ELSEVIER

COGNITION

Cognition 79 (2001) 161–196

www.elsevier.com/locate/cognit

Introspective physicalism as an approach to the science of consciousness

Anthony I. Jack*, Tim Shallice

Department of Psychology, Institute of Cognitive Neuroscience, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, UK

Received 15 January 2000; accepted 27 September 2000

Abstract

Most 'theories of consciousness' are based on vague speculations about the properties of conscious experience. We aim to provide a more solid basis for a science of consciousness. We argue that a theory of consciousness should provide an account of the very processes that allow us to acquire and use information about our own mental states – the processes underlying introspection. This can be achieved through the construction of information-processing models that can account for 'Type-C' processes. Type-C processes can be specified experimentally by identifying paradigms in which awareness of the stimulus is necessary for an intentional action. The Shallice (1988b) framework is put forward as providing an initial account of Type-C processes, which can relate perceptual consciousness to consciously performed actions. Further, we suggest that this framework may be refined through the investigation of the functions of prefrontal cortex. The formulation of our approach requires us to consider fundamental conceptual and methodological issues associated with consciousness. The most significant of these issues concerns the scientific use of introspective evidence. We outline and justify a conservative methodological approach to the use of introspective evidence, with attention to the difficulties historically associated with its use in psychology.

Keywords: Consciousness; Awareness; Executive control; Intentional action; Subjective reports; Introspection; Prefrontal cortex

1. Problems with the science of consciousness

Thirty years ago the attempt to produce scientific accounts of consciousness was a

* Corresponding author. Fax: +44-20-7813-2835.
E-mail address: a.jack@ucl.ac.uk (A.I. Jack).

0010-0277/01/$ - see front matter
PII: S0010-0277(00)00128-1

somewhat disreputable exercise indulged in by just a few (Mandler, 1975; Posner & Klein, 1973; Shallice, 1972). Yet, the last few decades have seen a burgeoning of scientific interest in consciousness. Commentaries on consciousness arise from a bewildering variety of scientific and philosophical traditions. New journals have been created to accommodate the 'interdisciplinary' literature. More significantly, articles on consciousness have begun to appear in the flagship journals of mainstream science (e.g. Tononi & Edelman, 1998). Nonetheless, it will not have escaped the notice of those interested in the topic that we have, at present, nothing resembling a science of consciousness (see Section 2.1).

It is no simple matter to define what makes a field of enquiry a science. However, it is tempting to believe that it is possible to get some indication of what is involved by looking at cases which are generally acknowledged to constitute successful science. The discovery of the double-helix structure of DNA provides a classical example of a scientific identification between a physical entity and a theoretical entity: as Watson and Crick (1953) pointed out, the structure proposed suggested a mechanism for replication, a property clearly essential for the gene. Many scientific theories of consciousness (e.g. Baars, 1988; Crick, 1994; Hameroff & Penrose, 1996; Tononi & Edelman, 1998) seem attractive because they appear to follow a similar model. A physical structure or process is proposed (e.g. global workspace, 40 Hz oscillations, collapse of the quantum wave equation, dynamic core) which is thought to account for some essential properties of consciousness (e.g. availability of information to multiple processes, unity of perceptual experience, non-determinism and non-locality, 'integration' and 'differentiation' of conscious states). There is a gross flaw in the analogy. The basic properties of the gene were already clear. In contrast, the essential properties of conscious experience remain undecided. The very diversity of proposed solutions indicates a problem. And, for each of the theories mentioned, either the claimed properties of conscious experience, or the existence of the physical process which is postulated to account for it, may be called into question. Even when consensus emerges – in this volume there is some convergence on a global workspace model (see Dehaene & Naccache and Dennett) – there is ample room to doubt that it is built on a solid foundation (e.g. Chalmers, 1996).

The theoretical and methodological difficulties facing a science of consciousness run deep. From time to time, a precarious consensus may emerge and cause these difficulties to fade from view. At such times, there is a temptation to forge ahead with experimental and theoretical work – to take advantage of the temporary suspension of critical impediments. Yet, there are eminently practical reasons for attending to the difficulties. Unless they are dealt with explicitly, they are likely to resurface, throwing much previous work into doubt.

2. A history of controversy

A boom and bust cycle of consensus and controversy is evident throughout the history of scientific investigations of consciousness. At the turn of the last century, the founding schools of psychology were confident in the conviction that experience should form the basic subject matter of a scientific psychology. Then intractable disagreements emerged between schools over the measurement and fundamental nature of experience (for a review, see Humphrey, 1951). These paved the way for a very different conception of psychology. Behaviourism (Watson, 1913) determinedly separated scientific accounts from the mental world seemingly known through experience (Wilkes, 1988). Even after the arrival of information processing (e.g. Broadbent, 1958; Miller, 1956), and the subsequent increase in confidence that those mistakes have been placed behind us, there has been a "widespread underestimation of the legacy of behaviourism" (Bisiach, 1988, p. 101). In the last half-century experimental psychology has dogmatically resisted any widespread use of verbal reports as data (Ericsson & Simon, 1993).

