscious experience. However, as in Libet's experiments (see sect. 5.6), the time-scale of the cutaneous rabbit is too slow for the comparator model: determination of the contents of consciousness can take up to a second or so (Dennett & Kinsbourne 1992, p. 186) rather than the tenth of a second suggested as the maximum by the comparator model. Dennett & Kinsbourne (1992) have interpreted such data as showing the impossibility of a "Cartesian theatre," that is, a single, unified locus in the brain (even if a ramified one, as discussed in sects. 5.7 and 3) at which neuronal activity becomes conscious. In distinction to their position, the comparator model preserves the notion of a Cartesian theatre, consisting in the comparator circuitry. Thus, the phenomena cited by Dennett & Kinsbourne as supporting their position, and especially the time-scale over which these phenomena operate, constitute an important objection to the comparator model. The account of Libet's experimental results (in terms of a delay prior to the onset of the comparison process), tentatively proposed at the end of section 5.3, cannot apply in this case, since the critical interval is interposed between different stimuli that give rise to the conscious experience.

5.10. Consciousness is highly selective

This well-known aspect of conscious experience flows directly from the comparator model. This specifies (Gray 1982a; 1982b) two routes to selectivity. First, inputs to the subicular comparator from sensory systems are determined by selective processes guided by the content of the last output from the comparator and the currently active prediction; in this way, continuity is provided to link successive events related, for example, by a common motor program. Second, any event that has a previous association with punishment or nonreward or any novel event is given priority for processing over others; this selectivity is accomplished within the hippocampal formation, a critical role being played by the monoaminergic afferents to this structure from the locus coeruleus and raphe nuclei.

5.11. Neuronal events operate in parallel; consciousness operates serially

This feature of consciousness also flows directly from the model, which supposes a succession of match/mismatch decisions by the comparator.

5.12. Conscious experience is closely linked to memory, both in that current experience can be distorted by past experience and in that episodic memory is a major criterion by which we infer that a conscious experience has indeed occurred

Again, this feature of consciousness flows directly from the model. The formation of a prediction is based explicitly upon a combination of current input and stored regularities involving similar stimuli and/or responses; and the outcome of the current match/mismatch decision is used to update memory stores (Gray 1982a; 1982b; Gray & Rawlins 1986).

5.13. In cases of sensory neglect not only is there lack of consciousness of the neglected stimuli, there is also no awareness that anything is missing (Jeannerod 1987)

As pointed out by Kinsbourne (in Marsh 1993, p. 258), this can be explained if consciousness is based upon a comparator system. Suppose that the damage to the brain underlying the neglect has removed not only the capacity to detect the neglected input, but also the capacity to set up the prediction that such an input is to be expected: no prediction, no mismatch, and so no awareness of anything missing.

5.14. Novelty and familiarity

The comparator model distinguishes between match and mismatch decisions, corresponding (if the decisions are correct) to novel or familiar events. Does this distinction imply a similar distinction between two modes of experience? It is clear that novel events have privileged access to conscious experience (see sect. 5.10). But it is equally clear that these are not the only events that have access to consciousness.

Consider, for example, the case of listening to a familiar piece of music: without conscious experience, this would be a rather pointless exercise. This is a particularly interesting example: my introspections reveal a definite predictive process, in which the next expected note is often generated in my head just before it enters again through my ears. There also appears to be a connection between the occurrence of this kind of predictive process and the ability to remember the music. The first time one hears a new piece of sufficient complexity it is difficult either to grasp its musical structure or to recall it. With repetition, it becomes both more comprehensible during listening and more easily recalled. Both these changes appear to be related to the increasing capacity to predict the next event in the musical sequence. Completely unpredicted events, in contrast, while they at once gain access to consciousness, are often difficult to recall (cf. Baars 1988, p. 181). This relationship between familiarity and memorability is to be expected on the comparator model: a "match" decision automatically provides a list of attributes on which the matching has been made and which can now be confirmed in memory stores.

Thus, both match and mismatch outcomes of the comparator can enter conscious experience (each perhaps, as noted above, with slightly different properties); and mismatch outcomes appear always to do so. A difficulty arises, however, when we ask the question: do all match outcomes register in consciousness? Consider a well-practiced motor program (the usual example is driving while simultaneously carrying on a conversation; e.g., O'Keefe 1985). The comparator system must continue to operate during the execution of such a program, since any mismatch (e.g., the front offside wheel crossing over the middle of the road) is rapidly detected. Yet everyone is familiar with the sudden realisation that one has just driven several miles, apparently without having paid any conscious attention to the process of driving. Schneider and Shiffrin (1977) coined the term "automatic processing" for this kind of capacity, suggesting that it arises as a result of prolonged practice on a task. The present hypothesis has no principled way of accounting for the difference between such apparently unconscious "match" outcomes and the conscious variety that occur in the music example.

5.15. The contents of consciousness are multimodal (cf. O'Keefe 1985, p. 90)

This follows directly from the multimodal nature of the sensory information that reaches the hippocampal formation from the neocortex via the temporal lobe (see the discussion of the unity of awareness in sect. 3).

5.16. What about simple single stimuli?

"The idea that conscious contents arise only in connection with match-mismatch decisions of a comparator seems difficult to apply to simple cases of awareness, for example that of a light touch to a finger" (B. Libet, personal communication, June 17, 1993). This objection can perhaps be met by the proposal that the onset of any stimulus event immediately gives rise to a predictive process, based upon the most relevant previous experiences of similar stimulus events, as to the continuation (with or without change) of that same event. We have recently found that including a postulate along these lines is of value in formulating a mathematical model (Schmajuk et al., in preparation) of latent inhibition, a phenomenon that is central to applying the general neuropsychological model to schizophrenia (see sect. 2).

5.17. Anxiety and schizophrenia

Since the new hypothesis is an extension of previous models of anxiety and schizophrenia (as outlined in sects. 1 and 2), it cannot be counted directly in its favour that it is applicable to these conditions. It is nonetheless worth indicating just how certain features of the subjectively experienced symptoms of anxiety and psychosis fit with the hypothesis.

A well-known feature of anxiety is that anxiogenic stimuli dominate conscious experience to the exclusion of other experiences, giving rise, for example, to worry, rumination, and failure to concentrate on other matters in hand. This feature of anxiety follows naturally from the priority given to threat stimuli within the comparator system (Gray 1982a; 1982b; sect. 1, above).

The model of schizophrenia (Gray et al. 1991a; 1991b) within which the new hypothesis is embedded treats positive psychotic symptoms as arising because stimuli that ought, if processing were proceeding normally, to be treated as "expected/familiar" are in fact treated as "unexpected/novel." This argument leads to a natural account of the way in which apparently trivial stimuli are able to force themselves upon the awareness of the schizophrenic (Hemsley 1987; 1993). Disruption in the Kamin blocking effect, which, as a model of positive psychotic symptoms, has many of the same credentials as does disrupted latent inhibition (Jones et al. 1992), provides an equally natural account of the manner in which schizophrenics form spurious delusional associations (Hemsley 1993), although not of the specific contents of such delusions. Frith (1987; Frith & Done 1989) has proposed a related hypothesis, according to which the neural circuitry normally responsible for tagging actions as "willed" (in our model, the link from prefrontal cortex to entorhinal cortex) is faulty in schizophrenia, giving rise to a variety of symptoms (including verbal auditory hallucinations) in which the patient experiences his own acts as alien. A further interesting possibility emerges from the new hypothesis, as formulated here. Matussek (1952, p. 92) describes a patient who was aware of: "a lack of continuity of his perceptions in both space and over time. He saw the environment only in fragments. There was no appreciation of the whole. He saw only details against a meaningless background." This is the kind of phenomenology that one might expect from a breakdown in the capacity of the comparator system (making use of contextually derived conceptual structure) to hold in register with each other (Jackendoff 1987, p. 300; sect. 3 above) the different inputs that arrive there simultaneously.

There are also a number of more general observations in psychopathology that support a role for temporal-lobe structures in subjective experience, including, for example, links between abnormal temporal-lobe function and epileptic "auras," between lesions and disruption of episodic memory and between electrical stimulation and forced remembering (Lishman 1987). A detailed analysis of these phenomena, however, would take us too far afield.

The arguments considered in sections 5.1–5.17 all apply the new hypothesis to existing data or to generally available introspective evidence. The hypothesis would, of course, be

much more valuable if it led to new predictions open to experimental test. The fact that I am unable to generate such predictions may indicate a basic weakness in the hypothesis relative to other hypotheses that are possible in the current state of knowledge; or it may indicate a general difficulty while we lack a more general, transparent theory of consciousness (Nagel, in Marsh 1993, p. 4).

6. Limitations of the hypothesis

In this final section I wish to consider what kind of advance the new hypothesis would represent, assuming that it turns out to be essentially correct. Even though this eventuality has a low probability, the exercise is worthwhile, I believe, since it can be used to consider anew the features that a truly successful theory of consciousness will need to possess. To reiterate material in the Introduction, such a theory would need to explain: (1) how subjective experiences evolved; (2) how they confer survival value on organisms possessing them (the evolutionary questions); (3) how they arise out of brain events; and (4) how they alter behaviour (the mechanistic questions; Gray 1971, p. 251).

From the biological point of view, it is the last of these questions which is in many ways crucial. If we knew how consciousness alters the behaviour that it accompanies, we would be able to see what survival value (question 2) consciousness confers, and so how it might have evolved (question 1). Unfortunately, our hypothesis has little to offer in this respect. The fault lies, however, more in the data than in the hypothesis itself. As noted in section 5.3, these suggest that most important psychological functions are completed before consciousness has time to have anything to do about them (Velmans 1991). This feature of the data is nicely accommodated by a hypothesis that links the contents of consciousness to the outcomes of a monitoring process. But the fit between hypothesis and data still leaves a void: if consciousness is a product of Darwinian evolution, it *must* confer survival value and therefore it *must* affect behaviour.

Rather than abandon this biological perspective, we would surely wish to continue to search for such a behavioural function. One suggestion arises from the original aim of the comparator hypothesis, which was to provide a neuropsychological account of anxiety (Gray 1982a; 1982b). From that point of view, the role of the comparator was to identify motor programs as faulty (in that they lead to unpredicted events, to punishment, or to the failure to obtain reward) followed by a search for an alternative, more successful program. If consciousness is tightly linked to the operation of the comparator, as our hypothesis suggests, then perhaps it is here that we should seek its function: not in transactions with the environment as they actually happen, but in the modification of such transactions for future use. To be sure, this proposal flies in the face of a stubborn intuition that consciousness has to do with the here and now. It is difficult, for example, to suppose that the only handicap associated with the lack of full visual experience in Weiskrantz's (1986) "blindsight" cases is a failure to adapt to *future* possibilities of visual input into the scotoma. [See also Compion et al.: "Is Blindsight an Effect of Scattered Light, Spared Cortex, and Near-Threshold Vision?" *BBS* 6(3) 1983.] Nonetheless, stubborn intuitions have proved to be wrong before; so this possibility is perhaps worth following up. An alternative to the evolutionary questions (1 and 2, above) would be to start on a taxonomic quest, by classifying species, for example, in terms of the development of the particular neuronal machinery (hippocampus, subiculum etc.) posited as underlying the formation of the contents of consciousness; and, in parallel, in terms of the behaviour patterns that depend upon the integrity of these structures. Such an exercise might in turn suggest possible answers to the evolutionary questions. But note that the absence of the relevant neuronal structures could not be taken to indicate absence of consciousness (since, in other cases, it is known that quite different structures can achieve the same functions; compare, e.g., the mammalian retina with insects' compound eyes).

In these ways, then, while the hypothesis does not fare at all well by the standards of questions 1, 2, and 4, it may nonetheless have some limited heuristic value. But the question to which it was primarily addressed was the third: how does subjective experience arise out of brain events? How does it fare by this standard?

The hypothesis does, I think, go beyond brute correlation: that is, beyond the mere statement that conscious experience is related to brain events of such-and-such a kind in such-and-such a place. It does so by proposing ways in which specific features of the relevant brain events give rise to specific features of the contents of consciousness. Note, however, that the critical features of the proposed brain events from which the relevant derivations are drawn are only secondarily *neuronal* features (impulses travelling into and out from the subiculum, for example). The critical features, rather, are specified in terms of information-processing activities: prediction, comparison, match-mismatch decisions, and so on. It is from these notions that it is possible to offer an account of the fact, for example, that one is conscious of the outcomes of motor programs, not of the motor program itself (sect. 5.2). Thus it is a *neuropsychological* hypothesis that is proposed, not a purely neural one. In this way, therefore, the hypothesis is perhaps just as congenial to those, such as Dennett (1991), who seek to explain (or explain away) consciousness in functionalist (information-processing) terms as to those, such as Searle (1980; 1993), who expect that we shall one day be able to point to specific features of the physical activity of the brain as giving rise to consciousness.

It is just this neutrality between these two points of view which exposes the limitations of the hypothesis. This can be seen as setting up a three-way set of equivalences: activity in a particular neural circuit (hippocampus, subiculum, etc.) = information-processing of a particular kind (the comparator function) = generation of the contents of consciousness. But the hypothesis has nothing to say about the following questions. (i) Suppose we changed the circuitry while retaining the information processing: would the contents of consciousness remain the same? (ii) Suppose we changed the information processes performed by the circuitry: would the same neuronal activity still generate conscious contents? (iii) Suppose the answers to questions (i) and (ii) are that one must preserve *both* the neural machinery *and* the information-processing functions in order to generate conscious experience: why does this particular combination produce *any* kind of conscious experience (as distinct from the particular contents of consciousness for which the hypothesis does begin to offer a genuine account) rather than none?

By normal scientific standards, then, it would appear that

we are still a long way from having a transparent theory (Nagel, in Marsh 1993, p. 4), that is, one able to predict the occurrence of conscious events, given relevant facts about behaviour and/or brain events (whether the latter are conceived in neuronal terms, in information-processing terms, or both). Nor does it appear likely that such a theory will emerge simply from the gathering of new facts of the same kind (useful as this exercise is likely to be for the eventual construction of a successful theory). The new hypothesis proposed here does, however, take us a small step forward, in that it attempts to account for the form taken by specific features of the contents of consciousness. How can we go beyond this? If it is correct (Marsh 1993, pp. 77–78) that we are waiting for a new kind of theory, then it is likely to be as difficult to answer this question now as it would have been to think about hunting for bosons in 1900.

NOTES

1. In general terms, this enterprise is similar to one undertaken previously by Edelman (1989). In more particular terms, it resembles most closely the model of consciousness proposed by O'Keefe (1985). [See also multiple book review of O'Keefe & Nadel's *The Hippocampus As A Cognitive Map BBS* 2(4) 1979.] This model too attributes central roles to the hippocampus, the hippocampal theta rhythm, and match/mismatch decisions. Despite this family resemblance, however, the two models have been developed independently and differ in many details.

2. Rawlins's work also shows that there is a projection to n. accumbens from the entorhinal cortex as well as from the subiculum, and that, for blockade of latent inhibition, destruction of the subicular region alone is insufficient. Thus it may be necessary to reformulate the critical output station from the hippocampal formation to the accumbens system as the "retrohippocampal" rather than the "subicular" area. This change of detail does not affect the arguments pursued here.

3. Nor, it should be added, is there evidence implicating other localised brain regions in the requisite manner. To be sure, there are regions in the mid-brain whose destruction abolishes consciousness. However, damage of this kind has much wider effects than merely eliminating the properties of subjective experience. For example, such damage also eliminates all waking behaviour, much of which appears not to depend upon conscious experience.

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Segmentalized consciousness in schizophrenia

Andrew Crider

Department of Psychology, Williams College, Williamstown, MA 01267
acrider@williams.edu

Abstract: Segmentalized consciousness in schizophrenia reflects a loss of the normal Gestalt organization and contextualization of perception. Gray's model explains such segmentalization in terms of septo-hippocampal dysfunction, which is consistent with known neuropsychological impairment in schizophrenia. However, other considerations suggest that everyday perception and its failure in schizophrenia also involve prefrontal executive mechanisms, which are only minimally elaborated by Gray.

On Gray's hypothesis, the contents of consciousness consist of multimodal Gestalts derived from a comparison between predicted and actually perceived states of the world. This hypothesis is based in part on an earlier neuropsychological formulation of schizophrenic symptom development (Gray et al. 1991a); I would accordingly like to return to the example of schizophrenia as a means of evaluating Gray's thesis.

In schizophrenia, the organized, Gestalt qualities of ordinary conscious experience are often lost or degraded, resulting in what Shakow (1963) described as *segmentalized* consciousness, or "an increased awareness of, and preoccupation with, the ordinarily disregarded details of existence . . . which normal people spontaneously forget, train themselves, or are trained rigorously to disregard." Segmentalized consciousness tends to arise in periods of delusional mood and is often a precursor to the development of primary delusional beliefs. In Sass's (1992) terms, segmentalized consciousness includes experiences of *unreality* and *fragmentation*. In *unreality*, sensory impressions are separated from their normal pragmatic and affective contexts, so that the world appears one-dimensional, static, and inauthentic. *Fragmentation* involves a splitting of normal perceptual Gestalts into component details that appear isolated and meaningless. The following account from Sechehaye (1970) complements the example cited by Gray (sect. 5, para. 22):

> In these disturbing circumstances I sensed again the atmosphere of unreality. During class, in the quiet of the work period, I heard the street noises – a trolley passing, people talking, a horse neighing, a horn sounding, each detached, immovable, separated from its source, without meaning. Around me, the other children . . . were robots or puppets, moved by an invisible mechanism . . . . On the platform, the teacher, too, talking, gesticulating, rising to write on the blackboard, was a grotesque jack-in-the-box. (p. 24)

The loss of contextualization and perceptual Gestalts in segmentalized consciousness is broadly consistent with the tendency of schizophrenia patients to substitute superficial for in-depth processing of stimuli (Anscombe 1987; Swerdlow et al. 1994). Gray's formulation permits a neuropsychological understanding of this processing failure in terms of a dysfunction of septo-hippocampal activity. Thus, *unreality* can be seen as secondary to a malfunctioning subicular comparator, which for Gray (after Hemsley 1993) produces a weakening of the influence of past experience on current perception (sect. 2, para. 13). That is to say, segmentalization in schizophrenia represents an improverishment of perceptual experience due to the loss of the significance normally supplied by access to memories of similar experiences.

Fragmentation is implicitly discussed by Gray under the rubric of the unity of awareness (sect. 3), which he regards as an achievement of comparator function as instantiated in hippocampal activity. Fragmented perception in schizophrenia can thus be attributed to a failure of the normal hippocampally mediated construction of perceptual Gestalts from disparate sensory inputs. Needless to say, there is ample evidence for impairments of hippocampal structure and function in schizophrenia (e.g., Friston et al. 1992; Suddath et al. 1990). Gray's formulation therefore provides a cogent explanation for central features of the segmentalization process in schizophrenia and, by extension, for the role of septo-hippocampal function in organized Gestalt perception.

On the other hand, the mechanisms by which Shakow's "normally disregarded details of existence" arise into consciousness are somewhat less apparent. Gray's account centers on comparator mismatch signals and the consequent activation of a behavioral inhibition system that interrupts motor programs and directs attention to the source of novelty. In schizophrenia, mismatch would hypothetically result from impaired access to stored regularities of experience upon which a prediction of actuality is normally made. In accord with Gray's formulation, segmentalized consciousness in schizophrenia is typically accompanied by a

posture of behavioral immobility and attentional fixation (Sass 1992), a posture compatible with the output of a behavioral inhibition system. The clinical observation that a return from segmentalized to more normal modes of perception can often be effected by a kind of disinhibitory engagement of the patient in familiar behavioral routines (Fleminger 1992; Sass 1992) is also clearly compatible with the relation drawn by Gray between motor programming and the contents of consciousness.

However, other considerations suggest that anomalous consciousness in schizophrenia involves a failure of a higher level executive in addition, if not as an alternative, to Gray's comparator mechanism. Normal individuals are quite capable of perceiving the details of their surroundings in a decontextualized and fragmented manner if they choose to do so by adopting certain facilitating attitudes (Sass 1992). In everyday perception, attention to such details is inhibited in favor of a focus on the Gestalt characteristics and thematic features of events (Cutting 1985; Stevens & Gold 1991). In Anscombe's (1987) analysis, segmentalized consciousness in schizophrenia betrays an inability to sustain an intentional focus to attention, which is by default captured by incidental features of the patient's environment. Anscombe refers to this condition as an "ataxia" of attention, which implies the failure of an executive properly to supervise a lower-level response selection system (e.g., Shallice 1988). In neural terms, an ataxia of attention can be related to a failure of prefrontal mechanisms to sustain goal-oriented attentional sets or to resist distraction by strong but incidental stimuli (Fuster 1989; Shallice 1988). There is also good evidence for prefrontal hypoactivity in schizophrenia, particularly on tasks requiring executive control over responding (Andreasen 1992).

Unfortunately, Gray's model is relatively silent regarding executive function, beyond the fairly cursory and unelaborated assignment of a systems coordinating role to prefrontal cortex (see also Gray et al. 1991a). The model would benefit from a fuller consideration of prefrontally mediated cognitive functions in interaction with limbic and striatal systems in the generation of conscious experience. An impairment of such higher level functions as intentionality and resistance to distraction appear to allow for the emergence of segmentalized consciousness in schizophrenia. By implication, normal conscious experience must require similar executive functions acting in concert with the motor programming and comparator processes central to Gray's model.

Overworking the hippocampus

Daniel C. Dennett

Center for Cognitive Studies, Tufts University, Medford MA 02155.
ddennett@emerald.tufts.edu

Abstract: The postulated hippocampal comparator, like any other subsystem, must rely on "syntactic" patterns in its "input," and hence could not have the extraordinary powers Gray supposes. It may play a more modest role, but it is not the place "where it all comes together" for consciousness.

Gray mistakenly thinks I have rejected the *sort* of theoretical enterprise he is undertaking, because, according to him, I think that "more data" is all that is needed to resolve all the issues. Not at all. My stalking horse was the bizarre (often pathetic) claim that no amount of empirical, "third-person point-of-view" science (data *plus theory*) could ever reduce the residue of mystery about consciousness to zero. This "New Mysterianism" (Flanagan 1991) is a school of "thought" that he should want to combat as vigorously as I have done.

I am all in favor of devising and investigating models of the sort Gray attempts, and for the reasons he gives. Gray's model, in fact, provides an excellent opportunity to clarify the major theoretical claims Kinsbourne and I have made (Dennett 1991; Dennett & Kinsbourne 1992), because it exhibits, in crisp detail, the intersection of two central misunderstandings. First, Gray has been seduced by one of the pet themes of the Mysterians: consciousness is all-or-nothing (Searle 1992, p. 83) and (hence) cannot be composed of lots of nonconscious (or quasiconscious) elements distributed in a large system. This leads Gray to overlook the possibility that a sufficiently powerful combination of merely "syntactic" (and unconscious) phenomena, like those he illustrates by DNA protein-synthesis, might account for the semantic competence that is, surely, our best hallmark of consciousness. Second, Gray is still an adherent of the "wasp-waist" (Kinsbourne 1993) or "bottleneck" (Dennett 1991) vision of consciousness as a property that gets conferred on the contents of various cerebral vehicles when they arrive somewhere special. (The idea is that arrival at this *inner* observer is straightforwardly analogous to such macroscopic events as the arrival of sound waves at the ears of whole observers.) Putting the two bad ideas together inspires Gray to try to promote his model of a subicular comparator system, which in its own right has many excellent features, into a job it could not possibly handle: the seat of consciousness.

Let us suppose that Gray has made the case for the existence, location, and general physico-chemical nature of a comparator system, straddling various perceptual (and other) pathways, and let us suppose moreover that he is right about its function of providing a "familiarity signal" (Dennett 1979, p. 101)[1] when it matches current activity with activity in a previous "moment." Now the hippocampal mechanism that accomplishes this postulated comparison task must operate on some "syntactic" principle: the brute physical pattern of activity in the perceptual stream will trigger the alarm or it won't, depending on whatever features of those local patterns the comparator is sensitive to. Such a syntactic comparison will have been designed (by an evolutionary process, as Gray says) to track or mirror whatever semantic regularities the brain must most reliably respond to. The most interesting – if tacit – claim made by Gray's model, I think, is that this comparator task is so important that this is a place in the brain to look for major convergence of the syntax and semantics of the whole system. The claim, in effect, is this: just as the protein-specifying semantics of DNA codons is the key to unlocking the mysteries of inheritance, so the perceptual-feature-specifying semantics of syntactic patterns detectable in the hippocampal comparator is a key to unlocking the mysteries of intentionality!

Now if Gray is wrong about this – if the syntactic patterns to which his hippocampal comparator is sensitive are relatively crude in what they track – then he has at best discovered a bit-player in the drama of consciousness. But even if he were right that the comparator is a major locus of semantic competence, this would still give it only a contributory role in the whole task of content discernment (and appreciation) that is the work of consciousness (agreeing with Gray that if consciousness did no work, as some mysterians suppose, then it could not have evolved).

The truly hopeless Cartesian alternative, of course, would be to postulate that the resources of an entire mind or comprehending intelligence are somehow brought to bear in this local region, rendering it capable of *directly* recognizing (appreciating, discriminating) the purely *semantic* similarities and differences in the stream. Gray seems to lapse into this view when he supposes "a second set of events – conscious events – as occurring besides, and in some as yet unknown way linked to, the neural events that constitute the subicular comparison process." He does not address the crucial question directly. He supposes that "multi-modal and highly elaborated perceptual description" together with information on other relevant topics (e.g., "information concerning motivational and reinforcing events" are "circulated" through the subicular area, but of course unless the subicular area can *understand* all the information available in its "input," its homuncular role ("seen this, done that") must be limited to syntactic comparisons. But Gray contrasts the power of the subicular comparator with the power of other cognitive machinery, claiming for it that "*semantically interpreted content* [emphasis in original] of representations plays a causal role in the unfolding of events." At that point, by concentrating all the power of mentality into his subicu-

lar subsystem, he forces it to be composed of wonder tissue; but this is a fatal misstep from which he could gracefully retreat.

"Merely syntactic" comparator operations are not nothing – they are, indeed, the very elements out of which any non-miraculous theory of consciousness or intelligence *must* be composed – but it is precisely this limited role that must prevent the subicular area (or any other neuro-anatomically restricted region) from being "the place where it all comes together" *for consciousness*. There can be bottlenecks – and we may suppose for the sake of argument that Gray has located a functionally important one – but not *appreciating* bottlenecks. And since the dimly envisaged capacity to appreciate the contents is well-nigh definitional of consciousness, one could never motivate the claim to have discovered the seat of consciousness (as opposed to the seat of some more minor functional component of mentality) without showing how arrival at that proposed seat led to or accomplished that appreciation.

Gray sees that this is a problem, and adjusts his hypothesis several times to make it at least not obvious that he couldn't solve it, but the adjustments have the effect of conceding, not rebutting, the main lines of our objections. It is not just the obvious problem that is pointed out by Libet in his communication with Gray (see sect. 3, para. 5, of target article): the privileged location of the activity necessary and sufficient for consciousness implies that "destruction of the subicular area" should obtund conscious experience, a prediction that Gray himself describes as apparently false – see also the commentary by Kinsbourne in this issue. It is the larger theoretical problem of trying to localize in space and time phenomena which by their very nature involve global competences that could not (except in ill-envisaged miracles) be the responsibility of a single module.

Might the hippocampal comparator then be the gateway into consciousness, the porter's lodge if not the Combination Room? (Gray speaks of the comparator's "feedback" as "contributing in a more nuanced manner to the description of the perceived world that finally *enters* [emphasis added] consciousness".) It might be *a* gateway, one among many, but what matters in any case for consciousness of some "message" is not the order or time of arrival at any gateway, but the way the message is eventually dealt with by subsequent processes, in whatever order.

NOTE

1. In this article I postulated just such a system as part of a purely speculative – and neuroanatomically uninformed – account of how *déjà vu* might be caused; many are the grounds for looking for not just one but many comparator systems.

Possible roles for a predictor plus comparator mechanism in human episodic recognition memory and imitative learning

Simon Dennis and Michael Humphreys

Department of Psychology, The University of Queensland, Australia, 4072.
mav@psy.uq.oz.au *and* **mh@psy.uq.oz.au**

Abstract: This commentary is divided into two parts. The first considers a possible role for Gray's predictor plus comparator mechanism in human episodic recognition memory. It draws on the computational specifications of recognition outlined in Humphreys et al. (1994) to demonstrate how the logically necessary components of recognition tasks might be mapped onto the mechanism. The second part demonstrates how the mechanism outlined by Gray might be implicated in a form of imitative learning suitable for the acquisition of complex tasks.

1. Human episodic recognition memory. Recognition memory is a phenomenon which might be expected to sit on the boundary of conscious and unconscious processing. Unlike cued and free recall, recognition decisions do not seem to require prior operations or to involve much in the way of retrieval strategies (Humphreys & Bain 1983). In conditions where recall performance is limited, the feeling that the item occurred in the specified context seems to be immediately available. According to Gray's comparator model, then, one might expect such a decision to be made in or near the subiculum.

In addition to the subjective evidence, however, the functional components of the comparator model provide a good fit to a possible computation-level specification of long term episodic recognition tasks. (Note that Rawlins 1985 and Gray 1985 suggest, on the basis of evidence from Sidman et al. 1968, that short term verbal memory is not implemented in Gray's architecture.) In list-specific item recognition (LSIR), subjects study a number of lists of words. After each list they are required to determine whether each of a set of test words occurred in that list or a prior list or not at all. The cues are the word itself and the appropriate list context. Humphreys et al. (1994) provide two alternative specifications of the LSIR task.

Alternative 1

$$\text{LSIR(LIST,P,L,M)} = \text{NotEmpty(Retrieve(Compatible(P,L),M)} \cap \text{LIST)}$$

Alternative 2

$$\text{LSIR(LIST,P,L,M)} = \text{NotEmpty(Retrieve(LIST,M)} \cap \text{Compatible(P,L))}$$

The first specification (Alternative 1) suggests that after being mapped to a central representation by the Compatible function, the presented word is used to query memory. The result of this Retrieve operation will be a set of items which should include the appropriate list context if the word appeared in that context. This set of items is intersected with the list context provided and the NotEmpty function is applied to determine whether the list context is in the retrieval set.

In contrast, the second specification (Alternative 2) uses the list context as a retrieval cue. If the word was studied in the list context it will be included in the retrieval set. The intersection and NotEmpty functions are used to determine whether the central representation of the word is in the retrieval set.

Neither functionally nor experimentally is it easy to distinguish these possibilities. The comparator architecture, however, would seem to be more compatible with the second alternative. The Generator of Predictions (GoP) prior to the presentation of a word would have access to the list context. Its prediction of the next stimulus would be equivalent to the Retrieve operation with the list cue, that is, it would predict any word that occurred in the appropriate context. When the word is actually input, the comparator could subsume the intersection and NotEmpty functions to produce an indication of match or mismatch on which the recognition response could be based. This match/mismatch signal must then be mapped onto an appropriate response (Humphreys & Dennis 1994). It is more difficult to see how the first alternative could be implemented, since it is the list contexts which are matched in that case, not the next input stimulus directly. Gray's model and the associated neuropsychological evidence, then, provide evidence for deciding an issue which evades computational, behavioural, and algorithmic analyses.

This analysis suggests two modifications to Gray's analysis. First, in general, it is not possible to predict what will come next even when there have been appropriate prior learning episodes. What will be predicted is a set of possible next states. Second, Gray's architecture can be used to implement an episodic recognition memory; however, a mechanism must exist to attach the output of the comparator to a response. Such a comparator and response attachment mechanism might be expected to confer substantial comparative advantage to the organism, since the ability to respond to familiar stimuli would aid many behaviours, including the interactions between individuals from the same group.

2. Imitative learning. The mathematical analysis of learning algorithms suggests that the learning of complex tasks, which could include language comprehension and production, can be more efficient if the learner has access to the correct response represented as a vector of targets rather than just a scalar feedback as to whether or not the response was correct (Williams 1986). In an autonomous system such as the human, this raises two related problems. First, the mapping from an intention to the appropriate motor commands must be learned. Second, the mapping from inputs (including the memorial state) to intentions must be acquired. In neither case are the correct outputs directly available.

Jordan and Rumelhart (1992) address the first of these problems by introducing the notion of a forward model (see also Jordan 1990; Jordan & Jacobs 1990) that maps from motor commands to predictions of the results of these commands in the environment (i.e., distal variables). The current *predicted* distal variables are compared against *actual* distal variables and the error information is used to adjust the parameters of the forward model. Subsequently, error information derived by comparing the *predicted* distal variables against the *intended* distal variables can be backpropagated through the forward model into the network responsible for formulating the motor commands (the control network) to ensure that these commands lead to the appropriate distal outcomes. In this way the control network can be trained without having access to the correct motor commands.

Dennis (1995) makes the point that an autonomous system only has access to variables resident in the brain or impinging upon the sensory apparatus (i.e., proximal variables). However, provided the distal variables give rise to proximal variables in an appropriate way, the proximal variables can be used to train a forward model and a control network (Dennis 1995). In Dennis's approach, instead of predicting the distal variables, the forward model predicts the next sensory input, as is the case in Gray's Generator of Predictions. The error signal generated by comparing the *actual* sensory input with the *predicted* sensory input (the output of Gray's comparator) can be backpropagated to optimise both the forward model and the control network. This mechanism gives rise to a form of imitative learning.

As an example, consider the task of learning the visual referent of a word. Suppose our system must learn to say "dog" in the presence of a dog. As a consequence of random articulation and babbling, the system forms a mapping between the articulation commands that it produces and the auditory input it receives as a consequence, that is, it learns a forward model. When the visual stimulus representing a dog appears, the system may or may not hear the auditory stimulus "dog," depending on whether an adult is present and whether they choose to respond. In a system that is unable to affect its environment, this indeterminacy would prevent the mapping from being optimised. The typical learner, however, can affect its environment by producing motor commands. If the error information is backpropagated into a control network capable of producing sounds, the system can reduce the prediction error by producing the auditory "dog" stimulus itself. That is, it can fulfil its own prophecy. The result is that the system comes to imitate other intentional agents in its environment. Dennis (1995) provides the control diagram of the imitative learning system and applies it to a simple robot arm targeting problem. The advantage of such a technique is that it relies on vector error feedback rather than a scalar reinforcement signal and therefore, at least in theory, it can be more efficient. Furthermore, the calculation of this error signal requires only proximal variables.

The two key components necessary for the implementation of imitative learning are a predictor of the next sensory input and a mechanism for comparing the prediction against the actual input. Both of these components are central to Gray's model.

3. Conclusion. The computational arguments derived from episodic recognition and imitative learning paradigms provide evidence from a very different style of analysis for the comparator and predictor mechanisms in Gray's model. While they do not answer why consciousness has been selected and there is substantial work to be done to demonstrate that such a system is implemented in the brain, they do suggest why these components may have conferred a comparative advantage.

Hunting for consciousness in the brain: What is (the name of) the game?

José-Luis Díaz

Centro de Neurobiología, Universidad Nacional Autónoma de México, Instituto Mexicano de Psiquiatría, and Cognitive Science Program, The University of Arizona, Tucson, AZ 85721. **jldiaz@ccit.arizona.edu**

Abstract: Robust theories concerning the connection between consciousness and brain function should derive not only from empirical evidence but also from a well grounded mind-body ontology. In the case of the comparator hypothesis, Gray develops his ideas relying extensively on empirical evidence, but he bounces irresolutely among logically incompatible metaphysical theses which, in turn, leads him to excessively skeptical conclusions concerning the naturalization of consciousness.

Gray is surely right that we need specific, heuristic, and plausible neuropsychological theories "linking" consciousness and brain. Moreover, he has made a commendable effort in developing an interesting and promising cognitive-informational model of consciousness (the comparator hypothesis) and a bold (but probably flawed) attempt at locating the comparator in the brain (the feedback activity between the subiculum and perceptual systems). My impression is that the location attempt is flawed, not because it rests almost exclusively on empirical data but because the theory behind it lacks a correspondingly well developed epistemology, and especially because it depicts a disconcerting mind-body ontology. As I will try to show in this commentary, in order to engender plausible theories concerning the mysterious connection between consciousness and brain activity, a well-adjusted marriage between philosophy and empirical science is not only convenient but necessary.

In the first sentence of the text, Gray's choice of the word "generates" suggests a metaphysical preference, since consciousness must be *something other* resulting from brain activity. Such an emergence notion resurfaces soon in propositions such as "consciousness is in some way a product of the brain." At this point I am a satisfied reader, because I believe there is certainly nothing (terribly) wrong with espousing an emergence notion of consciousness. In fact, despite its difficulties, emergentism is a reasonably coherent idea born in the scientific spheres of early Darwinism, developed in the layered world conception of the General System Theories, and specifically applied to the problem of consciousness in several materialist and mentalist modes (Díaz 1989; O'Connor 1994). Moreover, philosophers have developed a similar notion of their own called "supervenience," and two of them (Horgan 1993; Kim 1993) have meticulously worked out several possible interpretations of this notion, the difficult mental causality problems involved, and its relations with emergence. Thus, I assume a neural emergence account of consciousness until, soon after the opening statements, Gray selects the case of wave-particle complementarity and its elegant formalization in quantum mechanics as the best analogy of the consciousness-brain duality to guide his theory. Now, mind-brain notions based on Bohr's complementary principle are strongly connected to an ontological solution very different from that of supervenience and emergentism, namely, dual aspect, neutral monist theses (Globus 1973), where awareness and certain brain activities are considered to be either two sides of the same physical process (Lockwood 1989) or two aspects of a third psychological reality (Bohm 1986). I am sure there are promising ways to reconcile supervenience and dual aspect, by proposing, for example, that the highest emergent brain function is a psychophysical process with both neural and

mental aspects, but one soon realizes that Gray does not much care for reconciling metaphysical theories.

We seem to be back on track at the beginning of section 3, when conscious contents are conceived to be *outputs* of the subiculum neuro-cognitive comparator system, except that "output" indicates something more like physiological or perhaps psychological products (and therefore inert epiphenomena) rather than mereological emergents. Nevertheless, since Gray's theory demands that conscious contents be causally efficient in order to influence behavior, I recollect Kim's (1993) supervenience conundrum of mental causality: What should be the nature of these emergent properties so that they can be causally efficient, and why should a system evolve two efficient causes (neurophysiological and mental) of physical events? As I am pondering these questions, Gray presents a new surprise and uses the word *equivalence* when he establishes links between the neural and the subjective "levels" (sect. 1, para. 7; sect. 3, para. 2). To speak of "equivalents" is to endorse a psychophysical identity theory of sorts but surely not in this case, because the very architecture of the comparator system proposed in the first section encompasses four *levels* of analysis (behavioral, neural, cognitive, and experiential) that cannot be identical or equivalent to one another. Moreover, if there are levels, there are layers, so I had better stick to a complex system notion in which the hierarchy and emergence relationships among them needs to be carefully worked out. But this is not quite right either, because if I consider Gray's next four levels (neuroanatomical-neurochemical-cognitive processes-symptoms; Fig. 7) according to the text, the relationships between these layers are causal and no emergence or supervenience is posed. At this point one feels the urge for a clarification, but instead there are more surprises in store, because, in section 4, Gray leans toward functionalism with his genetic code analogy of the multiple instantiation thesis, combined with an apparently strong monist notion of a single level of (physicochemical) reality. This line of reasoning seems to head inexorably towards a form of property dualism so dear to cognitive scientists, but I am wrong again, because consciousness is considered in the very same section a *second level* of reality occurring besides the neural events. "Besides" faces one with psychophysical parallelism, and the argument given for such a dualism is that "we know that the level of conscious events exists" (sect. 4, para. 11). I do know that consciousness exists by simple introspection, but how do I know that it is a *level*?