The persistence of unresolved difficulties throughout the last 50 years is evident from another boom and bust cycle. Dixon (1971) provides an early history of research in perception without awareness. He shows that belief in the hypothesis of perception without awareness was widespread prior to the influential critique of Ericksen (1960), after which confidence in the hypothesis was slow to recover. He did not foresee that shortly after the publication of his second book on subliminal perception (Dixon, 1981), similar concerns would again throw the field into controversy (Campion, Latto, & Smith, 1983; Holender, 1986; Merikle, 1984; Schacter, 1989; Shanks & St. John, 1994). A third wave of confidence has arrived. It seems very clear to the contributors to this volume that there is good evidence for perception without awareness (e.g. see Dehaene & Naccache, Kanwisher and Merikle, Smilek, & Eastwood). Yet, this confidence does not derive from a theoretical resolution of earlier difficulties. For instance, the Merikle (1984) critique of subjective measures of awareness introduced the influential notion of discrimination as an 'objective' or 'bias-free' measure of awareness. Merikle et al. argue that "subjective measures should be the preferred means for assessing the presence or absence of awareness". Yet, they do not attempt to elucidate the conditions, if any, under which subjective measures might suffer from the problem of bias. If, as we argue, introspective evidence is essential for the investigation of subjective phenomena, then it must be a priority to clearly establish its methodological limitations.[1]

It seems unlikely that there will be yet another collapse of confidence in the hypothesis of perception without awareness, in part because of the influential contribution of neuropsychological evidence from blindsight (Weiskrantz, 1986), as well as other syndromes (Driver & Mattingley, 1998; Milner & Goodale, 1995; Shallice, 1988a). Similarly, the rigorous methodological treatment of verbal protocol procedures given by Ericsson and Simon (1993) is likely to ensure that verbal reports are increasingly recognized as a valuable source of evidence (e.g. Goel, Grafman, Tajik, Gana, & Danto, 1997). Still, these important advances provide only the very first steps towards the formulation of a method for studying consciousness. Numerous conceptual difficulties remain (see Section 1).

[1] In fact, we have argued that the charge of 'bias' is misconceived, arising from a misunderstanding about the relevance of Signal Detection Theory to measurements of awareness (Jack, 1998, Ch. 2). Other methodological issues are addressed in Section 5.

2.1. Do we already have a method for studying consciousness?

Much attention has centred recently on the search for the 'neural correlates of consciousness' or NCC. Can this approach of looking for neural activity associated with conscious mental representations elucidate the mechanisms of consciousness? It cannot, but it can still provide useful data.

Crick and Koch (1998) are strong proponents of the NCC approach, boldly stating its motivating assumptions and reviewing relevant experimental research. For them, the strongest experimental model comes from the pioneering work of Logothetis and colleagues (Leopold & Logothetis, 1996; Logothetis & Schall, 1989; Sheinberg & Logothetis, 1997) using binocular rivalry. Two distinct stimuli are presented to each eye. Although the input remains constant, the conscious percept gradually alternates between the two images. Sheinberg and Logothetis (1997) demonstrate a close association between the conscious percept and the activity of 90% of single neurones (sensitive to one of the stimuli when presented alone but not the other) in inferior temporal cortex (IT) and superior temporal sulcus (STS). Many fewer neurones (~35% or less) in earlier visual areas (e.g. V1/V2, V4, V5/MT) show this close association with awareness (Leopold & Logothetis, 1996; Logothetis & Schall, 1989). These experiments suggest a special role for the processing accomplished by neurones in IT and STS in the formation of a conscious percept. Plausibly, this processing is necessary for awareness. However, the NCC approach offers little insight into the nature of this processing. Furthermore, it is implausible that this processing alone could be *sufficient* for awareness. Crick and Koch (1998) put forward the hypothesis that consciousness arises from an exchange of information between prefrontal cortex and other areas. The NCC approach neither suggested nor is capable of testing this hypothesis.

Frith, Perry, and Lumer (1999) propose extending the NCC approach to include cases such as when the stimulus changes yet subjective experience remains constant. This may provide evidence about the anatomical location of the neural processing sufficient for awareness. Yet, it is again clear that this *correlational* method can offer little insight into the mechanisms operating in those areas. Frith et al. (1999) note that the approach can only address "the *association* between consciousness and neural activity and not the more difficult question of how consciousness arises from neural activity". This is well illustrated in Kanwisher's clear discussion of research using the NCC. She uses these findings to argue that neural activity associated with the contents of consciousness is located in modality specific regions of the brain. Yet, she is forced to turn elsewhere in search of an answer to the question of what further activity or processing, over and above mere strength of activation, would be sufficient for awareness.

Most other experimental research has only addressed the question of consciousness indirectly. Psychological and neuropsychological investigations have tended to be concerned with processes that can occur in the absence of awareness (Dehaene et al., 1998; Driver & Mattingley, 1998; Marcel, 1983; Reber, 1997; Weiskrantz, 1986). This work can give us important data about the processes that are not associated with awareness – the functions that are not specific to consciousness. Yet, it does not address the processes that make a particular representation conscious – nor identify the function of consciousness.

3. The function of consciousness

The essential properties of the gene were clear because its *function* was clear. The gene was posited to explain the inheritance of biological characteristics from one generation to the next. To perform that function, the gene needed (a) to encode a large amount of information and (b) to replicate that information in order to pass it on during reproduction. The function of consciousness is not clear (e.g. Block, 1995). This provides room for the diverging views about its essential properties mentioned in Section 2.

One reason why we do not know the function of consciousness is that there is no clear agreement about what needs to be explained. In other words, there is no coherent body of established empirical data on the phenomenon – equivalent to the findings that existed on inherited characteristics. This paper aims to resolve that problem by outlining a strategy for the collection of data relevant to consciousness (the focus of Sections 4 and 5). In addition, we will suggest methods for advancing our understanding of how these findings can be explained (see Sections 6 and 7).

First, we need to know what we are looking for. We take the view that in trying to locate or find some grounding for the concept of 'consciousness', there is little point in considering organisms, mechanisms or mental processes whose status is a matter of dispute. In the absence of a strong theory, the attribution of 'consciousness' to such borderline cases remains entirely speculative. Instead, a good place to start is with folk-psychological attributions that we use in everyday conversation to describe our own mental states (see Section 8). The term 'conscious' is usually applied in two ways that appear to convey meaningful information. We may say that we are 'conscious of' something, or that we have 'consciously' performed an action.[2]

The question about the function of consciousness can be understood as one about the functional difference between the cases in which we describe ourselves as conscious, and those in which we do not. First, consider the case of being 'conscious of' some information. Driver and Vuilleumier (this volume) are concerned with investigating the processes that determine whether we are 'conscious of' a perceptual stimulus – in other words, the processes underlying attentional selection (see also Merikle et al.). The question of function is different – it is 'What does attention

[2] We will attempt to avoid confusion about these two senses of 'conscious' by referring to (perceptual) awareness and intentional action. We regard these as synonyms for the two senses. Are 'consciously' performed actions just actions that we are 'conscious of'? We do not believe so. The phenomenology of perception and action have long been thought of as distinct (e.g. James, 1890). For example, one can be 'conscious of' performing a non-intentional or automatic action. There also appear to be some unusual cases in which subjects lack 'consciousness of' distinctly intentional actions (see Section 5, which discusses Siegler & Stern, 1998).

select for?' In other words, what processes are carried out on this information, which differ from the processes carried out on non-conscious information, and over and above the processes responsible for selection itself? In order to answer this question, we need to focus on characterizing the ways in which conscious information influences thought and behaviour.

Second, consider the case where we perform an action 'consciously' (also 'deliberately' or 'intentionally', or referred to as 'volitional' or 'willed' action) as opposed to 'non-consciously' ('automatically' or in an 'ideo-motor' fashion). The phenomenological distinction between these two cases has been argued to be relevant to the understanding of impairments caused by injury to prefrontal cortex (Norman & Shallice, 1986; Shallice, 1988a). The investigation of the processes underlying consciously performed action, such as those involved in planning, problem solving, inhibition of pre-potent response, and response to novelty, has become a major topic of research. This is variously described as research into 'executive function' and 'control processes'. Whilst the exact characterization of control processes remains a topic for further investigation (see Section 7), it is generally agreed that intentional actions engage processes different to those engaged by less effortful automatic actions.

We will now argue that there is a close conceptual linkage between these two senses of 'conscious', which is highly relevant to the question of the function of consciousness.

There is a fundamental principle that pervades work on perceptual awareness. This is the principle that *awareness is necessary for intentional action*. In some cases, this principle is explicitly stated in one of a variety of different forms. For instance, Van Gulick (1994) writes "Information needs to be presented to us phenomenally for it to play a role in the choice, initiation, or direction of the intentional action." Crick and Koch (1995, 1998) claim that the function of visual consciousness is "to produce the best current interpretation of the visual scene … and to make this interpretation directly available … to the parts of the brain that contemplate and plan voluntary motor output". In other cases, experimental evidence is interpreted as directly supporting this principle (e.g. Merikle et al.). However, by far the most significant role of this principle has been as an assumption underlying methodological approaches to the investigation of awareness. In particular, it is the foundation of the Process Dissociation Procedure (Jacoby, 1991, 1998), on which a large amount of empirical work on perception, memory and learning with and without awareness is now based. The details of this procedure are not relevant here. However, an important point is its use of tasks that place two processes in opposition: a 'conscious' and an 'automatic' process. An example of this sort of task is the Jacoby exclusion task, discussed by Merikle et al. (see also Dehaene & Naccache and Dennett). In the task, subjects are first shown a masked stimulus (e.g. 'table'), and then they are asked to complete a stem (e.g. 'tab…') with any word other than the stimulus. Awareness of a stimulus is inferred from the ability to 'consciously' avoid giving that stimulus as a completion of the stem, whereas non-conscious processing of the stimulus is evidenced by the 'automatic' tendency to repeat the stimulus presented.

We view the principle that awareness is necessary for intentional action as a contingent claim that is *broadly*[3] supported by a large and growing body of empirical work (e.g. see Dehaene & Naccache, this volume; Jacoby, 1998; Weiskrantz, 1997). Similarly to others, we adopt the principle as a working hypothesis. This principle motivates the approach we suggest for the collection of data on the function of consciousness. However, we do not use this principle in the manner suggested by Jacoby and others. We do not assume that the principle holds and use it as a basis for the measurement of conscious and unconscious information. Nor do we regard the statement of this principle, as it stands, to constitute an adequate specification of the function of consciousness. Instead, we view the principle itself as the focus of investigation. The aim of our investigation is to provide a thorough scientific characterization of this principle, or (as the philosophers might say) to 'cash it out' in information-processing terms.