One approaches the end of the target article somewhat shaken and, when Gray concludes that in the present state of affairs we are not ready to develop transparent theories about consciousness, brain, and behavior, one has the strong impression that he is suffering the disheartening consequence of a bewildering dance with several of the mutually exclusive psycho-physical theories (dashing but jealous and deceiving muses). Finally, when I read his brave and insightful idea that such a theory will not appear from the simple gathering of more empirical data, I cannot avoid concluding that if he had proceeded to assemble his metaphysics with the same zest and dedication invested in his neurophysiological models, he would have spun his limbic conjecture into a more coherent, robust, and personally satisfying neuropsychological theory.

ACKNOWLEDGMENTS

I thank the National Autonomous University of Mexico for a sabbatical fellowship and a share from grant DGAPA-IN602491.

Consciousness, memory, and the hippocampal system: What kind of connections can we make?

Howard Eichenbaum[a] and Neal J. Cohen[b]

[a]Center for Behavioral Neuroscience, State University of New York at Stony Brook, Stony Brook, NY 11794-2575; [b]Beckman Institute and Department of Psychology, University of Illinois, Urbana, IL 61801.

howard.eichenbaum@sunysb.edu *and* **njc@uiuc.edu**

Abstract: Gray's account is remarkable in its depth and scope but too little attention is paid to poor correspondences with the literature on hippocampal/subicular damage, the theta rhythm, and novelty detection. An alternative account, focusing on hippocampal involvement in organizing memories in a way that makes them accessible to conscious recollection but not in access to consciousness per se, avoids each of these limitations.

Gray is to be applauded for having the courage to take on consciousness and give it a serious and scholarly neuropsychological treatment. The scope of his account is remarkable: philosophical and cognitive considerations, the neuroanatomy and physiology of memory, attention, emotion, and motor programs, and the neuropharmacology and psychopathology of anxiety and schizophrenia. Gray's sweeping integration is too great a challenge for us to address in full in this brief commentary. Instead, our comments are confined to our shared interest in the hippocampus and the central role it plays in conscious experience.

We wish to focus on three areas in which the data on hippocampal function are inconsistent with Gray's hypothesis. Although he raises each of these problems himself, Gray dispenses with them in a way that is unsatisfying, leaving us to wonder whether any evidence can be sufficient to disprove his hypothesis. At the very least, these issues merit more serious treatment.

First, a wealth of data on the behavioral effects of hippocampal damage conflict with Gray's notion that "*the contents of consciousness consist of the outputs of the subicular comparator* (sect. 3, para. 1)." Large lesions of the hippocampal region, including most of the subiculum, have no effect on consciousness in humans and no effect on "conscious" reactions to stimuli in animals, to the extent that such reactions can be determined. Gray's obvious point that natural and experimental lesions of the hippocampus are seldom complete pales against the many experiments (including some very good ones by Gray and his colleagues) demonstrating that hippocampal damage results in profound impairments of memory but total sparing of other perceptual, motor, and cognitive capacities, including consciousness. Is Gray really rejecting the validity of neuropsychological dissociation in order to save his theory? Moreover, his assertion that we lack the ability to even detect alterations of consciousness in animals that might be produced by hippocampal/subicular damage renders the entire enterprise beyond the province of current empirical science.

Second, the duration of the hippocampal processing cycle (the theta rhythm) does not match the timing of neural events associated with consciousness (sect. 5.3). Not all is lost, however; at least the theta cycle is shorter than the period required for what Gray calls an "instant" of consciousness. But, other than that, do the data on theta put any constraints on Gray's theory, or is the theory again outside the province of empirical falsifiability?

Third, the critical novelty/mismatch processing attributed to the subicular comparator is not the only kind of event that has access to consciousness. Frankly, little of conscious experience is concerned with novelty detection. Although Gray seems to concede this point (sect. 5.14), he offers no explanation of why, then, the contents of consciousness should be so dependent on his subicular comparator.

We raise these concerns not to belittle Gray's effort, but rather to point to areas that are not well addressed by the current hypothesis but that can be accommodated within a somewhat different view of the role of the hippocampal system in conscious experience. In our own view of the hippocampal system, declarative memory (explicit remembering), and conscious experience, we have proposed (Cohen & Eichenbaum 1993; Eichenbaum et al. 1994) that the hippocampal system (including the hippocampus, subiculum, and immediately surrounding cortex) mediates the organization of memory representations in widespread neocortical regions. Our view is similar to Gray's in suggesting that the hippocampus supports comparisons among items and events. However, we part company from Gray in our view of how these comparisons are used and what role the hippocampus plays in conscious experience. He proposes that the hippocampus acts as a gateway to various behavioral and cognitive subsystems as stimuli

seek access to conscious experience. By contrast, we have suggested that the hippocampal system uses the outcomes of item comparisons to create and update networks of cortical memory representations to capture important relationships among the items, and that these cortical memory representations can be used by whatever (not yet fully specified) extra-hippocampal brain systems are involved in conscious recollection and cognitively mediated processing.

Hippocampally mediated declarative memories are, on our account, fundamentally relational, as described above, and flexible, in that they can be expressed in a variety of ways using forms of expression not necessarily used during the learning event. Thus, in humans, the brain's verbal systems can access, manipulate, and express declarative memories of even wholly nonverbal events. Conscious recollection is another very powerful example of access to and manipulations within relational networks by high-order cognitive mechanisms and systems in humans and in animals. Thus, in our model, the hippocampal system plays only an indirect role in consciousness – it organizes the database, so to speak, on which other brain systems may operate and, in so doing, it determines the structure and range of conscious recollections. Theta-paced hippocampal comparator mechanisms only come into play during new encodings, not during access to consciousness, thereby obviating the above stated concerns about the theta timing cycle and novelty detection. Also, consistent with the human and animal neuropsychological literature, the importance of the hippocampal system in conscious recollection is limited: Hippocampal damage prevents the establishment of new memory representations that are subject to conscious access but it does not prevent access to memory representations consolidated prior to the damage. Because the hippocampus has a time-limited role in the establishment and updating of cortical relational representations, the cortical networks can eventually be accessed without hippocampal support. Accordingly, in our view, the hippocampal system is neither the place where consciousness occurs nor the system critical for its occurrence. Instead, it has evolved to support a particular kind of memory (declarative memory) that, while essential for conscious recollection, is not equivalent to the "contents of consciousness."

Context and consciousness

Colin G. Ellard

Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada, N6A 5C2. **cellard@watarts.uwaterloo.ca**

Abstract: The commentary argues (1) that we cannot be sure that human consciousness has survival value and (2) that in order to understand the origins and, perhaps, the function of consciousness, we should examine the behavioural and neural precursors to consciousness in nonhumans. An example is given of research on the role of context in decisions regarding fleeing from probable predators in the Mongolian gerbil.

Gray's target article sets an agenda for the production of a transparent theory of consciousness that includes four separate but related issues: the (1) evolution, (2) survival value, (3) neural basis of consciousness, and (4) the mechanism by which consciousness alters behaviour. Gray's hypothesis relates mostly to (3) and, as he admits, sheds little light on any of the other points. I would like to offer one brief criticism of Gray's arguments and some further comments with regard to the subiculum's proposed function as a brain structure that compares predicted with perceived events.

I take issue with Gray's strong assertion that consciousness *must* affect behaviour in the sense of contributing to the Darwinian fitness of the possessor. Although it is not a very attractive possibility, one could argue that consciousness has arisen as an exaptive consequence of some other unknown property of neural tissue. If this is the case, it is no longer mysterious that conscious awareness of events appears to happen too late to have immediate behavioural consequences, or that we know nothing of the mechanism by which consciousness might alter behaviour, since no such mechanism need exist.

This criticism aside, my main interest in this article and in some of Gray's previous work was the assertion that much of the septo-hippocampal system, and the subiculum in particular, is devoted to an ongoing comparison of the perceived state of the world with the predicted state of the world. In other words, Gray argues that consciousness has arisen from a mechanism whose main function is to analyze context, and to separate the novel from the familiar.

My own work, on risk assessment and predator avoidance, illustrates another way of seeing the adaptiveness of contextual analysis. In an ongoing series of experiments, I have demonstrated that Mongolian gerbils will flee from transient, overhead visual stimuli in a laboratory setting, and that the configurational properties of the stimuli have little or no influence over the nature or the strength of such responses (Ellard & Chapman 1991). On the other hand, contextual variables, such as the familiarity of the environment in which overhead transients are presented, are of paramount importance. For example, gerbils that have habituated to repetitive presentation of a visual transient cannot be dishabituated by changing the visual properties of the stimulus, but can be dishabituated by changes in the spatial context in which the stimulus is presented (Ellard, submitted). I have suggested that gerbils (and probably many other prey animals) respond to some sensory stimuli as either threatening or innocuous on the basis of the circumstances that accompany the presentation of the stimulus, rather than on the basis of any local or configurational properties of the stimulus itself. In other words, gerbils respond to *all* stimuli as threatening or *no* stimuli as threatening, depending on their familiarity with the context in which the stimulus is presented. The obvious adaptive advantage of such an arrangement is that it pushes the time-consuming and computationally expensive problem of stimulus recognition to a point in time that actually *precedes* stimulus onset. By the time the stimulus appears, the animal is primed to treat it in an appropriate way.

From the point of view of Gray's conjecture about consciousness, my studies of predator "recognition" recommend themselves in two ways. First they may shed some light on the annoying feature of conscious awareness that it is always a few milliseconds too late to do any good. Although it is very unlikely, imagine that the same visual transients that elicit fleeing responses in gerbils also give rise to gerbil qualia. From the standpoint of the experiments that I have described, the qualia associated with the stimulus would be functionally irrelevant with respect to the fleeing response. Most of the relevant sensory context that produced the response could have been in place prior to stimulus presentation.

Second, the findings that I have described suggest another role for feedback connections to sensory areas that several authors including Gray in the target article (Humphreys 1992; Jackendoff 1987) have identified as being somehow related to consciousness. Perhaps, in addition to reinstating the stimulus events that gave rise to qualia (Humphreys 1992) or providing the modality-specificity of conscious experience (Gray, target article; Jackendoff 1987), these feedback connections serve to condition the responsiveness of sensory areas to incoming information on the basis of "top-down" influences related to context and expectations. Classical findings in neurophysiology (MacDonnell & Flynn 1966) suggest that receptive field properties of sensory neurons only one or two synapses removed from the receptor can be influenced by the motivational context in which the stimulus is presented.

Gray's hypothesis is attractive in that the neural architecture he describes has properties that suggest a relationship with self-awareness, but like all accounts of consciousness so far, he cannot offer a convincing account of the origins or the functions of consciousness. Although I remain unconvinced that consciousness has survival value, I agree with Gray that a fruitful approach to the origins of consciousness must include a consideration of the role of

brain regions in animals that are homologous with those that are thought to produce conscious experience in humans. Not only this, but careful attention to the environmental problems that an animal's nervous system has evolved to solve may shed light on aspects of animal behaviour that bear precursor relationships to human consciousness. In the case of the gerbil, processes that are analogous to those described by Gray in conscious humans are used in a novel way to reduce the computational load associated with responses in rapidly moving, overhead predators. Although gerbils need not be conscious to do this, there is little question that using such a mechanism increases their Darwinian fitness.

On seeking the mythical fountain of consciousness

Jeffrey Foss

Department of Philosophy, University of Victoria, Victoria, British Columbia, Canada V8W 3P4. **june19@uvvm.uvic.ca**

Abstract: Because consciousness has an organizational, or functional, center, Gray supposes that there must be a corresponding physical center in the brain. He proposes further that since this center generates consciousness, ablating it would eliminate consciousness, while leaving behavior intact. But the center of consciousness is simply the product of the functional linkages among sensory input, memory, inner speech, and so on, and behavior.

When still a small boy, I discovered what made the pictures in our TV, namely, the vacuum tube the repairman replaced, since with it everything was fine, while without it the sound was okay but the screen was blank. Using the same logic, but weaker evidence, Gray concludes that it is the "subicular comparator" that makes consciousness.

Empirical evidence can lead us away from the realization that it is the brain and person as a whole that is conscious, since parts of the brain can be removed or damaged and consciousness remain. Because the eyes, for example, can be removed and still leave one conscious, we cannot conceive of consciousness as being generated in the eyes, and so imagine it lodged instead deeper in the central nervous system. This sifting logic eliminates in turn the optic nerves, chiasm, lateral geniculate, and even the visual cortex. As Gray points out in his reply to Libet, only the elimination of (unspecified) mid-brain structures has the effect of removing consciousness – but sadly this also "eliminates all waking behavior." (note 3). Undaunted, Gray holds out the hope that ablation of the subicular comparator would remove consciousness, presumably while leaving behavior intact. As a bonus, this gives him respite from quick disproof, given the unlikelihood of its ablation due to its "complex spatial organisation."

Gray takes consciousness to be an all or nothing affair. But just as cutting away bits of the TV may reduce, but not eliminate, its performance, ablations of the brain or peripheral nervous system may reduce consciousness. That consciousness is reduced when one loses one's sight should be obvious enough, and yet the stubborn idea remains that the blind are, for all that, still completely conscious. Well, sure, but only as a reduced mass is still a mass. Nor, indeed, does consciousness vary simply in degree on a one-dimensional scale. For example, intensity is different from focus. Anyone who has known the pleasures of intoxicated love knows that consciousness may be intense while incoherent, a collection of bright fragments which defy discursive unification or clear memory. On the other hand, one may narrowly focus on a problem even as consciousness dwindles into sleep. Sometimes the internal voice dominates, commenting noisily on what is seen or sensed, while at other times the beauty or violence of what is seen silences the voice altogether. Usually we are somewhere in between in all of these dimensions, loosely focused on the day's business while wool-gathering on distant topics, intermittently aware of a dawning headache, but aware that, yes, the fridge has been making a strange noise, only when it is pointed out. Consciousness as we know it is the shifting intertwining of various strings of information, sometimes in subservience to a focal intention, sometimes in mere lyrical insouciance.

To his credit, Gray sees that to grasp what consciousness is one must come to grips with its organization: he offers a theory of the *contents* of consciousness, noting its "unity," its "disunity," its "level," and its "smeared" temporality, and wildly overemphasizing the comparison of expectation and observation. But things go badly awry when he tries to find physical, rather than functional, counterparts to these organizational details. Consider unity. Note, first, that not all consciousness is unified. Certainly my consciousness is poorly unified sometimes, as in the seconds after the alarm clock sounds, and I think the consciousness of small babies and hyperactive children is similarly disunified. Nevertheless, the sort of consciousness aspired to, and sometimes achieved, by some contemporary adults is well-integrated: it has a center. This center is functional, not physical. But Gray has form follow function, with the center of consciousness matched by a center in the brain: the subicular comparator. Anything of which we may be conscious must come to this area, which then generates "'consciousness-instigating' feedback." Lo, the fountain of consciousness!

However, some real centers are not physical. Consider the focal point of a lens: it is not another material thing beside the lens, nor something inside the lens – yet a lens without a focal point is just a chunk of glass. Likewise, the center of consciousness is not a spiritual thing beside the brain, nor is it some structure inside the brain. It is instead a functional consequence of the ideal organization of the various information streams available to consciousness. Your current perception of variously shaped inkstains on the page is informed by memory traces induced by past reading lessons, as well as your current reflections on what I am saying, integrated with yet other things, to yield your present state of consciousness. This integration is imperfect: there are relevant bits of information in your memory which you should recall, but do not, to mention just one flaw.

With training, skill, and luck, the integration is good enough to squeeze a sort of unity out of the many bits it encompasses – though one is often at odds with oneself, of two minds, making excuses for eating the doughnut while conjuring up images of clogged arteries to urge its shunning. One may be unable to recall driving home, and, like Gray, mistakenly judging that one did so "without having paid any conscious attention to the process of driving." Perhaps a little effort, hypnotic prompting, or sodium amytal, might jog one's memory, but even if it does not, one can take comfort in the reflection that the imperfect integration of one's consciousness of the drive home with one's present consciousness is not proof that one was unconscious of the road.

Suppose Gray's mooted test were satisfied beyond his wildest dreams, and one were able to turn the subicular comparator on or off like a light. After a behaviorally normal day at the office, discovering one had carelessly left it switched off all day, one flips it on, only to find oneself saying, 'My god! I've gone the whole day without consciousness!' Would this not be evidence of the fragmentation of one's consciousness, rather than its day-long absence?

Consciousness is for other people

Chris Frith

Wellcome Department of Cognitive Neurology, Institute of Neurology, % Cyclotron Unit, Clinical Sciences Centre, Hammersmith Hospital, London W12 0HS, England. **cfrith@cu.rpms.ac.uk**

Abstract: Gray has expanded his account of schizophrenia to explain consciousness as well. His theory explains neither phenomenon adequately because he treats individual minds (and brains) in isolation. The primary function of consciousness is to permit high level interactions with

other conscious beings. The key symptoms of schizophrenia reflect a failure of this mechanism.

1. The relevance of schizophrenia to theories of consciousness. Gray and his colleagues (Gray et al. 1990) previously presented an elaborate account of schizophrenia which attempted to link psychotic symptoms with brain function via a description at the cognitive level. This proposal is now extended to provide an account of consciousness in the normal case. I agree with Gray that studies of schizophrenia throw important light on the normal functioning of consciousness. This is because the major symptoms of schizophrenia (e.g., thought insertion and delusions of alien control) are essentially disorders of consciousness (Frith 1979).

I am not convinced, however, that Gray's account explains such symptoms. The theory can certainly explain a disorder in which the sufferer is unable to distinguish between relevant and irrelevant stimuli. This might well lead to elaborate accounts (delusions) as to why certain irrelevant events were important. But this disorder is not schizophrenia. The symptoms of schizophrenia are far more specific. In the case of auditory hallucinations, for example, the patient will hear voices that are experienced as coming from a powerful entity that is trying to monitor and control his or her behaviour (Chadwick & Birchwood 1994). For other symptoms too, (e.g., delusions of alien control, paranoid delusions, and delusions of reference) an essential component concerns the intentions and actions of other beings (see Frith 1992, Chap. 7).

Likewise, Gray's account does not capture the essence of consciousness. As stated by the author himself the theory does not explain *why* the brain should generate conscious experience. In what follows, I shall very briefly sketch a conjecture that resolves both of these problems.

2. Cognitive accounts of consciousness. Gray's proposal closely resembles a number of previous accounts (e.g., the Supervisory Attentional System of Shallice 1988, Chap. 14). These accounts contrast unconscious, automatic processes with conscious, strategic processes. For example, consciousness is said to have a "trouble shooting" role which is brought into play in novel situations when automatic processes fail. The problem with this family of accounts is that no one has yet put forward a good reason why this type of information processing requires phenomenological consciousness.

3. The contents of consciousness. These accounts of consciousness start by enumerating the contents of consciousness and then suggest that all the different contents involve one particular type of information processing. However, in order to make such a link it is crucial to contrast the things that are in consciousness with the things that are not. This side of the equation is usually omitted. Bridgeman (1992) makes a very important observation about the difference between information that is conscious and information that is not by considering the visual system. In this system (as in others) different kinds of information about the same scene are represented in different brain locations (Zeki 1978). An important classification of these different kinds of information makes a distinction between *motor* and *cognitive* vision (Bridgeman's terms). The motor representation provides an absolute egocentric calibration of visual space, while the cognitive representation is independent of egocentric space. The motor representation allows an intimate connection between vision and movement permitting grasping, for example. Information in the motor system is not conscious, while information in the cognitive system is. A double dissociation of these two systems can be observed in neurological patients (e.g., Goodale & Milner 1992).

I propose, as a general principle, that consciousness contains only those representations that are coded independently of egocentric coordinates. Thus the model of the world that is represented in our consciousness is, as far as this is possible, independent of our own point of view.

4. The purpose of consciousness. Of all the representations held in the brain, that which is coded in non-egocentric coordinates will most closely resemble that held in the brain of another. It is these representations that will best enable prediction of the behaviour of another creature in the current situation. Such a representational system can eventually develop the capacity to recognise agents (things which act under their own power, have goals, and have intentions) and predict the behaviour of these agents on the basis of the beliefs and wishes of these agents (having an intentional stance). Phenomenological consciousness is necessary for taking an intentional stance towards other agents. To do this we have to be able to treat ourselves as agents also. For this purpose, awareness of our own intentions and the various possible actions between which we appear to be choosing is very important. These abilities confer enormous advantages when interacting with other creatures. Shareable knowledge (which I equate with the contents of consciousness) is the necessary basis for the development of language and communication. In this account, the major mistake of most theories of consciousness is to try to develop an explanation in terms of an isolated organism.

5. Consciousness and schizophrenia. In this account consciousness plays a critical role in interpreting the behaviour of others in terms of wishes and intentions. It is precisely this ability that goes awry in schizophrenia. Studying the neural basis of the symptoms of schizophrenia will provide clues to the neural basis of consciousness.

Psychopathology and the discontinuity of conscious experience

David R. Hemsley

Psychology Department, Institute of Psychiatry, London, SE5 8AF, England

Abstract: It is accepted that "primary awareness" may emerge from the integration of two classes of information. It is unclear, however, why this cannot take place within the comparator rather than in conjunction with feedback to the perceptual systems. The model has plausibility in relation to the continuity of conscious experience in the normal waking state and may be extended to encompass certain aspects of the "sense of self" which are frequently disrupted in psychotic patients.

Gray's argument that "primary awareness" emerges, in an as yet unspecified way, from the integration of two (at least) classes of information – one generated from stored material and one impinging on the sense organs – appears plausible. (It is not obvious, however, how the model is to be applied to the deliberate retrieval of material into conscious awareness.) The conscious experience is thus equated with the "operation" of schemata on sensory input. The problem for the organism is to ensure that contextual influences "determine the appropriate schema, and to match the present occurrences with the frame provided for them" (Norman & Bobrow 1976, p. 119).

It is not clear why this integration, and its conscious equivalent, cannot take place within the comparator, as is implied in section 3, rather than in conjunction with feedback to the perceptual systems (sect. 3, para. 10). One can accept that the model must take account of the constructivist positions of Neisser (1976) and Jackendoff (1987), but is there any reason why the comparator could not both signal novelty and "contribute in a more nuanced manner" to awareness? All of the requisite information would appear to be available. In defence of the latter formulation, it could be argued that it is better able to deal with such phenomena as hallucinatory experiences occurring within a single modality.

Gray's target article suggestion that a function of consciousness should be sought in "the modification of such transactions for future use" (sect. 6, para. 3) is appealing. One possibility is that it acts as a guide to what is most usefully stored and employed in the generation of future predictions. "Match outcomes" (cf. sect. 5.14, para. 3) may only enter conscious experience when the reinforcement systems are activated.

A feature of consciousness emphasized by James (1890; 1892) is its essential continuity in the normal waking state. To the extent that aspects of preceding sensory input are preserved within the

"expectancy" which is generated, and which is integrated with subsequent input, the model has a certain "transparency" with respect to this feature of consciousness. James proposed that the transitive parts of conscious experience result in its "distinctive stream like attributes," and that they are composed of vague impressions of cognitive activity or "feelings of tendency." It is tempting to equate these with the unconscious expectancies of Gray's model (sect. 1, para. 13). As James pointed out, where the stream of consciousness is disrupted, psychopathology results, corresponding to such phenomena as "thought block." It is of course a malfunction of the "prediction generator" which is at the core of the cognitive model of schizophrenia (Gray et al. 1991a; 1991b; Hemsley 1993; 1994).

Gray excludes self awareness from his model (Introduction). Collicutt and Hemsley (1985), however, have emphasized James's distinction between the "empirical self," which can be the object of thought, and the "pure ego" which gives active thought its own personal identity and which may indeed be identical with the stream of thought. I have argued (Hemsley 1994) that this aspect of the sense of self corresponds to the consistent manner in which stored material operates on sensory input. James noted that we are aware of responding to even a novel stimulus in a way that has a certain familiarity. The idea that there is a fundamental disturbance in the "sense of self" in schizophrenia has played a central role in the development of theories of the disorder. On the above formulation it is implicit in the cognitive model.

In the normal subject, a novel stimulus is considered to elicit the "mismatch" signal and the associated increase in arousal and attentional allocation because no "stored regularity" is available. In schizophrenia, the model proposes that repeated "mismatch" signals result from a failure within the circuit contributing to the moment by moment predictions of subsequent sensory input. It is therefore reassuring that anxiety is prominent in the early stages of schizophrenia (Chapman 1966) and is associated with positive symptomatology (Penn et al. 1994).

Schizophrenia clearly encompasses a set of phenomena where the characteristics of normal conscious experience are particularly disrupted. However, although the Gray et al. (1991a; 1991b) model can plausibly account for a range of disturbances, it does, as the authors acknowledge, have difficulties with hallucinations, the most striking of the disturbances of the stream of conscious experience. It is unclear why repeated "mismatch" signals should result in such phenomena. It is possible that they provide a "window of opportunity" for the intrusion of material from long term storage. Gray's (1982; see multiple book review: "The Neuropsychology of Anxiety" *BBS* 5(3) 1982) model of anxiety suggested that the storage of those regularities upon which expectancies are based may be within the temporal lobes, and that the direct projection from temporal cortex to subiculum (Van Hoesen et al. 1979) may therefore be relevant. Normal input to subiculum is conceivably required to inhibit the direct flow of information from the temporal lobes. I have argued elsewhere (e.g., Hemsley 1993) that the abnormal perceptual experiences of normal subjects in conditions of unstructured sensory input (no "predictions" possible) provide a model for such phenomena.

It appears likely that the model will require considerable extension/modification to deal with the abnormalities of conscious experience characteristic of the later stages of schizophrenia. Anxiety/arousal is not normally prominent, whether indexed behaviourally or in terms of electrodermal responsiveness; and the clinical picture is frequently of such negative symptoms as "poverty of thought." As Anscombe (1987, p. 254) puts it, "less and less the subject *forms* his own impressions and more and more he is impinged upon by his environment" (my emphasis). This conceivably corresponds to a failure to generate *any* predictions, or ones so vague as to fail to elicit the "mismatch" signal. This in turn raises a further issue the model must eventually address: What factors affect the "limited set of attributes" (sect. 5.8) upon which matching is to be based and which in turn determine the content of conscious experience?

Perspective, reflection, transparent explanation, and other minds

S. L. Hurley

Department of Politics, University of Warwick, Coventry CV4 7AL, England

Abstract: Perspective and reflection (whether involving conceptual or nonconceptual content) have each been considered in some way basic to phenomenal consciousness. Each has possible evolutionary value, though neither seems sufficient for consciousness. Consider an account of consciousness in terms of the combination of perspective and reflection, its relationship to the problem of other minds, and its capacity to inherit evolutionary explanation from its components.

There are two features that have seemed to many in some way important in understanding phenomenal consciousness. These are perspective and reflection. By perspective, I mean the point of view that we attribute to creatures who are agents and perceivers, who move actively through their environment and interact with it while perceiving it. The challenge here is to say something about what it is to occupy a perspective that does not simply presuppose consciousness. By reflection, I mean that exercise of a capacity to sustain states with contents (be they conceptual or nonconceptual) that are reflexively higher-order in some way, that is, are in some way about their possessor or its contentful states, among other things. Though "reflection" may suggest conceptual content and thought, what is needed here may be reflexively higher-order content with no commitment to its conceptual character, if one wishes leave open whether creatures without concepts may be conscious. Of course, reflexive contents, along with other contents, need not be conscious; I may have an unconscious thought about my own thoughts. Without this proviso, the appeal to reflection would move in a very small circle.

I take Gray's hypothesis to relate to the challenge to say something that doesn't simply presuppose consciousness about the perspectival aspects of consciousness. There is a generic similarity among various current moves in this area, all of which involve an appeal to feedback and something like a comparator system. I describe the essential structure here as that of a complex adaptive feedback system (Hurley, work in progress). The general idea implicit in these various moves, as I see it, is this: a creature with perspective (that is, a creature that is an agent and a perceiver) keeps an ongoing record of the relationships between motor output and sensory input, and compares current motor output with current sensory input for match or mismatch with the recorded relationships. Sensory input is a function of both motor output and environment, and keeping running track of this function via multiple feedback channels and a dynamic comparator system is essential to both agency and perception, and thus to the capacity for perspective. This is an important idea, because it corrects the dominant tendency to think of sentient creatures exclusively the other way around: to regard motor output as a function of sensory input, and to consider how any cognitive or rational capacities may interface between the two.

Gray regards his hypothesis as falling short of meeting the demand for a transparent explanation of consciousness pressed by Nagel and others: while it goes beyond mere brute correlation to provide some structural understanding of consciousness, it still leaves mysterious why there should be any such thing as consciousness at all. The point could be generalized to the appeal to perspective at large: even if the above general ideas about perspective are correct, why should a creature with perspective in this general sense be conscious? Why couldn't the whole dynamic system operate without consciousness, and what makes the difference? What is the added value of consciousness, in evolutionary or other terms? Of course, there is no similar problem about why perspective understood as above might have evolved – there is plenty to say about that. It is just that perspective, it seems, might obtain without consciousness – so why should it obtain with consciousness? The essential ingredient still seems to be missing; the demand for transparent explanation is not met.

Similar points have been made about the appeal to reflexively higher-order content in explaining consciousness. In a nutshell, it seems quite possible for there to be unconscious contentful states and unconscious higher-order contentful states; so it seems possible that the entity to which we attribute reflective states not be conscious – unless, that is, we have built consciousness into the idea of reflection from the beginning. Could there not be an automaton or a computer with the capacity for reflexively higher-order contentful states? Surely it would not therefore necessarily be conscious! Again, the essential ingredient seems to be missing, and the demand for transparent explanation is not met. While we can understand why a capacity for higher-order contentful states per se might be of evolutionary value (consider false alarm calls and the like), why couldn't any such evolutionary work be done just as well without consciousness?

The above is intended to summarize very roughly points that are already familiar in the literature. The question I want to go on to raise is this: even if we grant these familiar points when they are made separately about the appeals to perspective and to reflection, what happens when we combine these two ideas? That is, suppose instead of trying to account for consciousness either in terms of perspective or in terms of reflection separately, we appeal to both at once? (Are they really as independent of one another as they first appear to be? – a deep issue, which I don't pursue here.) Consider the possibility that neither alone is sufficient for consciousness, that neither alone provides the missing ingredient, but that they are both necessary and are jointly sufficient for consciousness. A centipede or a sleepwalker may have perspective, but presumably lack reflection; a computer may have higher-order states, but it lacks the full-blooded perspective of agency and perception. The combined account doesn't rule out of consciousness creatures incapable of deploying concepts, since it allows for nonconceptual reflexive states. Thus it may meet worries, which attach to pure higher-order thought accounts, about whether higher-order thought is necessary for consciousness. (Davies & Humphreys 1993, pp. 25–26). But is the combination sufficient for consciousness? If an entity has both perspective and reflection, is there a further question about whether it is conscious?

A bad response to this question is this: How could these factors together make for consciousness, when neither does separately? This just assumes without justification a kind of separability, that is, that perspective and reflection could not together amount to something qualitatively different from what they amount to separately. A different response is more enlightening: it amounts to invoking a version of the problem of other minds. After all, philosophers are familiar with the sceptical hypothesis that, for all we know, the other human beings around us, who feature both perspective and reflection par excellence, might really be robots or automata, not conscious at all. To the extent that it is conceivable that creatures with these features lack phenomenal consciousness, appeal to these features cannot provide a fully satisfying explanation (cf. Davies & Humphreys 1993, p. 35). If satisfying both criteria doesn't quiet such worries even in the case of human beings where the reflection in question involves conceptual content, how could the combined account explain consciousness? And how could the combination then possibly suffice for consciousness in other animals, when even less is required, that is, where nonconceptual reflexive states are involved?

But in reply we should ask: Does the demand for transparent explanation, when applied to the combined appeal to perspective and conceptual reflection, amount to any more than pressing a form of the problem of other minds? Some philosophers still take this problem seriously, but others do not. At least, if the need for transparent explanation of consciousness in such cases could be connected with a form of the problem of other minds, many might find it less urgent. It is far from unintelligible that creatures with perspective and conceptual reflection are conscious; rather, it is natural to assume they are. To the degree this assumption is philosophically acceptable, the explanatory gap may be reduced: that is, the need may be reduced for transparent explanation of why creatures with these features should be conscious (though not for empirical explanation of these features themselves, as in Gray's target article). The next question would then be: In the context of a combined account, is the move from conceptual to nonconceptual reflexive states critical? A difficulty here is to leave logical space for a solution to occupy: we may want to allow the possibility that creatures without concepts may be conscious, but want also to allow the possibility that creatures with perspective and reflection are not conscious. The danger is that, for any account that doesn't simply help itself to consciousness directly, we want to insist both that creatures who fail to satisfy it may be conscious and that creatures who do satisfy it may not be. But nothing could meet this specification.

Finally, notice a byproduct of the combined appeal to perspective and reflection in explaining why there should be such a thing as consciousness at all: the combined account inherits any evolutionary work done by perspective and by reflection separately. We needn't offer an evolutionary account of why there should be consciousness per se if we already have accounts of how each of the two elements of consciousness might separately be of evolutionary value. (Of course, I have not provided such accounts, merely gestured in their direction.)

ACKNOWLEDGMENT

I am grateful to Nicholas Rawlins and Anita Avramides for discussion of these issues, and to the McDonnell-Pew Centre for Cognitive Neuroscience in Oxford for its support of interdisciplinary work between philosophy and neuroscience.

Mind – your head!

R. P. Ingvaldsen[a] and H. T. A. Whiting[b]

[a]Department of Sport Sciences, University of Trondheim, N-7055 Dragvoll, Trondheim, Norway; [b]Department of Psychology, University of York, Heslington, York, YO1 5DX, United Kingdom. **rolf.ingvaldsen@no.unit.avh** *and* **htaw1@tower.york.ac.uk**

Abstract: Gray takes an information-processing paradigm as his departure point, invoking a comparator as part of the system. He concludes that consciousness is to be found "in" the comparator but is unable to point to how the comparison takes place. Thus, the comparator turns out not to be an entity arising out of brain research per se, but out of the logic of the paradigm. In this way, Gray both reinvents dualism and remains trapped in the language game of his own model – ending up dealing with the unknowable.

1. Mind – your head! The essential and necessary element in any theory that rests upon a cybernetic way of thinking is a comparator. Robb (1972), for example, in her exposition of an information-processing approach to skill acquisition, sees the function of such a comparator to be the comparison of mental images of the intended action with the actual outcome – the difference being the information of value and the reason for further action. Gray goes beyond this basic requirement in equating the output of just such a comparator with the phenomenon of consciousness, at the same time leaving it unclear as to why the brain should generate conscious experience at all.

In an attempt to clarify his own position, Gray resurrects some of the old body-mind ghosts for our attention. In declaring Cartesian dualism to be an obsolete position, he puts forward an alternative which he explains in terms of levels of organization – consciousness being postulated as the fourth level in a comparator model that encompasses three interlinked levels of analysis: behavioural, neural, and cognitive. Gray is at pains to assure us that his scientific approach (using as an analogy the development of quantum mechanics) is not just a question of "dealing with the unknowable or a bad language habit" but represents a real step in the direction of a scientific theory of consciousness.

Given the paradigm selected by Gray, and his conviction that

"consciousness is in some way a product of the brain, a product intimately connected with the brain's role in behaviour and the processing of information," this is a position which it is very easy to understand. We are in no way convinced, however, that Gray has solved any of the fundamental problems arising from the position of Cartesian dualism (e.g., by introducing the concept of "brain events and conscious experience" or by getting rid of the language problems involved in the selected paradigm. On the contrary, we wish to claim that Gray both *reinvents* dualism in his model and remains trapped in the language-game – ending up dealing with the unknowable.

2. Dualism and/or duality. Gray wishes to claim that, in scientific terms, one's own conscious experience is a datum that "stands in need of explanation," whereas for others, consciousness only functions as an hypothesis by which to explain behaviour. In making this claim, Gray contradicts Descartes' *cogito ergo sum* in a radical way. After all, for Descartes, consciousness was the only thing that could not be doubted and it accordingly formed the cornerstone for all other truths – as such, it was not a matter for science but served as a premise on which to build logic (and science). On the other hand, Gray, like Descartes, puts forward the individual experience of consciousness as an independent datum to be trusted and insists upon consciousness influencing behaviour, that is, "steering" the physical matter. For Descartes, the connection between consciousness and the physical world was guaranteed by God; for Gray, by Darwin. We would like to make two comments in relation to this standpoint:

As Wittgenstein (1976) points out, as our conscious experience is private, we cannot know it as we do other "things." It will remain, therefore, as Gray also points out, a kind of hypothesis that can explain the behaviour of others. But this is not the crux of the problem! As Wittgenstein (1976, p. 293) points out, the root problem is that we cannot take our experience as an independent datum because there are no criteria by which it can be evaluated. An attempt can be made to rescue the situation by appealing, as does Gray, to the Darwinian logic of this datum, but to do this one must reason with someone, that is, the private datum of consciousness rests on a collective discourse, thereby questioning the original argument of consciousness as an independent datum.

Gray also insists upon consciousness as having some kind of effect on behaviour; "we" are not just observing our own experience. He also admits that there are no (neurophysiological) clues as to *how* the connection could occur, just some indication of *where* in the brain it is likely to happen. In this respect, Gray is on the same level as Descartes – they both think (hope) to know the whereabouts of consciousness. For Descartes, the problem was how to connect two kinds of "reality"; for Gray it is how to understand the relation between different levels of organization. But, as the relation between consciousness and organization on the level below are not known, this constitutes a kind of methodological dualism – arising from the levels of analysis approach implicit in the paradigm adopted. However, as the relation between the lower levels is also obscure, we would prefer to say that we are here dealing with a set of dualisms – there are several sets (levels) of data without any known connection even though they are presupposed to be interlinked, arising, as they do, from looking at the same organism. This is, as we see it, a problem inherent in any interpretation in which "levels" are given a kind of independent, ontological, status. One must look for the connections – each unknown connection giving rise to its own dualism (e.g., body vs. mind, neurophysiology vs. information processing, etc.). The problem can only be resolved by investigating the model invoked (the paradigm) and not by looking into the head. Haken (1988), from a Synergetics perspective, does make an attempt to resolve the problem by trying to find rules which would connect the different levels but this, in itself, does not remove the fundamental problem implicit in any levels of analysis approach.