That a scientific characterization of the principle is essential can be seen by examining the problems that have plagued research in perception without awareness (see Section 2). Here we believe the principle has been misapplied (Jack, 1998). For many years, research in perception without awareness has been hindered by disagreement over the measurement of awareness. It has long been known that subjects are affected by stimuli which they claim not to see (e.g. Ericksen, 1960). However, doubt has been cast on this evidence on the grounds that subjective measures are prone to a number of methodological difficulties. Two methodological issues have been of particular importance. Firstly, there has been a concern that it is not clear what subjects *mean* when they claim not to be aware of stimuli. Watson (1920) drew attention to this when he wrote about 'the problem of reference'. More recently it has been referred to as the problem of establishing a 'criterion for awareness' (Ericksen, 1960; Reingold & Merikle, 1990). Heavy masking of stimuli can give rise to anomalous visual impressions, which are difficult to characterize (e.g. see Kanwisher). Perhaps subjects say they are not aware of stimuli even when they are aware of partial information which is sufficient to allow them to identify the stimulus (Fuhrer & Ericksen, 1960). Secondly, subjects may not report being aware because, although they have a fleeting visual impression of the stimulus, they lack *confidence* that this impression is veridical (Merikle, 1984; Shanks & St. John, 1994).

These methodological problems are important (see Section 5). However, they are only problems *in principle* for experiments purporting to show perception without awareness – there is no empirical evidence indicating that they actually occurred (see Merikle et al.). Furthermore, they might be avoided by the adoption of more

[3] As mentioned previously, a great deal of work has simply assumed that this principle holds, and used it as a basis for deciding either that the subject lacks awareness or that a particular sort of action is intentional. Experimental findings may only be read as supporting the principle when there are independent sources of evidence establishing (a) that the subject is perceptually aware (or unaware) and (b) that the action is intentional (or automatic). In general, only evidence relevant to (a) is collected. However, enough is known about the qualitative differences between different types of action in terms of their behavioural properties (e.g. susceptibility to dual task interference) to infer whether an action is intentional or automatic, and thus provide broad support for the principle (see Section 6).

rigorous methods for questioning subjects (e.g. see Dixon, 1971, 1981). By themselves, these possible methodological problems should not have been sufficient to cast doubt on the large number of studies showing perception without awareness. In fact, the most influential critiques of perception without awareness (Ericksen, 1960; Holender, 1986; Merikle, 1984) all contained an extra element. All of these reviews gave bite to their methodological criticisms by using evidence from other, objective, measures that appeared to indicate that when subjects said they were not aware of the stimuli, they actually were aware of them (e.g. discrimination performance in Ericksen, 1960; Merikle, 1984; as well as other measures in Holender, 1986).

Why was this objective evidence interpreted as providing evidence about the subjective state of awareness? The answer is explicit in the original critique of Ericksen (1960). He assumed that discrimination performance provided a measure of awareness because the task appeared to involve an intentional response directly concerning the identity of the stimulus. In summary, the history of perception without awareness shows that methodological concerns about subjective evidence were lent weight because of the overly crude application of the principle that awareness is necessary for intentional action. By employing the principle, experimenters attempted to collect objective rather than subjective evidence of awareness. Yet, subjects viewed themselves as ‘guessing’ (e.g. Marcel, 1983), which is clearly phenomenologically distinct from full intentional action. Evidently, subjective measures of awareness are more reliable than experimenters’ intuitions about whether an action is intentional or not. It is a mistake, historically motivated by behaviourism, to suppose that objective evidence is needed to validate subjective measures of awareness (e.g. as argued by Merikle, 1992; Merikle, Joordens, & Stolz, 1995). Rather, subjective evidence is needed to validate the claim that an objective measure serves as a measure of *awareness*. If the principle (that awareness is necessary for intentional action) is correct, then the fact that subjects report no awareness of stimuli that they can discriminate at above chance levels must indicate that discrimination is not an intentional act. Accordingly, we believe that the task of discriminating between a small set of known stimulus alternatives – a fast low-effort task involving a one-to-one stimulus-response mapping – can be carried out largely automatically.

If the principle that awareness is necessary for intentional action is to be preserved, and to be of practical use, it needs to be made more precise. We identify three ways in which this principle needs be better characterized. Progress in all three is, we believe, essential for a scientific understanding of the function of consciousness. In order to understand why this principle needs further clarification, it is important to realize that it is not, in the first instance, a principle couched in information-processing terms. This should be clear, since there is at present no definitive information-processing account of intentional action, and no account had even been attempted when Ericksen (1960) explicitly appealed to the principle and mistakenly assumed that awareness is necessary for discrimination.

We regard this principle as being based primarily on introspective evidence. In other words, the terms used in the principle (i.e. ‘awareness’ and ‘intentional’) are terms which we come to understand through consideration of, and abstraction from, the phenomenology associated with our own mental states (see Section 8). A *scientific* specification of the function of consciousness requires a restatement of this principle in information-processing terms. However, in order that this restatement should count as a scientific specification of the function of *consciousness*, it is essential that we formulate the information-processing account in such a way that it coheres with our phenomenological understanding of the principle. Therefore, we proceed in two directions. Firstly, we attempt to provide an account of how the principle should be grounded in phenomenology. Secondly, we suggest two ways in which we may generate and/or refine an understanding of the principle in information-processing terms: (i) by providing a more precise specification of the relation between perceptual awareness and intentional action; and (ii) through the search for concepts which can be used to explain intentional action. We discuss next the grounding of the principle in phenomenology, and implications for the collection of data on consciousness. The relation between awareness and intentional action (i) is further discussed in Section 6. There we introduce the Supervisory Attentional System model (Norman & Shallice, 1986; Shallice, 1988a). This model also provides a framework for understanding intentional action (ii). Section 7 discusses how this framework may be further refined.