3. Traps in the language of information-processing. Any theory that takes an information-processing paradigm as its departure point must invoke a comparator as part of the system, the comparator forming the link between input and output leading to a controllable system (Ingvaldsen & Whiting 1993). If one decides, like Gray, to incorporate consciousness into the controlling system one is ultimately faced with the question about the whereabouts of the seat of consciousness. From the paradigm Gray embraces, there can be only one answer – in the comparator! But, even if we locate this "seat," we are still left with the problem of explaining how the comparison takes place and how the levels are interconnected. This, as Gray admits, we do not know! From his theoretical standpoint, we only know that the result is in terms of match/mismatch decisions which make up the content of consciousness. Thus, it seems clear that the comparator is not an "entity" arising out of brain research per se but out of the logic of the paradigm. In this way, Gray remains trapped in his own language game, with the consequent risk of trying to "know" the "unknowable"!

Information synthesis in cortical areas as an important link in brain mechanisms of mind

Alexei M. Ivanitsky

Institute of Higher Nervous Activity and Neurophysiology, Russian Academy of Sciences, GSP-7, Moscow 117865, Russia. **ivanit@ivnd.msk.su**

Abstract: To explore the mechanism of sensation correlations between EP (evoked potential) component amplitude and signal detection indices (d′ and criterion) were studied. The time of sensation coincided with the peak latency of those EP components that showed a correlation with both indices. The components presumably reflected information synthesis in projection cortical neurons. A mechanism providing the synthesis process is proposed.

Gray proposes an attractive brain mechanism of consciousness. His ideas are rather close to our hypothesis about the brain mechanism of sensation based on experimental work (Ivanitsky & Strelets 1977; Ivanitsky et al. 1989). I returned to these ideas in my recent publications (Ivanitsky 1993; 1994). These data have also been summarized in two monographs published in Russian (Ivanitsky 1976; Ivanitsky et al. 1984). Since that time some important data have appeared that are likewise in good correspondence with our views.

The idea of our experiments was as follows. In one series, the subjects had to discriminate the intensity of visual stimuli and in another series that of somatosensory stimuli, responding via button press. Visual and somatosensory cortical evoked potentials in response to stimuli were recorded. The psychophysical indices of perception of the same stimuli were also measured using signal-detection theoretic methods. Thus, we have the characteristics of stimulus processing referring to two levels of events: brain events in the form of evoked potential amplitudes and mental events presented by two psychophysical indices – the detectability index d′ and the response criterion index. Correlations were calculated between these two rows of events. These correlations were about equal for visual and somatosensory modalities: the amplitude of the first EP components correlated with d′ values and the amplitude of the late EP components, especially the P300 wave, correlated with the criterion index. These correlations were quite reasonable, as first components actually reflect the sensory input to the cortical projection areas and the late components were related to the evaluation of stimulus significance. Most important was the double correlation of both the two psychophysical indices and some intermediate components of the projection cortex EPs. These components were the N140 somatosensory EP and P180 wave visual EP. It was concluded that these EP components reflected the synthesis of two kinds of information: sensory information about the stimulus physical parameters and memory information about stimulus significance.

A hypothetical brain mechanism underlying the information synthesis was proposed: There is first information flow from

projection to associative cortex, then to limbico-hippocampal area, the subcortical centers of emotion and motivation, with a subsequent return of the excitation to the cortex where this inflow was superimposed on traces of sensory excitation in the projection cortex neurons. It is noteworthy that a similar excitation run was described later by Mishkin (1993), who recorded neuronal activity in the monkey brain.

It is of special interest that the peak latencies of those evoked potential components that showed the double correlation with psychophysical indices coincided rather precisely with the sensation time measured in psychophysical experiments (Baars 1993; Blumenthal 1977; Pieron 1960). One may also note that this time is quite compatible with the duration of the human theta-rhythm cycle, which Gray considers a main timer in brain circuits underlying mental events.

The idea of the excitatory circuits as a brain basis of consciousness has also been developed by Edelman (1989). It is noteworthy that the time of 150 msec taken for the reentrant process in Edelman's scheme corresponds precisely with the visual sensation time predicted in our experiments (180 msec). This time consisted of the requisite 30 msec for nerve impulse arrival to the visual cortex and 150 msec for a reentrant cycle.

The idea of information synthesis as an important link in the brain basis of mind was later developed in our works on the brain mechanisms of the thinking process. Special cortical formations – interaction foci – have been described for the synthesis process performance in associative cortical areas (Ivanitsky 1993).

Thus, we now see that rather similar networks have been proposed by a number of scientists working in different countries and using different methods. This fact could perhaps be considered a sign that our direction is correct and we are actually approaching an understanding of the real mechanism of mind.

ACKNOWLEDGMENT

The research described in this commentary was supported in part by grant N M9Q000 from the International Science Foundation.

Septohippocampal comparator: Consciousness generator or attention feedback loop?

Marcel Kinsbourne

Center for Cognitive Studies, Tufts University, Medford, MA 02155.

Abstract: As Gray insists, his comparator model proposes a brute correlation only – of consciousness with septohippocampal output. I suggest that the comparator straddles a feedback loop that boosts the activation of *novel* representations, thus helping them feature in present or recollected experience. Such a role in organizing conscious contents would transcend correlation and help explain how consciousness emerges from brain function.

The target article embeds within an intuition that there is a Hard Question about consciousness, a model which, to the author's regret, merely addresses the Easy Question. But perhaps addressing the Easy Question will make the Hard One go away.

1. Intuitions about consciousness. Gray's intuition is that we should be able to do better in explaining consciousness than to rely upon "brute correlations between brain events and . . . conscious experiences" (Intro., para. 3). This is true. But to call for a paradigm shift to explain consciousness may be overkill. I believe he underestimates the potential for explanation that resides in studying the relationship between individual experiences and the corresponding neural states. If such work results in simplifying formulations that have predictive power, then this will have rendered conscious states no less "transparent" than any theory in physics does its target problem. The intuition that consciousness is so special that it deserves more, has, since Gray (1971), been enthusiastically restated by well-known others. This intuitive meme at least motivates its carrier to strenuous efforts to stay alive (and while alive, to spread the meme). However, intuitions are equal opportunity enterprises, at the mercy of conflicting *intuitions,* such as my own. In any case, major discoveries are necessarily counterintuitive and opaque to common sense. Otherwise, they would already have been discovered.

2. The comparator. Being conscious is what it is like for a brain to be in any one of a set of functional states. How does it achieve such functional states? Gray suggests a brute correlation between the subicular output of the septohippocampal (SH) system and being conscious, but not how this comes about or how it relates to the economy of the brain. Can we extend this suggestion beyond correlation so as to align it with a functional model? If cell assemblies entrain the representations they embody into the totality of what the person is conscious of at the time, how might the comparator influence these cell assemblies?

3. What does the comparator compare? Gray's model embodies Sokolov's suggestion that the brain compares observed input to anticipated input, a mismatch indicating novelty, in the sense of unexpectedness. Since we do not first experience an event and then its unexpectedness, the match/mismatch must be preconscious. Gray plausibly concludes (contra Searle) that fully elaborated and meaningful but as yet unconscious percepts queue up for the comparator treatment. Does SH matching encompass a convergence of coded information about all inputs mediated by all sensory systems, classifying (tagging) them new or old?

My argument against massive sensory convergence (of which Gray disapproves), was aimed against just this kind of homuncular heresy. Of course sensory systems project to the hippocampal formation. But given "sparse coding" in the hippocampus it is implausible that the full richness of ongoing experience is shoehorned through SH for "consciousness treatment." "It is inconceivable that any single region of the brain might integrate spatially all the fragments of sensory and motor activity necessary to define a set of unique events" (Damasio 1989, p. 38). Moreover, hippocampal destruction does not abolish consciousness. But the comparator might still *contribute* to the selection of contents for consciousness, by intervening in the continual competition between cortical representations for prominence in experience.

4. The comparator's role in selective attending and remembering. To remember (consciously re-experience) an episode, one needs a retrieval cue (in any modality) that uniquely characterizes the episode, so that the memorandum can be singled out from countless generically similar events (Kinsbourne 1989). Clearly such a cue must have elicited a "mismatch" signal when the corresponding experience originally occurred. By virtue of its novelty it would at that time have elicited selective attention by activating the well documented back-projections from SH, through subiculum (Swanson & Cowan 1977) directly as well as via thalamic nuclei, to the cerebral cortex (Rosene & Van Hoesen 1977). So Gray's comparator straddles multiple loops that might feed back to the novel representations/cell assemblies so as to raise their level of activation. By means of this positive feedback, their novelty would lend them a competitive edge relative to other cell assemblies for entrainment in the "dominant focus" of cerebral activity. Also, such multiple representations that are simultaneously boosted would, by virtue of joining the dominant focus, mediate relations among stimuli (Hirsh 1980). I have suggested that a "dominant focus" of brain activity, consisting of entrained representations and continually varying in its composition, underlies ongoing experience and intention (Kinsbourne 1988; in press a). Having a particular experience is what it is like for a brain to incorporate a particular dominant focus.

The hippocampus' essential role in episodic remembering may not be the direct result of selection pressures but a by-product of this hypothesized boosting of the activation of distinctive representations for further selective attention. When reactivated (cued) at some later time, cortical cell assemblies that had been selectively activated by feedback from the SH loop trigger the reinstatement in awareness, by the vector completion property of the

neural network, of crucial elements of the remembered (re-experienced) events. At that moment, the dominant focus represents a prior event rather than what is happening at the time.

5. Which representations undergo comparator treatment? Gray confronts my proposal that any representation could conceivably contribute to awareness with Jackendoff's observation that it is "intermediary" representations that do so. But these claims are not incompatible. Those cell assemblies that cohere more than momentarily more often represent finished percepts than earlier stages in their microgenesis. But when processing sequences are disrupted, other representations may also qualify and generate altered states of consciousness in patients with brain damage or psychopathology (Kinsbourne, in press b).

6. The comparator's significance for consciousness. Although SH probably does not output the totality of conscious content, as Gray suggests, it may largely determine which figure emerges against the ground of competing inputs. This would be no brute correlation but a step toward a general theory of how consciousness works. If so, there would be no need to scrutinize the output from subiculum for elite consciousness-generating neurophysiology. Consciousness remains the subjective aspect of the integrated activity of varying subsets of a wide range of cortical cell assemblies (rather than of a unique module). We have no reasons to attribute unique properties to this circuitry, as distinct from minimal necessary organizational characteristics that are yet to be discovered. Easy and Hard Questions merge.

Correlating mind and body

T. J. Lloyd-Jones, N. Donnelly, and B. Weekes

Department of Psychology, University of Kent, Canterbury, Kent CT2 7LZ, U.K. **t.j.lloyd-jones@ukc.ac.uk**

Abstract: Gray's integration of the different levels of description and explanation in his theory is problematic: (1) The introduction of consciousness into his theorising consists of the mind-brain identity assumption, which tells us nothing new. (2) There need not be correlations between levels of description. (3) Gray's account does not extend beyond "brute" correlation. Integration must be achieved in a principled, mutually constraining way.

Gray provides an impressive overview of his research, concentrating especially on his concerns with anxiety and schizophrenia. The research, as he explains, addresses and attempts to integrate theories at four levels of description: neuroanatomical, neurochemical, information processing, and subjective symptomatology. Considered within these levels, the outlined research, we believe, is exemplary. However, despite a valiant effort, the attempt to integrate the different levels of description and explanation remains problematic.

Gray has set up a three-way set of equivalences where neural activity (neuroanatomy and neurochemistry) = information processing of a particular kind = the generation of the contents of consciousness (sect. 6.6). In this way, Gray presumes a form of mind-brain identity theory (cf. Smart 1959) with an intermediate level of description: that of information processing. Now, the debate as to the relative merit of identity theory or parallelism (and of whatever particular form of the theory one prefers, e.g., type or token) has been long and strenuously argued. Moreover, it is still as yet unresolved (see Bechtel 1988; Churchland 1984; Macdonald 1989). Nevertheless, the whole substance of Gray's claims to be able to introduce the contents of consciousness into his theorising consists of this assumption. A "new" theory that explains conscious awareness in terms of mind-brain identity tells us nothing new. Moreover, from this assumption arise two crucial issues which need to be addressed: (1) To what extent are there or need there be correlations between two or more of these levels? and (2) To what extent does Gray's account extend beyond "brute" correlation?

Let us address the first issue. Neuropsychology has in the past aimed to establish brain-behaviour relationships without in general exploring the richness of the computations that may underlie any particular faculty. According to this research strategy, it is possible, in principle, to discover a location in the brain which is responsible for consciousness (for example, the subiculum). Therefore, to the extent that correlations exist between consciousness and neural activity, these correlations will only point to a mapping between a brain location and conscious awareness. In contrast, however, we believe (with Gray) that to understand the nature and complexity of consciousness it will be necessary to develop functional models. These, we hope, will specify the component processes which function so as to give rise to conscious activity. However, it is important to acknowledge that the functional level of description may, in principle, be instantiated in an infinite number of ways at the neural level. Furthermore, it may also be the case that consciousness can be instantiated within a variety of different functional architectures. For example, Baddeley (1993) suggests that conscious awareness may be a function of the central executive in his working memory model. Jackendoff (1987) suggests that the functional component responsible for visual conscious awareness is the 2.5D sketch as delineated by Marr (1982). Thus, in contrast to the neuropsychological approach, no simple correlations may exist between the functional level of description and either the neural or the conscious levels; indeed, there need not be *any* correlation between functional processing and conscious awareness levels of description and explanation. At best, information gained from the neural level and theorising at the neural level can constrain theorising at the functional level.

With this in mind, we turn to point 2: Gray has incorporated an intermediate functional level to facilitate the link between neural and conscious levels of description. From the paragraph above the question arises as to the extent to which information at the neural level constrains: (1) the structure and function of the comparator (producing match/mismatch decisions) and (2) the involvement of schemas in the matching of stored knowledge to incoming sensory data in the interaction of top-down and bottom-up information processing. As far as we can ascertain, although Gray suggests interesting possibilities for a range of information (device) – neural level activity parallels – at no point does this analysis move beyond that of brute correlation. No architectural constraints at the neural level govern postulated functions at the information processing level. Hypothesising more specific units within levels of description nevertheless fails to go beyond correlation and towards explanation. It is possible, however, to define such constraining architectural features for at least some psychological systems (see Kosslyn et al. 1990 for an example from vision research). To illustrate, recordings of cells within the visual pathways have suggested at least two functional systems: One begins in the magnocellular layers of the lateral geniculate nucleus (LGN) and projects upwards to the parietal cortex. This pathway seems relatively specialised for the analysis of retinal motion and of stereopsis; it is concerned with the location of objects in 3D space. The second pathway begins primarily in the parvocellular layers of the LGN and projects to the temporal cortex. Here the various streams process colour and spatial detail and seem concerned primarily with the recognition of objects (see Zeki 1993).

In summary, although attempts to integrate across levels of description have immense value, the integration must be achieved in a principled manner. Theorising is at its most powerful when computational theory, neurological theory, and neurological and behavioural evidence are used in mutually constraining way, and in so doing drive research.

Human consciousness: One of a kind

R. E. Lubow

Department of Psychology, Tel Aviv University, Ramat Aviv, 69978, Israel.
lubow@freud.tau.ac.il

Abstract: To avoid teleological interpretations, it is important to make a distinction between functions and uses of consciousness, and to address questions concerning the consequences of consciousness. Assumptions about the phylogenetic distribution of consciousness are examined. It is concluded that there is some value in identifying consciousness an exclusively human attribute.

Gray's treatment of consciousness, while intriguing and thought-provoking, at times takes a disturbing teleological turn, as when he asks about its function. Although this may reflect nothing more than careless phrasing, a few elaborations are in order. To begin with, the distinction must be drawn between "function" and "use." Consciousness has no function, in the sense that it does not exist in order to accomplish something; it is not goal oriented. On the other hand, granted that consciousness exists, it is, indeed, useful. The distinction becomes obvious when one looks at physical as opposed to biological phenomena. Gravity, for example, does not exist in order to perform some function. Nevertheless, it is extremely useful, in an endless number of unconnected ways, from keeping coffee in my cup to shaping the evolution of inorganic materials and organic life forms. Given that gravity exists, there are certain inevitable consequences. But, surely we would not want to identify that state of affairs with notions concerning the functions of gravity.

When Gray does describe the consequences of consciousness, he focuses on a very limited aspect of it: the momentary awareness that is a byproduct of selective attention processes initiated by exteroceptive stimulation. Surely the important advantages of consciousness are not to be found in the coffee cup, but in some larger arena. Indeed, the most salient feature of consciousness is that it promotes an escape from the immediate present. With consciousness, we actively reorganize the past as well as shape the future. It is consciousness that allows us to cross the boundaries of time and space, to build cathedrals and spaceships, to write histories and, perhaps most to the point, to rewrite them.

These uniquely human, awe-inspiring consequences of consciousness bring me to the next point. Although it has long been fashionable, at least in some academic circles, and certainly in Gray's paper, to treat consciousness as a product of evolution, in the sense that consciousness is (or was) present in some degree in organisms other than humans, there is good reason not to accept such a hypothesis. To begin with, consciousness is a subjective state which cannot be reported on by past or present nonverbal organisms. However, that by itself is insufficient reason for refusing to accept the hypothesis. There are many other such subjective states which do not require introspective linguistic reports to verify their existence, for example, vision. No one doubts that a pigeon sees, although its visual experience may in no way resemble our own; nor can one ever determine whether it does. This, despite the fact that one can empirically demonstrate that human and pigeon have similar visual acuity, similar spectral sensitivity curves, and so forth. It is these latter observations, and hundred of more obvious ones, both casual and formal, that allow us to conclude that, yes, the pigeon does have vision. It is instructive that no one would seriously debate that proposition. But what comparable observations can be made that would allow us to assess consciousness in nonhuman animals?

The early behaviorists, for reasons already noted, clearly accepted the proposition that consciousness was not the proper domain for the scientific study of animal behavior. However, at a metatheoretical level, they endorsed concepts such as gradualism, continuity, and progress in regard to relating animal to man. Consequently, in the interest of consistency, they were obliged to delegitimize the study of consciousness in humans. The behaviorists avoided the phenomenon of consciousness in order to emphasize the animal-like qualities in humans (and, indeed, at that time, there was much to be gained from this tactic).

On the other hand, contemporary cognitive scientists, not sharing with the early behaviorists their investment in animal learning theory, focus almost exclusively on human behavior. Nevertheless, they still accept traditional Darwinian notions of continuity, and consequently they tend to see consciousness everywhere (Lubow 1981). Gray is no exception. By proposing an animal model of schizophrenia (Gray et al. 1991), and using such a model in his discourse on consciousness (present paper), he is surely embracing the untestable notion that animals have consciousness.

We appear, then, to have two alternatives, both serving the interests of phylogenetic consistency: consciousness is either nowhere or everywhere. However, there is a third hypothesis, neither more nor less valid than that of continuity, one that allows for a reconciliation between the demand for dealing with observables and the impossibility of studying consciousness at the subhuman level: consciousness is exclusively a property of the human brain. Somewhere in the long line of hominid evolution, a critical mass of some type(s) of neurological tissue was reached, and consciousness appeared, as though full-blown from the head of Zeus. This, of course, is the familiar idea of an emergent property, in regard to which Jaynes (1976) argued that consciousness appeared as recently as the first millennium B.C. Although one need not take so strong a position, if one maintains that for prehistoric humans, as in lower animals, we can only infer the presence of consciousness from the material artifacts which survived their existence, then, indeed, consciousness would be a recent, and exclusively human, acquisition.

In summary, it is futile to make assumptions about the distribution of consciousness that are either all-inclusive or all-exclusive. It would appear more reasonable to assume that consciousness is not a mere extension of what preceded it, neither a longer nor a stronger limb, but rather like language, which may be inextricably linked with it, a qualitative change from prior conditions. And, just as it would not be the best of tactics to study human language by examining the grunts and snorts of lower animals, so the study of consciousness would seem to be best confined to ourselves. There may be some elemental truth in the tautological statement that human consciousness is one of a kind.

Comparators, functions, and experiences

Harold Merskey

Professor Emeritus of Psychiatry, University of Western Ontario, London Psychiatric Hospital, London, Ontario N6A 4H1, Canada

Abstract: The comparator model is insufficient for three reasons. First, consciousness is involved in the process of comparison as well as in the output. Second, we still do not have enough neurophysiological information to match the events of consciousness, although such knowledge is growing. Third, the anatomical localisation proposed can be damaged bilaterally but consciousness will persist.

The theme is difficult and Gray has put together an attractive theoretical schema. I think it fails to be acceptable for three reasons.

First, there is something different between the outputs of a comparator and an intuitive notion of consciousness. At least part of consciousness appears to function in the middle of human processes of evaluation, appreciation, or decision-making. Such activities would be the computer/comparator's process and not its results. But states of consciousness or outputs are also results and not only steps, gears, agents (or Gods) in the middle of the machine. Thus, the comparator model is in difficulties. What it is thought to do does not comprehend the whole of cognition and awareness.

The example of planning a journey challenges Gray's views. Here consciousness is, at least in part, planning agent and ticket clerk, partly visualizing where to go, while itemizing a sequence. This must apply to a number of visualized actions. Such actions then appear as outcomes rather than only as integral parts of comparisons. This is not the pure function of a comparator.

Second, we can be very optimistic about a somewhat millennialist notion that, given time, the brain functions or changes which correspond to consciousness will emerge and be evident on the basis of continuing neurophysiological discovery. For that to happen we would have to be able to match conscious awareness with contemporary, physiological reports, but that does not seem to be inherently improbable, particularly in the days of modern imaging. Until then we may be wiser to wait, and time should suffice for this outcome.

The third main reason for the failure of the theory is that it depends upon some questionable anatomical propositions. Presumably, if the subiculum on both sides is not working, the comparator system will not be active and consciousness will not be available. If both temporal lobes are significantly damaged, the individual does not lose consciousness but suffers a significant impairment in memory (Scoville & Milner 1957). If the function of the subiculum is held to be an essential component of consciousness, then the theory is probably wrong in that particular detail. Bilateral lesions of the hippocampus are not a favoured condition, but have occurred once or twice in the past, and the descriptions of such cases indicate continuing consciousness and awareness, despite severe loss of anterograde memory. If the theory depends upon this specific location for its validity, it cannot survive this sort of factual information.

Pain provides another illustration of the anatomical problem. As a primitive reflex, nociception calls into being responses which may only involve consciousness secondarily. Pain is a very primitive phenomenon, and – in so far as we can talk about it by analogy in animals – it is associated with retreat from damaging or potentially damaging stimuli. Nociception produces a response which does not immediately require consciousness because the built-in reflexes for removal from the scene or for defensive fighting behaviour can be treated as automatic, occurring before the emergence of an emotional state. Reports of localization of pain to the cerebral cortex are quite rare for good reason since pain hardly requires the elaborate contribution of the cerebrum. The subjective states which we call pain may serve to amplify some initial emergency behaviour and to provide continuing adaptive behaviour. The best high level localization known for pain is probably the cingulate portion of the limbic system (Talbot et al. 1991). This does not seem to require the subiculum for its existence.

Gray makes a good point that consciousness is presumed to have an evolutionary value. Perhaps this is because we can conceptualize brain functions better in terms of the relationship between awareness and responsiveness than if we rely only on the latter. Consciousness may justify its biological pre-eminence simply because it also has some evolutionary benefits. It is tempting to suppose that these may include anticipatory visual formulations and testing of ideas, and later adaptations which subjective states will promote. Such benefits can be adaptive but again are not part of the function of a simple comparator.

Perhaps widespread or multiple comparators are required, some with process functions and others with experiential qualities which provide for drives. This goes beyond Gray's propositions that we can be grateful to him for offering a powerful complex stimulus in the form of his schema.

The control of consciousness via a neuropsychological feedback loop

Todd D. Nelson

Department of Psychology, Michigan State University, East Lansing, MI 48824. **tdnelson@msu.edu**

Abstract: Gray's neuropsychological model of consciousness uses a hierarchical feedback loop framework that has been extensively discussed by many others in psychology. This commentary therefore urges Gray to integrate with, or at least acknowledge previous models. It also points out flaws in his feedback model and suggests directions for further theoretical work.

Though many before him have discussed the structure of goal-directed behavior, no one until Gray, in this target article, has explored in such detail the possibility of conscious experience being linked to specific neurological structures, and for that alone, Gray's model is worthy of serious consideration. Does it represent an advance in our knowledge of how conscious experiences are formed? I believe, with the right modifications (which I present below), the model could present such an advance. I will restrict my remarks to a discussion of the self-regulation model at its core.

Gray's comparator is the generic comparator that is central to the models of Powers (1973), Carver and Scheier (1981), and Miller et al. (1960). However, Gray proposes that the comparator, at the neurological level only, seeks to determine whether there is a "match or mismatch" (sect. 1, para. 7) between the current and the predicted (or expected) state of the world. I find this a bit simplistic for a description of the function of any comparator system. Many researchers (including Gray) acknowledge that behavior is organized hierarchically (Nelson 1993) and that the functions of lower-level systems are controlled by the outputs of their next-higher systems. Within this hierarchy, at each level, there is a range of acceptable states of the world which will "fit" with one's expectations for the comparison at that level. There is also a range of perceptions which do not fit, or are unacceptable. Hence even at the neurological level, there is this range of acceptable and unacceptable comparisons between what is predicted and what is perceived. It is not enough to ask whether there is a "match" or "mismatch" between the present condition and what is expected. Rather, the question should be "If there is a mismatch, what *kind* of mismatch is it?" Specifically, many researchers (notably Powers and Carver & Scheier) contend that behavior is controlled by a series of hierarchically ordered *negative* feedback loops. That is, if there is a match between perceptions and expectations, everything is fine. If the perception *falls short* of expectations (a *negative* discrepancy), then the self-regulatory system interrupts behavior (as Gray suggests occurs in his description of the mismatch). What happens if the mismatch is one in which our perceptions *exceed* our expectations (the *positive* discrepancy)? Gray does not address this in his model, and he should. These perceptions may not interrupt the behavioral sequence, even at the neurological level. Some (e.g., Bandura 1982) have argued that positive discrepancies are essential to the self-regulatory system. Others agree, but stipulate that these positive feedback systems (or "discrepancy producing" systems) are inherently unstable, and that they are subordinate to a higher-level negative feedback ("discrepancy reducing") system (Scheier & Carver 1988).

I was struck by the failure to refer to a literature that seems to bear quite directly on the problem of how the neurological systems produce the contents of conscious experience. Powers (1973; 1989) has proposed a cybernetic approach to behavior with a hierarchical organization of self-regulatory feedback systems along nine levels of control, ranging from the most abstract (the ideal self-image) to a specific level of control (e.g., the neuronal level). Powers (1980) even discussed how his model could account for conscious experience. In his detailed model of the hierarchical organization of behavior, Powers (1973) suggested that we act in

order to change the environment so that it matches our expectations for that perception. In other words, behavior is simply the control of perception. Gray suggests that his model shares much with Neisser's (1976) perspective. Neisser writes, "The cognitive structures crucial for vision are the anticipatory schemata that prepare the perceiver to accept certain kinds of information rather than others and thus control the activity of looking. . . . it is these schemata . . . that determine what will be perceived" (as cited by Gray, sect. 3, para. 12). Again, it seems that Gray is re-inventing the wheel in stating almost exactly what Powers wrote over 20 years ago.

Gray (section 5.14) states that his model has "no principled way of accounting for the difference between such apparently unconscious 'match' outcomes and the conscious variety that occur." I am curious why Gray believes that the processes at the "automatic processing" (Schneider & Shiffrin 1977) level and the conscious processes should be different. Such a distinction seems unnecessary. The basic comparator processes that occur in conscious activity also function outside conscious awareness (see Carver & Scheier, 1981, for a detailed discussion of this point).

Although the notion that conscious experiences can be tied to specific neurological structures and processes is not new (contrary to Gray's assertions; see Powers 1973), Gray's model represents a laudable first step in the investigation of how consciousness is generated and how it is changed. An important next step for this model (or variants of it) is to specify the reciprocal influence of the feedback systems at the neurological level, and the higher levels (e.g., the behavior sequences or prescriptions of behavior, what Powers termed the "program" and "principle" levels, respectively) in determining the moment by moment changes in the contents of consciousness. In addition, it would be useful to examine how the neurological system deals with positive discrepancies between expectations and perception.

Reticular-thalamic activation of the cortex generates conscious contents

James Newman

Colorado Neurological Institute, Denver, CO 80218. **newmanjb@aol.com**

Abstract: Gray hypothesizes that the contents of consciousness correspond to the outputs of a subicular (hippocampal/temporal lobe) comparator that compares the current state of the organism's perceptual world with a predicted state. I argue that Gray has identified a key contributing system to conscious awareness, but that his model is inadequate for explaining how conscious contents are generated in the brain. An alternative model is offered.

1. Our work with Global Workspace (GW) theory and the neural substrates of conscious processes (Baars 1983; 1988; Newman 1990; 1995; Newman & Baars 1993; Newman et al., in press) has identified several key concepts and constraints in consciousness studies. In his target article, Gray speaks to a number of these issues. For example, he rightly concludes that a "unifying set of concepts" is necessary to a scientific account of the evolutionary and survival value of conscious behavior. The focus of Gray, however, is upon how "primary awareness" might arise out of brain events. His conjecture is that the "contents of consciousness consist of the outputs of the subicular comparator" (sect. 3, para. 1). His model has a number of strengths and weaknesses. To highlight these requires a brief presentation of an alternative model of how conscious contents are generated within the brain.

A basic premise of GW theory is that the study of conscious contents and processes must begin with a "contrastive analysis" of conscious and *un*conscious phenomenon. For example, what is the difference between my experience of writing this commentary and my memory of writing a letter the previous day? Certainly the latter exists in some recallable, yet nonconscious form. Or, what is the difference between my conscious awareness of the ideas I am conveying in this sentence and the multiple sub-tasks being automatically carried out by my brain (e.g., word selection, sentence parsing, typing, etc.), all with minimal-to-no awareness (Baars 1988)?

The second example illustrates two basic distinctions in GW theory. The first is that routine, overlearned cognitive tasks tend to be performed unconsciously, by arrays of specialized modular processors. In contrast, the "conscious system" consists of a global workspace, or monitoring, architecture that allocates processing resources based, first, upon biological contingencies of novelty, need, or potential threat and, second, cognitive schemas, purposes, and plans. The defining properties of perceptual contents that engage conscious attention (i.e., the global allocation of processing resources) are that they: (1) vary in some significant degree from current expectations; or (2) are congruent with the current, predominant intent/goal of the organism. In contrast, the processing of stimuli which are predictable or overlearned is automatically allocated to nonconscious, modular processors (Newman et al., in press).

2. These GW criteria dovetail nicely with Gray's comparator model. Conscious awareness requires an ongoing comparison of present and past experience (as well as current intentions and goals) to determine which of many competing stimuli are relevant to conscious requirements and which are not. In a sense, the organism is continually performing a contrastive analysis of its present perceptions and behavior. Gray's "comparator" is clearly central to this processes. But do its outputs, per se, generate primary awareness?

The neuropsychological evidence militates against the subicular region being the *sine qua non* for primary consciousness. The most dramatic (but hardly unique) example is the widely studied case of H. M., who underwent bilateral removal of the anterior two-thirds of his hippocampus and a goodly portion of both temporal lobes. His capacity for encoding new episodic memories since the 1953 operation has been profoundly impaired. He is quite incapable of comparing his current state of mind with his post-surgery experiences. Yet, no one has suggested that H. M.'s brain ceased producing conscious contents. Hippocampal lesioned patients simply have no *continuity* of awareness/memory beyond the moment (except pre-surgery LTM) (Edelman 1989; Rose 1992).

3. There are regions of the brain the destruction of which results in direct impairments of conscious awareness and attention. These are, in order of severity: the reticular formation of the brain stem, the intralaminar complex of the thalamus (ILC in Fig. 1, below), and the cortical areas most densely interconnected with this complex. Bilateral destruction of the reticular formation or thalamic complex produces coma. Unilateral lesions of various cortical areas or subcortical nuclei interconnected with the reticular-thalamic core produce attentional neglect syndromes (Bogen 1995; Mesulam 1985; Heilman et al. 1985).

Newman & Baars (1993) reviews the accumulated evidence of over forty years of research implicating the extended reticular-thalamic activation system (ERTAS) in the genesis of consciousness. A more recent paper focuses upon the central role of the thalamus in the generation of complex, distributed patterns of EEG activation. It is these recurring oscillatory patterns, especially in the 20–60 Hz range, which Newman (1995) posits as the neural analogs of conscious contents. Hippocampal-induced theta rhythms (4–12 Hz) are integral, but secondary, to these processes.

4. Gray's subicular comparator makes a vital contribution to the ERTAS system. I would argue, however, that subicular outputs produce *alterations in patterns of ERTAS activation*, not conscious contents per se (the subiculum's role in memory encoding and recall is probably much more content specific, but is beyond the limits of this commentary). Accepting Gray's "alternative" hypothesis (sect. 3), I posit that his comparator provides feedback to the ERTAS, "flagging" present perceptions as "expected/familiar" or "unexpected/novel." A novel/unexpected flag interrupts ongoing behavior, causing an ERTAS-mediated orienting response. Con-

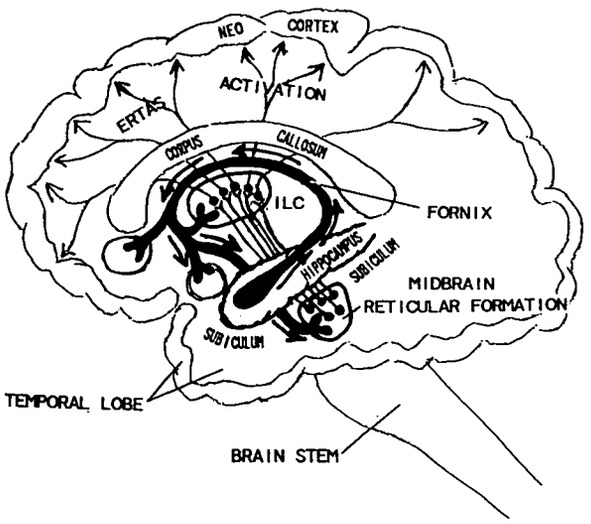

Figure 1 (Newman). Outputs from the hippocampus/subiculum travel via the fornix to the intralaminar complex (ILC) of the thalamus, the midbrain reticular formation, and limbic nuclei (not labeled). This activation effected via the fornix (theta rhythms) conveys "flagging" messages ("unexpected" or "expected") to the extended reticular-thalamic activation system (ERTAS), altering the cortical EEG.

versely, an expected/familiar flag would produce habituation or, in the case of goal-directed behavior, a shift in attention to the next step in the current program. These scenarios are consistent with what is known about the anatomy and physiology of the ERTAS and the comparator system Gray describes. The fornix, the main subcortical output pathway of the subiculum/hippocampus, projects first to the intralaminar complex (ILC) and, finally, to the midbrain reticular formation (see Fig. 1). The fornix is believed to be the "pacemaker" for the hippocampal theta rhythm (Diamond et al. 1985).

In conclusion, Gray articulately describes a complex limbic-striatal-cortical system that is undoubtedly key to arousal, motivation, motor programming, and memory processes. Our model views these functions of the comparator system, however, as *mediating between the unconscious and conscious systems* of the brain. This simply means that before any highly integrated perception that *could* become a conscious content *can* become one, it must undergo a comparator "test" which flags it as relevant, or not, to the organism's conscious needs/intentions. Edelman's (1989) theory of "primary consciousness" says essentially the same thing about the hippocampal system, describing it as "a selective link between parallel and sequential patterns of global mappings that . . . allow[s] temporal ordering of perceptual categorizations" (p. 132). In this context, Gray actually goes considerably farther towards elucidating the adaptive value of consciousness than he realizes.

The elusive quale

Howard Rachlin

Psychology Department, State University of New York, Stony Brook, NY 11794. **hrachlin@psych1.psy.sunysb.edu**

Abstract: If sensations were behaviorally conceived, as they should be, as complex functional patterns of interaction between overt behavior and the environment, there would be no point in searching for them as instantaneous psychic elements (qualia) within the brain or as internal products of the brain.

A friend of mine, a newly married heterosexual man, once said to me, "I'd give anything to know how a woman feels during sex." My immediate response was to ask him whether he thought he already knew how another man (I, for instance) felt during sex. On reflection, I should have asked him whether he thought he knew how he himself felt during sex, perhaps just an hour ago. He probably would have replied that he remembered, and so knew in the most direct possible way, how he himself felt. But in what way is remembering a conscious experience like having the experience itself? My friend's memory of his feeling during a sex act an hour ago would certainly not have been a repetition or even a watered-down sex experience (we were walking in the street at the time). That memory must have been qualitatively different (a different quale) from the original experience. My friend's memory could not possibly serve as a datum for the science (Gray envisions) that would explain, "experience of primitive sensations, that is, qualia." Perhaps, for Gray, a memory may itself be a quale, but it must certainly be a different quale from the sensation it is a memory of. Even if we grant that during sex my friend was having some sort of particular conscious experience (of the sort envisioned by Gray), there would be no way for him to know (in the sense envisioned by Gray) how he felt during sex in general; how could he compare that quale with a previous one? This is why, contrary to Gray, even "the whole of psychophysics," past, present, and future, will never flush out a quale. In psychophysical experiments people make judgments about the presence or intensity of environmental *stimuli*, not about conscious experience. We have neither internal organs to sense internal quales (as distinct from environmental stimuli) nor a vocabulary to describe them (as distinct from the vocabulary we use to describe those stimuli).

Gray is right in criticizing philosophical functionalists and materialists for identifying consciousness with the software or the hardware of the brain. But he makes the same error they do in supposing that consciousness is to be found in the brain in the first place. As Gray knows, the stimulus that gives rise to the sensation of red (let alone the experience of sex) consists of a very complex spatially and temporally extended series of events. Why then does he suppose that a sensation itself is a quale – a unitary internal conscious element? It would be better to conceive of a sensation (or any other conscious event) as an overt pattern of behavior; to have a sensation of red is to have in the past discriminated between red objects and not-red objects and at the moment to be making such a discrimination. The sensation is located where all mental events are located – not inside the body but in the functional interaction of the whole body and the environment in what Skinner (1953) called a "history of reinforcement" and Gibson (1950) called an "affordance."