How can our phenomenological understanding of 'awareness' and 'intentional action' be used to guide the search for the scientific formulation of the principle? This issue touches on a deep and long-standing methodological problem for psychology. Namely, how can we use introspective evidence to inform scientific accounts of mental processes? This issue is discussed in greater depth in Section 5. This discussion is central to the project we outline here, since we regard introspective evidence as essential for the generation of solid empirical results on the function of consciousness. In our proposal, introspective evidence is necessary to identify certain processes, which we shall call Type-C processes, that can only operate on information available for report. In addition, introspective evidence may help in the search for tasks involving intentional action. In our framework, the principle that awareness is necessary for intentional action can be more precisely stated as the hypothesis that tasks involving intentional action recruit Type-C processes, whereas automatic actions do not.[4]

We regard the use of introspective evidence to measure perceptual awareness as largely unproblematic, provided certain methodological precautions are taken. The reason that introspective reports concerning states of perceptual awareness are unproblematic is that these reports are closely related to objective judgements

[4] Could there be cases in which Type-C processes operate, and yet where there are no 'conscious contents'? We would allow such cases. Our claim is that the operation of a Type-C process on information is a necessary and sufficient condition for awareness of that information. However, there may be cases in which a Type-C process operates without operating on any information. In other words, there may be special types of 'unguided' intentional action for which perceptual awareness is not necessary (for an analogous concept, see Humphrey, 1951). This would be our interpretation of cases where subjects are asked to select an action (from a range of possible actions) without having any information on which to base the selection (as reviewed by Frith, 2000). For a related discussion of awareness and executive control see Badgaiyan (2000).

concerning stimuli presented. For instance, a judgement of awareness may be equivalent to a judgement about the presence of a stimulus (Cowey & Stoerig, 1995). Thus, judgements of perceptual awareness are well grounded – it is relatively easy to establish what subjects *mean* by their reports. The use of introspective evidence to identify instances of intentional action is much less straightforward – it is much harder to establish what subjects *mean* by the claim that an action was intentional (see Section 5). Nonetheless, we believe it is important to attempt to 'anchor' our understanding of intentional action, by identifying what we regard as the paradigm example of a conscious act. This is the act of introspection itself, i.e. the act of reflecting upon, imagining, or comparing between one's own mental states. Some acts of introspection involve making a judgement about one's own mental state, and result in the production of a response that indicates the judgement made (e.g. 'the coffee tastes more bitter than I remember the last time I tried it'; see Dennett, 1988). We regard this as the paradigm case in which awareness is necessary for intentional action. For in this case, it is clear that *the phenomenology associated with the perceptual experience plays a causal role in the production of response*. In our view, we will have a theory of consciousness when, and only when, we can provide a detailed information-processing account of all[5] the processes involved in making judgements of this sort.

At present, we are a long way from providing any such account. How are we going to get there? Our approach rests on the assumption that mental states that appear, introspectively, to be related are likely to be functionally related (see Section 8).

Some 'Higher Order Thought' theorists hold that mental states are conscious just if we are introspecting, i.e. only when we are having thoughts about that state (Rosenthal, 1986). However, Ryle (1949, p. 164) remarks that "introspection is an attentive operation and one which is only occasionally performed, whereas consciousness is supposed to be a constant element of all mental processes". Whilst we agree on the special theoretical status given to introspection, our view is different from Rosenthal's. We explain the impression that we are virtually constantly conscious by positing that the processes that underlie introspection are closely functionally related to the processes that operate during the performance of other intentional mental operations. According to our view, to be conscious of information, it is sufficient for any Type-C process to be effectively operating on that information.[6] As will become clear in Section 4, Type-C processes are involved in many judgements that only explicitly concern the world outside the subject (e.g. fine discriminations of colour shade). Thus, on our view, there is no requirement for

[5] It is important to appreciate that an account that merely serves to explain the origin of the difference in perceptual appearance, i.e. why the coffee seems more bitter, is not going to be sufficient. Critically, what we want is an account of the more general processes responsible for comparing any current experience with any past experience, and translating the resulting information into a form capable of guiding response. For some preliminary remarks, see Section 6.

[6] Note that it becomes meaningless to discuss the temporal aspects of phenomenal processes at a finer grain than the operation of individual Type-C processes (see Dennett, 1991).

the subject to be making a judgement ‘about’ their own mental state in order to be conscious.[7]

It is important to note that introspection, as understood here, cannot be defined purely behaviourally – for instance, as occurring whenever subjects make reports about their own mental states. Along with Ryle (see above), we regard introspection as an attentive activity, requiring mental effort and resources. We follow Ericsson and Simon (1993) in their claim that some forms of concurrent verbal protocol can be carried out largely automatically. This may occur (e.g. whilst solving an arithmetic problem) when subjects are simply required to verbalize their conscious contents as they naturally occur, and where the reported contents (e.g. numbers) are of a form that can easily be converted into language. Whilst these reports constitute a form of subjective evidence, in the sense that there is no *direct* method of verifying their accuracy, their production (after practice) does not appear to require subjects to make introspective judgements (for a closely related claim, see Weiskrantz, 1997, p. 75).