By way of questioning this sort of behavioristic program for studying the mind, Gray asks, "What is the difference between two awake individuals, one of them stone deaf, who are both sitting immobile in a room in which a record player is playing a Mozart string quartet?" (cited by Rachlin 1994b; Staddon 1993). Let us call this Gray's first question. A behaviorist could (and should) answer: "One individual is hearing the music and one is not." But what does it mean to hear and not to hear? To hear is to consistently discriminate between sound and silence by revealing over time what Gray would call a "brute correlation" between overt behavior and sound. For the hearing person this correlation is significantly different from zero; for the stone deaf person, the correlation is zero. That and that alone is what it means to hear. Gray's first question singles out a pair of corresponding (or intersecting) points on two correlations. But it is in the whole correlation over time, not any individual point, that the meaning of "hearing" and "stone deaf" hides. Certainly the brain contains the *mechanism* by which hearing works. But, as Gray's entire target article illustrates, no matter how precisely we examine this underlying mechanism, the quale still eludes us. There is every reason to continue to study the brain, just as there is every reason to pursue the science of chemistry. But we are no more likely to understand the mind through the workings of the brain than we are to understand the meaning of a painting through the chemistry of paint.

In further response to arguments like these, Gray (personal communication) has posed a second question: "What if both listeners are hearing the Mozart quartet but one is enjoying it and one is not?" This question progresses from (supposedly simple) perception to perception plus feeling or emotion. But the behavioral view still applies. What matters for mental states is not just present overt behavior but present overt behavior in the context of the past and future. If one listener listens to Mozart all the time and all his friends and relatives believe he loves Mozart then by definition he loves Mozart; if the other usually refuses to listen to Mozart and his friends and relatives all believe he hates Mozart then by definition he hates Mozart. The Mozart lover still loves Mozart even if he happens to be suffering from a broken leg at the moment (see Rachlin, 1985, for a discussion of pain as behavior). The Mozart-hater still hates Mozart even if he is in the process of digesting a pleasant meal. Patterns of behavior may overlap. What is the point of adding one or a series of love-Mozart or hate-Mozart quales to these patterns? Can you really hate Mozart and still repeatedly and consistently act as if you loved Mozart? Such a state of affairs is not just empirically impossible – it is logically impossible (Rachlin 1994a; 1994b).

Finally, granting that Gray's conception of consciousness (as an eventually understandable but not presently known internal state caused by, but not identical to, brain physiology) has motivated superb physiological research (by Gray and others) and perhaps led to some useful clinical techniques, why complain? The reason to complain is that Gray's approach sets consciousness up on a pedestal and at the same time trivializes it. According to Gray we should try to understand consciousness for the same reason that Hilary climbed Mt. Everest, "because it is there." In that case we could pursue physiology, ignore behavior, and safely wait for some psychological Hilary (or Einstein or Planck) to come alone to figure out how the brain gives birth to qualia.

In summary, even such relatively simple conscious events as sensations and perceptions are best conceived as complex functional patterns of interaction between overt behavior and environment. The reason the brain mechanism generates such patterns is that they maximize utility (Rachlin 1995) and, thereby, survival. Conscious events (properly understood as behavioral patterns) have no meaning at an instant of time. It is therefore futile to search for an instantaneous psychic element (a quale) either in the brain or as an internal product of the brain. It is *not* there.

Unitary consciousness requires distributed comparators and global mappings

George N. Reeke, Jr.

Laboratory of Biological Modelling, The Rockefeller University, New York, NY 10021. **reeke@lobimo.rockefeller.edu**

Abstract: Gray, like other recent authors, seeks a scientific approach to consciousness, but fails to provide a biologically convincing description, partly because he implicitly bases his model on a computationalist foundation that embeds the contents of thought in irreducible symbolic representations. When patterns of neural activity instantiating conscious thought are shorn of homuncular observers, it appears most likely that these patterns and the circuitry that compares them with memories and plans should be found distributed over large regions of neocortex.

1. Consciousness and computation. Gray promises a scientific approach to explaining the contents of consciousness (specifically "primary awareness": Edelman 1989; Jackendoff 1987), but he is only able to identify these contents rather loosely with the output of a neural circuit that compares current and predicted states of the perceptual world. After criticizing Dennett and Kinsbourne (1992), quite properly, for just asserting that awareness is an irreducible property of appropriately functioning neural assemblies, he fails to provide a sufficiently rich alternative. He recognizes that the output of a comparison is a simple match/mismatch decision, but then founders in trying to describe how this binary output might be augmented to represent thoughts. But a mere coded signal can never *be* a thought (Bickhard 1991). Gray makes this category error because he fails to free himself from the assumption that thought is a computation.

This assumption that the main function of the brain is to process information, (sect. 1, para. 6) that is, to compute, implies that the brain traffics in encoded representations. If that is all there is, awareness can indeed only be an epiphenomenon, as Dennett has recognized. But if the main function of the brain is instead to *create* information, then conscious thought can be understood as one form taken by the created information.

The computational analogy leads Gray to err by breaking conscious processes down into modules that communicate by exchanging messages, as in a parallel computer (see Reeke 1991 for a commentary on a similar assumption by Minsky 1985). One of these modules is the Cartesian theater, which Gray identifies with his comparator circuitry (sect. 5.9), although by Dennett's definition it should be the place where the output of the comparator is interpreted. It is curious that Gray makes a point of dismissing the arguments against the Cartesian theater, even though he accepts that it creates difficulties for his theory (sect. 5.9), when his theory does not require such a region. Gray does so to help explain the perceived unity of consciousness in the face of the separateness of the sensory modalities (Jackendoff 1987).

A better explanation for this apparent dichotomy can be found by focussing less on the comparator and more on what is to be compared. In this view, the contents of consciousness must comprise large portions of the responses of individual sensory areas, as well as responses in areas (e.g. the entorhinal cortex) that receive input from multiple sensory pathways. The perceived unity of these diverse elements is a consequence of their extensive reentrant interconnections, which give rise to what Edelman (1987) has called "global mappings." By postulating a single area where the output of his comparator is interpreted to generate conscious thought, Gray invests his theory with a not-so-hidden homunculus (Reeke & Edelman 1988), the functioning of which he does not even attempt to explain.

2. Comparator and hippocampus. If we were to set aside these arguments about the inadequacy of comparator output as the physical basis of thought and accept Gray's account of consciousness, we would still find several problems with the anatomical and physiological picture of the comparator itself that Gray has given us. Contrary to Gray, such a comparator could not be located in the hippocampal formation, but would have to be widely distributed throughout the cerebral cortex to be consistent with the data on behavior and memory in humans and animals with hippocampal lesions (reviewed in, e.g., Cohen & Eichenbaum 1994), particularly that regarding the celebrated case H. M. (Scoville & Milner 1957) who is perfectly conscious yet, as Gray himself admits, lacks most of the circuitry critical to the comparator theory. The theta rhythm can hardly be the main timer that regulates conscious experience in Gray's model, given that it is so inconspicuous as to be arguably absent in primates. Gray too readily dismisses the smoothness of conscious perception in the face of the chunkiness predicted by his theory based on a succession of discrete comparison events. Finally, the binding problem appears to be solved by phase entrainment of pulse trains in cortex, not by subicular activation as argued by Gray (Eckhorn et al. 1988; Gray & Singer 1989; Tononi et al. 1992).

Gray mentions (Introduction, para. 5), but does not describe, the consciousness theory of Edelman (1989), which Edelman works out in full detail, including a discussion quite different from Gray's (sect. 2) of how disorders such as schizophrenia might arise from disruptions of particular pathways in the conscious brain. Because Edelman's theory also involves a neural comparator in a central role, it is important to distinguish the two theories.[1] For Edelman, *categorization* rather than *comparison* is the central process of consciousness. Categorization is a property of even the most rudimentary nervous systems, because it is necessary for

discriminative behavior. The key added feature in conscious brains is the presence of massive reentrant loops in which categorizations of interoceptive stimuli, external stimuli, and comparisons of the two are continuously confronted with new categorizations of the world. Categorization of the comparison between external and internal stimuli (**C**[C(W)·C(I)] in Edelman's notation) is the basis for memory, which therefore emerges as a *recategorization* of correlated current and prior responses. Edelman avoids ascribing the contents of consciousness to the output of the comparator, or of any other single area, considering it rather to be a global property of the entire system.

3. Adaptive value of consciousness. Gray professes to be unable to find sufficient adaptive value in conscious awareness to support its emergence during evolution. Yet, if he is conscious the way I am, he must have noticed the ongoing contributions of conscious thought to problem solving and planning, both short- and long-term, without which survival would be incomparably more difficult. Gray's problem can only arise if conscious awareness is an epiphenomenon of an underlying computational process that is actually responsible for planning and problem solving. Gray accepts this point of view, citing arguments and data reviewed in Velmans (1991) and Libet (1993) to the effect that conscious awareness can be reported in certain situations only after behavior has been decided. Gray recognizes that this view requires an explanation for the efficacy of computation to control behavior, as does that of the spiritual mind in dualistic theories. He appeals for this to Pylyshyn (1984), who describes how the semantic content of neural signals can affect their interactions with the neuronal circuitry in which they exist.

However, an explanation for the puzzling data of Libet and others, one that preserves the efficacy of conscious thought, can be arrived at by dropping the assumption that conscious processes are unitary and occur in a single Cartesian theater. If awareness is distributed across most of the cortex, as suggested above, and unified by reentrant signalling, then it is quite possible for different aspects of awareness to occur in different sequences as perceived elsewhere in the system. The situation is analogous to the relativity of simultaneity in Einsteinian space-time, except that here the velocity of signal propagation is so much less than the velocity of light that apparent reversals in temporal sequence can readily be perceived, even within the brain of a single observer. (see Fig. 1).

4. Qualia. The participation of multiple, reentrantly interacting brain regions in conscious awareness provides one possible, tentative approach to the problem of qualia, one of the phenomena for which Gray would have us seek an account (Introduction, para. 4). Gray makes a beginning with his idea that we can deduce similarities in some aspects of the conscious experience of rodents and primates from the similar effects of certain drugs in the two orders (sect. 2, para. 19). However, this analysis is suspect, because the similarity of drug effects can be predicted from similarities in physiology without any reference to the contents of neural signals. Anaesthetics, paralytics, and more subtle drugs would be expected to work just as well in zombies as in humans.

I suggest that a better approach can be found by considering the nearly perfect continuity of conscious awareness across small changes in the environment and its consistency across multiple sensory domains. These facts practically demand a corresponding continuity and consistency in the underlying neural activity. Edelman & Reeke (1990) made a similar argument regarding people's ability (which we called "zoomability") to ponder aspects of the world at many different levels of detail, moving among these levels at will. Perhaps when we become able to enumerate some of the possible mappings of phenomena in the world onto patterns of neural response, it will be found that these requirements of continuity, consistency, and zoomability will impose severe constraints on the mappings that are admissible. These constraints would operate similarly in all individuals of a species, leading to similarities in their neural responses and possibly in the qualia they perceive.

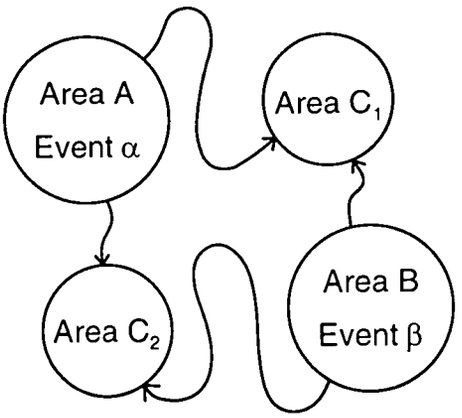

Figure 1 (Reeke). Because of the relatively slow transmission velocity of neural signals, particularly along polysynaptic pathways, two neural responses, α, occurring in region A, and β, occurring in region B of the cerebral cortex, can be perceived in opposite order in different areas constituting parts of the apparatus of conscious awareness, designated by C_1 and C_2. In this highly schematic example, β reaches C_1 before α, but α reaches C_2 before β. The percept depends on the nature of the signals and the past experience of the organism, although, in fact, α and β are simultaneous events.

NOTE

1. My description here is a severely abbreviated summary of Edelman's ideas which necessarily omits relevant qualifications and supporting evidence. See the references cited for details.

ACKNOWLEDGMENT

This work was supported by a grant from the Neurosciences Research Foundation.

Prospects for a cognitive neuroscience of consciousness

Antti Revonsuo

Department of Philosophy/Center for Cognitive Neuroscience, University of Turku, FIN-20500 Turku, Finland. **revonsuo@sara.utu.fi**

Abstract: In this commentary, I point out some weaknesses in Gray's target article and, in the light of that discussion, I attempt to delineate the kinds of problem a cognitive neuroscience of consciousness faces on its way to a scientific understanding of subjective experience.

Gray's approach to consciousness is reasonable. He admits that phenomenological experience really exists and that the tough problem is: What could provide us with strong theoretical understanding about the link between the mental and the neural reality? He advocates a multidisciplinary strategy; convergence of the research programs of philosophy, psychology, and neuroscience.

Given the merits of such a cognitive neuroscientific approach, why is it then that the contribution Gray makes to our understanding of consciousness remains relatively modest (as the author himself admits)? One of the main problems appears to be that there is currently no coherent field of consciousness research, drawing from a common database of empirical evidence. Gray's conjecture is based on models of anxiety and schizophrenia, but there are many other theories that draw on completely different sources. For example, Schacter's (1990) model is based on the explicit/implicit distinction in neuropsychology, Crick's and Koch's mostly on the neurobiology and psychology of vision (Crick 1994; Koch & Crick 1994), Llinás's on the neurophysiological properties of wakefulness and sleep (Llinás & Paré 1991; Llinás & Ribary

1993), and Baars's (1994; Baars & Newman 1994) mostly on cognitive and physiological psychology (for reviews, see Revonsuo 1993; 1994). It is encouraging that these models have also been explicitly compared (Churchland 1994; Crick 1994; Llinás & Ribary 1994). After all, they claim to investigate the very same phenomenon; how could they afford to ignore each other?

It seems however, that Gray largely ignores these other theories. He assigns a central role to the hippocampus and its theta rhythm, whereas most of the theories mentioned above emphasize the role of thalamocortical loops between primary sensory cortices and thalamic nuclei and the 40-Hz oscillations discovered in those systems. Gray does not give us any hint concerning whether his hypothesis is a competing or a complementary one in relation to the others. Also, there is an awkward lack of concordance between Gray and some others about the basic question of whether the hippocampus plays any part in bringing about conscious phenomena. Consider the comments of Churchland (1994, p. 14) in a recent review:

> the hippocampus might have seemed a likely candidate for a central role in consciousness because it is a region of tremendous convergence of fibres from diverse areas in the brain. We now know, however, that bilateral loss of the hippocampus, though it impairs the capacity to learn new things, does not entail loss of consciousness. At this stage, ruling something out is itself a valuable advance.

So, can we rule the hippocampus out or not? Are these disagreements due to interpreting the same data differently, or simply to some parties being unaware of some of the relevant data? If there is no agreement in the field on even the very basic questions, it is no wonder that consciousness research makes progress so agonizingly slow. The lesson is that we need interaction between the different theories of consciousness that draw from very different empirical sources.

An important characteristic of Gray's model is the attempt to integrate features of conscious experience with other levels of analysis. This endeavor reflects the realization that a theory of consciousness must not only grant that consciousness exists, but also provide us with a description of the *explanandum.* While Gray's attempt is certainly respectable, it seems to fall short of providing any genuine insights into the relation between consciousness and brain mechanisms. Why is this? As Gray (sect. 6, last paragraph) notes, we are "waiting for a new kind of theory." But what kind of theory is that? Neither the symbolic paradigm nor connectionism are entirely satisfactory, because they do not capture the relevant levels of organization in nature; there are levels of organization in the brain that we do not yet know about (Bechtel 1994).

If we are realists about consciousness, we grant that there is a level of analysis that includes the facts of consciousness, that is, facts about how things are *for* a subject (Nagel 1993). And this is the very phenomenon that is the starting point of the whole enterprise; the level that we actually want to understand in terms of the lower levels. The greatest obstacle here seems to be that we have no systematic account of the phenomenon *even on its own terms;* only haphazard lists of the features of consciousness, such as the one Gray (sect. 5) presents. I am afraid that not much progress can be made unless those descriptions are based on more methodical foundations.

Understanding consciousness on its own terms, describing the phenomenon accurately, treating it as a true level intrinsic to the mind-brain seems to require a systematic, empirical, scientific phenomenology. Only in case phenomenology can constitute an appropriate *level of description* are we able to treat it as an appropriate *level of explanation.* So far, not much of an empirical, systematic phenomenology is in sight. Perhaps the search for an adequate metaphor of consciousness – Cartesian theatre; multiple drafts (Dennett 1991); virtual reality (Revonsuo 1995) – is the first symptom of the need for an appropriate way of thinking about consciousness.

In sum, I consider Gray's target article an impressive feat of theorizing on consciousness, but at the same time I think it suffers from certain serious deficiencies, perhaps typical of the whole field of consciousness studies at the moment. There are two major problems: first, lack of common ground and interaction between different theories, and second, lack of a systematic description of the phenomenon of interest. The former disadvantage is not difficult to overcome, and I believe that the domain of consciousness studies will be much more integrated in the not too distant future. The latter problem may be more demanding. In any case, it is difficult to see how cognitive neuroscientific theories of consciousness such as Gray's could ever be adequately developed unless we get a better hold of the crucial level of description: the phenomenology of consciousness itself.

ACKNOWLEDGMENT

The author is financially supported by the Academy of Finland.

Communication and consciousness: A neural network conjecture

N. A. Schmajuk and E. Axelrad

Department of Psychology: Experimental, Duke University, Durham, NC 27708-0086. **nestoreacpub.duke.edu**

Abstract: The communicative aspects of the contents of consciousness are analyzed in the framework of a neural network model of animal communication. We discuss some issues raised by Gray, such as the control of the contents of consciousness, the adaptive value of consciousness, conscious and unconscious behaviors, and the nature of a model's consciousness.

The present commentary addresses some of the issues analyzed by Gray in the framework of a neural network model of operant conditioning applied to the description of animal communication. While Gray emphasizes that the contents of consciousness are controlled by novelty, we elaborate on the communicative processes that might control part of those contents.

Schmajuk (1994) presented a neural network model of operant conditioning that describes escape and avoidance behaviors as moment-to-moment phenomena. The model includes (a) a classical conditioning block that builds a predictive internal model of the environment and (b) an operant conditioning block that selects from alternative responses (R). As in Gray's comparator model (sect. 1, para. 7 and Fig. 3), mismatches between predicted and actual environmental events modify (a) the internal model and (b) the behavioral strategies. The model learns to avoid a shock unconditioned stimulus (US) when a conditioned stimulus (CS) is presented by generating the correct avoidance response (Ra). A simulated animal demonstrates many aspects of escape and avoidance behavior shown by real animals.

Schmajuk and Axelrad (1995) modeled animal communication by assuming that two simulated animals interact with the environment and with each other in a "language loop" (Smith 1994) or "verbal community" (Skinner 1957). Figure 1 shows that in the communication model, each simulated animal selects through operant conditioning, not only Rs but also Calls that are input to the other animal. Both Rs and Calls are also fed back to the animal that generates them. The model incorporates a social facilitation property (Thorpe 1963) by which simulated animals will tend to answer a Call with the same Call. Simulated animals can model some animal communication paradigms, such as discriminative avoidance for different alarm calls, imitative learning, and superstitious calling.

It has been suggested (Egger 1904; Piaget 1967) that an important part of the contents of consciousness is a form of "inner speech." This self-communicative property of consciousness is modeled by the feedback Calls of the simulated animals. As is the case for all responses, these Calls are selected by the operant conditioning block.

In the context of the communication model depicted in Figure 1, we now consider some of the issues raised by Gray:

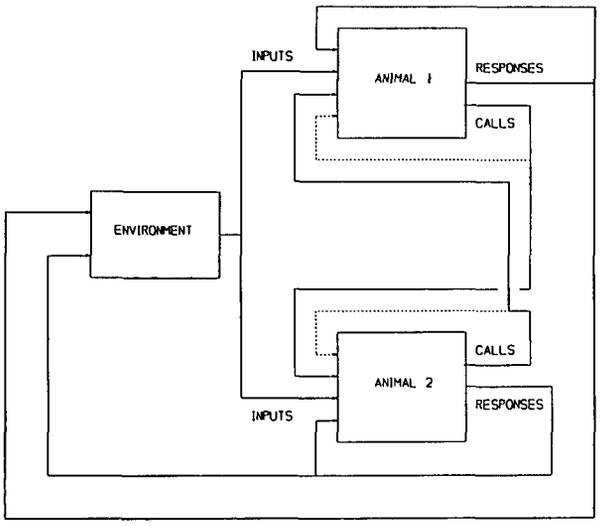

Figure 1 (Schmajuk). Two neural networks of operant conditioning are applied to the description of animal communication. It is conjectured that part of the contents of consciousness are the feedback Calls (dashed lines) within each network. The laws governing conscious processes are assumed to be the same as those ruling communication and, therefore controlled by reinforcement rather than novelty.

1. Contents of consciousness. Gray (sect. 4, last para.) writes that "[the contents of consciousness] consist of multimodal perceptual descriptions derived by comparison between predicted and actual perceived states of the world." In contrast to Gray's conjecture that the contents of consciousness are controlled by novelty, in the communication model, part of these contents are governed by the same reinforcing principles as operant conditioning. This view does not exclude the fact that, as suggested by Schmajuk, Lam & Gray (submitted), novelty strongly modulates the formation of CS-Call associations.

2. Evolutionary advantage of conscious processing. Gray (sect. 1, last para.) argues that "consciousness must make a behavioural difference . . . subject to Darwinian selection and evolution." If part of the contents of consciousness is Call feedback, then the evolutionary pressure selecting for communication also selects for this aspect of consciousness.

The communication model shown in Figure 1 was used to test for an advantage to Call feedback in a cooperative task for the simulated animals to solve with or without Call feedback. In this task, when a CS is presented to both simulated animals, they learn to avoid the US by converging to the same R (cooperation). Simulated animals communicate through arbitrary Calls with no direct effect on the delivery of the US. Figure 2 shows that animals demonstrate more cooperation and, therefore, more successful avoidance when they have Call feedback regardless of whether they communicate or not. Simulations show that the advantage of detecting Calls (whether as feedback or communication) is that Calls amplify the effect of environmental input on the selection of correct responses by each simulated animal, that is, a CS activates the correct response directly through CS-R associations and indirectly through CS-Call and Call-R associations.

3. Conscious and unconscious processing. Gray (sect. 1, last para.) writes "it is precisely the fact that we may leave the status – conscious or unconscious – of such processes undefined . . . that most clearly demonstrates our lack of progress in understanding consciousness." Again, we use the communication model depicted in Figure 1 to better understand the interaction between conscious and unconscious processing in a cooperative avoidance task. During avoidance learning, in the operant block CS1, Call1 feedback, and R1 feedback each contribute to the selection of the

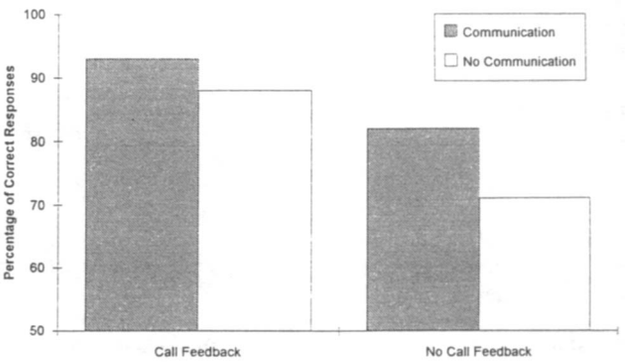

Figure 2 (Schmajuk). Simulated results for different combinations of communication and Call feedback for two animals in a cooperative avoidance task. Percentage of correct responses over 800 trials, for simulations that include (a) Call feedback and Communication; (b) Call feedback and no Communication; (c) no Call feedback and Communication; and (d) no Call feedback and no Communication.

avoidance response R1. Changes in the contents of consciousness are modeled by introducing a second CS2, strongly associated with Call2, which "changes" feedback Call1 to feedback Call2. CS2 can be regarded as a "distractor."

Early in training, both the CS1-R1 and the Call1-R1 associations contribute to successful avoidance. When distractor CS2 is introduced and Call2 is elicited, the absence of Call1 impairs avoidance. Therefore, conscious processing is important at this stage for amplification of the effect of CS1 on the generation of R1. Late in training, the association between CS1 and R1 is sufficient for successful avoidance. When distractor CS2 is introduced, the absence of Call1 does not impair avoidance. Therefore, conscious processing is not important at this stage for the elicitation of R1. The avoidance response is "unconsciously" or "automatically" maintained.

4. Modelling of conscious processes. Gray (sect. 1, last para.) suggests that "if we were to model [comparisons between predicted and actual states of the world] in a computer program. . . . we would surely suppose [the program] to lack conscious components." Since we define Call feedback as a part of conscious processing, and each model generates and detects its own Calls, then, for that fleeting moment of simulation, the model does indeed have one component of consciousness. More generally, just as learning models certainly learn, and models of attention attend, a complete model of consciousness would be conscious.

In sum, some of the issues analyzed by Gray are addressed in the framework of a neural network model of animal communication. Based on the idea that the contents of consciousness and animal communication are intimately intertwined, it is suggested that (1) the contents of consciousness are specified by reinforcement and modulated by novelty, (2) conscious processing is evolutionarily advantageous in cooperative tasks, (3) behavior can be consciously or unconsciously controlled at different stages of training, and (4) models of consciousness might have conscious components.

ACKNOWLEDGMENTS

We are grateful to John Staddon and Owen Flanagan for their comments on an early version of the manuscript. This project was supported in part by contracts from the Office of Naval Research and the Air Force Office of Scientific Research.

Consciousness beyond the comparator

Victor A. Shames and Timothy L. Hubbard
Department of Psychology, University of Arizona, Tucson, AZ 85721.
victor@ccit.arizona.edu

Abstract: Gray's comparator model fails to provide an adequate explanation of consciousness for two reasons. First, it is based on a narrow definition of consciousness that excludes basic phenomenology and active functions of consciousness. Second, match/mismatch decisions can be made without producing an experience of consciousness. The model thus violates the sufficiency criterion.

Although we are intrigued by Gray's conjecture about the relationship between consciousness and a comparator located in the subicular area of the hippocampal formation, we are concerned about two shortcomings of this approach. The first is the unnecessarily narrow definition of consciousness upon which this model is based and the second is the failure of the model to establish that the activity of a comparator is a sufficient condition for conscious experience.

Gray fails to consider two features of consciousness in his definition:

1. Basic phenomenology. Gray states that "the contents of consciousness consist of activity in the subicular comparator" (sect. 3, para. 10). However, he has failed to show the connection between the activity of the comparator and the phenomenology of consciousness. As Jackendoff (1987, p. 14) points out, the effectiveness of neurological models is determined by their ability to show how consciousness "as you or I experience it arises from what our brains are doing." We wonder whether any understanding of the subjective aspects of consciousness is provided by a model such as Gray's that addresses the neurological issues but fails to make an adequate link to the phenomenology of subjective experience.

2. Active as opposed to reactive functions. Gray maintains that "consciousness occurs too late to affect the outcomes of the processes to which it is apparently linked" (sect. 5.3), suggesting that the role of consciousness is merely a reactive or passive one. While such a notion is consistent with the common observation that we do not seem to have any privileged access to the workings of our own cognitive processes, it appears too broad a generalization. Even if consciousness occurs too late to affect perceptual or motor processing, it is clear that it must play a more active role in higher cognitive processes such as planning and problem-solving. We fail to see how Gray's comparator model accounts for awareness of the nonperceptual representations involved in these as well as other abstract thought processes.

In addition to overlooking these aspects of consciousness, Gray has produced a model that fails to accomplish his initial objective of explaining how conscious experiences "arise out of brain events" (sect. 1, para. 4). This model fails as an explanation because the activity of the comparator proposed by Gray is not a sufficient condition for the experience of consciousness. Furthermore, failure to meet this sufficiency criterion occurs regardless of whether Gray's model is treated as an explanatory-causal model or an identity model.

Evidence for the violation of the sufficiency criterion can be observed in studies of implicit memory (for a review, see Schacter 1987). Kihlstrom (1990) defines implicit memory as the effect on experience, thought, and action of events that cannot be consciously remembered. Implicit tests of memory such as stem or fragment completions do not refer directly to a prior study episode; therefore, performance on such tests involves matching the test stimulus with a memory trace that may be nonconscious. To account for implicit memory data, the comparator in Gray's model would have to execute match/mismatch decisions on implicit tests without the subject necessarily being conscious of the output.

When a fragment completion task is used as an implicit test of memory, subjects are instructed to study a list that includes words such as ASSASSIN; they are then asked to complete word fragments such as A____A____I__ to form legal English words (see Tulving et al. 1982). The instructions for the fragment completion task do not refer directly to the study list. Even amnesic patients, who are virtually unable to recall or recognize the previously studied words, show a greater probability of successfully completing fragments that match words on the study list than fragments that do not.

A different type of implicit test involving pictures rather than words was used by Tranel and Damasio (1985) in a paradigm involving prosopagnosic patients. Patients with prosopagnosia lose the ability to recognize previously familiar faces. In this study, patients were shown photographs of familiar and unfamiliar faces. Although they were unable to recognize the familiar faces consciously, the patients demonstrated more frequent as well as larger skin conductance responses to familiar than to unfamiliar faces.

In these two examples, words and faces that cannot be consciously remembered exert an effect on two very different implicit measures. Successful completion of word fragments in one case and increases in skin conductance in the other case can be attributed in large degree to the matching of a test stimulus with a nonconscious memory trace. Although task performance in both examples depends on the execution of a match/mismatch decision by the comparator, the output of this decision cannot result in consciousness for amnesic or prosopagnosic patients and need not result in consciousness for nonpatient populations. Because the output of this match/mismatch decision does not necessarily produce an experience of consciousness, the activity of the comparator is not a sufficient condition for consciousness.

There are two reasons why the comparator model proposed by Gray fails to provide an adequate explanation of consciousness. One is that the aspects of consciousness that Gray takes into consideration fail to include important features such as the phenomenology and active functions of consciousness. The other is that the explanatory link he has proposed between the comparator and consciousness requires that an experience of consciousness occur whenever a match/mismatch decision is made. As we have shown in our examples from the implicit memory literature, this is clearly not the case.

The homunculus at home

J. David Smith
Department of Psychology and Center for Cognitive Science, State University of New York at Buffalo, Amherst, NY 14260.
psysmith@ubvms.cc.buffalo.edu

Abstract: In Gray's conjecture, mismatches in the subicular comparator (needing problem resolution) and matches (during appetitive approach) have equal prominence in consciousness. In rival cognitive views novelty and difficulty (i.e., information-processing mismatches) especially elicit more conscious modes of cognition and higher levels of self-regulation. The mismatch between Gray's conjecture and these views is discussed.

Gray's target article seeks a heuristic and reasonable alliance between consciousness and a monitoring function that oversees the organism's daily coping. This monitoring function is linked to the mismatches and matches encountered by the subicular comparator. I was reading this reasonable proposal easily, paying scant attention, when suddenly I became conscious of a fundamental mismatch. The mismatch I became conscious of was that consciousness is supposed to be about mismatches, but Gray only half endorses this view. This half endorsement could mean that the conjecture is wholly wrong.

A hundred years of cognitive science has urged the view that mismatches, novelty, difficulty, and thwartings foster more con-

scious modes of cognition. In other cases, Dewey (1934/1980, p. 59) noted that "the behavioral impulses will be too smooth and well-oiled to admit of consciousness." James (1890/1952, p. 93) added that consciousness "is only intense when nerve processes are hesitant." Tolman (1932/1967, p. 217) noted that "conscious awareness and ideation tend to arise primarily at moments of conflicting sign gestalts, conflicting practical differentiations and predictions."

Half of Gray's conjecture restates these claims, though none of the ancestors are credited. Gray acknowledges that mismatches in the comparator create a unique cascade of neurological and behavioral events. This cascade (behavioral inhibition, increased attention, exploratory behaviors, extra information uptake, etc.) represents the Behavioral Inhibition System (BIS) in control mode, and Gray ascribes the phobic symptoms of human anxiety to this system.

If the organism's regulation in response to *mismatches* were selectively linked to consciousness, then Gray's view would converge with that of James, Dewey, Tolman, and my own (Smith et al. 1995). But this is not what Gray intends. For he links *matches* in the comparator equally strongly to consciousness. Indeed, a third of the paper incorporates the Behavioral Approach System (BAS) into the model. The BAS is activated by rewarding stimuli, or those associated with the termination of punishment. It allows the organism to move up the spatiotemporal gradient towards reward, perhaps using the waypoints provided by intermediate, conditioned, appetitive stimuli.

The problem is that these approach mechanisms may well be too well-oiled to admit of consciousness; they may need no "extraneous help" (James 1890/1952, p. 93) from consciousness to run off smoothly. They may become, and arguably should become, automatic in Shiffrin and Schneider's (1977) terminology. Including them in consciousness is a striking mismatch to many other views of its contents.

Conscious of this mismatch, I made a Gedanken sandwich. In one scenario, the peanut-butter jar was full, there was a clean spoon, the bread was fresh and whole-wheat. I moved up the peanut-butter gradient smoothly, automatically, and unconsciously, though using Gray's BAS extensively, and lighting thousands of matches in the subicular comparator. In a second scenario, the jar had been misplaced, the spoons were dirty, the bread was stale and rye. Consciousness was acute through all these trying mismatches, and muttering was heard above the exploratory, regulatory roar of the BIS.

Thus, conscious problem-resolution (mismatches), and automatic approach (matches), have an equal role in Gray's conjectured consciousness. He notes this equity in a brief passage in sect. 5.4, last para., having nothing principled to say about either it or the inequity that strongly favors mismatches in consciousness. But the trouble for Gray's conjecture is worse than its silence on this point. The problem is structural, and only by changing conjectures could it be remedied. Once Gray says (sect. 3, para. 1) that "the contents of consciousness consist of the outputs of the subicular comparator," this curious equity is ensured. For the comparator is happily outputting, all the time, both thwarting mismatches and easily flowing matches. Both are in the comparator, and both will equally color consciousness. Game, set, and mismatch with nearly all other ideas about consciousness.

Gray might possibly benefit by changing conjectures. He might have a monitoring function watching the screen of the comparator, poised to jump in with the BIS when problematic mismatches need resolution or close calls need a referee. He might have this same monitor be able to choose to watch the screen of the comparator when familiar beauty is present and elegant, desired matches are unfolding (see his discussion of music in sect. 5.14, para. 2). This new conjecture would still grant to Gray that the subicular comparator is a very interesting brain region. It might really be the big screen, multimodal, multimedia entertainment and information system of the mind/brain.

However, on this view, consciousness would not be in the comparator. Consciousness would be sitting off unknown and shadowed somewhere, choosing to watch the comparator's suspense dramas and its beautiful concerts, possibly an occasional news program or info-mercial, but not the hackneyed, matching reruns. Possibly consciousness, whom we haven't found yet on this view, could even be holding a remote and channel surfing.

Ultimate differences

G. Lynn Stephens and George Graham

Department of Philosophy, University of Alabama at Birmingham, Birmingham, AL 35294. **arhu006@uabdpo.bitnet**

Abstract: Gray unwisely melds together two distinguishable contributions of consciousness: one to epistemology, the other to evolution. He also renders consciousness needlessly invisible behaviorally.

Gray's interesting target article is about how consciousness arises from neurology. In our judgment there is much else in the article worthy of comment. One such topic will focus our commentary.

Gray writes: "Consciousness must make a behavioral difference, otherwise it could not have been the subject of Darwinian selection and evolution" (sect. 1, last para.) He also remarks: "If we knew how consciousness alters the behaviour that it accompanies, we would be able to see what survival value . . . consciousness confers, and so how it might have evolved" (sect. 6, para. 2).

There are two contexts in which consciousness (subjective experience) may make a behavioral difference. The first is Darwinian or evolutionary. To be explained as a product of natural selection, consciousness must alter behavior in ways that confer survival value. The second is epistemological. To justify our belief that organisms (other than ourselves) are conscious, consciousness must make some distinctive or characteristic alteration in an organism's behavior. The alteration may not be criterial for consciousness, but it should at least offer good reason to believe that consciousness is responsible for behavior.

Judging from remarks like those mentioned above, Gray seems to think that the evolutionary and epistemological difference must be one and the same. The behavioral effects that explain the evolutionary difference made by consciousness must be identical to the behavioral effects that constitute the epistemological difference.

Gray takes a gloomy view of the prospects for identifying the epistemological difference made by consciousness. He writes: "Nothing we know so far about behaviour . . . is such that the hypothesis of consciousness would arise" (Introduction, para. 7). What sorts of behavior are distinctive of conscious as opposed to unconscious processes? For Gray, behavior seems silent. Nonconscious information processing can explain "conscious" behavior just as well.

Is there need for gloom? We think not. In our discipline of philosophy, there is now a hefty literature on whether consciousness has behavioral effects. Much of the literature consists of comparisons and contrasts between behavioral performance with and without consciousness (Flanagan 1992; Lahav 1993; Van Gulick 1995). One plausible hypothesis is that the presence of consciousness improves performance even when performance occurs without consciousness. I may unconsciously identify a predator but not as well or effectively as when perceptually conscious of the predator. The "improvement" may consist in one or more of several qualities of behavior, such as alacrity, flexibility, and readiness for recall. Or to put the hypothesis differently, consciousness may be essential to the effectiveness of behavior without playing an essential role in its occurrence.

According to Gray, skepticism about epistemology produces skepticism about whether we can learn of Mother Nature's function for consciousness. In equating the epistemological difference with the evolutionary one, Gray is also gloomy about discovering how consciousness enhances survival.

We will not try to allay fears about whether the biological function of consciousness is discoverable. However, even if the function is undiscoverable, we should not be impelled to equate the evolutionary difference with the epistemological one. The two may well be different. First off, the behavioral difference which warrants our hypothesis that an organism is conscious may play no role in explaining the biological function of consciousness. Suppose that consciousness has a biological function, for example, enhancing predator identification. It is coherent to suppose that in addition to predator identification there is a set of behavioral effects of consciousness (e.g., head scratching) which are completely devoid of survival value. We may be warranted in believing that an organism is conscious because, say, it scratches its head, even though predator identification rather than head scratching improves survival.

In the second place, consciousness may have a biological function without making an epistemological difference. Suppose, for example, that predator identification is essential for survival and is performed consciously. Suppose also that, unknown to us, this same behavior could not be performed unconsciously. Alas, however, suppose that we cannot tell whether the behavior is performed consciously. In such a gray and gloomy epistemological circumstance, consciousness has a function without making an epistemological difference.

There are two tasks which should not be fused from the outset in constructing a theory of consciousness. One (the epistemological) consists in determining which behavior is performed consciously and perhaps needs to be performed consciously. The other (the evolutionary) consists in determining whether consciousness admits of Darwinian explanation and enhances a species' chances for survival. If it turned out that these two tasks are one and the same, that would be an interesting and perhaps unexpectedly Panglossian result. However, contrary to Gray, the result is not required by a successful theory of consciousness. A successful theory can make do with ultimate differences.

Don't leave the "un" off "consciousness"

Neal R. Swerdlow

Department of Psychiatry, University of California–San Diego, La Jolla, CA 92093-0804. **nswerdlow@ucsd.edu**

Abstract: Gray extrapolates from circuit models of psychopathology to propose neural substrates for the contents of consciousness. I raise three concerns: (1) knowledge of synaptic arrangements may be inadequate to fully support his model; (2) latent inhibition deficits in schizophrenia, a focus of this and related models, are complex and deserve replication; and (3) this conjecture omits discussion of the neuropsychological basis for the contents of the unconscious.

Several models suggest that the brain substrates of movement, affect, and thought are structured in parallel cortico-striato-pallido-thalamic (CSPT) circuits, and that consequently movement, affect, and thought are regulated by similar cortical and subcortical interactions (Baxter et al. 1992; Gray, Feldon et al. 1991; Modell et al. 1989; Nauta 1989; Swerdlow & Koob 1987). Extrapolating from these models, Gray proposes specific substrates for the contents of consciousness that include, in a broad sense, the extended CSPT substrates described in his previous BBS target article (Gray, Feldon et al. 1991) focusing on the subicular "comparator" unit as a key circuit element.

That the contents of consciousness are regulated by this extended CSPT circuitry is implied in Gray's previous proposals (Gray et al. 1991a): The substrates that provide order and disorder to thought in schizophrenic patients must be critical for determining the structure of thought in nonschizophrenics. I raise three areas where this model may be challenged.

1. Synaptology. Many elements of Gray's circuitries, particularly the microcircuit interactions, extend beyond established anatomy and physiology. One example is the claim that the subiculum "interrupts all accumbens function relating to motor steps, enabling instead the activation of exploratory behaviour outputs." What neural events might cause such an "interruption"? Some data suggest that the subicular output facilitates DA release via presynaptic activation of DA terminals in the nucleus accumbens (NAC) (Imperato et al. 1990; Wang 1991); other evidence suggests a direct termination of excitatory limbic cortical inputs onto accumbens cell bodies (Bouyer et al. 1984; Sesack & Pickel 1990; Smith & Bolam 1990), which would presumably oppose the effects of presynaptic DA release. The way we "view" this single circuit element – whether NAC glutamate and dopamine function in concert, opposition, or both – could drastically change models for the subiculo-accumbens regulation of behavior. Dopamine and glutamate might interact differently within different accumbens subregions (shell vs. core; see Gerfen et al. 1987; 1991; Groenewegen et al. 1987; Heimer et al. 1991) or functional units (striasome vs. matrix; see Graybiel et al. 1981; Graybiel & Ragsdale 1983). We are struggling to understand how such complex synaptic arrangements regulate simple rat behaviors (Pulvirenti 1991; Swerdlow et al. 1992; Wan et al. 1995). Are we ready to apply such shaky synaptology to questions of consciousness?

2. Latent inhibition (LI). Central to Gray's model for subiculo-accumbens regulation of conscious experience is his interpretation of studies of LI. LI occurs when "a past regularity (CS with no consequence) adversely affects the current learning of a new association." Preclinical studies link disturbances in subiculo-accumbens activity, and accumbens hyperdopaminergia, with the loss of LI. The critical implication vis-à-vis the present proposal is that changes in subiculo-accumbens activity change the conscious experience, that is, the salience of the preexposed CS.

Supporting this theory are two published reports that "acute schizophrenics in the preexposed condition of the LI paradigm learn this association faster than do preexposed normal controls." These important findings deserve careful examination. The total published sample of preexposed acute schizophrenia patients who exhibit LI deficits is, to my knowledge, 22 (Baruch et al., 1988, report 23 subjects divided between preexposed and nonpreexposed groups; N. Gray et al., 1992a, report 9 subjects in the preexposed group). Both studies use a similar audiotape-based LI task; is the phenomenon task-specific? Only patients defined as "acute schizophrenics" have deficient LI, but these patients average 7 years of illness, 3.5 previous hospitalizations, and 500 mg/day equivalents of chlorpromazine (approximately). Can we attribute the loss of LI in these "acute" patients to acute hyperdopaminergia – a critical linkage for Gray's model – since most subjects were taking antipsychotics at the time of testing? Finally, data suggest that schizophrenics learn more slowly than controls in the *nonpreexposed* (NPE) condition (N. Gray et al. 1992a); this contributes to the significant interaction of condition × group, on which claims of "deficient LI in acute schizophrenia" are based. These concerns are raised with full knowledge of how difficult it is to complete such work: It is no surprise that there are just two LI studies in schizophrenics yet dozens of published animal studies that "model" this observation. Still, with critical sample size of 22, in a between-subject design, with nonparametric comparisons of bimodally distributed data, nonautomated data collected with nonblind testing conditions, and possible between-group differences in nonpreexposed performance, we might wish to be cautious in formulating global hypotheses based on this finding.

3. The unconscious. Gray's model of conscious content doesn't provide a mechanism in series or parallel that accounts for the contents of the unconscious. Just as "hardly anyone doubts that consciousness is . . . a product of the brain," so most psychiatrists and psychologists would not doubt the existence of unconscious processes that direct behavior and thought. A hypothesis of the contents of consciousness must accommodate at least the existence, if not the precise neural substrates, of unconscious processes.

Gray argues that "the level of intentionality is the same as that of

conscious experience," and that the contents of consciousness "bear a reasonably close relationship to states in the real environment, since they derive from a process of selection." But, is a "close relationship to real states" necessarily favored in selection? Creativity, fantasy, denial, rationalization, and other products of unconscious processing offer selective advantages. Individuals who cannot effectively utilize these processes suffer negative societal and interpersonal consequences and negative health events, all of which reduce genetic viability. The brain accommodates this genetic pressure by sustaining neural substrates capable of unconscious processing and by sustaining substrates that allow us to recognize, and gain insight into, unconscious processes.

"Preset" expectations – developmentally or genetically determined – provide default parameters for the "match/mismatch" comparator. Such "presettings" guide action that is not based on the "true" comparator readout from the external world. This might occur when our actions are reactions to past situations (e.g., transference, projection, displacement) and, in some individuals, paranoia. The content of many forms of anxiety has *unconscious* origins (e.g., horrific image of infant evisceration in patients with OCD), and unconscious "presettings" appear throughout our conscious experience – from parapraxis to prejudice. Superstition, hope, fantasy, guilt, conscience: all yield "action plans" guided by information that does not consciously enter the comparator. What is the substrate for preset expectations?

The striatal Spiny I matrix may be one substrate capable of "gating" cortical content, repressing unwanted ("ego dystonic") information, and promoting/amplifying desired ("ego syntonic") information. Within this matrix, complex interactions determine which cortical events will be promoted and which will be dampened. The particular cortico-striatal dynamics – the relationship of limbic cortical inputs (source of the "primordial reservoir") to striasome or matrix, and the subsequent efferent targets of these striatal cells (Albin et al. 1992; Dure et al. 1992; Flaherty & Graybiel 1994; Giménez-Amaya & Graybiel 1990; Penney & Young 1986; Ragsdale & Graybiel 1990), the spatial relationships of ascending dopamine terminals and descending cortical glutamate terminals projecting onto the medium spiny cell dendrites (Bouyer et al. 1984; Sesack & Pickel 1990; Smith & Bolam 1990) and onto each other (Imperato et al. 1990; Wang 1991), the distributions of the numerous modulatory neuropeptides (Voorn et al. 1989; Gerfen et al. 1991), the spatial relationships of the matrisomes and their association with tonically active neurons (TANs) (Aosaki et al. 1994), among other unknown elements of striatal organization – are certainly influenced by genetic and developmental forces (Graybiel & Ragsdale 1980). They must have profound interindividual, cultural, racial, and geographically based differences, and these differences must be reflected in templates of "presettings." While the neural substrates of the unconscious are probably distributed throughout CSPT circuitry (at least), connections between different portions of the limbic cortex and the Spiny I matrix are perfectly suited to provide sources for limbic "primary process" information and unconscious cognitive censorship and promotion – key structural elements of the psychic apparatus.

Gray leaves room for striatal censorship in his discussion of consciousness as a "monitoring process." But if not the unconscious brain, "who is watching the monitor?" In his treatment of the "binding problem," Gray notes that "neuronal events that code the . . . perceptual experience are themselves nonconscious . . . [but lead to] a prediction which constitutes the basis of the eventual conscious experience." The characteristics of these neuronal events determine, within certain constraints, the nature of the conscious experience. How we bind information, the structure of "the set that makes up a particular content of consciousness," is a function of our unconscious brain.

Gray's conjecture addresses the *contents* of consciousness, rather than consciousness *per se,* and thus it need not describe the entirety of consciousness or experience, just as a list of ingredients does not fully describe a cake. In the end, we are still left with the mystery of how the parts make the whole. I suggest that a key ingredient – the unconscious – has been omitted, and, without it, the cake cannot rise. Fortunately, much of Gray's proposal comfortably accommodates an unconscious, as if, at an unconscious level, this was his intention.

ACKNOWLEDGEMENTS

The commentator was supported by NIMH R37-MH42228 and R29-MH 48381.

On giving a more active and selective role to consciousness

Frederick Toates

Biology Department, The Open University, Milton Keynes MK7 6AA, England. **f.toates@open.ac.uk**

Abstract: An active role for conscious processes in the production of behaviour is proposed, involving top level controls in a hierarchy of behavioural control. It is suggested that by inhibiting or sensitizing lower levels in the hierarchy conscious processes can play a role in the organization of ongoing behaviour. Conscious control can be more or less evident, according to prevailing circumstances.

I congratulate Gray on his insightful article. There is much to agree with (e.g., the idea of consciousness having to do with monitoring disparity between expected and actual, a check on the efficacy of actions, cf. Baars 1988). However, I would also argue for a more active role of consciousness in organizing ongoing behaviour. I suspect that Gray's long adherence to a behavioural inhibition system model of hippocampal function has delivered a somewhat restricted view of consciousness (vital as an inhibition function is, amongst other roles). Based on rat studies, I would see the hippocampus as mediating the use of cognitive, declarative knowledge in the control of on-going behaviour (Toates 1994a; 1994b; 1994c). I would see it as a high level input in a hierarchy of behavioural control (Gallistel 1980), which in humans is associated with conscious experience (Baars 1988). This fits well with subjective experience or, as Gray puts it, "stubborn intuitions" (such intuitions must surely be allowed to carry considerable weight in the present exercise in theory building, given Gray's partly experiential database).

I would see a slow, serial, cognitive (and, in humans, conscious), mode of control having at its disposal lower, parallel, and faster levels in the hierarchy (Baars, 1988), to which the higher level can allocate greater weighting as a task becomes familiar (Toates 1994a,b,c). [See also Toates "Homeostasis and Drinking" *BBS* vol. 2(1) 1979.] Modulation of the sensitivity of the lower level controls by the higher might indeed at times involve inhibition on their activity so that physically present stimuli that might otherwise trigger behaviour no longer do so. However, in other circumstances, the top-down control might equally involve sensitization of lower level controls. (Maybe this is what Gray would attribute to an anatomically distinct BAS system). I believe that cognitive (hippocampally mediated) control over on-going behaviour is exerted when (a) the current sensory input offers no behavioural solution based upon lower level procedural controls (an addition to Gray's assumptions) and (b) when procedural control fails (i.e., mismatch occurs between the expected and actual consequences of behavior, in agreement with Gray). At a high level in a hierarchy of control, disparity between an expected and actual state of the world would seem to be an ideal candidate to play a role in the *instigation* of on-going behaviour. Indeed, from a control-theoretic perspective (Miller et al. 1960; Powers 1973; Toates 1994b), I find it difficult to imagine how purposive behaviour could possibly be organized in other than such a negative feedback mode.

How then do we answer the puzzle that consciousness appears to be too slow to mediate on-going behaviour? In one sense, this also fits intuition: highly practised tasks (e.g., tying a tie) are often

slowed up and disrupted by paying attention to them. Switching to full cognitive control can indeed mean slowing up, which is the inevitable price (along with the cost of engaging specialized processing capacity) of giving flexibility to behaviour. However, both intuition and the law of the land suggest a role for consciousness even in rapid responses. Possibly the solution is to see consciousness as acting in the role of overall modulator of lower levels in a hierarchy (cf. Gallistel 1980). Thus, conscious processes might not necessarily mediate, say, the finger triggering a gun but a prior aggressive motivational content of consciousness might exert a role in sensitizing aggression-related lower level controls which are unconsciously activated by the sight of a potent target. Similarly, for the most part, Gray might well be justified in his statement of approval of Velmans's (1991) ideas: "we are aware even of what we are saying only after we have said it (sect. 5.2)." However, (a) I believe that I consciously select emotionally and cognitively biased verbal *strategies* and goals to which the actual words are subservient and (b) I know that, when I really want to, I can choose my words carefully (i.e., go into full cognitive, conscious control over a lower level in the hierarchy) but at a cost in speed of presentation.

Similarly, I would want to give consciousness a role in pain over and above simply noting what felt bad for future reference (as important as that role is). The sufferer can sometimes find creative here-and-now solutions to minimizing nociceptive input by monitoring the conscious sensation of pain and what actions lead to its reduction (e.g., try lying on your left side in bed and then gently squeeze the back muscles). Gray is doubtless right in claiming that pain "occurs too late actually to affect responses to the noxious stimulus that gives rise to it" (sect. 5.4) in the case of quickly pulling the foot from a thorn but not in the case of chronic pain.

Gray wonders why we are unaware of our actions in driving during a period over which a good match to expectation occurs, whereas we are aware of listening to, say, a thoroughly familiar record of Tchaikovsky's *Nutcracker*. I would say that consciousness will be alerted by mismatch but, being an active process, it can choose to focus upon even a perfect match if one happens to be motivated to do so and one exerts selectivity within conscious control.

Consciousness does not seem to be linked to a single neural mechanism

Carlo Umiltà[a] and Marco Zorzi[b]

[a]Dipartimento di Psicologia Generale, Università di Padova, 35139 Padova, Italy, [b]Dipartimento di Psicologia, Università di Trieste, 34123 Trieste, Italy.
umilta@ipdunivx.unipd.it

Abstract: On the basis of neuropsychological evidence, it is clear that attention should be given a role in any model (or conjecture) of consciousness. What is known about the many instances of dissociation between explicit and implicit knowledge after brain damage suggests that conscious experience might not be linked to a restricted area of the brain. Even if it were true that there is a single brain area devoted to consciousness, the subicular area would seem to be an unlikely possibility.

1. Attention and consciousness. Patients with unilateral neglect after (right) parietal lesion ignore the affected (left) side of space (e.g., papers in Robertson & Marshall 1993). They behave as if the left half of the world had ceased to exist at a conscious level. Neglect is associated with damage to the system(s) that direct attention to the left side and, perhaps, to the vigilance system in the right hemisphere (Posner & Dehaene 1994; Posner & Peterson 1990). Normal observers too may neglect stimuli in one hemifield when attention is covertly directed to the periphery of the other hemifield (Graves & Jones 1992).

There is ample evidence, however, that neglect patients show normal processing of neglected stimuli in the affected, left side (e.g., Robertson & Marshall 1993). For example, implicit knowledge has been documented by Làdavas et al. (1993). Their patient could not read aloud words presented to the left hemifield, nor could he judge their lexical status or semantic content. He could not even detect the presence of strings of letters in that hemifield. However, response to a word in the intact hemifield was faster when the word was preceded by a brief presentation of an associated word in the neglected hemifield. Similarly, Berti and Rizzolatti (1992) found implicit knowledge in patients with severe neglect. In their study, a prime object presented to the neglected hemifield facilitated responses to target objects presented to the intact hemifield. This effect was also present when the prime and the target were physically different but belonged to the same category.

In conclusion, it would seem that neglect patients do process information originating from the neglected side of space. They ignore this information simply because they are unaware of it. In other words, in the absence of attention, normally processed information does not have access to consciousness. Thus, the role of attention (and that of the neural mechanisms mediating its deployment) must be taken into consideration in proposing a neuropsychological model of consciousness.

2. A single neural mechanism for consciousness. Variants of the dissociation between normal (or near normal) implicit knowledge and severely impaired (or even absent) explicit knowledge have been observed in a number of other neuropsychological syndromes, besides neglect (see, e.g., Schacter et al. 1988; and papers in Umiltà & Moscovitch 1994). Among such phenomena are "blindsight," in which patients with damage to the occipital cortex lack the conscious experience of sight but can nonetheless respond to visual stimuli; memory without awareness in patients who are profoundly amnesic in tests that require conscious recollection; and perception without awareness in prosopagnosic patients who cannot recognize faces at a conscious level but can do so in indirect tests of knowledge.

Basically, three explanations were offered. The first is that intact modules are disconnected from the higher level system(s) that are necessary for consciousness. The second is that damaged modules transmit degraded outputs to the consciousness system(s). The third is that damaged modules are functionally replaced by other modules that do not have access to the consciousness system(s). Explicit/implicit dissociations are domain-specific, in the sense that they always occur in one single domain only. This strongly suggests that consciousness is not subserved by a unitary nervous structure. If there were this unitary system, then lesions to it should produce patients showing the explicit/implicit dissociations across a number of different domains. No such patient has been described so far.

The explicit/implicit dissociation was shown to occur in lesioned neural networks (e.g., Farah et al. 1993, for face recognition). In a neural network there is no mechanism assigned to consciousness, that is, one that is separate from the mechanisms that subserve the (distributed) representations needed to perform specific tasks. Nonetheless, the dissociation between overt and covert recognition of faces showed by prosopagnosic patients was simulated without the assumption of two (at least partly) dissociated mechanisms responsible for overt and covert recognition, respectively. On the basis of Farah et al.'s study, one should conclude that consciousness is linked to a certain minimal quality of information representation within the visual system for processing faces, rather than to a specific neural mechanism.

3. The subicular area. There are at least two patients that invalidate Gray's hypothesis of the subicular area as the localized neural mechanism responsible for consciousness. Patient D. R. B. (Damasio et al. 1985) suffered extensive bilateral damage to the temporal lobe and basal forebrain. The mesial temporal lobes were bilaterally destroyed, so that the lesion encompassed the hippocampus, the parahippocampal and cingulate gyri, and other components of the limbic system (like the amygdala). On the basis of Gray's hypothesis, one would expect that the patient displayed a "nonconscious" behavior. In fact, the patient, though densely

amnesic, was alert, co-operative, behaved intelligently in a variety of situations, and his speech was fluent and well articulated.

A second patient with lesions involving the anatomical structures considered by Gray to be the "hardware" of consciousness is H. M. (Scoville & Milner 1957). This patient had a bilateral lesion to the hippocampal formation, parahippocampal gyrus, and some nuclei of the amygdala, as a result of bilateral ablation of most of the medial temporal cortices. This patient was also densely amnesic, but his consciousness did not seem to be even slightly impaired.

It is interesting to note that patient D. R. B. "formulated no plans for the future and showed no anticipation of events or responses" (Damasio et al. 1985, p. 256). This observation supports Gray's suggestion that the subicular area is a "comparator," that is, a system that predicts the expected events and compares them with the actual (perceived) events. It is quite possible that the results of the computations performed by the comparator system (and possibly the computations themselves) are explicit (i.e., they contribute to awareness). This fact, however, does not imply that these representations are the contents of consciousness *per se*.

In conclusion, if one is committed to the "localized mechanism" hypothesis, the spared cognitive abilities of D. R. B. (and H. M.) rule out the possibility that consciousness is linked to the neural activity in the subicular area.

ACKNOWLEDGMENT

Preparation of this paper was supported in part by grant 9300752-PF41 from CNR.

The limits of neurophysiological models of consciousness

Max Velmans

Department of Psychology, Goldsmiths, University of London, New Cross, London SE14 6NW, England. **mlv@gold.ac.uk**

Abstract: This commentary elaborates on Gray's conclusion that his neurophysiological model of consciousness might explain how consciousness arises from the brain, but does not address how consciousness evolved, affects behaviour or confers survival value. The commentary argues that such limitations apply to all neurophysiological or other third-person perspective models. To approach such questions the first-person nature of consciousness needs to be taken seriously in combination with third-person models of the brain.

What would a theory of consciousness need to explain? According to Gray it would need to explain (1) how consciousness evolved, (2) how it confers survival value, (3) how it arises out of brain events, and (4) how it alters behaviour (Introduction, para. 5). Much of Gray's target article deals with question (3), focusing both on how consciousness relates to the neurophysiological structure of the brain and to the brain viewed as an information processing system. The research literature in this area is extensive (cf. Farthing 1992 and reviews in Velmans 1996) – but Gray's closely argued attempt to relate consciousness to the output of a subicular comparator approaches the issues from an unusual direction in that it draws on the blocking of latent inhibition in laboratory rats and studies of schizophrenia in humans, whereas it is more common in this area to draw on experimentally induced contrasts between conscious and pre- or nonconscious processing in normally functioning human adults (cf. Baars 1988; Velmans 1991) or on clinical dissociations between consciousness and nonconscious functioning within neuropsychology (cf. readings in Milner & Rugg 1992).

Gray's conclusions about the kind of information processing that might support consciousness nevertheless in some respects converge with those of other theorists. It is generally accepted, for example, that any theory that relates consciousness to human information processing needs to deal not only with the diversity of the contents of consciousness, but also with how those diverse contents are constructed into a coherent experience, already integrated, assessed for its novelty or importance, and served up in a way that allows adaptive interaction with the world. Whether this is achieved primarily by a "comparator system," as Gray suggests, or is best thought of in another way depends heavily on the experimental phenomena one seeks to explain. Students of preconscious versus conscious input analysis, for example, usually allocate such functions to either "pre-attentive" or "focal-attentive" processing, whereas those concerned with conscious versus nonconscious control of action tend to speak of a central "executive," "monitoring," or "control" system. Gray's suggestion that input stimuli take around 50 msec to become conscious (sect. 5.3) is also debatable, as there is converging evidence that the first 200 msec or so of input processing is preconscious, which would allow time for analysis, identification, selection, and integration of attended-to input before they enter consciousness (Libet 1993; Neeley 1977; Posner & Snyder 1975). Many would also take issue with Gray's association of consciousness with neurophysiological activity in a *fixed* location. But I will leave the fuller discussion of such details to other commentators – and turn to the wider issues.

Although the bulk of Gray's target article is concerned with neurophysiological modelling, he has a fine grasp of the limits of neurophysiological accounts of consciousness. As he notes (sect. 4), biological forms are embodied in physicochemical processes, but the evolution of biological forms cannot be fully understood in terms of physics or chemistry. Rather, the combinatorial options in the genetic code have a syntax and semantics that can only be understood in terms of selection by consequences, following biological laws of Darwinian survival and Mendelian genetics. In similar fashion, emergent patterns of neural activity can only be understood in terms of their wider adaptive consequences. The syntax and semantics of language, for example, may be embodied in neurophysiological activity, but can only be understood in terms of the constraints required for successful communication between individuals. In short, there is a nonmysterious sense in which genetics does not reduce to chemistry and the higher-order patterning of neural activity does not reduce to neurophysiology.

But in what way do such emergent properties relate to consciousness? As Gray notes, one can give "a perfectly good materialist account of brain events, whether considered under a physicochemical or a syntactic (computational) description . . . without recourse to the *hypothesis* of consciousness." The main reason for bringing in consciousness is because *it is there* (sect. 4, para. 6). However, he argues, there is one further level of organization in the brain which is the same as the level of conscious experience. At this level, "the semantically interpreted content of representations plays a causal role in the unfolding of events." Following Pylyshyn (1986), such semantically interpreted content is also "cognitively penetrable." As with other levels of organization, this "conscious" level does not reduce to lower levels, but again for nonmysterious reasons.

There are difficulties in linking consciousness exclusively to cognitive penetrability, however, as this would imply that sensations which are not cognitively penetrable (alterable by cognition) are not conscious – which would seem to rule out bee-stings, the smell of camphor, and the roar of a jet engine. But these sensations clearly *are* conscious. And there is a more fundamental problem: even if there *were* a distinct level of neural organization to which consciousness corresponds, one would not need to refer to consciousness to explain that organization. Cognitive penetrability, for example, can be explained in information processing terms, without reference to consciousness. In short, although Gray comes closer than most other theorists to accepting the sense in which consciousness seems to be tangential to information processing or neural accounts, he shies away from the logical conclusion that it *is* tangential to such accounts (cf. Velmans 1991).

The reasons, of course, have to do with the "dire" theoretical consequences. The attempt to relate consciousness to neuro-

physiological activity or information processing in the brain is unquestionably important because this will eventually reveal something about the necessary and sufficient conditions (within the brain) for conscious experience, thereby providing an answer to question 3. But if we can explain how such processing fulfills its adaptive purposes *without reference to consciousness* we would seem to be in trouble with questions 1, 2, and 4. And this would produce an awkward gap in evolutionary theory. Gray is acutely aware of the problem. As he notes, "if consciousness is a product of Darwinian evolution, it *must* confer survival value and therefore *must* affect behaviour." Rather than abandon this biological perspective, he proposes to search on for what this influence might be (sect. 6).

In Velmans (1991), I suggest that the function to which consciousness is most closely linked is information dissemination, a late-arising stage of focal-attentive processing that enables the brain to respond to the most pertinent, selected information in an integrated way. Absence of information dissemination produces dissociations of functioning, for example, in blindsight (where part of the system has the ability to make visual discriminations but the system as a whole does not "know that it knows"). Similar proposals have been made by Baars (1988), Navon (1991), and Van Gulick (1991). However, even if this *were* the most closely associated function, the problem outlined above would not go away. Information dissemination, like cognitive penetrability, can be explained in purely information processing terms without reference to consciousness – and the same would be true of *any function that can be described in information processing terms.* If so, a prolonged search for consciousness within an information processing model of the brain is doomed to failure.

Yet, as Gray rightly points out, *consciousness exists* – and there are aspects of consciousness (qualia, what it is like to be something, how things appear from a first-person perspective) that do not seem reducible to either a physical or a functional state of the brain (see discussions between Dennett, Fenwick, Gray, Harnad, Humphrey, Libet, Lockwood, Marcel, Nagel, Searle, Shoemaker, Singer, Van Gulick, Velmans, and Williams in Marsh 1993). Given this, it is time for radical measures.

My suggestion is that we should stop trying to reduce consciousness to a physical or functional state of the brain and start to take consciousness *in the form that we normally experience it* seriously. That is, we need to reincorporate the first-person perspective into psychological science in order that we may properly come to understand its relation to traditional third-person perspective science. There are many arguments in favour of this, and the consequences which follow for our understanding of how consciousness relates to the brain and physical world are radical. I cannot go into these matters in the limited space available for commentaries (but see Velmans 1990; 1991; 1993a; 1993b). In passing, however, it is interesting to note that Gray's questions 1, 2, and 4, which seem to be unapproachable from a purely third-person perspective, seem to be more tractable when they are simultaneously approached from a first-person perspective. For example, if one accepts that given conscious states have correlates that (viewed from a third-person perspective) take the form of given neural representational states, then as neural representations evolve, their conscious correlates will also evolve (question 1). From a first-person perspective, what we perceive, think, feel, and so on has obvious effects on what we do – that is, conscious states alter behaviour in an indefinitely large number of ways (question 4). From a first-person perspective consciousness also has an obvious "survival value," for without it, few human beings would *wish* to survive (question 2). How such first-person facts relate to some expanded evolutionary theory presents a new challenge for science.

Kuhn (1970) has described normal science as a strenuous and devoted attempt to force nature into the conceptual boxes supplied by the existing paradigm. Our understanding of consciousness has been blocked by the attempt to squeeze its manifest first-person nature into a third-person information processing model of the brain. It is time for the squeezing to stop.

Author's Response

Consciousness and its (dis)contents

Jeffrey A. Gray

Department of Psychology, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, England. **spjtjag@ucl.ac.uk**

Abstract: The first claim in the target article was that there is as yet no transparent, causal account of the relations between consciousness and brain-and-behaviour. That claim remains firm. The second claim was that the contents of consciousness consist, psychologically, of the outputs of a comparator system; the third consisted of a description of the brain mechanisms proposed to instantiate the comparator. In order to defend these claims against criticism, it has been necessary to clarify the distinction between consciousness-as-such and the contents of consciousness, to widen the description of the neural machinery instantiating the comparator system, and to clarify the relationship between the contents of consciousness in the here-and-now and episodic memory.

The target article considered essentially three issues: the lack of a scientific, causal ('transparent') theory of the links between consciousness, on the one hand, and brain-and-behaviour on the other; a psychological hypothesis attributing the contents of consciousness to the outputs of a comparator system; and a neurological hypothesis as to the brain substrate of this system. This reply will deal with these issues roughly in that order (Table 1); although, as in the target article, they are closely intertwined.

R1. Metaphysical issues. The target article commenced with the confident assertion that "the main debate (Marsh 1993) now centres on just how to go about determining the way in which consciousness *in fact* relates to brain function." Judging from the commentaries, this was my first mistake. Consciousness continues to ignite considerable controversy of a more philosophical, indeed metaphysical, kind.

The most general complaint was made by **Díaz**, who believes that one has to get the epistemology of consciousness right before tackling the science. He sees the target article as being mired in metaphysical muddle, hovering dangerously between an emergence notion of consciousness, dual-aspect neutral monism, epiphenomenalism, psychophysical identity, functionalism, and property dualism (and possibly all or none of the above). Díaz notes that I have not given these issues careful thought and that, sadly, I am not much concerned about reconciling metaphysical theories. He is right, but I am unrepentant. The history of thought teaches us that metaphysics follows physics, rather than the other way round: consider, as but one example, the influence of Einstein's theory of relativity upon the metaphysics of space and time. This is not to say that conceptual confusion is not to be deplored. But the confusions Díaz points to (especially in the terms used to depict the relationship between brain function and conscious experience, which range from the former "producing" the latter to the two being two different "levels" of functioning of the same system) arise, I believe, not (or not only) because I am confused but because *I don't know* (nor does anyone else) how brain function and conscious experience are linked, only that they are. This state of ignorance,

Table 1

Section	Commentators
R1. Metaphysical issues	**Dennett; Díaz; Ellard; Hurley; Ingvaldsen & Whiting; Kinsbourne; Lloyd-Jones et al.; Lubow; Merskey; Rachlin; Schmajuk & Axelrad; Shames & Hubbard; Swerdlow; Velmans**
R2. The survival value of consciousness	**Dennis & Humphreys; Ellard; Frith; Hemsley; Hurley; Lubow; Merskey; Reeke; Schmajuk & Axelrad; Shames & Hubbard; Stephens & Graham; Swerdlow; Toates; Velmans**
R3. The contents of consciousness	**Eichenbaum & Cohen; Frith; Hemsley; Ivanitsky; Kinsbourne; Merskey; Newman; Reeke; Revonsuo; Schmajuk & Axelrad; Shames & Hubbard; Smith; Stephens & Graham; Toates; Umiltà & Zorzi; Velmans**
R4. The comparator	**Dennett; Dennis & Humphreys; Frith; Hurley; Ingvaldsen & Whiting; Kinsbourne; Merskey; Nelson; Newman; Shames & Hubbard; Toates; Umiltà & Zorzi**
R5. Localized vs. distributed processes	**Dennett; Foss; Kinsbourne; Reeke; Smith; Swerdlow; Umiltà & Zorzi**
R6. Effects of temporal lobe lesions	**Dennett; Eichenbaum & Cohen; Foss; Kinsbourne; Merskey; Newman; Reeke; Revonsuo; Umiltà & Zorzi**
R7. Episodic memory	**Dennis & Humphreys; Eichenbaum & Cohen; Kinsbourne; Merskey; Nelson; Newman; Umiltà & Zorzi**
R8. Anatomy of the comparator system	**Hemsley; Ivanitsky; Kinsbourne; Merskey; Newman; Reeke; Revonsuo; Swerdlow; Umiltà & Zorzi**
R9. Timing	**Eichenbaum & Cohen; Hemsley; Ivanitsky; Reeke; Revonsuo; Velmans**
R10. Episodic memory revisited: Clive Wearing	**Eichenbaum & Cohen; Hurley; Kinsbourne; Merskey; Nelson; Newman; Umiltà & Zorzi**
R11. Psychopathology	**Crider; Frith; Hemsley; Swerdlow**
R12. Conclusion	**Dennett; Kinsbourne; Newman**

indeed, was one of the main messages the target article tried to communicate. Only when it is dispelled – and this is likely to arise from theoretical and empirical advances within science, not from developments in metaphysics – shall we be able to develop a language adequate properly to depict the relationship between brain function and consciousness. Till then, maybe we should invent some suitable nonsense word to clothe our ignorance (the brain "cons" consciousness, perhaps?)

Of the metaphysical options that **Díaz** sees my terminology as leaving open, the one I am keenest to reject is that of epiphenomenalism (Harnad 1994). **Ellard**, in contrast, accepts this possibility (though he finds it unattractive), namely, that consciousness might have arisen as a by-product of some other property of neural tissue, but without consequences of its own; both **Hurley** and **Velmans** likewise seem prepared to consider this option. I reject it, not on philosophical but on scientific grounds: As stated in the target article, given Darwinian theory (the cornerstone of modern biology), consciousness *must* be a product of evolution, so it *must* confer survival value, and therefore it *must* affect behaviour. As Hurley puts it, I am claiming that consciousness must confer "added value" in evolutionary terms. According to S. Harnad (personal communication, Feb. 23, 1995), this position is dangerously dualist (and more unpalatable even than epiphenomenalism): By allowing consciousness to possess causal properties (acting upon and via behaviour) additional to those of the brain functions to which they are linked (by which they are conned?), I open the door to "an extra causal force in nature." However, until we know the details of how the causal properties of conscious events and brain processes, respectively, link up with each other, the metaphysical implications of accepting the existence of the former remain uncertain. Among the commentators, only **Ingvaldsen & Whiting** charge me with dualism. The charge, however, is one of misdemeanour rather than felony: *methodological* dualism, that is, the position that there are different sets of data (those of consciousness and the rest) requiring analysis in different ways, and they have no known connection. To this charge I plead guilty, since such methodological dualism simply reflects, once again, the state of ignorance repeatedly affirmed in the target article. The polar opposite of dualism is usually taken to be mind–brain identity theory (Smart 1959); **Lloyd-Jones et al.** regard the view presented in the target article to be an example. If so, it is not a very good one, since (as noted above) the argument from evolution leads me to suppose, in opposition to the usual form taken by mind–brain identity theory (Gray 1971), that consciousness possesses causal properties of its own. I suppose that being classified in this way as simultaneously a dualist and a mind–brain identity theorist shows once more that I am in a terrible metaphysical muddle. However, as indicated in my reply to Díaz, I am not too upset about this.

The view that there is no Hard Question (**Kinsbourne**) about consciousness, or at any rate none that simple tricks won't make go away, is to be found in a variety of forms among the commentators.

Thus, **Velmans** advises us to talk about things in a first-person way alongside the third-person way, and to accept that both are legitimate. This is not only a reasonable suggestion, but also what most people do most of the time. However, it begs the question – the Hard Question – how

and why are organisms so constructed that both ways *are* legitimate? That staunch defender of radical behaviourism (or, in its most recent incarnation, "teleological" behaviourism; Rachlin 1995), **Rachlin**, so totally ignores Velmans's advice that he manages to talk about even his own experiences in an entirely third-person manner. Thus: "to hear is to consistently discriminate between sound and silence by revealing over time . . . a correlation between overt behavior and sound." Well, yes, that's the way I would determine whether another person or a rat can hear. And, if that's all there were, we would have no problem of consciousness. But, as stated in the target article (and endorsed by Velmans), that problem arises because conscious experiences exist, for each one of us individually (and as reported by each of us to others). Moreover, such experiences are characteristically clearly located in time and space: They are in the here and now. Rachlin, in contrast, affirms that "conscious events (properly understood as behavioral patterns) have no meaning at an instant of time." This startling remark makes me wonder whether what-it-is-like-to-be a Radical Behaviourist might be radically different from what-it-is-like-to-be, well, me for one. Rachlin is, however, at least consistent in his behaviourism: Conscious events are "behavioral patterns" for people as well as for animals. **Lubow**, in contrast, occupies a curiously mixed position. Having allowed that one can empirically demonstrate that human beings and pigeons have similar visual acuity, spectral sensitivity curves, and so forth, he concludes that both have "vision," but that only people have consciousness. Why, from the same behavioural observations, should one make different inferences for pigeon and person; and what is "vision" anyway if it is nonconscious?

The philosophical counterpart of **Rachlin's** position is stated by **Ingvaldsen & Whiting**, who appeal for authority to Wittgenstein (1976) for the assertion that we cannot take our own experience as an independent datum because we lack criteria by which to evaluate it. As they admit, however, this situation can be rescued by way of "collective discourse." Indeed it can, and, at the start of the target article, I pointed to the edifice of psychophysics as testimony to this fact. **Rachlin**, however, is unconvinced by this testimony, since in psychophysical experiments people make "judgements about the presence or intensity of environmental *stimuli*, not about conscious experience." That is a very curious way of describing, for example, colour aftereffects or the waterfall illusion. The natural way of stating what is going on in these experiments is indeed that "collective discourse" (standardised into experimental procedures) is used to validate verbal reports of individual conscious experiences. Starting from the *stimulus*, no Martian would be likely to predict the waterfall illusion; though he might very well have come to predict it (as **Dennett** would surely hasten to point out) by appropriate observations on the brain (cf. Newsome & Salzman 1993). So I remain unconvinced that the privacy of conscious experience constitutes a real problem either scientifically, experimentally, or philosophically (despite the authority of Wittgenstein). Experimentally, there is ample evidence that, under similar conditions, conscious experiences are broadly similar in different individuals; this should be sufficient also to settle the problem of other minds (no longer, according to **Hurley**, taken seriously by many) within philosophy.

As startling as **Rachlin's** statement of Radical Behaviourism is **Schmajuk & Axelrad's** formulation of Strong AI (Searle 1980). They discuss a neural network model of interanimal communication (involving Calls and Call Feedback), including a simulation of self-communications; they take these to correspond to a form of inner speech, and the latter in turn to constitute a part of the contents of consciousness. This allows them to define Call Feedback as a part of conscious processing. They then go on to make the remarkable claim that, as the model generates and detects its own calls, then "for that fleeting moment of simulation" the model has one component of consciousness. "More generally," they write, "just as learning models certainly learn, and models of attention attend, a complete model of consciousness would be conscious." Thus the Hard Question of consciousness is neatly sidestepped in two major respects. First, without the benefit of any experimental study or theoretical justification, it is taken for granted that consciousness goes along with the information processing aspects of brain function, not with its neural aspects, its links to behaviour, or its transactions with an environment (see, e.g., the final paragraphs of the target article). Second, no account is taken of the well-known problem (restated by **Hurley** and **Velmans**, as well as at several points in the target article) that, at least in the current state of our understanding, any function to which we attribute consciousness can in principle be accomplished by a system that lacks consciousness. So there is no reason to suppose that Schmajuk & Axelrad's neural network system does the job of communication with conscious concomitants (nor, conversely, any way of telling that it does not). In this important respect, being conscious is quite unlike learning or attending (so long as the latter terms are themselves understood as referring to information processing minus any conscious concomitants which, were we the learning or attending organisms concerned, we might experience). Unfortunately, sidestepping the Hard Question does not resolve it.