4. Type-C processes

In this section we provide an outline for a cognitive research project. As mentioned in Section 3, the aim of this project is to provide data in need of explanation by a theory of consciousness. Type-C processes are defined as processes that can only operate effectively on information when normal subjects report awareness of that information. The aim is to identify these Type-C processes by providing examples of tasks which are well specified in two senses: (i) the Type-C process should play a role in the production of responses which can be experimentally isolated; and (ii) the task should reliably recruit the Type-C process (i.e. it should not be possible to perform the task accurately except by recruiting the Type-C process). Candidate processes may be identified, and then tested, by applying the principle that there must be no situations in which a Type-C process can occur in the absence of reported awareness of the relevant content.[8] Later in the section, we provide an initial list of seven Type-C processes. Only four of these, listed as ‘experimentally characterized Type-C processes’ (Section 4.2), are well specified in the two senses outlined above. The other three ‘pre-experimentally characterized

[7] It may turn out, on closer inspection, that all Type-C processes do in fact involve the subject making a judgement about their own mental state. According to Rosenthal, we might overlook this because we can fail to notice that we are having a thought about our own mental state. On his scheme, noticing that we have these thoughts would require us to have a thought about a thought about our own mental state. We allow for Rosenthal’s position, but we do not assume it is correct.

[8] In theory, it may be possible to interfere specifically with the Type-C processes underlying introspective report but not with other Type-C processes. Thus, the finding that a process can operate when subjects are unable to report the relevant content does not logically entail that the process is not Type-C. In practice, any such finding in normal subjects must be taken to indicate the process is not Type-C unless a strong a priori case can be made for the interference being specific to report. Otherwise, the hypothesis that a process is Type-C would be unfalsifiable.

Type-C processes' (Section 4.1) are included because of their central theoretical importance.

Although often overlooked, it is self-evident that a full theory of consciousness should give an account of the processes which specifically operate on reportable information, and underlie its various behavioural effects – in particular the processes involved in the actual production of reports (see also Dennett, 1991, p. 255; where he calls this the 'Hard Question'). However, opinions differ as to the theoretical status of such an account. Some theorists (e.g. Crick & Koch, 1990) regard these processes as subsidiary to the processes that actually give rise to or constitute a state of awareness. In the terminology of Weiskrantz (1997, p. 203), this would be the same as the view that Type-C processes merely *enable* the subject to use or communicate conscious information. The alternative view would be that Type-C processes themselves *endow* awareness. Weiskrantz (1997, p. 76) suggests that "it is the very achieving of the ability to make a commentary of any particular event that is what gives rise to awareness". Similarly, our view is that all Type-C processes (including, of course, those involved in making commentaries or introspective reports) share some basic information-processing operations, and that those operations actually give rise to awareness. Consequently, we regard the project of identifying Type-C processes as essential for the collection of further empirical data on consciousness. Through further investigations of Type-C processes, we aim to get a 'fix' on these basic information-processing operations. Nonetheless, the 'endowing' view is not assumed by the project outlined in this section. This view would need to be abandoned if, for instance, it is shown that there is no single functional distinction that distinguishes between Type-C processes and other processes (for a related possibility, see Allport, 1988).

A theoretically important subset of Type-C processes are those that actually involve making an introspective judgement. One way of attempting to identify tasks of this sort is by looking for tasks that appear, introspectively, to involve thoughts about one's own mental states. However, how can objective evidence be used to identify tasks involving an introspective judgement? In certain cases, a type of task may already have been extensively investigated behaviourally. In this case, the hypothesis that the task involves an introspective judgement may be supported because it helps to explain patterns of behavioural data in a number of experiments. A strong example of this sort is provided by Koriat and Goldsmith (1996). They present an argument for the existence of and importance of recognizing metamemory processes in free recall. Koriat and Goldsmith's claim, which they support empirically, is that, in free recall tasks as opposed to forced-choice recognition tasks, subjects control the production of items on the basis of judgements they make concerning their own recall accuracy. Experiments involving attributions of perceptual fluency (Jacoby & Whitehouse, 1989; Mandler, Nakamura, & Van Zandt, 1987; Whittlesea, 1993) provide a second example where the hypothesis that some tasks involve an introspective judgement helps to explain patterns of data across many experiments (Bornstein & D'Agostino, 1992; Jack, 1998).

Even where evidence from a large number of behavioural experiments is not available, it is sometimes possible to make a strong case for the claim that a task involves an introspective judgement. Consider the Jacoby exclusion task discussed in Section 3. It is possible to make a case, supported by objective evidence, that subjects typically understand the instructions for this task as involving a judgement about their own state of awareness. In other words, subjects understand the instructions to be 'If you are aware of the masked word, do not give it as a completion to the stem. Otherwise, give the first word that comes to mind.' (Jack, 1998).[9]

We now provide an initial list of Type-C processes.