The target article started from the position that there *is* a Hard Question about consciousness: namely, why should brain process + information process + behaviour + environment (in whatever combination of specific tokens of these general elements) come together to produce any conscious experience at all rather than none? I took **Dennett** and **Kinsbourne** to be champions of the opposing view – that there is no such Hard Question. I am relieved (and apologetic) that I got Dennett's position wrong. The target article attributed to him the view (supported here by **Merskey**) that the problem of consciousness will be resolved by the mere gathering of more data. Not so, Dennett's commentary clarifies: he is concerned rather to attack the view that no amount of scientific investigation ("data *plus theory*" Dennett's italics) could ever eliminate the mystery of consciousness. I share with Dennett the view that although one cannot be sure of solving any scientific problem in advance, and although the problem of consciousness is unlikely to yield easily to analysis, there is nonetheless no reason in principle why we should not expect eventually to solve it; and we should certainly not yet give up trying. Furthermore, the addition above of the italicised phrase ("plus theory") closes the gap between our respective views as to how the problem of consciousness will eventually be solved. What is not clear, however, is the nature of the new theory that will be required. I do not myself see how this can come about just from developing more and more detailed theories of particular phenomena,

each deploying already familiar theoretical principles (as implied by Kinsbourne's comment that Easy and Hard Questions will simply merge). Though the target article itself offered just such a minitheory, it also indicated its sharp limitations as a way of approaching the Hard Question. (My acknowledgement of these limitations was ignored by **Ingvaldsen & Whiting, Lloyd-Jones et al.** and **Shames & Hubbard**, who complain that I did not go beyond brute correlations, nor demonstrate any *principled* connections between the different levels of analysis employed – brain, information processing, conscious experience, and so on, but I did not claim to have done so.) However, only time will tell what kind of theorising will eventually prove successful.

Swerdlow turns the Hard Question on its head. He is concerned that the target article provides no mechanism to account for the "contents of the *un*conscious", given that most psychiatrists and psychologists suppose the existence of unconscious processes that direct behaviour and thought. Certainly, I too make the latter supposition: Most of behaviour appears to be controlled by processes of which we have no conscious awareness. But this does not seem to me to pose any particular scientific problem, since we can understand in principle, and indeed in increasingly detailed fact, how the brain does the trick. Swerdlow's commentary, however, implies a particular analysis of these "unconscious processes" – the psychodynamic version – which would perhaps pose a problem, if it were correct, but for which there is little empirical support (Eysenck 1985) or logical necessity [See also BBS multiple books review of Grunhairen's *Foundations of Psychoanalysis BBS* 9(2) 1986]. In this psychodynamic version, the unconscious is endowed with "everything the conscious has, *only minus consciousness*," in Searle's (1992, p. 169) succinct phrase. Searle gives an excellent account of the difficulties of formulating this view coherently. His account distinguishes between nonconscious, nonmental processes (e.g., myelination of axons), unconscious mental processes (e.g., the belief, not at present consciously entertained, that Paris is in France), and conscious mental processes (e.g., the same belief, consciously entertained). His conclusion is that the contents of the unconscious – its "ontology" – "is strictly the ontology of a neurophysiology capable of generating the conscious" (1992, p. 172). [See also Searle: "Consciousness, Explanatory Inversion, and Cognitive Science" *BBS* 13(4) 1990.] This dispositional account of the unconscious is in essential agreement with the view I expressed in the target article that, beyond the level of neural processes, there is only one other level of brain/mind events that needs to be taken into account, namely, that of conscious processes. To the extent that nonconscious neurophysiological processes, short of becoming conscious, gain semantics or intentionality (whether seen through the distorting prism of psychodynamics or more neutrally), this is due to selection by consequences during transactions with the environment, in the relatively unmysterious manner outlined in section 4 of the target article.

R2. The survival value of consciousness. Apart from the few commentators who accepted the possibility of epiphenomenalism (discussed in the previous section), there was general acceptance that consciousness must have some survival value (indeed, some were surprised that this could be questioned), and a variety of suggestions were made as to where to find it. **Stephens & Graham** make a valuable distinction between two contexts for the behavioural difference that consciousness makes: the evolutionary (i.e., the behavioural contribution made by consciousness to survival) and the epistemological (i.e., the behavioural signs by which one would be warranted to infer that another is conscious). As they say, there is no guarantee that the solutions to these two problems will be identical, though it would help if they were. At present, a solution to either would be very welcome.

Lubow is a clear exception to the general consensus that consciousness would be expected to have survival value. Indeed, he sees the target article as taking a "disturbingly teleological turn." Since I was merely applying general Darwinian theory to consciousness, I find this charge puzzling; still more so when Lubow dismisses the notion that consciousness might have a "function" by likening it to such nonbiological phenomena as gravity. It is in application to the *biological* world that the term "function" acquires meaning, namely, the carrying out of some process or other that aids the survival of a replicating organism. No one would seek a function, in this sense, for gravity; biologists would, however, generally seek one for any major characteristic of a species and, whatever consciousness is, it surely fits that description. This surprisingly unbiological position is apparent again when Lubow supposes that consciousness emerged, fully fledged and with no prehuman precursors, when a critical mass of some type of neural tissue was reached in our own species, "as though full-blown from the head of Zeus." Although I cannot refute this possibility, there is no parallel (so far as I am aware) in the rest of biology for such a sudden and massive evolutionary leap. Once consciousness has "emerged" from neural tissue in this dramatic way, it then –according to Lubow – permits many good things, like the building of cathedrals and the writing of books. But these, and here Lubow and I agree, are not the stuff of Darwinian survival. From this, Lubow infers that consciousness is not a product of Darwinian selection, whereas I infer that we need to find something earlier and more basic than cathedral-building to provide the missing survival value.

A rather high-level function for consciousness, that of understanding and communicating with other animals, is suggested in somewhat different ways by **Frith** and **Schmajuk & Axelrad**.

For **Frith**, the survival value of conscious experience is that it enables one to take an intentional stance toward other agents, thus allowing better prediction of their behaviour in the current situation, a point of view previously espoused by Humphrey (1983). It is very likely that conscious awareness is indeed a necessary prerequisite for the taking of an intentional stance. But it is difficult to see how this advantage (though undoubtedly conducive to survival) could, any more than the building of cathedrals, provide the *initial* spur to the evolution of consciousness (one of the reasons that caused Humphrey, 1992, to abandon his earlier position). For evolution to work, there has to be a genetic change upon which Darwinian selection can act. Suppose that this change is sudden (e.g., a mutation) and that, as required by Frith's hypothesis, the consciousness to which it gives rise confers the ability to predict the intentions of others (by way of inferring or empathising with *their* states of consciousness). Useful as this "mind-reading" capacity might be, it can only be deployed if there are other

conscious minds to read. But *this* condition requires that the genetic mutation responsible for the new capacity should simultaneously strike a sufficient number of other animals to cash in its survival value. This is such an unlikely scenario that it can, I think, be safely rejected. (A similar argument, by the way, precludes saltatory evolution of language capacity; see Pinker & Bloom, 1990.) Thus, we need to seek a prior survival value for consciousness: only when this has spread to large numbers of other animals can the additional value of mind-reading do any work.

For **Schmajuk & Axelrad**, the basic building block (instantiated in a neural network model) is the kind of communication that might occur when one animal utters an alarm call that is perceived and reacted to by another; this part of interanimal communication is seen as involving nonconscious processes. They then proceed to postulate that animals are sensitive to feedback from their own calls, and demonstrate that, in their model, this further capacity confers a behavioural advantage in a simulated cooperative avoidance task. (**Dennis & Humphreys** describe a neural network model of a very different kind of behaviour – learning the visual referent of a word. Interestingly, their model too shows a clear behavioural advantage when the model is equipped with the capacity to detect feedback from its own vocal outputs.) On the basis that an important part of conscious experience consists of inner speech, Schmajuk & Axelrad finally link their model to consciousness by supposing that the feedback from an animal's own calls achieves awareness. If so, they argue, the behavioural advantage of sensitivity to call feedback, demonstrated in the neural network model, can provide the survival value needed for consciousness to evolve. This argument probably avoids my criticism of **Frith's** position, that is, that simultaneous evolution would be needed in many organisms, provided that (as would no doubt be readily testable within the model) behavioural advantage accrues when only one of two cooperating animals is sensitive to call feedback. What the model fails to do – a failure that applies, *mutatis mutandis*, to all current hypotheses (as discussed in **Hurley's** and **Velmans's** commentaries), including my own – is to explain why sensitivity to call feedback should require awareness. This postulate is simply imported from introspective human knowledge that inner speech is an important part of the contents of consciousness. But, even on this basis, the postulate is curiously arbitrary, since we are also aware of other people's speech (not to mention a lot of other things). So why not declare everything (or nothing) in the Schmajuk & Axelrad model to be conscious?

I am not sure why **Schmajuk & Axelrad** select inner speech as the route by which to approach the problem of consciousness; perhaps they regard this as relating primarily to those things in our own awareness that others don't know about (unless we tell them), as distinct from entities in the external world of which others too are aware. **Frith**, in contrast, is explicit in treating the contents of consciousness as consisting above all of "shareable knowledge," which he further defines as representations coded in allocentric rather than egocentric coordinates. Let us call these contrasting aspects of consciousness "internalized" and "externalized," respectively. In agreement with Frith's position, I believe it to be highly unlikely that such internalized processes as hearing one's own voice in subvocal speech, imagining a previously observed visual scene, and so on, could have developed without the prior capacity to consciously appreciate the externalized auditory, visual, and other characteristics of the world in which we move.

Several other commentators, however, follow a strategy similar to **Schmajuk & Axelrad's**, focusing in their search for the function of consciousness upon internalized processes. **Shames & Hubbard** and **Reeke**, for example, emphasise the contributions of conscious thought to problem solving and planning. Planning is also picked out by **Merskey**, who comments on the role played by the visualized outcomes of actions in this process. I see no problem in accepting that, in these ways, conscious processes do influence behaviour, and that this influence is beneficial to survival (Reeke). Indeed, the target article itself suggests that we should seek the function of consciousness not in transactions with the environment as they actually happen, but in the modification of such transactions for future use. But, as indicated above, I do not see how such internalized uses of conscious awareness could have arisen without there first being qualia that are referred to the external world. And these are not the consequence of any planning (not, at any rate, of the conscious variety). On the contrary, they just seem to occur, but – and this is where the problem of survival value arises – they occur too late to be involved in the processes of perception and action to which they are apparently linked (Velmans 1991). This problem cannot be completely solved by appeal to behavioural advantages that accrue slowly and late as the result of internalized conscious processes (as in Merskey's example of arrival at a destination as the result of planning the imagined journey), though such behavioural advantages may ultimately provide part of the solution.

Swerdlow goes even further in favouring internalized over externalized aspects as providing the survival value of consciousness. The target article (section 4) takes the view that the contents of consciousness "must in general bear a reasonably close relationship to states in the real environment, since they derive from a process of selection by consequences." Swerdlow wonders, however, whether such a close relationship to real states would necessarily be favoured by evolutionary selection, suggesting in contrast that such products of unconscious processing as creativity, fantasy, denial, and rationalization may offer selective advantages despite, or indeed because of, their divergence from reality. Such advantages, Swerdlow suggests, may be found in improved social and interpersonal functioning and reduced negative health events. While these advantages may be real to some degree and/or in some instances, they are not incompatible with the view that the contents of consciousness must *in general* reflect rather accurately the state of the external environment, or none of us would survive for very long. One who is creative, fantastical, denying, or rationalizing to the extent that his perceptual world differs radically from that of others is normally hospitalized as psychotic and usually has substantially reduced chances of procreation. Swerdlow also points to the role of "preset expectations," genetically determined or arising from early experience, in determining the outputs of the comparator. The existence of such biasses is undeniable. A good example is to be found in Hebb's (1946) observation of laboratory-reared chimpanzees who respond to a shaking rope as though it were a snake. But such examples imply, precisely, that Darwinian selection has built in a "snake template" that is well suited to the recognition of real snakes, albeit with sufficient potency

also to announce "snakes" when none are there. Such exceptions to the rule are only possible given the rule, namely, that perceptions are normally more-or-less well adapted to reality.

The suggestion that the survival value of conscious experience lies in future transactions with the environment finds favour with **Hemsley**, and **Ellard** and **Toates** each suggest how this could be mediated by a (conscious) contextual set, influencing subsequent rapid (nonconscious) reactions to specific stimuli. (**Swerdlow** makes just the opposite point, that unconscious contextual biasses may determine the contents of consciousness. The two suggestions are not in conflict; both types of biassing, with their mysterious comings and goings between conscious and unconscious processing almost certainly occur.) Ellard's suggestion is of particular interest, since it concerns just that type of behaviour – risk assessment – in which the comparator hypothesis set out in the target article itself originated (Gray 1982a; 1982b). **Stephens & Graham**, in a similar vein, suggest that, though consciousness may not be essential for the occurrence of a particular form of behaviour, it may nonetheless increase its effectiveness (though Toates reminds us of the many instances in which conscious attention to skilled performance can ruin it). These suggestions have the merit that one can see how they might permit gradualist evolution (*pace* **Lubow**, who dislikes this notion) of consciousness – if we could only see how to get it into the system in the first place.

R3. The contents of consciousness. The central hypothesis proposed in the target article concerned not the nature of consciousness itself, but the contents of consciousness: namely, that these consist in the outputs of a comparator system. **Revonsuo** points out that, in order to evaluate this or any comparable proposal, it would help to have a systematic empirical phenomenology; **Velmans** similarly urges us to take the phenomenology of conscious experience more seriously. It is hard to disagree in principle with these suggestions, but also hard to forget the scientific morass that was the last wave of introspectionist psychology at the turn of the century. So guidance is awaited as to how to establish the missing systematic phenomenology. I was under the impression that I had in any case taken the phenomenology of conscious experience seriously, by testing the comparator hypothesis against a range of features of the contents of consciousness that were in many cases derived from introspection; I am therefore puzzled by the accusation (**Shames & Hubbard**) that I primarily addressed the neurological issues but failed to make a link between them and the phenomenology of subjective experience. Certainly, I concur that such a link will be an essential feature of any eventually successful theory.

The general notion that the contents of consciousness consist of the outputs of a comparator was welcomed by **Toates**, who agrees that consciousness has to do with monitoring the disparity between expected and actual events; by **Newman**, who points to the considerable similarities (acknowledged in the target article; see section 3) between this concept and "global workspace" theory (Baars 1988; Newman & Baars 1993); by **Ivanitsky**, who proposes a rather similar neural underpinning for the comparator; and by **Hemsley. Kinsbourne, Smith & Hall**, and **Umiltà & Zorzi** give more guarded approval to the target article's view of the contents of consciousness, commenting that the comparator might *contribute* to the selection of contents for consciousness. **Reeke**, however, contrasts my application of the comparator concept unfavourably with Edelman's (1989) application to categorisation rather than comparison. **Merskey** suggests that the contents of consciousness appear to be the comparator's processing rather than its outputs. But, in making this point, he is referring to the internalized aspect of consciousness, that is, processes of evaluation, appreciation, and decision making. As pointed out above, these processes are probably derivatives from the primary role of consciousness in reflecting the external world; it is this role that the target article addressed.

Frith puts forward a view of the contents of consciousness that he sees as different from that contained in the target article, but that I see as very similar. He emphasises the value of "contrastive analysis" (as do **Newman** and **Stephens & Graham**), namely, establishing contrasts between information that is conscious and information that is not. (Although perhaps not made explicit, this approach was important also in developing the hypothesis proposed in the target article; see, e.g., section 5.2.) On this basis, Frith distinguishes between *motor* and *cognitive* representations: the former being coded in egocentric coordinates, closely linked to motor output, and unconscious; the latter being coded in allocentric coordinates, conscious, and the basis of knowledge that can be shared with others. The latter, therefore, constitute the contents of consciousness, along with awareness of possible actions that the subject can undertake with regard to them. This view is highly compatible with that of the target article. According to this, the chief function of the comparator system is to check that the outcomes of motor programs are as expected (section 2; and see Gray, 1994). Frith's distinction between motor and cognitive representations appears to encapsulate essentially the same point as the formulation in section 5.2 of the target article, namely, that "we are not normally aware of either the planning or the execution of movements as such, but only of the end points (which may include kinaesthetic and proprioceptive feedback making up the 'feel' of the movement) that constitute the successive sub-goals of these movements." Frith's point that such subgoals are primarily coded in allocentric coordinates is a valuable extension of this formulation (although the elements contained in the bracketed phrase also form part of the overall story). His emphasis on allocentric coding is, in addition, consistent with the role allotted in the target article to the hippocampal system, given the well-established role of this system in the analysis of allocentrically organised spatial information (O'Keefe & Nadel 1978).

Several authors comment upon the role of reinforcement in determining the contents of consciousness. **Hemsley** takes up the difficulty I have in accounting, within the theory, for the difference between those "match" outcomes of the comparator that enter consciousness and those that do not (see the discussion of the driving and music examples in section 5.14); he suggests that such outcomes may only enter conscious experience when reinforcement systems are activated. In **Schmajuk & Axelrad's** model, the contents of consciousness are said to be sensitive to both reinforcement and novelty. Also addressing the same difficulty, **Toates** proposes that one can choose to focus conscious awareness even upon a perfectly familiar stimulus if one is motivated to do so (which may simply be another way

of saying "given sufficient positive reinforcement"). I welcome these attempts to get me out of my difficulty but doubt their effectiveness. Within the conceptual framework provided by reinforcement theory, a motor program that is running successfully is, by definition, one that is in receipt of positive reinforcement (for approach behaviour) or negative reinforcement (for escape or avoidance behaviour; see Gray, 1975). Thus, one cannot account for the difference between those outcomes of such a program that do or do not enter consciousness by appeal to reinforcement, as in **Hemsley's** suggestion. There is little point, therefore, in translating Toates's proposal into the language of reinforcement theory, as I did above. Toates himself does not seem in any case to have intended it in that way. Rather, he stresses the active nature of consciousness, which can itself "choose" to focus upon a familiar piece of music. Although I recognise, introspectively, the experience **Toates** describes in this way, I am puzzled by the logic of an attempt to explain a feature of consciousness by having consciousness itself create the feature. **Smith** propose more radical surgery to deal with the problem: In line with what he correctly claims to be a long tradition (see his references), he suggests that only mismatch outcomes from the comparator reach consciousness (**Eichenbaum & Cohen**, however, depart from this tradition and take exactly the opposite view, namely, that "little of conscious experience is concerned with novelty detection".) But the Smith move simply lands us onto the other horn of the dilemma, as instanced in the familiar musical example (target article, section 5.14). So I am still left with the difficulty of providing a principled account of the difference between those predicted outcomes of motor programs of which one is or is not consciously aware.

R4. The comparator. **Nelson** too takes up the problem posed by the difference between match outcomes of which one is or is not aware, but believes it to be a pseudoproblem. Why, he asks, do I suppose that conscious and nonconscious comparator processes should differ anyway?

Both **Nelson** and **Shames & Hubbard** point out that there are plenty of basic comparator processes that function outside of conscious awareness (for some good examples, see Shames & Hubbard and **Umiltà & Zorzi**), and the brain certainly contains multiple comparator systems (**Dennett, Merskey**). Furthermore, as Shames & Hubbard remark, I have not demonstrated that the activity of the comparator proposed in the target article is a sufficient condition for the occurrence of conscious experiences. **Hurley**, too, makes this point. Unlike (apparently) Shames & Hubbard, however, she realises that, not only did I not claim to have made any such demonstration, but (like her) I do not see why *any* such "complex adaptive feedback system" *should* generate conscious experiences. That, precisely, is (a version of) the Hard Question about consciousness. So, even assuming that the target article is right in assuming that a comparator process contributes in some way to the contents of consciousness, there is nonetheless no understanding of why the outputs of this comparator (but not others) should achieve consciousness. We do not even know whether the difference between comparator outputs that do and do not achieve consciousness lies in the comparator process itself (a possibility that Nelson implicitly rejects) or is due to other factors. Even within the comparator process postulated in the target article, part of it (the outputs) achieves consciousness, but part of it (the comparison process) does not. Nor is this just an arbitrary theoretical move. As **Kinsbourne** points out, supporting this feature of the model, we do not first experience an event and then its unexpectedness, so the match/mismatch decision must itself be unconscious. Thus, in answer to Nelson's question, we know, empirically, that there *is* a difference: Some comparator processes achieve consciousness, and others do not; even for those types of comparator processes that do achieve consciousness, part of them does and part does not. Clearly, these differences stand in need of *some* explanation. Whether that explanation will turn out to call upon differences in the underlying comparator systems themselves is not yet known. This hypothesis cannot at present be rejected, but, like Nelson, I doubt it.

It is accepted, then, that there is nothing about comparator processes that requires consciousness. Nonetheless, the notion that a comparator process might have *something* to do with consciousness met with some approval. **Hurley** and **Nelson** point out, indeed, that it is quite common in current moves in this area to appeal to feedback devices and comparator systems; **Dennis & Humphreys** indicate the role played by a comparator in their computational model of episodic recognition; and **Newman** points out that global workspace theory, like the target article, treats perceptual enitities that engage conscious attention as varying in some degree from current expectations and as being relevant to current goals.

More generally, **Nelson** takes me to task for not acknowledging my debt to the overall cybernetic literature in which the concept of the comparator has its origin (a debt pointed out also by **Ingvaldsen & Whiting**), and especially to the work of Powers (1973), whose relevance is indicated also by **Toates**. I freely acknowledge that debt. Indeed, control theory has been integral to my own thinking since as an undergraduate I was exposed, even before Powers's work, to the papers of Kenneth Craik (1943); it influenced my first amateur attempt at quantitative theorising (Gray & Smith 1969). Powers's view of behaviour as the control of perception, cited approvingly by Nelson (and echoed by **Hurley**), is highly congruent with the stance adopted in the target article: It provides a rationale for the fact that the contents of consciousness contain the goals, but not the programming, of action – a key phenomenon that the comparator hypothesis attempted to address. Also from the perspective of cybernetics, Nelson suggests that I paid insufficient attention to the fact that the brain consists of a *hierarchy* of self-regulatory feedback systems, while Toates proposes that consciousness is associated with a high level of such a hierarchy. Within a framework of this kind, Nelson points out, the match–mismatch process would not be a simple binary output, but rather would consist of a range of discrepancies with respect to specifications at a variety of levels of the hierarchy. These are all points I accept. In addition, Nelson asks how the system I described deals with the case of *positive* discrepancy, that is, when perception exceeds expectation. I find this question intelligible only if one supposes the excess to have a positive valence in motivational terms, for example, more food or money than was expected. I have discussed the behavioural and emotional consequences of such positive discrepancies elsewhere (Gray 1975), but I do not see them as throwing any particular light upon the problems addressed in the target article.

The comparator hypothesis is seen in especially interesting ways by **Hurley** and **Dennett**. As Hurley comments, it is widely agreed that a key feature of conscious experience is perspective, that is, a point of view held by creatures who are "agents and perceivers," actively moving through and interacting with an environment. In its emphasis upon the keeping of an ongoing record of the relationships between motor output and perceptual feedback, the notion of a comparator is able to capture some of the essence of this point of view. Hurley also discusses a second key feature of consciousness, "reflection," that is, the capacity to sustain states that are in some way about their possessor. This capacity is outside the scope of the comparator model, at least as so far framed (though perhaps not beyond the model advocated by **Frith**). Nonetheless, it seems possible that further analysis of the comparator hypothesis may serve to act as a bridge between philosophical, computational, and physiological approaches to the problem of consciousness. Dennett's comment is couched quite differently from Hurley's, but points to the same possibility: "the most interesting – if tacit – claim made by Gray's model . . . is that this comparator task is so important that this is a place in the brain to look for major convergence of the syntax and semantics of the whole system." I am content that Dennett has acted as midwife to this, indeed tacit, feature of the model, and delighted to claim (joint?) parenthood. Between them, then, these two commentaries see the comparator hypothesis as potentially valuable in the approach to two of the hallmarks of consciousness: perspective and semantic competence.

Dennett takes me to task, however, for ignoring the possibility that in the functioning of the comparator process (1) consciousness might be composed of lots of unconscious elements, and (2) a sufficiently powerful combination of "merely syntactic" phenomena might account for semantic competence. In fact, I find both of these possibilities – which Dennett implies, and I find likely (target article, section 4), to be actually one and the same – perfectly acceptable, so long as we do indeed regard them as possibilities, not as certainties. That is to say, I can see no reason *in principle* why consciousness and semantics should not arise out of the operation of a comparator system composed of nonconscious entities that instantiate a syntax. Nobody, however, yet knows how the trick is done, so it is risky to suppose that, when we do know, it will fit this description.

R5. Localized vs. distributed processes. The target article proposed not only that the contents of consciousness consist in outputs from a comparator system, but also that this system consists in a number of specific structures within the brain. This attempt at localizing some of the processes critical to conscious experience met with a number of objections, even unto the dreaded bogeyword (**Smith**): homunculus! For some reason, the attribution of a special role in consciousness to only part of the brain rather than the whole is seen by some as tantamount to inviting the little chap in to listen to what the chosen part is saying. (**Swerdlow**, in an interesting twist, suggests that it is the unconscious brain that is doing the listening.) Thus **Dennett** remarks that, by concentrating all the power of mentality into the subicular system, I force this to be composed of "wonder tissue." **Reeke**, similarly, states that, by postulating a single area where the outputs of the comparator are interpreted to generate conscious thought, I invest the theory with a "not-so-hidden homunculus." (Reeke himself, in trying to account for some of Libet's curious experimental findings, goes on to propose that different aspects of awareness might occur in different sequences "as perceived elsewhere in the system." If I wanted to trade bogeyword for bogeyword, this is homunculism in spades!). But one might as well level the same charges about the brain as a whole. Somewhere there *is* some wonder tissue, that is, tissue that gives rise to conscious experience: Is the degree of wonder inversely proportional to the amount of tissue? If so, maybe we should seek for the physical substrate of consciousness in the body, because it's bigger than just the brain. There is nothing in principle that requires that an account of conscious experience should appeal to "the brain and the person as a whole" (**Foss**) or to global properties of the entire system (Dennett, Reeke), though it may turn out that way. Foss sweeps this whole issue away by appealing to the analogy of the focal point of a lens: just as this is not additional to the glass of which the lens is constructed, so the "centre of consciousness" is just "a functional consequence" of the organization of information streams available to consciousness. He may be right, but again we must wait until we know how the trick is done. In short, the homuncular insult is best ignored till it becomes obsolete (as it will when we know what wonder tissue *does*).

The choice between the whole brain and some brain subsystem(s) or other as "the seat of consciousness" (**Dennett**) has to be made, I believe, upon empirical, not philosophical, grounds (for an interesting recent example to do just this, see Crick & Koch 1995). In the target article (section 3), I argued against Kinsbourne's (1993) ecumenical view that any representation can contribute to consciousness, partly on the basis of Jackendoff's (1987) analysis of the contents of consciousness as always being intermediary constructions arising from interactions between bottom-up and top-down processing. **Kinsbourne** sees this Jackendoffian analysis as compatible with his own view, namely, that entry into consciousness is a function merely of the strength and coherence of cell assemblies momentarily in competition with each other to form a dominant focus of neural activity, since those cell assemblies that cohere well will usually represent finished percepts. [See Amit: "The Hebbian Paradigm Reintegrated" *BBS* 18(4) 1995.] Although this is reasonable, it does not address the other problem posed (target article, section 3) for the ecumenical view, namely, the fact that the contents of consciousness almost entirely exclude motor processes (although, of course, they include the outcomes of motor programs). There are surely intense, enduring, and coherent cell assemblies shuttling round the motor cortex and basal ganglia during, say, a game of tennis: Why do they not produce some awareness of the signals that are being sent to muscles? [See Jeannerod: "The Representing Brain" *BBS* 17(2) 1994.] Facts like these lead inescapably to the conclusion that the brain's consciousness is not egalitarian: it attaches to some cell assemblies, but not others; and it does so, not randomly or meritocratically (in terms of mere quantity of firing), but systematically and qualitatively. This conclusion does not, of course, entail the further inference that conscious experience is linked to brain activity in only one region or system; the privileged parts (while still not encompassing the whole brain) could be widely scattered (and even independent of each other, a position defended

by **Umiltà & Zorzi**, but which I find unattractively complicated).

R6. The effects of temporal-lobe lesions. The weakness of **Kinsbourne's** and **Dennett's** ecumenical view, however, does not necessarily imply that the *processes* responsible for consciousness are located in the regions where the privileged cell assemblies circulate. It remains an open question whether the "wonder tissue" is there, somewhere else, or everywhere. But it was not the purpose of the target article to propose a hypothesis about the location or nature of such tissue. That would be a hypothesis about consciousness-as-such; such a hypothesis, able to meet Nagel's (in Marsh, 1993) requirement of transparency, has yet to be proposed. The hypothesis in the target article, rather, was one about *the contents of consciousness*.

The distinction between these two types of hypothesis is especially important when one comes to consider the effects of lesions to those brain regions to which the comparator function is attributed. The target article (section 3) cites Libet's objection (personal communication, June 17, 1993) that the neuroanatomy attributed to the comparator predicts that "destruction of the subicular area should abolish the contents of consciousness" and admits that there is no evidence to support this prediction. Libet's point was enthusiastically echoed by the commentators (**Dennett**, **Eichenbaum & Cohen**, **Merskey**, **Newman**, **Reeke**, **Revonsuo**, **Umiltà & Zorzi**). Upon rereading the section of the target article in which I put Libet's objection, I see that I did not myself clearly distinguish between the effects one might expect after damage to (1) a region critical for consciousness as such, or (2) a region responsible for organising the contents of consciousness. It is not surprising, therefore, that some of the commentators (but not **Foss**) also blur this distinction. Thus, Eichenbaum & Cohen, Merskey, Reeke, and Revonsuo all point out that bilateral loss of the hippocampus does not entail "loss of consciousness" (Churchland 1994); Newman similarly comments that the neuropsychological evidence rules out the subicular region as "the *sine qua non* for primary consciousness." I accept these facts and the conclusion drawn from them, but deny that the target article predicts anything different (though regretting that this point was not made sufficiently clear before).

R7. Episodic memory. A move that several of the commentators (**Eichenbaum & Cohen**, **Kinsbourne**, **Merskey**, **Newman**, **Umiltà & Zorzi**) urge upon me is to suppose that damage to the comparator system eliminates only the capacity to enter conscious experiences into episodic memory for subsequent recognition or recall. A move of this kind is supported by the effects upon this type of memory of hippocampal and other temporal-lobe damage.

Eichenbaum & Cohen, for example, agree with my approach to hippocampal function in supposing that this is concerned with comparisons between items and events, but they suggest that this affects only memorial processes, namely those creating and updating networks of cortical memory representations. From a different point of view, **Dennis & Humphreys** draw out some interesting parallels between their neural-network model of recognition memory and certain functions (the generation of predictions, the match/mismatch operation, and the prediction of the next perceived input as analysed by the brain's sensory systems) ascribed in the target article to the comparator system. **Kinsbourne** focuses upon episodic memory (a category to which Dennis & Humphrey's recognition memory and Eichenbaum & Cohen's declarative memory also belong), defined as the conscious reexperiencing of an episode. This commentary is particularly valuable, as it fills in an important, and hitherto missing, link between the comparator hypothesis and this form of memory. As Kinsbourne puts it, to remember an episode one needs a retrieval cue that uniquely characterizes it, so that "the memorandum can be singled out from countless generically similar events." Consider, therefore, a particular episode, say, walking along a familiar street to carry out a familiar shopping routine. Even if you have performed the routine hundreds of times before, so that the comparator system is correctly predicting most of the stimuli you observe along the way, there is also bound to be much that has never before been put together in quite the same way as now: There will be a particular combination of weather, leaves on the ground, passing cars, your emotional state, and so forth, that is unique to this occasion. The overall output of the comparator system will be a moment-by-moment multidimensional description of the passing scene, each component tagged according to the degree and nature (**Nelson**) of match/mismatch. As in Eichenbaum et al.'s (1994) *BBS* treatment of the functions of the hippocampal system, this output is used to update cortical memory stores, and, as suggested by Kinsbourne, those parts of it that are unique (carry a mismatch tag) are able to act as retrieval cues for episodic (conscious) memory.

R8. The anatomy of the comparator system. An issue raised by several of the commentators concerned the anatomy of the postulated comparator system and its relations with other brain regions that have been linked to consciousness.

Hemsley wonders why there is any need to include feedback from the comparator to cortical perceptual systems, since all the necessary information is already present in the comparator itself. **Kinsbourne** provides the answer to this question: Although there are indeed projections that feed into the hippocampus from all sensory and other relevant systems, it is implausible that "any single region of the brain might integrate spatially all the fragments of sensory and motor activity necessary to define a set of unique events" (cited by Kinsbourne from Damasio, 1989). Thus, as argued in the target article (section 3), it is necessary to include feedback from the comparator to those sets of neurons in perceptual systems that have just provided input to the comparator in respect of the current process of comparison. **Reeke** finds such feedback insufficient, claiming (1) that the comparator could not be located in the hippocampal formation, and (2) that it would have to be widely distributed throughout the cerebral cortex. The first part of this statement is based upon the data from patients with hippocampal lesions considered above (section on temporal lobe lesions). I am not sure whether Reeke has independent grounds for the second part of the statement; they are perhaps those summarised by **Umiltà & Zorzi**, namely that lesions to specific cortical areas give rise to domain-specific losses of conscious awareness. Finally, questions were asked about the connection between the comparator hypothesis, with its emphasis on the hippocampal system, and work by others on (1) the extended reticular–thalamic activation system (ERTAS), lesions to which at either the reticular or the thalamic level produce

coma (**Newman**); and (2) the thalamocortical loops linking primary sensory cortices and thalamic relay nuclei (**Revonsuo**). It is possible, I believe, to answer these questions, while at the same time (1) satisfying Reeke's requirement for proper inclusion of cortical structures in the comparator; (2) providing a mechanism by which output from the comparator system is able to intervene in Kinsbourne's competition between cell assemblies for "prominence in experience"; (3) clarifying further the vital distinction between consciousness-as-such and the contents of consciousness; and (4) linking up with the various other scientists whose work Revonsuo accused me of having ignored.

In part, the answer is provided by **Newman** himself, who points to the projections from the hippocampus (via the subiculum and fornix) to the thalamus as the route by which the comparator process attributed to the hippocampal system can influence the ERTAS. However, an important additional route has recently become apparent from work in Grace's laboratory. Lavin and Grace (1994) have studied what happens to the outputs from the nucleus (n.) accumbens (to which the subiculum also projects via the fornix) further downstream. Using electrophysiological and tract-tracing techniques, these workers have demonstrated that the inhibitory GABA-ergic output from the n. accumbens synapses, in the ventral pallidum, upon further GABA-ergic inhibitory neurons that project to the nucleus reticularis thalami (NRT). The NRT is unusual among thalamic nuclei in that it consists mainly of inhibitory GABA-ergic neurons; these project to a number of the surrounding thalamic nuclei whose job is to relay impulses originating in peripheral sense organs to the appropriate sensory regions of the cerebral cortex. The possible role of the NRT in the selection of stimuli for attention and conscious processing was first pointed out by Crick (1984) and has been incorporated into a neural-network model by Taylor (1992). Note that, since the pallidal output to these neurons is itself inhibitory, its activation has the effect of disinhibiting these sensory relay pathways, that is, increasing the entry to the cerebral cortex of those stimuli that are currently engaging the thalamocortical loops. Figure 1 presents this circuitry in diagrammatic form (a similar hypothesis, though different in detail, has been applied to schizophrenia also by Carlsson & Carlsson, 1990).

Let us consider how the circuitry of Figure R1 would be likely to work under the conditions of an experiment in which an indirect dopamine (DA) agonist, such as amphetamine or nicotine, is used to block latent inhibition (LI) by causing DA release in n. accumbens (see target article, sections 2 and 5.17, in which the relevance of such experiments is outlined, and Gray et al., 1991a, and in press). The basic phenomenon of LI consists in the fact that a preexposed conditioned stimulus (CS) is slow to enter into an association with a Pavlovian unconditioned stimulus (UCS). In line with the general argument pursued in the target article, we can interpret this as reflecting a lack of access to conscious processing by the preexposed CS. If, however, presentation of this CS is accompanied by enhanced DA release in n. accumbens (as induced pharmacologically, by activation of the retrohippocampal input to n. accumbens, or during acute psychosis), LI is overcome, indicating *ex hypothesi* that the preexposed CS has regained the capacity to engage conscious processing. The circuitry of Figure R1 constitutes a mechanism by which this effect can be produced. DA release within n. accumbens inhibits (by acting on DA D2 receptors; Robertson & Jian 1995) the GABA-ergic pathway to the ventral pallidum, thus disinhibiting the pallidal GABA-ergic pathway to NRT, which in turn inhibits the GABA-ergic projections from NRT to the ascending thalamocortical sensory relay projections, thus disinhibiting the latter. In this way, accumbal DA release should lead to an intensification of processing *in whichever thalamocortical sensory relay projections were already operative in the prior instant of time.* In the LI experiment, this intensification of sensory processing will allow the preexposed CS (which otherwise would not have been fully processed) to enter more readily into association with the UCS.

Given **Revonsuo's** request to relate the present model to the work of others, I note that the proposal illustrated in Figure R1 is similar to some of the ideas expressed, but in a more general form, by Damasio (1989). Gaffan (1994), too, proposes a similar hypothesis, but in relation to episodic memory for scenes, or "snapshot memory," and utilising the pathway returning from hippocampus to cortex via the mammillary bodies and the anterior thalamus rather than the route via n. accumbens; these differences apart, Gaffan's approach has much in common with the one suggested here. I hope also that **Swerdlow** will see Figure R1 (adapted as it is from one of his own; Swerdlow & Koob 1987) as somewhat mitigating his charge that many elements of the circuitry proposed in the target article, and the earlier papers cited therein (Gray 1982a; 1982b; Gray et al. 1991a), reach beyond the established anatomy and physiology.

Swerdlow's charge is to a considerable degree justified, though I have tried to take account of such features of the relevant anatomy and physiology as *are* established. In my defence, if theory does not attempt to extend beyond established facts, it is unlikely to be of much use in the search for new ones. Swerdlow particularly asks how the subicular input to the accumbens interacts with the mesolimbic dopaminergic input. This is indeed, as he points out, a difficult issue, since there appear to be numerous different such modes of interaction (see also Gray et al. 1991a; 1991b). One possibility, however, is the following: we have shown that intrasubicular injections of excitatory amino-acids, or of an antagonist (bicuculline) to the inhibitory transmitter, GABA, give rise to DA release in the n. accumbens; and that this effect can be blocked by section of the fimbria-fornix, which carries the subicular projection to n. accumbens (S. Mitchell et al., in preparation). This effect could be mediated by either or both of the routes that Swerdlow sees as potentially in conflict with one another: presynaptic activation of DA terminals in n. accumbens, or a direct excitatory termination on accumbal cell bodies. The second of these routes could involve (see Figures 5 and 6 in the target article): (1) glutamatergic subicular activation of accumbal GABA-ergic cells; (2) accumbal GABA-ergic inhibition of GABA-ergic cells in the ventral pallidum, with (3) consequent disinhibition of the A 10 dopaminergic cells that receive the GABA-ergic pallidal output; and (4) increased DA release from A 10 terminals in n. accumbens. The further consequences of elevated intra-accumbal DA release are specified in Figure 1. "Are we ready to apply such shaky synaptology to questions of consciousness?" Swerdlow asks. Perhaps not. But we are readier, I think, by the year; and readier than we are to apply it to the psychodynamic concepts (e.g., unconscious cognitive censorship) chosen by Swerdlow himself for similar treatment.

Despite **Swerdlow's** cautionary remarks, I propose (as

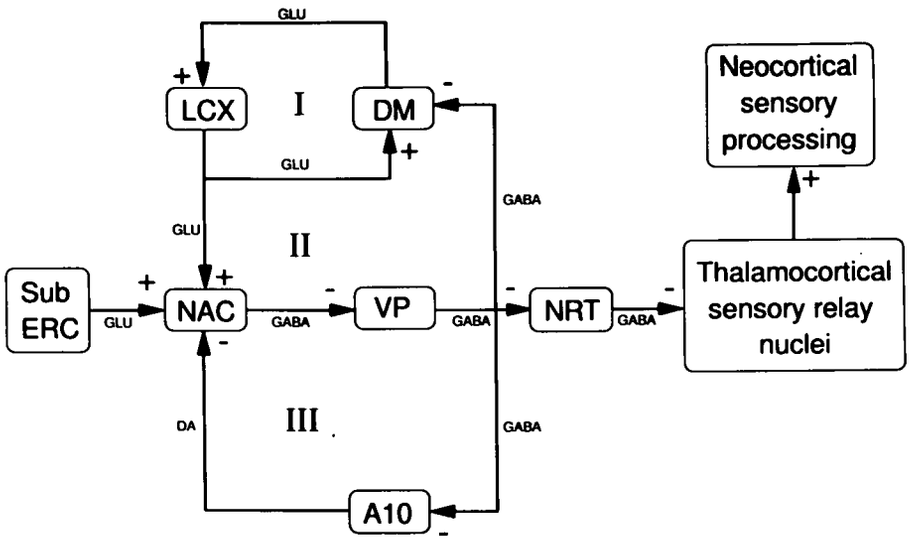

Figure R1. Connections from the subiculum (sub) and entorhinal cortex (ERC) to the n. accumbens (NAC) component of the motor system (see Figures 5 and 6 in the target article), and from that system to the nucleus reticularis thalami (NRT) and thalamocortical sensory pathways. LCX: limbic cortex, including prefrontal and cingulate areas. DM: dorsomedial thalamic nucleus. NAC: nucleus accumbens (ventral striatum). VP: ventral pallidum. A 10: dopaminergic nucleus A 10 in the ventral tegmental area. GLU, GABA and DA: the neurotransmitters, glutamate, gamma-aminobutyric acid and dopamine. + −: excitation and inhibition. I, II, III: feedback loops, the first two positive, the third negative.

an extension of the hypothesis proposed in the target article) that intensification of cortical sensory processing, consequent upon the activity of the hippocampal system as fed back to the thalamus via n. accumbens, constitutes the neural machinery by which the outputs of the comparator determine the contents of consciousness. (**Ivanitsky** makes a similar, though less detailed, proposal.) In this way, **Reeke's** requirement that the contents of consciousness must comprise large portions of the activity of individual sensory areas is met, and a neural basis is provided for the selective facilitation of **Kinsbourne's** cell assemblies in the competition for "prominence in experience." Furthermore, the comparator theory is now able, as required by **Newman** (to whose thoughtful commentary the extended hypothesis owes much) and **Revonsuo**, to make proper contact with the reticulo-thalamic core that (judging from the coma caused by lesions thereto) provides the substrate for the *consciousness* into which the comparator system inserts (cons) the *contents*. This extended hypothesis can also provide an answer to **Merskey's** question concerning the anatomical location at which pain affects the comparator system. Given the known involvement of the thalamus in nociception (Melzack & Wall 1983), this is a region in which pain could provide a signal to interrupt the ongoing comparator process (target article, section 5.4). Even supposing, however, that the extended hypothesis is correct, it would, of course, still fall short of a transparent theory of how the brain cons consciousness; but it may at least serve to unify two unsolved problems into one.

R9. Timing. As in the target article, the extended hypothesis takes an "instant" (i.e., the time taken for the output of the comparator system to be computed and fed back to thalamocortical sensory processing pathways) to be of the order of 100 msec, consistent with the proposed overall neural machinery and especially with the period of a hippocampal theta wave. This "timing" aspect of the model met with a mixed response. **Hemsley** comments that the linkage between the immediately preceding output of the comparator machinery (used to generate an expectancy), its current output (a comparison involving that expectancy), and the subsequent output (to be based on a further generated expectancy) is consistent with James's (1890; 1892) emphasis upon the continuity of the stream of consciousness. **Reeke**, however, makes just the opposite inference from the theory – that it predicts a "chunkiness" in the stream of consciousness, based upon the succession of discrete comparator outputs (a point made also in the target article, section 5.5). This may be an issue that would repay careful investigation with the methods of modern experimental cognitive psychology. It may also be possible to investigate it by studying types of psychopathology in which the stream of consciousness appears to be interrupted (Hemsley).

Reeke argues that the lack of prominence of the theta rhythm in primates rules it out as the main timing system. But, though spectral analysis is needed to demonstrate it clearly, theta is certainly present in the human EEG; the reason it stands out clearly in recordings from, for example, the rat, probably reflects morphological and experimental factors rather than fundamental differences between the rodent and primate brain (Gray 1982a). **Ivanitsky** presents interesting theoretical arguments and data (psychophysical and evoked-potential) consistent with the duration of an instant proposed both by Edelman (1989) and in the target article. **Velmans**, in contrast, believes that I have set its duration too short; **Eichenbaum & Cohen** see the theta cycle (oddly) as being shorter than the proposed theoretical instant. However, the frequency of theta, though commonly about 7–8 Hz, can be as low as 4 and as high as 14 Hz, giving "instants" ranging from 70 to 250 msec, centred on about 130 msec. The phrase "of the order of 100 msec" was not meant to be more precise than this; the known moment-to-moment variation in theta frequency (Gray 1982a) in any case rules out a fixed value to an instant. The issue is further blurred by fact that the moment of occurrence of any

particular stimulus varies with respect to the point within the current timing cycle, thus also affecting the relations between the timing of two successive stimuli. The major contrast of the "theta" hypothesis is with the view (**Reeke, Revonsuo**) that consciousness critically depends upon the faster, c. 40-Hz rhythms favoured by Crick (1994) and others (C. M. Gray & Singer 1989). While I accept that these faster rhythms may play an essential role in binding the different features of a single perceptual entity together, they imply an instant (c. 25 msec) that is too fast to readily account for entry into awareness. In contrast, as an order of magnitude, the timing given by theta is about right. The extended hypothesis, proposed above, can also account for the much longer value of an instant (500 msec) suggested by Libet et al.'s (1991; target article, section 5.3) experiments: the thalamic stimulation in those experiments would first need to establish appropriate reverberatory activity in thalamocortical loops as a candidate content of consciousness before this activity could be conned into consciousness.

R10. Episodic memory revisited: Clive Wearing. As we saw, several commentators (**Eichenbaum & Cohen, Merskey, Newman, Umiltà & Zorzi**) are willing to see the comparator hypothesis applied to episodic memory, but draw the line at conscious experience. Yet the difference between conscious experience in the (apparent) here-and-now and episodic memory may lie only in the duration of the interval that elapses between the initial impact of environmental stimulation and perception or memory respectively. As illustrated in Figure R1, the comparator hypothesis supposes that, approximately 100 msec after activity triggered by sensory stimulation has initially reached thalamocortical sensory processing pathways, a further signal arrives from the comparator system reflecting the degree and nature (**Nelson**) of computed match/mismatch relevant to the triggered cortical activity; and that this match/mismatch signal, combined with the existing pattern of cortical activity, becomes (is conned into being) the next content of consciousness. The lateness with which the signal from the comparator reaches the thalamocortical pathways implies that conscious experience is already, in Edelman's (1989) term, the remembered present. Later on, a pattern of activity, similar to the one originally conned into consciousness, can be reinstated into episodic memory by the occurrence of a feature (or features) of the original input (initially novel, and so uniquely able to define an episode; **Kinsbourne**). If the comparator were damaged, then, one might expect to see alterations in both episodic memory and in at least some aspects of conscious experience.

A dramatic single case study (whose implications were drawn to my attention by Nicholas Rawlins) supports this view. This patient (Clive Wearing) was the subject of a 1986 U.K. television programme ("Prisoner of Consciousness," Equinox, Channel 4; and see Wilson et al. 1995; Wilson & Wearing, in press). Magnetic resonance imaging shows extensive bilateral damage (the result of herpes simplex encephalitis) to both hippocampal formations, both amygdalae, the substantia innominata on both sides, both temporal poles, the left fornix, the left inferior temporal gyrus, the anterior portion of the left middle temporal gyrus, the anterior portion of the left superior temporal gyrus, and the left insula. In addition, there may be damage to both mammillary bodies; the thalamus, however, is intact. Clive Wearing has extremely severe anterograde and retrograde amnesia, with essentially total loss of episodic memory. A highly talented and successful professional musician, he has extraordinarily intact musical skills, is capable of reading and playing music and conducting a choir, has fluent, well-articulated speech and writing, and is even still able to translate from Latin to English.

It is Clive Wearing's own description of his condition that is remarkable. He feels constantly that he has "just woken up" and keeps a diary in which this feeling is repeatedly recorded, at intervals of hours or even minutes. Upon so "waking up," he regularly remarks upon the absence of conscious experience in the preceding period of time and the sudden reemergence of conscious experience in the present moment (the following statements are transcribed from the television programme): "I have been blind, deaf, and dumb for so long"; "suddenly, I can see in colour"; "I can hear traffic sounds now"; "today is the first time I've actually been conscious of anything at all"; "none of my senses are working at all"; "I have no sense of taste"; "I have no sense of touch or smell"; "it's like being dead" (this particular remark is repeated many times). Clearly, Clive has consciousness in the sense that he is not in a coma and enters a waking state that contrasts with the state of sleep. Clearly, too, he retains enough of the normal contents of consciousness to be able, from time to time and fleetingly, to comment upon the loss of these contents of consciousness during the periods between his "wakings up." But, if we are to take him at his own words, equally clearly, he has suffered a remarkable encroachment upon the normal continuous richness of the contents of consciousness. Furthermore, this encroachment is not modality specific, thus providing a counterexample to the cases described by **Umiltà & Zorzi**, in which impairments of conscious processes were always domain specific.

Wilson and Wearing (in press) describe an aspect of Clive Wearing's experience that fits particularly well with the argument advanced here.

> Clive was not able to retain impressions for more than the briefest moment. The effect was of course that the environment appeared to be in a state of flux. On many occasions he would comment, "You weren't wearing blue just now. No, no, it was yellow or pink but it certainly wasn't blue." In response to the constantly new appearance of the room Clive would keep asking, "How do they do that?" One day he put this phenomenon to the test. He held a chocolate in one hand and repeatedly covered and uncovered it with the other. He could feel that the chocolate never moved, yet each time he uncovered it, it appeared to be a new chocolate, however quickly he looked (Wilson & Wearing, in press).

This account suggests that damage to a comparator system located in the temporal lobe prevents Clive from forming a "neuronal model of the stimulus" (Sokolov 1960), with the consequence, first, that his conscious experiences of the here-and-now are radically altered; second, that those experiences are forgotten so rapidly that each new moment presents itself to him as a fresh awakening from a period of total loss of consciousness; and, third, that he has complete loss of episodic memory. Thus, contrary to the view expressed by many commentators (see above, section R6), it is premature to rule out a contribution of this region of the brain to the nature of conscious experience (though not to consciousness-as-such).

Umiltà & Zorzi consider in some detail another single

patient ("DRB"; Damasio et al. 1985), who had bilateral damage to the hippocampus, the parahippocampal and cingulate gyri, and other components of the limbic system. They take (wrongly; see above, section R6) the comparator hypothesis as predicting that such a patient should display "nonconscious" behaviour, and present against this prediction the fact that DRB, though densely amnesic, was "alert, co-operative, behaved intelligently in a variety of situations, and his speech was fluent and well articulated." This shows, indeed, that DRB was not asleep or in a coma. But what does it tell us about the subjective quality of his experience? It is at this point, confronted by someone with deeply compromised behavioural capacities and a severely damaged brain, that we have to take seriously, I believe, the "zombie" (**Hurley**) possibility, usually explored only as an abstract theoretical gambit (Harnad 1994). It is a matter of common experience that one can carry out highly complex intelligent behaviour, such as driving (target article, section 5.14), for quite some time without conscious memory of there having been any subjective experience accompanying it. This is also true of speech, which can occur even during behavioural sleep; I once observed this in a roommate, a native English speaker, who regularly spoke German in his sleep but had no knowledge of this when I woke him up. So, given DRB's dense amnesia, the degree to which he has qualitatively normal conscious experiences while manifesting "intelligent behaviour" must remain in doubt.

Clive Wearing's condition is again revealing in this connection. As noted, prior to his illness he was a gifted professional musician. After the illness, despite his dense amnesia, he was still able to conduct, sight-reading, a choir singing a highly complex classical piece. But in the 1986 television programme, immediately after the piece has finished, he is shown to lack any understanding of what has just taken place, saying to the choir "when I waved, you sang." Thus, the alteration in Clive's conscious experience is closely linked to an impaired grasp of intentionality. He does not grasp that his waving his arms is used by the choir to guide their singing, behaving in fact like the man in the "Chinese room" in Searle's (1980) famous *gedankenexperiment*, translating from notes to arm-waving without understanding the meaning of what he is doing. Note, in this context, that the playing of music is typical of behaviour that runs too fast for consciousness to be involved in the moment-to-moment organisation of the sequence of steps in the motor program. Thus, despite the fact that, more than any other human activity, musical behaviour is incomprehensible without qualia, it appears that Clive Wearing was probably conducting the choir in automatic mode.

R11. Psychopathology. The theory presented in the target article drew upon a previous attempt to model the positive psychotic symptoms of acute schizophrenia (Gray et al. 1991a; 1991b). It is interesting to note, therefore, that several of Clive Wearing's symptoms resemble those sometimes seen in schizophrenia, including perceiving the familiar as novel (e.g., the episode with the chocolate, described above), hallucinations of hearing his own music played, and delusional beliefs about how he comes to be in his present condition (Wilson & Wearing, in press). (The converse is also true, in that schizophrenics show disturbances in memory; McKenna et al., 1990.) **Hemsley** extends the model to the "sense of self," suggesting that this corresponds to the consistent manner in which stored material normally operates on sensory input; a breakdown in the functioning of the comparator system would then lead to the disturbance in the sense of self often seen as characteristic of schizophrenia. **Crider** describes a number of further schizophrenic symptoms (segmentalized consciousness involving experiences of unreality and fragmentation) that he sees as falling naturally out of the account of the contents of consciousness given in the target article. He also notes that these symptoms may be accompanied by a posture of behavioural immobility and fixed attention, outputs that one might reasonably expect if the behavioural inhibition system (of which the proposed comparator system forms part) were strongly activated by perceived mismatch. These convergences are encouraging. However, they generate the problem of then differentiating between the mechanism that produces positive psychotic symptoms (thought to occur, as pointed out by Hemsley, when mismatch signals arise from malfunctioning of the comparator circuitry) and the one that produces anxiety (which occurs when the environment gives rise to genuine mismatch signals). As at least a partial solution to this problem, Hemsley cites data showing that anxiety is, as one might therefore predict, prominent in the early stages of schizophrenia and associated with positive symptomatology (Penn et al. 1994). Thus, differentiation between schizophrenic symptoms and anxiety may only become clear-cut as the schizophrenic illness proceeds (as outlined by Hemsley).

Swerdlow advises caution in relation to the Gray et al. (1991a; 1991b) model of schizophrenia. As he points out, a large theoretical edifice has been built on the basis of a very small number of patients who have served in experiments on latent inhibition (although data obtained with the Kamin blocking effect are also relevant to the model; Jones et al. 1992). However, there are now three (not two) replications of the basic effect, that is, absence of LI in acute schizophrenics, provided either that they are in the first two weeks of current neuroleptic medication (Baruch et al. 1988; N. S. Gray, Hemsley et al. 1992) or that they are in the first year of a current unmedicated episode (N. S. Gray et al., in press). Furthermore, the fact that, in the preexposed condition, the acute schizophrenics actually outperform normal controls provides strong evidence for a genuine loss of the LI effect, even if it is sometimes also the case that nonpreexposed patients learn less well than controls. Nonetheless, further replication in laboratories other than our own clearly remains of great importance. Swerdlow also asks whether the difference between the performance of patients in the first two weeks of a medicated episode is properly attributed to dopaminergic hyperactivity, given that neuroleptics should be counteracting this. There is, of course, a general problem in understanding how and why dopamine receptor blockers produce their delayed effect on psychotic symptoms, of which the pattern of results with LI is a further instance. But it is not the only experimental paradigm in which this pattern is found. Both the Kamin blocking effect (Jones et al. 1992) and negative priming (A. Pickering, personal communication) show a similar disruption in the first two weeks of a current medicated psychotic episode with normalisation later. This recurrent pattern must reflect some underlying neural mechanism; at present it seems likely that this includes in some way activity in the mesolimbic dopaminergic system.

Frith finds even less of value in the Gray et al. (1991a) model of schizophrenia. In accord with his view of con-

sciousness as mediating high-level interactions with other human beings (see above, section R2), he sees the aberrant conscious experiences of the schizophrenic as reflecting a breakdown in the mechanisms subserving such interactions. Thus, whereas one might be able to account for hallucinations *per se* along the lines of the comparator model (**Hemsley**, indeed, sketches out one such account), Frith claims that it would not be applicable to the content of such hallucinations, since it would miss out an essential component concerning the intentions and actions of other people, for example, that hallucinated voices are emanating from a powerful, controlling being. I regret seeing Frith distance his current approach to schizophrenia (Frith 1992) from both ours and his own earlier approach (Frith 1987); the latter were in many ways compatible with each other. **Crider** comments on the lack of emphasis in our model upon frontal executive function, especially in the light of evidence for frontal hypofunction in schizophrenia. In part, our relative silence on this issue arose because we saw Frith (1987) as having provided an account of the frontal contribution to schizophrenic cognitive abnormality that was complementary to our own (see discussion in Gray et al., 1991a). There is certainly no contradiction between our emphasis upon overactivity in dopaminergic transmission in n. accumbens and, for instance, Weinberger's (Weinberger et al. 1986) upon frontal hypofunction, since it has been shown in the rat that interventions in the temporal lobe that enhance dopaminergic transmission in n. accumbens decrease such transmission in frontal cortex (Lipska et al. 1992; 1993; Pycock 1980).

R12. Conclusion. The first claim made in the target article was that there is an unsolved Hard Question (**Kinsbourne**) about consciousness: namely, how to create a transparent (Nagel 1993, p. 4), scientific, causal theory of the links between consciousness and brain-and-behaviour. The commentaries provide no reason to depart from that view; nor, I fear, have we made any progress towards answering the Hard Question (though I hope we have clarified some of the issues that surround it). I was sorry to see no one address the issue, raised in the penultimate paragraph of the target article, of how one might attempt at least to choose between the two quite different stances adopted by those who believe that the answer to the Hard Question will lie in the realm of information-processing or in the realm of brain science, respectively. A successful turn at this choice point would serve better to focus future efforts (a suggestion as to how to go about choosing is made by Gray, in press).

The second claim in the target article was that, psychologically speaking, the contents of consciousness might consist of the outputs of a comparator system; the third consisted in a description of the neurology instantiating the comparator system. Though these claims are logically distinct, a number of issues have been clarified as the result of putting them up for commentary together. Most important, the distinction between consciousness *per se* and the contents of consciousness has been drawn more sharply, and I have accepted that the neural machinery proposed to instantiate the comparator process probably does not also underlie the creation (connation?) of consciousness-as-such. Drawing upon Lavin and Grace's (1994) recent experimental work, I have proposed an extension (Figure R1) to the model contained in the target article that links the outputs of the comparator system to the reticulo-thalamic core which, as set out in particular by **Newman**, seems likely to underlie the generation of consciousness-as-such. Also, I hope, clarified (especially in the light of **Kinsbourne's** commentary and the description of the patient, Clive Wearing, by Wilson et al., 1995, and Wilson & Wearing, in press) is the relationship between the role of the comparator system in generating the contents of consciousness, on the one hand, and episodic memory, on the other. These extensions to the model as initially outlined in the target article have considerably widened the regions of brain involved in the overall system for generating both consciousness-as-such and its contents: This now includes, besides much of the limbic system and basal ganglia, the reticulo-thalamic core and the thalamocortical sensory processing systems. Nonetheless, it is still clear that only some brain processes end up represented in consciousness and that the selection of those that do is not simply based upon quantitative considerations such as intensity or coherence of firing (contra Kinsbourne). Thus, there is still a bottleneck of some kind, even though the neck is rather wide. It is unclear whether this is a homuncular "appreciating bottleneck" (**Dennett**) or not, but the Hard Question about consciousness is no easier if it is the whole brain, rather than part of it, that does the appreciating.

References

Letters a and r appearing before author's intials refer to target article and response respectively.

Albin, R. L., Makowiecz, R. L., Hollingsworth, Z. R., Dure, L. S., Penney, J. B. & Young, A. B. (1992) Excitatory amino acid binding sites in basal ganglia of the rat: A quantitative autoradiographic study. *Neuroscience* 46:35–48. [NRS]

Amsel, A. (1962) Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review* 69:306–28. [aJAG]

(1992) *Frustration theory.* Cambridge University Press. [aJAG]

Andreasen, N. C., Rezai, K., Alliger, R., Swayze, V. W., Flaum, M., Kirchner, P., Cohen, G. & O'Leary, D. S. (1992) Hypofrontality in neuroleptic-naive patients and in patients with chronic schizophrenia. *Archives of General Psychiatry* 49:943–58. [AC]

Anscombe, R. (1987) The disorder of consciousness in schizophrenia. *Schizophrenia Bulletin* 11:241–60. [AC, DRH]

Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A. M. & Kimura, M. (1994) Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *Journal of Neuroscience* 14(6):3969–84. [NRS]

Baars, B. J. (1983) How does a serial, integrated and very limited stream of consciousness emerge out of a nervous system that is mostly unconscious, distributed, and of enormous capacity? In: *CIBA symposium on experimental and theoretical studies of conciousness,* ed. G. R. Brock & J. Marsh. Wiley. [JN]

(1988) A cognitive theory of consciousness. Cambridge University Press. [arJAG, JN, FT, MV]

(1993) *A cognitive theory of consciousness.* Cambridge University Press. [AMI]

(1994) A global workspace theory of conscious experience. In: *Consciousness in philosophy and cognitive neuroscience,* ed. A. Revonsuo & M. Kamppinen. Erlbaum. [AR]

Baars, B. J. & Newman, J. (1994) A neurobiological interpretation of global workspace theory. In: *Consciousness in philosophy and cognitive neuroscience,* ed. A. Revonsuo & M. Kamppinen. Erlbaum. [AR]

Baddeley, A. D. (1993) Working memory and conscious awareness. In: *Theories of memory,* ed. A. F. Collins, S. E. Gathercole, M. A. Conway & P. E. Morris. Erlbaum. [TJL-J]

Bandura, A. (1982) Self-efficacy mechanism in human agency. *American Psychologist* 37:122–47. [TDN]

Baruch, I. (1988) Differential performance of acute and chronic schizophrenics

in a latent inhibition task and its relevance for the dopamine hypothesis and the dimensional view of psychosis. Ph.D. thesis, London University. [aJAG]

Baruch, I., Hemsley, D. R. & Gray, J. A. (1988) Differential performance of acute and chronic schizophrenics in a latent inhibition task. *Journal of Nervous and Mental Disease* 176:598–606. [arJAG, NRS]

Baxter, L. R., Schwartz, J. M., Bergman, K. S., Szuba, M. P., Guze, B. H., Mazziotta, J. C., Akazraju, A., Selin, C. E., Ferng, H.-K., Munford, P. & Phelps, M. E. (1992) Caudate glucose metabolic rate changes with both drug and behavior therapy for obsessive-compulsive disorder. *Archives of General Psychiatry* 49:681–89. [NRS]

Bechtel, W. (1988) *Philosophy of mind: An overview for cognitive science.* Erlbaum. [TJL-J]

(1994) Levels of description and explanation in cognitive science. *Minds and Machines* 4:1–25. [AR]

Berti, A. & Rizzolatti, G. (1992) Visual processing without awareness: Evidence from unilateral neglect. *Journal of Cognitive Neuroscience* 4:345–51. [CU]

Bickhard, M. H. (1991) Cognitive representation in the brain. *Encyclopedia of Human Biology* 2:547–58. [GNR]

Blumenthal, A. L. (1977) *The process of cognition.* Prentice-Hall. [AMI]

Bogen, J. E. (1995) On the neurophysiology of consciousness: 1. An overview. *Consciousness and Cognition* 4(2):137–58. [JN]

Bohm, D. J. (1986) A new theory of the relationship of mind and matter. *The Journal of the American Society for Psychical Research* 80:113–35. [J-LD]

Borst, C. V., ed. (1970) *The mind-body identity theory.*. Macmillan. [aJAG]

Boulenguez, P., Joseph, M. H., Gray, J. A. & Mitchell, S. N. (1994) Dopamine release in the nucleus accumbens after systemic and intrahippocampal injections of 5HT-1 agonists which differentially affect attention in the rat. *British Journal of Pharmacology* 111[Suppl]:153P. [aJAG]

Bouyer, J. J., Park, D. H., Joh, T. H. & Pickel, V. M. (1984) Chemical and structural analysis of the relation between cortical inputs and tyrosine hydroxylase-containing terminals in the rat neostriatum. *Brain Research* 302:267–76. [NRS]

Brazell, M. P., Mitchell, S. N., Joseph, M. H. & Gray, J. A. (1990) Acute administration of nicotine increases the in vivo extracellular levels of dopamine, 3,4-dihydroxyphenylacetic acid and ascorbic acid preferentially in the nucleus accumbens of the rat: Comparison with caudate-putamen. *Neuropharmacology* 29:1177–85. [aJAG]

Bridgeman, B. (1992) Conscious vs. unconcious processes. The case of vision. *Theory & Psychology* 2:73–88. [CF]

Carlson, M & Carlsson, A. (1990) Interactions between glutamatergic and monoamine systems within the basal ganglia - implications for schizophrenia and Parkinson's disease. *Trends in Neurosciences* 13:272–76. [aJAG]

Carver, C. S. & Scheier, M. F. (1981) *Attention and self-regulation: A control-theory approach to human behavior.* Springer-Verlag. [TDN]

Cassaday, H. J., Mitchell, S. N., Williams, J. H. & Gray, J. A. (1993) 5,7-Dihydroxytryptamine lesions in the fornix-fimbria attenuate latent inhibition. *Behavioral & Neural Biology* 59:194–207. [aJAG]

Cassaday, H. J., Hodges, H. & Gray, J. A. (1993) The effects of ritanserin, RU 24969 and 8-OH-DPAT on latent inhibition in the rat. *Journal of Psychopharmacology* 7:63–71. [aJAG]

Chadwick, P. & Birchwood, M. J. (1994) The omnipotence of voices: A cognitive approach to auditory hallucinations. *British Journal of Psychiatry* 164:190–201. [CF]

Chapman, J. (1966) The early symptoms of schizophrenia. *British Journal of Psychiatry* 112:225–51. [DRH]

Churchland, P. M. (1984) *Matter and consciousness: A contemporary introduction to the philosophy of mind.* MIT Press/Bradford books. [TJL-J]

Churchland, P. S. (1994) Can neurobiology teach us anything about consciousness? In: *The mind, the brain, and complex adaptive systems* [SFI Studies in the Sciences of Complexity, Proc. vol. 20.], ed. H. Morowitz & J. Singer. Addison-Wesley. [AR, rJAG]

Cohen, N. J. & Eichenbaum, H. (1993) Memory, amnesia, and the hippocampal system. MIT. Press. [HE, GNR]

Collicutt, J. R. & Hemsley, D. R. (1985) Schizophrenia: A disruption of the stream of thought. Unpublished manuscript. [DRH]

Cooper, L. A. & Shepard, R. N. (1973) Chronometric studies of the rotation of mental images. In: *Visual information processing,* ed. W. G. Chase. Academic Press. [aJAG]

Craik, K. J. W. (1943) *The nature of explanation.* Cambridge University Press. [rJAG]

Crick, F. (1984) The function of the thalamic reticular complex: The searchlight hypothesis. *Proceeding of the National Academy of Science, USA* 81:4586–90 [rJAG]

Crick, F. (1994) *The astonishing hypothesis: The scientific search for the soul.* Scribner. [AR, rJAG]

Crick, F. & Koch, C. (1985) Are we aware of neural activity in primary visual cortex? *Nature* 375:121–23. [rJAG]

Crow, T. J. (1980) Positive and negative schizophrenic symptoms and the role of dopamine. *British Journal of Psychiatry* 137:383–86. [aJAG]

Cutting, J. (1985) *The psychology of schizophrenia.* Churchill Livingstone. [AC]

Damasio, A. R. (1989) Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33:25–62. [MK, rJAG]

Damasio, A. R., Eslinger, P. J., Damasio, H., Van Hoesen, G. W. & Cornell, S. (1985) Multimodal amnesic syndrome following bilateral temporal and basal forebrain damage. *Archives of Neurology* 42:252–59. [CU, rJAG]

Davies, M. & Humphreys, G. W. (1993) Introduction. In: *Consciousness: Psychological and philosophical essays,* ed. M. Davies & G. W. Humphreys. Blackwell. [SLH]

Dennett, D. C. (1979) On the absence of phenomenology. In: *Body, mind and method: Essays in honor of Virgil C. Aldrich,* ed. D. Gustafson and B. Tapscott. Dordrecht: D. Reidel. [DCD]

(1991) *Consciousness explained.* Little, Brown. [aJAG, DCD, AR]

Dennett, D. C. & Kinsbourne, M. (1992) Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences* 15:183–247. [aJAG, DCD, GNR]

Dennis, S. (1995) Sources of error in connectionist models of cognitive processes. *Proceedings of the Sixth Australian Conference on Neural Networks.* School of Electrical Engineering, Sydney University, Australia. [SD]

Desmedt, J. E. (1981) Scalp-recorded cerebral event-related potentials in man as point of entry into the analysis of cognitive processing. In: *The organization of the cerebral cortex,* ed. F. O. Schmitt & F. G. Worden. M.I.T. Press.

Deutsch, J. A. (1964) *The structural basis of behaviour.* Cambridge University Press. [aJAG]

Dewey, J. (1934/1980) *Art as experience.* Perigee Books. [JDS]

Diamond, M. C., Schiebel, A. B. & Elson, L. M. (1985) *The human brain coloring book.* Barnes & Noble. [JN]

Díaz, J. L. (1989) *Psicobiología y conducta. Rutas de una indagación.* Mexico City: Fondo de Cultura Económica. [J-LD]

Dunn, L. A., Atwater, G. E. & Kilts, C. D. (1993) Effects of antipsychotic drugs on latent inhibition: Sensitivity and specificity of an animal behavioral model of clinical drug action. *Psychopharmacology* 112:315–23. [aJAG]

Dure, L. S., Young, A. B. & Penney, J. B. (1992) Compartmentalization of excitatory amino acid receptors in human striatum. *Proceedings of the National Academy of Sciences USA* 89:7688–92. [NRS]

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. & Reitboeck, H. J. (1988) Coherent oscillations: A mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. *Biological Cybernetics* 60:121–30. [GNR]

Edelman, G. M. (1987) *Neural Darwinism: The theory of neuronal group selection.* Basic Books. [GNR]

(1989) *The remembered present: A biological theory of consciousness.* Basic Books. [arJAG, AMI, JN, GNR]

Edelman, G. M. & Reeke, G. N., Jr. (1990) Is it possible to construct a perception machine? *Proceedings of the American Philosophical Society* 134:36–73. [GNR]

Egger, V. (1904) *La parole intérieure.* Paris: Alcan. [NAS]

Eichenbaum, H. T., Otto, N. & Cohen, N. J. (1994) Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences* 17:449–518. [arJAG, HE]

Ellard, C. G. (submitted) Habituation and dishabituation of responding to overhead threats in the Mongolian gerbil: Dissociating effects of stimulus configuration and environmental context. [CGE]

Ellard, C. G. & Chapman, D. G. (1991) The effects of posterior cortical lesions on responses to visual threats in the Mongolian gerbil (*Meriones unguiculatus*). *Behavioural Brain Research* 44:163–67. [CGE]

Emrich, H. M. (1992) Subjectivity, error correction capacity and the pathogenesis of delusions of reference. In: *Phenomenology, language and schizophrenia,* eds. M. Spitzer, F. A. Uehlein, M. A. Schwartz & C. Mundt. Springer-Verlag. [aJAG]

Eysenck, H. J. (1985) *Decline and fall of the Freudian empire.* Viking. [rJAG]

Farah, M. J., O'Reilly, R. C. & Vecera S. P. (1993) Dissociated overt and covert recognition as an emergent property of a lesioned network. *Psychological Review* 100:571–88. [CU]

Farthing, G. W. (1992) *The psychology of consciousness.* Prentice-Hall. [MV]

Feldon, J. & Weiner, I. (1991) The latent inhibition model of schizophrenic attention disorder: Haloperidol and sulpiride enhance rats' ability to ignore irrelevant stimuli. *Biological Psychiatry* 29:635–46. [NRS]

Flaherty, A. W. & Graybiel, A. M. (1994) Input-output organization of the sensorimotor striatum in the squirrel monkey. *Journal of Neuroscience* 14:599–610. [NRS]

Flanagan, O. (1991) *The science of the mind,* 2d ed. MIT Press. [DCD]

(1992) *Consciousness reconsidered.* MIT Press. [GLS]

Fleminger, S. (1992) Seeing is believing: The role of 'preconscious' perceptual processing in delusional misidentification. *British Journal of Psychiatry* 160:293–303. [AC]

Freedman, B. J. (1974) The subjective experience of perceptual and cognitive disturbance in schizophrenia. *Archives of General Psychiatry* 30:333–40. [aJAG]

Friston, K. J., Liddle, P. F., Frith, C. D., Hirsch, S. R. & Frackowiak, R. S. J. (1992) The left medial temporal region and schizophrenia. *Brain* 115: 367–82. [AC]

Frith, C. D. (1979) Conciousness, information processing and schizophrenia. *British Journal of Psychiatry* 134:225–35. [CF]

(1987) The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine* 17:631–48. [arJAG]

(1992) *The cognitive neuropsychology of schizophrenia.* Erlbaum. [CF, rJAG]

Frith, C. D. & Done, D. J. (1989) Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine* 19:359–63. [aJAG]

Fuster, J. M. (1989) *The prefrontal cortex.* Raven. [AC]

Gaffan, D. (1994) Scene-specific memory for objects: A model of episodic memory impairment in monkeys with fornix transection. *Journal of Cognitive Neuroscience* 6:305–20. [rJAG]

Gallistel, C. R. (1980) *The organization of action: A new synthesis.* Erlbaum. [FT]

Geldard, F. A. & Sherrick, C. E. (1972) The cutaneous rabbit: A perceptual illusion. *Science* 178:178–79. [aJAG]

Gerfen, C. R., Herkenham, M. & Thibault, J. (1987) The neostriatal mosaic: 2. Patch- and matrix-directed mesostriatal dopaminergic and non-dopaminergic systems. *Journal of Neuroscience* 7(12):3915–34. [NRS]

Gerfen, C. R., McGinty, J. F & Young, W. S. (1991) Dopamine differentially regulates dynorphin, substance P, and enkephalin expression in striatal neurons: In situ hybridization histochemical analysis. *Journal of Neuroscience* 11:1016–31. [NRS]

Gibson, J. J. (1950) *The perception of the visual world.* Houghton Mifflin. [HR]

Giménez-Amaya, J. M. & Graybiel, A. M. (1990) Compartmental origins of the striatopallidal projection in the primate. *Neuroscience* 34(1):111–26. [NRS]

Globus, G. G. (1973) Consciousness and brain. *Archives of General Psychiatry* 29:153–76. [J-LD]

Good, M. & Honey, R. (1993) Selective hippocampus lesions abolish contextual specificity of latent inhibition and conditioning. *Behavioral Neuroscience* 107:23–33. [aJAG]

Goodale, M. A. & Milner, A. D. (1992) Separate visual pathways for perception and action. *Trends in the Neurosciences* 15:20–25. [CF]

Graves, R. E. & Jones, B. S. (1992) Conscious visual perceptual awareness vs. non-conscious visual spatial localisation examined with normal subjects using possible analogues of blindsight and neglect. *Cognitive Neuropsychology* 9:487–508. [CU]

Gray, C. M. & Singer, W. (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences USA* 86:1698–1702. [GNR, rJAG]

Gray, J. A. (1971) The mind-brain identity theory as a scientific hypothesis. *Philosophical Quarterly* 21:247–52. [arJAG, MK]

(1975) *Elements of a two-process theory of learning.* Academic Press. [arJAG]

(1977) Drug effects on fear and frustration: Possible limbic site of action of minor tranquillizers. In: *Handbook of psychopharmacology,* vol. 8, ed. L. L. Iversen, S. D. Iversen & S. H. Snyder. Plenum. [aJAG]

(1982a) *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system.* Oxford University Press. [arJAG, DRH]

(1982b) Précis of *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system. Behavioral and Brain Sciences* 5:469–94. [arJAG]

(1984) The hippocampus as an interface between cognition and emotion. In: *Animal cognition,* ed. H. L. Roitblat, T. G. Bever & H. S. Terrace. Erlbaum. [aJAG]

(1985) Memory buffer and comparator can share the same circuitry. *Behavioral and Brain Sciences* 8:501. [SD]

(1987) The mind-brain identity theory as a acientific hypothesis: A second look. In: *Mindwaves,* ed. C. Blakemore & S. Greenfield. Blackwell. [aJAG]

(1993) Consciousness, schizophrenia and scientific theory. In: *Experimental and theoretical studies of consciousness,* ed. J. March. Ciba Foundation Symposium 174. Wiley. [aJAG]

(1994) The neuropsychology of anxiety and schizophrenia. In: *The cognitive neurosciences,* ed. M. J. Gazzaniga. MIT Press. [arJAG]

Gray, J. A., Feldon, J., Rawlins, J. N. P., Hemsley, D. R. & Smith, A. D. (1991) The neuropsychology of schizophrenia. *Behavioral and Brain Sciences* 14:1–20. [arJAG, AC, CF, DRH, REL, NRS]