4.1. Pre-experimentally characterized Type-C processes

1. 'Conscious reflection' – the process which occurs when we reflect upon the nature of an experience, and which underlies the ability to make judgements based on the nature of that experience (e.g. judgements of familiarity and perceptual clarity; Whittlesea, Jacoby, & Girard, 1990). This process is held to underlie the ability to discriminate between mental states, as well as discriminations concerning the world external to the subject which require a careful consideration of the phenomenology associated with perception (e.g. fine discriminations of colour). A subset of cases will further involve 'meta-awareness' – the process that occurs when we have the thought that we have experienced a particular conscious mental state (e.g. the thought 'I was aware of x'). This process is held to underlie the ability to categorize one's own states of awareness.
2. The process which underlies the ability to freely report the identity of an unanticipated but known stimulus at the time of presentation, and which occurs when we have the subjective sense of spontaneously recognizing or 'noticing' (Bowers, 1984) a stimulus. We take this to be a pre-experimental process as it is presumed to be the same process whatever type of stimulus is being recognized.
3. The process underlying the re-experiencing of a past event held in memory. This is a process of 'autonoetic consciousness' involving 'ecphory', in the terminology of Tulving (1983), and is held to be the basis of 'remember' as opposed to 'know' judgements (Gardiner, Ramponi, & Richardson-Klavehn, 1998). In addition, the processes underlying the use of information from episodic memory for the strategic regulation of performance (Koriat & Goldsmith, 1996), checking the veridicality of recalled information (Burgess & Shallice, 1996a), or for the planning of action (Schank, 1982).

[9] Jack (1998) shows that when subjects are given instructions that stress the need to avoid repeating the word presented, regardless of whether or not they consciously see it, they adopt additional strategies. For instance, when they cannot see the masked word, they complete the stem with unusually long and infrequent words that are unlikely to match the high-frequency words presented. This illustrates a distinction between understanding the task instructions as purely objective (always try to avoid giving the masked word), as opposed to understanding them as having a subjective component (only avoid repeating the masked word if you are aware of it).

4.2. Experimentally characterized Type-C processes

4. The process involved in encoding material into episodic memory – the process, occurring at the time of stimulus presentation, which enables the later process of retrieval to occur, in which we have the subjective sense of recollecting the perceptual event (autonoetic consciousness; Tulving, 1983). This process is held to underlie the ability to retrospectively report the identity of the earlier stimulus in, say, a free recall task. This process is also a prerequisite for above chance performance on *some* recognition tasks (e.g. Mandler et al., 1987), although it is not for others. On the two-process theory of recognition (Mandler, 1980), the critical factor would be whether the subject is willing to make their response purely on a feeling of familiarity. Familiarity can be evoked by perceptual fluency (see process 6 below and Merikle & Reingold, 1991).
5. The process of 'exclusion' involved in the Jacoby exclusion task, discussed in previous sections (see also Merikle et al.).
6. The process underlying the discounting of perceptual fluency due to prior exposure of a stimulus. This process is held to underlie the abolition of various 'perceptual fluency' effects, which have been shown to influence judgements of familiarity, preference, perceptual clarity, brightness and darkness (e.g. Mandler et al., 1987; Whittlesea et al., 1990). For instance, Jacoby and Whitehouse (1989) show that subjects are more likely to judge that a word has been presented in a previous study episode if it is presented, heavily masked, just prior to the judgement being made. Subjects do not show the same bias when the stimulus is lightly masked and clearly visible. In the second case, awareness of the word presented allows them to discount the effect of fluency on their familiarity judgement. Bornstein and D'Agostino (1994) and Whittlesea (1993) provide experimental evidence for the generalization of this effect to other task contexts.
7. The process underlying the addition of stimuli to a discriminatory response set. Jack (1998) investigated the situation where subjects have to identify single letters in a perceptual masking experiment. Subjects were told that four different stimuli were to be presented; however, they were only familiar with the identity of three of these stimuli. The fourth stimulus was initially presented only heavy masked, but later in the experiment it was also presented under light masking conditions. Subjects were able to discriminate the three familiar stimuli well above chance under all masking conditions throughout the experiment. However, they were only able to discriminate the fourth unanticipated stimulus once it had been consciously identified in the lightly masked condition. Incorporation of the stimulus into the response set required conscious identification.

If the current approach is correct, the tasks listed above must all involve at least one component process, such that awareness of a particular content is necessary for the operation of the process. For each of the seven examples, there are no situations known where the *Type-C* process can occur *without* the relevant content being available for report. This is critical. A basic assumption underlying the approach is: if on some occasions subjects carry out a task in the absence of awareness of particular information (e.g. the identity of a masked word), then we conclude that *Type-C* processes are not necessary for processing the information in that manner (e.g. semantic priming); any awareness of the relevant content on other occasions is taken to be due to the operation, on those occasions, of one or more additional processes. Awareness of a word involves processes over and above those mediating semantic priming effects, as shown for instance by Marcel (1983).

We have not included in our list one process hypothesized by Merikle et al. (this volume) (see also Cheesman & Merikle, 1986; Merikle et al., 1995) to require awareness. This is the process held responsible for the facilitation of reaction time due to stimulus redundancy in their modified Stroop task. Other experiments indicate that closely related effects occur in the absence of awareness (Jack, 1998; Miller, 1987; Shanks & Johnstone, 1997). Clearly, it is important to make an effort to investigate any suggestion that a candidate process can operate when the relevant information is not available for report. The historical example of discrimination indicates that tasks that initially appear to involve an intentional action may still not qualify as Type-C processes (see Section 3).