Gray, J. A., Hemsley, D. R., Feldon, J., Gray, N. S. & Rawlins, J. N. P. (1991) Schiz bits: Misses, mysteries and hits. *Behavioral and Brain Sciences* 14:56–84. [arJAG, DRH]

Gray, J. A. & Rawlins, J. N. P. (1986) Comparator and buffer memory: An attempt to integrate two models of hippocampal function. In: *The hippocampus,* vol. 4, ed. R. L. Isaacson & K. H. Pribram. Plenum. [aJAG]

Gray, J. A. & Smith, P. T. (1969) An arousal-decision model for partial reinforcement and discrimination learning. In: *Animal discrimination learning,* ed. R. Gilbert & N. S. Sutherland. Academic Press. [rJAG]

Gray, J. A., Joseph, M. H., Hemsley, D. R., Young, A. M. J., Warburton, E. C., boulenguez, P. , Grigoryan, G. A. , Peters, S. L., Rawlins, J. N. P., Tai, C.-T., Yee, B. K., Cassady, H., Weiner, I., Gal, G., Gusak, O., Joel, D., Shadach, E. Shalev, U., Tarrasch, R., & Feldon, J. (in press) The role of mesolimbic dopaminergic and retrohippocampal afferents to the nucleus accumbens in latent inhibition: Implications for schizophrenia. *Behavioral Brain Research.* [rJAG]

Gray, J. A., Williams, S., Nunn, J. & Baron-Cohen, S. (in press) The possible implications of synaesthesia for the Hard Question of consciousness. In: *Synaesthesia,* ed. S. Baron-Cohen & J. Harrison. Blackwell. [rJAG]

Gray, N. S., Hemsley, D. R. & Gray, J. A. (1992) Abolition of latent inhibition in acute, but not chronic, schizophrenics. *Neurology, Psychiatry and Brain Research* 1:93–89. [arJAG, NRS]

Gray, N. S., Pickering, A. D., Hemsley, D. R., Dawling, S. & Gray, J. A. (1992) Abolition of latent inhibition by a single 5 mg dose of amphetamine in man. *Psychopharmacology* 107:425–30. [arJAG]

Gray, N. S., Pilowsky, L. S., Gray, J. A. & Kerwin, R. W. (in press). Latent inhibition in drug naive schizophrenics: Relationship to duration of illness and dopamine D2 binding using SPET. *Schizophrenia Research.* [arJAG]

Graybiel, A. M. & Ragsdale, C. W. (1980) Clumping of acetylcholinesterase activity in the developing striatum of the human fetus and young infant. *Proceedings of the National Academy of Sciences USA* 77:1214–18. [NRS]

(1983) Biochemical anatomy of the striatum. In: *Chemical neuroanatomy,* ed. P. C. Emson. Raven. [NRS]

Graybiel, A. M., Ragsdale, C. W., Yoneoka, E. S. & Elde, R. P. (1981) An immunohistochemical study of enkephalins and other neuropeptides in the striatum of the cat with evidence that the opiate peptides are arranged to form mosaic patterns in register with the striosomal compartments visible by acetylcholinesterase staining. *Neuroscience* 6:377–97. [NRS]

Groenewegen, H. J., Vermeulen-Van der Zee, E., Te Kortschot, A. & Witter, M. P. (1987) Organization of the projections from the subiculum to the ventral striatum in the rat. A study using anterograde transport of *Phaseolus vulgaris* leucoagglutinin. *Neuroscience* 23:103–20. [NRS]

Groves, P. M. (1983) A theory of the functional organization of the neostriatum and the neostriatal control of voluntary movement. *Brain Research Reviews* 5:109–32. [aJAG]

Haken, H. (1983) *Synergetics, an introduction: Non-equilibrium phase transitions and self-organisation in physics, chemistry and biology,* 3d ed. Springer-Verlag. [RI]

Harnad, S. (1994) Why and how we are not zombies. *Journal of Conciousness Studies* 1:164–67. [rJAG]

Hebb, D. O. (1946) Emotion in man and animal: An analysis of the intuitive processes of recognition. *Psychological Review* 53:88–106. [rJAG]

Heilman, K. M., Watson, R. T. & Valenstein, E. (1985) Neglect and related disorders. In: *Clinical neuropsychology,* ed. K. M. Heilman & E. Valenstein. Oxford University Press. [JN]

Heimer, L., Zahm, D. S., Churchill, L., Kalivas, P. W. & Wohltmann, C. (1991) Specificity in the projection patterns of accumbal core and shell in the rat. *Neuroscience* 41:89–125. [NRS]

Hemsley, D. R. (1987) An experimental psychological model for schizophrenia. In: *Search for the causes of schizophrenia,* ed. H. Hafner, W. F. Gattaz & W. Janzavik. Springer-Verlag. [aJAG]

(1993) A simple (or simplistic?) cognitive model for schizophrenia. *Behaviour Research and Therapy* 31:633–45. [aJAG, AC, DRH]

(1994) Cognitive disturbance as the link between schizophrenic symptoms and their biological bases. *Neurology, Psychiatry and Brain Research* 2:163–70. [DRH]

Hirsh, R. (1974) The hippocampus and contextual retrieval of information from memory: A theory. *Behavioural Biology* 12:421–44. [aJAG]

(1980) The hippocampus, conditional operations and cognition. *Physiological Psychology* 8:175–82. [MK]

Horgan, T. (1993) From supervenience to superdupervenience: Meeting the demands of a material world. *Mind* 102:555–86. [J-LD]

Humphreys, M. S. & Bain, J. D. (1983) Recognition memory: A cue and informational analysis. *Memory and Cognition* 11(6):583–600. [SD]

Humphreys, M. S. & Dennis, S. (1994) Going from task descriptions to memory structures. *Behavioral and Brain Sciences* 17:483. [SD]

Humphreys, M. S., Wiles, J. & Dennis, S. (1994) Towards a theory of human

memory: Data structures and access processes. *Behavioral and Brain Sciences* 17:655–92. [SD]
Humphrey, N. (1983) *Consciousness regained.* Oxford University Press. [rJAG]
Humphrey, N. (1992) *A history of the mind.* Chatto & Windus. [CGE, rJAG]
Hurley, S. L. (in press) *The reappearing self.* Harvard University Press. [SLH]
Imperato, A., Honore, T. & Jensen, L. H. (1990) Dopamine release in the nucleus caudatus and in the nucleus accumbens is under glutamatergic control through non-NMDA receptors: A study in freely moving rats. *Brain Research* 530:233–228. [NRS]
Ingvaldsen, R. P. & Whiting, H. T. A. (1993) The two faces of motor (skill) learning. In: *Learning motor skills,* ed. C. A. M. Doorenbosch. Amsterdam: Free University Press. [RI]
Ivanitsky, A. M. (1976) *The brain mechanisms of signal evaluation.* Medicina Publishing House (in Russian). [AMI]
(1993) Consciousness: criteria and possible mechanisms. *International Journal of Psychophysiology* 14:179–87. [AMI]
(1994) Brain mechanisms of the mind: Information synthesis in dominant cortical areas. In: *Seventh International Congress of Psychophysiology Abstracts* 53. [AMI]
Ivanitsky, A. M., Korsakov, I. A. & Strelets, V. B. (1989) Perception as a result of synthesis of the sensory and the emotional. In: *Systems approach in physiology: vol. 2. Emotions and behaviour: A systems approach,* ed. K. V. Sudakov. Gordon and Breach. [AMI]
Ivanitsky, A. M. & Strelets, V. B. (1977) Brain evoked potentials and some mechanisms of perception. *Electroencephalography and Clinical Neurophysiology.* 43:397–403. [AMI]
Ivanitsky, A. M., Strelets, V. B. & Korsakov, I. A. (1984) *Brain information processing and mental activity.* Moscow: Nauka (in Russian). [AMI]
Jackendoff, R. (1987) *Consciousness and the computational mind.* MIT Press. [aJAG, CGE, DRH, TJL-J, GNR, VAS]
James, W. (1890) *The principles of psychology.* Macmillan. [DRH, rJAG]
(1890/1952) *The principles of psychology* [vol. 53 of Great Books of the Western World]. University of Chicago Press. [JDS]
(1892) *Textbook of psychology.* Macmillan. [DRH, rJAG]
Jaynes, J. (1976) *Origins of consciousness in the breakdown of the bicameral mind.* Houghton-Mifflin. [REL]
Jeannerod, M., ed. (1987) *Neurophysiological and neuropsychological aspects of spatial neglect.* Amsterdam: North Holland. [aJAG]
Johnston, W. A., Hawley, K. J., Plewe, S. H., Elliott, J. M. G. & De Witt, M. J. (1990) Attention capture by novel stimuli. *Journal of Experimental Psychology* 119:397–411. [aJAG]
Jones, S. H., Gray, J. A. & Hemsley, D. R. (1992) Loss of the Kamin blocking effect in acute but not chronic schizophrenics. *Biological Psychiatry* 32:739–55. [arJAG]
Jordan, M. I. (1990) Motor learning and the degrees of freedom problem. In: *Attention and performance,* ed. 13th, ed. M. Jeannerod. Erlbaum. [SD]
Jordan, M. I. & Jacobs, R. A. (1990) Learning to control an unstable system with forward modelling. In: *Advances in neural information processing systems 2,* ed. D. S. Touretsky. Morgan Kaufmann. [SD]
Jordan, M. I. & Rumelhart, D. E. (1992) Forward models: Supervised learning with a distal teacher. *Cognitive Science* 16:307–54. [SD]
Joseph, M. A., Peters, S. L. & Gray, J. A. (1993) Nicotine blocks latent inhibition in rats: Evidence for a critical role of increased functional activity of dopamine in the mesolimbic system at conditioning rather than pre-exposure. *Psychopharmacology* 110:187–92. [aJAG]
Kelly, P. K., Seviour, P. W. & Iverson, S. D. (1975) Amphetamine and apomorphine responses in the rat following 6-OHDA lesions of the nucleus accumbens septi and corpus striatum. *Brain Research* 94:507–22. [aJAG]
Kihlstrom, J. F. (1990) The psychological unconscious. In: *Handbook of personality: Theory and research,* ed. L. A. Pervin. Guilford. [VAS]
Killcross, A. S. & Robbins, T. W. (1993) Differential effects of intra-accumbens and systemic amphetamine on latent inhibition using an on-baseline, within-subject conditioned suppression paradigm. *Psychopharmacology* 110:479–89. [aJAG]
Kim, J. (1993) *Supervenience and mind.* Cambridge University Press. [J-LD]
Kinsbourne, M. (1988) Integrated field theory of consciousness. In: *The concept of consciousness in contemporary science,* ed. A. J. Marcel & E. Bisiach. Oxford University Press. [MK]
(1989) The boundaries of episodic remembering: A commentary. In: *Varieties of memory and consciousness: Essays in honour of Endel Tulving,* ed. F. I. M. Craik & H. L. Roediger. Erlbaum. [MK, rJAG]
(1993) Integrated cortical field model of consciousness. In: *Experimental and theoretical studies of consciousness,* ed. J. Marsh. Ciba Foundation Symposium 174. Wiley. [arJAG, DCD]
(in press a) What qualifies a representation for a role in consciousness? In: *Scientific approaches to the question of consciousness,* ed. J.D. Cohen & J.W. Schooler. Erlbaum. [MK]
(in press b) Representations in consciousness and the neuropsychology of insight. In: *Insight and psychosis,* ed. X. F. Amador & A David. Oxford University Press. [MK]
Koch, C. & Crick, F. (1994) Some further ideas regarding the neuronal basis of awareness. In: *Large-scale neuronal theories of the brain,* ed. C. Koch & J. L. Davis. MIT Press. [AR]
Kosslyn, S. M., Flynn, R. A., Amsterdam, J. B. & Wang, G. (1990) Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes. *Cognition* 34:203–77. [TJL-J]
Kuhn, T. S. (1970) *The structure of scientific revolutions.* University of Chicago Press. [MV]
Làdavas, E., Paladini, R. & Cubelli, R. (1993) Implicit associative priming in a patient with left visual neglect. *Neuropsychologia* 31:1307–20. [CU]
Lahav, R. (1993) What neuropsychology tells us about consciousness. *Philosophy of Science* 60:67–85. [GLS]
Lashley, K. S. (1956) Cerebral organization and behavior. In: *The brain and human behavior,* ed. H. Solomon, S. Cobb & W. Penfield. Williams & Wilkins. [aJAG]
Lavin, A. & Grace, A. A. (1994) Modulation of dorsal thalamic cell activity by the ventral pallidum: Its role in the regulation of thalamocortical activity by the basal ganglia. *Synapse* 18:104–27. [rJAG]
Libet, B. (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* 8:529–66. [aJAG]
(1993) The neural time factor in conscious and unconscious events. In: *Experimental and theoretical studies of consciousness,* ed. J. Marsh. Ciba Foundation Symposium 174. Wiley. [aJAG, GNR, MV]
Libet, B., Pearl, D. K., Morledge, D. E., Gleason, C. A., Hosobuchi, Y. & Barbaro, N. M. (1991) Control of the transition from sensory detection to sensory awareness in man by the duration of a thalamic stimulus. *Brain 114*:1731–57. [arJAG]
Lipp, O. V. & Vaitl, D. (1992) Latent inhibition in human Pavlovian differential conditioning: Effect of additional stimulation after preexposure and relation to schizotypal traits. *Personality and Individual Differences* 13:1003–12. [aJAG]
Lipska, B. K., Jaskiw, G. E., Chrapusta, S., Karoum, F. & Weinberger, D. R. (1992) Ibotenic acid lesion of the ventral hippocampus differentially affects dopamine and its metabolites in the nucleus accembens and prefrontal cortex in the rat. *Brain Research* 585:1–6. [rJAG]
Lipska, B. K., Jaskiw, G. E. & Weinberger, D. R. (1993) Postpubertal emergence of hyperresponsiveness to stress and to amphetamine after neonatal excitotoxic hippocampal damage: A potential animal model of schizophrenia. *Neuropsychopharmacology* 9:67–75. [rJAG]
Lishman, A. (1987) *Organic psychiatry.* Blackwell. [aJAG]
Llinás, R. & Paré, D. (1991) Of dreaming and wakefulness. *Neuroscience* 44:521–35. [AR]
Llinás, R. & Ribary, U. (1993) Coherent 40-Hz oscillation characterizes dream state in humans. *Proceedings of the National Academy of Sciences USA* 90:2078–81. [AR]
(1994) Perception as an Oneiric-like state modulated by the senses. In: *Large-scale neuronal theories of the brain,* ed. C. Koch & J. L. Davis. MIT Press. [AR]
Lockwood, M. L. (1989) *Mind, brain & the quantum.* Blackwell. [J-LD]
Lubow, R. E. (1973) Latent inhibition. *Psychological Bulletin* 79:398–407. [aJAG]
(1981) On animal analogies to human behavior and the biological bases of value systems. *The Journal of Mind and Behavior* 2:195–207. [REL]
(1989) Latent inhibition and conditioned attention theory. Cambridge University Press. [aJAG]
Lubow, R. E. & Josman, Z. E. (1995) Latent inhibition defects in hyperactive children. *Journal of Child Psychology and Psychiatry.* [aJAG]
Lubow, R. E., Ingberg-Sachs, Y., Zalstein-Orda, N. & Gewirtz, J. C. (1992) Latent inhibition in low and high "psychotic-prone" normal subjects. *Personality and Individual Differences* 13:563–72. [aJAG]
Macdonald, C. (1989) *Mind-body identity theories.* Routledge. [TJL-J]
MacDonnell, M. J. & Flynn, J. P. (1966) Control of sensory fields by stimulation of the hypothalamus. *Science* 152(3727):1406–8. [CGE]
Mandler, G. (1984) *Mind and body.* Norton. [aJAG]
Marcel, A. J. & Bisiach, E., eds. (1988) *Consciousness in contemporary science.* Clarendon. [aJAG]
Marr, D, (1982) *Vision.* Freeman. [aJAG, TJL-J]
Marsh, J., ed. (1993) *Experimental and theoretical studies of consciousness.* Ciba Foundation Symposium 174. Wiley. [arJAG, MV]
Matussek, P. (1952) Studies in delusional perception. *Psychiatrie und Zeitschrift Neurologie* 189:279–318. [Transl., 1987, in: *The clinical roots of the schizophrenia concept,* ed. J. Cutting & M. Shepherd. Cambridge University Press.] [aJAG]
McKenna, P. J., Tamlyn, D., Lund, C. E., Mortimer, A. M., Hammond, S. & Baddeley, A. D. (1990) Amnesic syndrome in schizophrenia. *Psychological Medicine* 20:967–72. [rJAG]

Meehl, P. E. (1966) The compleat autocerebroscopist: A thought-experiment on Professor Feigl's mind-body identity thesis. In: *Mind, matter and method*, eds. P. K. Feyerabend & G. Maxwell. University of Minnesota Press. [aJAG]

Melzack, R. & Wall, P. D. (1983) *The challenge of pain*. Basic Books. [rJAG]

Mesulam, M.-M. (1985) *Principles of behavioral neurology*. F. A. Davis. [JN]

Miller, G. A., Galanter, E. & Pribram, K. H. (1960) *Plans and the structure of behavior*. Holt, Rinehart & Winston. [TDN, FT]

Milner, A. D. & Rugg, M. D., eds. (1992) *The neuropsychology of consciousness*. Academic Press. [MV]

Minksy, M. (1985) *The society of mind*. Simon & Schuster. [GNR]

Mishkin, M. (1993) What is recognizing memory and what neural circuits are involved? In: *Thirty-second Congress of Physiological Sciences* (Glasgow, Scotland, August 1–6, 1993). Abstract 27.1/O:42–43. [AMI]

Mitchell, S. N., Yee, B. R., Feldon, J., Gray, J. A. & Rawlins, J. N. P. (in preparation) Activation of the retrohippocampal region in the rat causes dopamine release in nucleus accumbens; disruption by fornix section.

Modell, J. G., Mountz, J. M., Curtis, G. C. et al. (1989) Neurophysiological dysfunction in basal ganglia/limbic striatal and thalamocortical circuits as a pathogenetic mechanism of obsessive-compulsive disorder. *Journal of Neuropsychiatry* 1:27–36. [NRS]

Mowrer, O. H. (1960) *Learning and behavior*. Wiley. [aJAG]

Nagel, T. (1993) What is the mind-body problem? In: *Experimental and theoretical studies on consciousness*, ed. G. R. Bock & J. Marsh. Wiley. [AR]

Nauta, W. J. H. (1989) Reciprocal links of the corpus striatum with the cerebral cortex and limbic system: A common substrate for movement and thought? In: *Neurology and psychiatry: A meeting of minds*, ed. J. Muleller. Karger. [NRS]

Navon, D. (1991) The function of consciousness or of information? *Behavioral and Brain Sciences* 14(4):690–91. [MV]

Neely, J. H.(1977) Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General* 106:226–54. [MV]

Neisser, U. (1976) *Cognition and reality*. Freeman. [aJAG, DRH, TDN]

Nelson, T. D. (1993) The hierarchical organization of behavior: A useful feedback model of self-regulation. *Current Directions in Psychological Science* 2:121–26. [TDN]

Newman, J. (1980) *Cognition and consciousness*. Gainesville, FL: Center for the Applications of Psychological Type. [JN]

(1995) Review: Thalamic contributions to attention and consciousness. *Conciousness and Cognition* 4(2):172–93. [JN]

Newman, J. & Baars, B. J. (1993) A neural attentional model for access to consciousness: A global workspace perspective. *Concepts in Neuroscience* 4(2):255–90. [JN, rJAG]

Newman, J., Baars, B. J. & Cho, S.-B. (in press) A neurocognitive model for attention and consciousness. In: *Advances in consciousness research*, ed. G. G. Globus & M. I. Stamenov. Amsterdam: John Benjamins. [JN]

Newsome, W. T. & Salzman, C. D. (1993) The neuronal basis of motion perception. In: *Experimental and theoretical studies of consciousness*, ed. J. Marsh. Ciba Foundation Symposium 174. Wiley. [arJAG]

Norman, D. A. & Bobrow, D. G. (1976) On the role of active memory processes in perception and cognition. In: *The structure of human memory*, ed. C. N. Cofer. Freeman. [DRH]

Oatley, K. (1988) On changing one's mind: A possible function of consciousness. In: *Consciousness in contemporary science*, ed. A. J. Marcel & E. Bisiach. Clarendon. [aJAG]

O'Connor, T. (1994) Emergent properties. *American Philosophical Quarterly* 31:91–104. [J-LD]

O'Keefe, J. (1985) Is consciousness the gateway to the hippocampal cognitive map? A speculative essay on the neural basis of mind. In: *Brain and mind*, ed. D. A. Oatley. Methuen. [aJAG]

O'Keefe, J. & Nadel, L. (1978) *The hippocampus as a cognitive map*. Clarendon Press.

Orbach, H. S. (1988) Monitoring electrical activity in rat cerebral cortex. In: *Spectroscopic membrane probes*, ed. L.M. Loew. Boca Raton, FL: CRC Press.

Papez, J. W. (1937) A proposed mechanism of emotion. *Archives of Neurology Psychiatry* 38:725–43. [aJAG]

Penn, D. L., Hope, D. A., Spaulding, W. & Kucera, J. (1994) Social anxiety in schizophrenia. *Schizophrenia Research* 11:277–84. [DRH, rJAG]

Penney, J. B. & Young, A. B. (1981) GABA as the pallidothalamic neurotransmitter: Implications for basal ganglia function. *Brain Research* 207:195–99. [aJAG]

(1986) Striatal inhomogeneities and basal ganglia function. *Movement Disorders* 1:3–15. [NRS]

Peters, S. L., Grigoryan, G. A., Joseph, M. H., Hodges, H. & Gray, J. A. (in preparation) Facilitation of latent inhibition by 6-OHDA lesions of the nucleus accumbens. [aJAG]

Piaget, J. (1967) *Six psychological studies*. Random House. [NAS]

Pieron , H. (1960) *La sensation*. Presse Université, France [AMI]

Pinker, S. & Bloom, P. (1990) Natural language and natural selection. *Behavioral and Brain Sciences* 13:707–84. [rJAG]

Polanyi, M. & Prosch, H. (1975) *Meaning*. University of Chicago Press. [aJAG]

Posner, M. I. & Dehaene, S. (1994) Attentional networks. *Trends in Neuroscience* 17:75–79. [CU]

Posner, M. I. & Peterson, S. E. (1990) The attention system of the human brain. *Annual Review of Neuroscience* 13:25–42. [CU]

Posner, M. I. & Snyder, C. R. R. (1975) Facilitation and inhibition in the processing of signals. In: *Attention and Performance* 5., eds. P. M. A. Rabbitt & S. Dornick. Academic Press. [MV]

Powers, W. T. (1973a) *Behavior: The control of perception*. Aldine de Gruyter. [TDN, rJAG]

(1973b) *Behaviour: The control of perception*. Wildwood House. [FT]

(1980) A systems approach to consciousness. In: *Psychology of consciousness*, ed. J. M. Davidson & R. J. Davidson. Plenum. [TDN]

(1989) *Living control systems*. Control Systems Group. [TDN]

Pulvirenti, L., Swerdlow, N. R. & Koob, G. F. (1991) Intra-accumbens infusion of an NMDA antagonist reverses the psychomotor effects of cocaine, heroin or intra-accumbens dopamine, but not caffeine. *Pharmacol Biochem Behav* 40:841–45. [NRS]

Pycock, C. J., Kerwin, R. W. & Carter, C. J. (1980) Effect of cortical dopamine terminals on subcortical dopamine receptors in rats. *Nature* 286:74–77. [rJAG]

Pylyshyn, Z. W. (1980) Cognition and computation: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences* 3:111–32. [aJAG]

(1984) *Computation and cognition: Toward a foundation for cognitive science*. MIT Press. [GNR]

(1986) *Computation and cognition*. MIT Press. [aJAG, MV]

Rachlin, H. (1985) Pain and behavior. *Behavioral and Brain Sciences* 8:43–83. [HR]

(1994a) *Behavior and mind*. Oxford University Press. [HR]

(1994b) From overt behavior to hypothetical behavior to memory: Inference in the wrong direction. *Behavioral and Brain Sciences* 17:147–48. [HR]

(1995) Self-control: Beyond commitment. *Behavioral and Brain Sciences* 18:109–21. [HR, rJAG]

Ragsdale, C. W. Jr. & Graybiel, A. M. (1990) A simple ordering of neocortical areas established by the compartmental organization of their striatal projections. *Neurobiology* 87:6196–99. [NRS]

Rawlins, J. N. P. (1985) Associations across time: The hippocampus as a temporary memory store. *Behavioral and Brain Sciences* 8:479–528. [aJAG, SD]

Reeke, G. N., Jr. (1991) Book review: Marvin Minsky, *The society of mind*. *Artificial Intelligence* 48:341–48. [GNR]

Reeke, G. N., Jr. & Edelman, G. M. (1988) Real brains and artificial intelligence. *Daedalus, Proceedings of the American Academy of Arts and Sciences* 117:143–73. [GNR]

Revonsuo, A. (1993) Cognitive models of consciousness. In: *Consciousness, cognitive schemata, and relativism*, ed. M. Kamppinen. Kluwer. [AR]

(1994) In search of the the science of consciousness. In: *Consciousness in philosophy and cognitive neuroscience*, ed. A. Revonsuo & M. Kamppinen. Erlbaum. [AR]

(1995) Conciousness, dreams, and virtual realities. *Philosophy Psychology* 8:35–38. [AR]

Robb, M. D. (1972) *The dynamics of motor-skill acquistion*. Prentice-Hall. [RI]

Robertson, G. S. & Jian, M. (1995) D_1 and D_2 dopamine receptors differentially increase fos-like immunoreactivity in accumbal projections to the ventral pallidum and midbrain. *Neuroscience* 64:1019–34. [rJAG]

Robertson, I. H. & Marshall, J. C., eds. (1993) Unilateral neglect: Clinical and experimental studies. Erlbaum. [CU]

Rolls, E. T. & Williams, G. V. (1987) Sensory and movement related neuronal activity in different regions of the primate striatum. In: *Basal ganglia and behavior: Sensory aspects and motor functioning*, ed. J. S. Schneider & T. I. Kidsky. Hans Huber. [aJAG]

Rose, S. (1993) *The making of memory*. Doubleday. [JN]

Rosene, D. L. & Van Hoesen, G. W. (1977) Hippocampal efferents reach widespread areas of the cerebral cortex in the monkey. *Science* 198: 315–17. [MK]

Sass, L. A. (1992) *Madness and modernism: Insanity in the light of modern art, literature, and thought*. Harvard University Press. [AC]

Schacter, D. L. (1987) Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:501–18. [VAS]

(1990) Toward a cognitive neuropsychology of awareness: Implicit knowledge and anosagnosia. *Journal of Clinical and Experimental Neuropsychology* 12:155–78. [AR]

Schacter, D. L., McAndrews, M. P. & Moscovitch, M. (1988) Access to consciousness: Dissociations between implicit and explicit knowledge in neuropsychological syndromes. In: *Thought without language*, ed. L. Weiskrantz. Oxford University Press. [CU]

Scheier, M. F. & Carver, C. S. (1988) A model of behavioral self-regulation: Translating intention into action. In: *Advances in experimental social psychology*, ed. L. Berkowitz. Academic Press. [TDN]

Schmajuk, N. A. (1994) Behavioral dynamics of escape and avoidance: A neural network approach. In: *From animals to animats 3*, ed. D. Cliff, P. Husbands, J.-A. Meyer & S. W. Wilson. MIT Press. [NAS]

Schmajuk, N. A. & Axelrad, E. T. (1995) *Animal communication: A neural network approach.* Presented at the 66th Annual Meeting of the Eastern Psychological Association, Boston, March 30–April 2. [NAS]

Schmajuk, N. A. & Christiansen, B. (in preparation) Haloperidol reinstates latent inhibition impaired by hippocampal lesions. [aJAG]

Schmajuk, N. A., Lam, Y.-W. & Gray, J. A. (in preparation) Latent inhibition: A neural network approach. [aJAG, NAS]

Schneider, W. & Shiffrin, R. M. (1977) Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review* 84:1–66. [aJAG, TDN]

Scoville, W. B. & Milner, B. (1957) Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurological and Neurosurgical Psychiatry* 20:11–12. [HM, GNR, CU]

Searle, J. R. (1980) Minds, brains and programs. *Behavioral and Brain Sciences* 3:417–57. [arJAG]

(1983) *Intentionality.* Cambridge University Press. [aJAG]

(1992) *The rediscovery of mind.* MIT Press. [DCD, rJAG]

(1993) The problem of consciousness. In: *Experimental and theoretical studies of consciousness*, ed. J. Marsh. Ciba Foundation Symposium 174. Wiley. [aJAG]

Sechehaye, M. 91970) *Autobiography of a schizophrenic girl.* New American Library. [AC]

Sesack, S. R. & Pickel, V. M. (1990) In the medial nucleus accumbens, hippocampal and catecholaminergic terminals converge on spiny neurons and are in apposition to each other. *Brain Research* 527:266–79. [NRS]

Shakow, D. (1963) Psychological deficit in schizophrenia. *Behavioral Science* 8:275–305. [AC]

Shallice, T. (1988) *From neuropsychology to mental structure.* Cambridge University Press. [AC, CF]

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: 2. Perceptual learning, automatic attending, and a general theory. *Psychological Review* 84:127–90. [JDS]

Sidman, M., Stoddard, L. T. & Mohr, J. P. (1968) Some additional quantitative observations of immediate memory in a patient with bilateral hippocampal lesions. *Neuropsychologia* 6:245–54. [SD]

Skinner, B. F. (1953) *Science and human behavior.* Macmillan. [HR]

(1957) *Verbal behavior.* Appleton-Century-Crofts. [NAS]

Smart, J. J. C. (1959) Sensations and brain processes. *Philosophical Review* 68:141–56. [TJL-J, rJAG]

Smith, A. D. & Bolam, J. P. (1990) The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurons. *Trends in Neuroscience* 13:259–65. [NRS]

Smith, B. H. (1994) *Doing without meaning* (adapted from *Belief and resistance: Dynamics of theoretical controversy*, a work in progress). Presented at Systems Theory and the Postmodern Condition, Indiana University, September 22–25, 1994. [NAS]

Smith, J. D. Schull, J., Strote, J., McGee K., Egnor, R. & Erb, L. (in press) The uncertain response in the bottlenosed dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology: General.* [JDS]

Sokolov, E. N. (1960) Neuronal models and the orienting reflex. In: *The central nervous system and behaviour* (transactions of the 3d conference), ed. M. A. B. Brazier. Josiah Macy Jr. Foundation. [rJAG]

Solomon, P. R., Crider, A., Winkelman, J. W., Turi, A., Kamer, R. M. & Kaplan, L. J. (1981) Disrupted latent inhibition in the rat with chronic amphetamine or haloperidol-induced supersensitivity: Relationship to schizophrenic attention disorder. *Biological Psychiatry* 16: 519–37. [aJAG]

Solomon, P. R. & Staton, D. M. (1982) Differential effects of microinjections of d-amphetamine into the nucleus accumbens or the caudate-putamen on the rat's ability to ignore an irrelevant stimulus. *Biological Psychiatry* 17:742–56. [aJAG]

Staddon, J. E. R. (1993) *Behaviorism: Mind, mechanism and society.* Duckworth. [HR]

Stevens, J. R. & Gold, J. M. (1991) What is schizophrenia? *Behavioral and Brain Sciences* 14:50–51. [AC]

Suddath, R. L., Christison, G. W., Torrey, E. F., Casanova, M. F. & Weinberger, D. R. (1990) Anatomical abnormalities in the brains of monozygotic twins discordant for schizophrenia. *New England Journal of Medicine* 322: 789–94. [AC]

Swanson, L. W. & Cowan, W. M. (1977) An autoradiographic study of the organization of the efferent connections of the hippocampal formation in the rat. *Journal of Comparative Neurology* 172:48–49. [MK]

Swerdlow, N. R., Braff, D. L., Taaid, N. & Geyer, M. A. (1994) Assessing the validity of an animal model of deficient sensorimotor gating in schizophrenic patients. *Archives of General Psychology* 51:139–54. [AC]

Swerdlow, N. R., Caine, S. B., Braff, D. L. & Geyer, M. A. (1992) Neural substrates of sensorimotor gating of the startle reflex: A review of recent findings and their implications. *Journal of Psychopharmacology* 6: 176–90. [NRS]

Swerdlow, N. R. & Koob, G. F. (1987) Dopamine, schizophrenia, mania and depression: Toward a unified hypothesis of cortico-striato-pallidothalamic function. *Behavioral and Brain Sciences* 10:197–245. [arJAG, NRS]

Tai, C.-T., Cassaday, H. J., Feldon, J. & Rawlins, J. N. P. (in press) Both electrolytic and excitotoxic lesions of nucleus accumbens disrupt latent inhibition in rats. *Behavioral and Neural Biology.*

Talbot, J. D., Marrett, S., Evans, A. C., Meyer, E., Bushnell, M. C. & Duncan, G. H. (1991) Multiple representations of pain in human cerebral cortex. *Science* 251:1355–58. [HM]

Tarrasch, R., Weiner, I., Rawlins, J. N. P. & Feldon, J. (1992) Disruption of latent inhibition by interrupting the subicular input to nucleus accumbens and its antagonism by haloperidol. *Psychopharmacology* (abstract) 6:111. [aJAG]

Thorpe, W. H. (1963) *Learning and instinct in animals.* Methuen. [NAS]

Toates, F. (1994a) Hierarchies of control: Changing weightings of levels. In: *Perceptual control theory*, ed. M. A. Rodrigues & M. H. Lee. The University of Wales, Aberystwyth. [FT]

(1994b) What is cognitive and what is *not* cognitive? In: *From animals to animats 3*, ed. D. Cliff, P. Husbands, J.-A. Meyer & S. Wilson. MIT Press. [FT]

(in press) Cognition and evolution: An organization of action perspective. *Behavioural Processes.* [FT]

Tolman, E. C. (1932/1967) *Purposive behavior in animals and men.* Appleton-Century-Crofts. [JDS]

Tononi, G., Sporns, O. & Edelman, G. M. (1992) Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex* 2:310–35. [GNR]

Tranel, D. & Damasio, A. R. (1985) Knowledge without awareness: An autonomic index of facial recognition by prosopagnosics. *Science* 228: 1453–54. [VAS]

Tulving, E., Schacter, D. L. & Stark, H. A. (1982) Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8:336–42. [VAS]

Umiltà, C. & Moscovitch, M., eds. (1994) *Attention and performance 15.* MIT/Bradford. [CU]

Van Gulick, R. (1991) Consciousness may still have a processing role to play. *Behavioral and Brain Sciences* 14(4):699–700. [MV]

(1995) Deficit studies and the function of phenomenal consciousness. In: *Philosophical psychopathology*, ed. G. Graham & G. L. Stephens. MIT Press. [GLS]

Van Hoesen, G. W., Rosene, D. L. & Mesulam, M. (1979) Subicular input from temporal cortex in the rhesus monkey. *Science* 205:608–10. [DRH]

Velmans, M. (1990) Consciousness, brain, and the physical world. *Philosophical Psychology* 3(1):77–99. [MV]

(1991) Is human information processing conscious? *Behavioral and Brain Sciences* 14:651–726. [arJAG, GNR, MV]

(1993a) A reflexive science of consciousness. In: *Experimental and theoretical studies of consciousness* , ed. J. Marsh. Ciba Foundation Symposium No. 174. Wiley. [MV]

(1993b) Consciousness, causality, and complementarity. *Behavioral and Brain Sciences* 16(2):404–16. [MV]

Velmans, M., ed. (in press) *The science of consciousness: Psychological, neuropsychological and clinical reviews.* Routledge. [MV]

Voorn, P., Gerfen, C. R. & Groenewegen, H. J. (1989) Compartmental organization of the ventral striatum of the rat: Immunohistochemical distribution of enkephalin, substance P, dopamine, and calcium-binding protein. *Journal of Comparative Neurology* 289:189–201. [NRS]

Wan, F. J., Geyer, M. A. & Swerdlow, N. R. (in press) Presynaptic dopamine-glutamate interactions in the nucleus accumbens regulate sensorimotor gating. *Psychopharmacology.* [NRS]

Wang, J. K. T. (1991) Presynaptic glutamate receptors modulate dopamine release from striatal synaptosomes. *Journal of Neurochemistry* 57:819–22. [NRS]

Weinberger, D. R., Berman, K. F. & Zec, R. F. (1986) Physiological dysfunction of dorsolateral prefrontal cortex in schizophrenia: 1. Regional cerebral blood flow evidence. *Archives of General Psychiatry* 43:114–25. [rJAG]

Weiner, I., Lubow, R. E. & Feldon, J. (1981) Chronic amphetamine and latent inhibition. *Behavioural Brain Research* 2:285–86. [aJAG]

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

(1988) Some contributions of neuropsychology of vision and memory to the problem of consciousness. In: *Consciousness in contemporary science,* ed. A. J. Marcel & E. Bisiach. Clarendon. [aJAG]

Williams, C. W. (1986) Learning and problem solving with multilayer connectionist systems. Doctoral dissertation, University of Massachusetts. [SD]

Williams, J. H., Wellman, N. A., Rawlins, J. N. P., Geaney, D. P., Cowen, P. J. & Feldon, J. (1994) Haloperidol increases latent inhibition in high schizotypal subjects. [Abstract of paper given at 6th Winter Workshop on Schizophrenia, Les Diablerets.] *Schizophrenia Research* 11:162. [aJAG]

Wilson, B. A., Baddeley, A. D. & Kapur, N. (1995) Dense amnesia in a professional musician following herpes simplex virus encephalitis. *Journal of Clinical and Experimental Neuropsychology* 17:1–14. [rJAG]

Wilson, B. A. & Wearing, D. (in press) Prisoner of consciousness: A state of just awakening following herpes simplex encephalitis. In: *Broken memories,* ed. R. Campbell & M. Conway. Blackwell. [rJAG]

Winocur, G. (1981) The amnesic syndrome: A deficit in cue utilization. In: *Human memory and amnesia,* ed. L. Cermak. Erlbaum. [aJAG]

Wittgenstein, L. (1976) *Philosophical investigations.* Blackwell. [RI, rJAG]

Yee, B. K. , Feldon, J. & Rawlins, J. N. P. (1995) Latent inhibition in rats is abolished by NMDA-induced neuronal loss in the retrohippocampal region but this lesion effect can be prevented by systemic haloperidol treatment. *Behavioral Neuroscience* 109:227–40. [aJAG]

Young, A. M. J., Joseph, M. H. & Gray, J. A. (1993) Latent inhibition of conditioned dopamine release in rat nucleus accumbens. *Neuroscience* 54:5–9. [aJAG]

Zeki, S. (1978) Functional specialisation in the visual cortex of the rhesus monkey. *Nature* 274:423–28. [aJAG, CF]

(1993) *A vision of the brain.* Blackwell. [TJL-J]