The division between experimental and pre-experimental Type-C processes is not simply a division between processes that are and are not engaged in common experimental paradigms. The three pre-experimental processes listed, and their variations, occur frequently whilst subjects are carrying out a wide range of experimental paradigms. In many cases, the operation of these pre-experimental processes may be the immediate precursor of response (e.g. process 1 in the experiments of Mandler et al., 1987). However, it is only for the tasks listed under experimental Type-C processes (Section 4.2) that an experimental manipulation affects whether or not the Type-C process operates, and where the operation of the Type-C process is clearly reflected in performance. Thus, the analysis of tasks that meet the criteria for experimental Type-C processes may inform hypotheses concerning the processes involved in introspection. An initial goal of our proposed project would be to extend the list of experimental Type-C processes through the identification of tasks that similarly isolate the pre-experimental Type-C processes (Section 4.1). In addition, it may be possible to isolate other Type-C processes using tasks that involve *problem solving*, *planning*, *reasoning*, *rule generation and verification*,[10] *inhibition of prepotent responses*, *correction of action slips*, *following instructions*, and *response to novel situations*.

5. Introspective evidence

In Section 4, we distinguished between our belief that Type-C processes *endow* consciousness and the view that Type-C processes merely *enable* the communica-

[10] Explicit, as opposed to implicit, learning may be accounted for by the operation of these processes, and the subsequent use of inferred information in the creation of schemas for action (see also Shanks & St. John, 1994; St. John & Shanks, 1997).

tion and/or use of information that has already reached consciousness. In either case, introspective evidence is needed to identify types of behaviour that can only occur when information is available for introspective report. Nonetheless, according to the 'endowing' view the primary goal of a theory of consciousness should be to account for particular sorts of objectively observable behaviour. More specifically, the theoretical framework we outline in Section 6 aims to provide an initial account of the processes that distinguish one set of objectively observable behaviours (which can only be carried out when information is available for report) from another (which can be carried out when information is not available for report).

In contrast, most theorists tend to assume the 'enabling' view, and adopt a strategy that relies on introspective evidence in a different way. This strategy involves the identification of a very limited set of specific properties of experience. A theory of consciousness is then proposed in which particular mechanisms or processes are claimed to account for these subjective properties. The scientific theories of consciousness listed in Section 1 all adopt this approach. For example, Tononi and Edelman (1998) put forward their dynamic core hypothesis partly on the basis that it accounts for the observation that "each conscious state comprises a single "scene" that cannot be decomposed into independent components".

Philosophers have also argued that experience has peculiar properties, although they frequently do so in order to argue that consciousness presents a special problem for scientific accounts. For example, Block (1995) argues that scientific attempts to account for the functional role of conscious information ('access-consciousness') do not address the phenomenological properties of conscious experience ('phenomenological-consciousness').

We are sceptical of accounts that place such a heavy burden on analyses of such individual properties of conscious experience. Our concern is that these properties are highly abstract and based on the consideration of introspective evidence alone. It is hazardous to place any reliance on generalizations derived from experience unless they can be validated by objective evidence. This is because the principal problem with the use of introspective evidence is that it is prone to misinterpretation. This is dramatically illustrated by the history of psychophysics.

In an important book, Laming (1997) reviews the history of the measurement of sensation, beginning with Fechner. He finds that unjustified interpretations of subjects' reports have caused the field to be mired in controversy right up to the present day. It was Fechner's conception of a 'physics of the mind' and search for psychophysical laws (relating physical dimensions of the stimulus to subjective dimensions) which led to the emergence of experimental psychology in the middle of the nineteenth century. Nonetheless, Laming argues that Fechner was fundamentally mistaken in "the implicit assumption that sensation admitted measurement on any kind of continuum at all". In a thorough analysis, Laming outlines where each of a series of psychophysical laws breaks down, from the Weber–Fechner law (which accounts for comparisons between stimuli) to Stevens' power law (which attempts to describe judgements of absolute magnitude). He concludes: "The evidence so far to hand does not support any intermediate continuum at the psychological level of description which might reasonably be labelled 'sensation'. While the underlying pattern of sensory neural activity is obviously germane to the perceptual process, not even that can be identified as 'sensation', essentially because there is no corresponding psychological entity. Although this rejoinder might seem no more than a philosophical quibble, it does matter in practice. Experiments by different investigators, seeking to measure the perception of that neural activity as sensation by different methods, have found no basis for agreement."

Laming advises scientists as follows, "…without independent corroboration, introspective evidence should not be taken at face value. Psychologists who disregard this dictum are liable to involve themselves in artificial arguments… [T]he seeming impossibility of such corroboration does not mean that scientists should proceed without it; it means, instead, that the question addressed lacks the empirical basis needed for an answer to be agreed and that scientists should not proceed at all."[11]

The philosopher David Chalmers has argued that conscious experience cannot be captured by the 'third-person' language of science (Chalmers, 1996; also see Dennett in this volume). He proposes instead that a different sort of language should be developed – one that captures our 'first-person' experience. According to Chalmers, these authentic first-person accounts might then be related to physical states, which he believes have a one-to-one correspondence with conscious states. Yet, the project that Chalmers proposes anew appears to be identical to Fechner's.

There is a simple and fundamental reason why all attempts to get at the 'raw data' of experience fail: introspective evidence always arrives already interpreted. In other words, all descriptions of experience, no matter how basic, carry implicit theoretical commitments of one sort or another. In order to understand and describe an experience, subjects need to employ concepts and categories (i.e. mental state concepts, see Section 7.1). Thus, introspective reports may be seen as the product of two factors: firstly the 'raw data', which the subject has access to via introspection, and secondly the conceptual framework, or 'model', which the subject uses to interpret that data. The extent to which subjects are correct depends on the validity of their model for interpreting the 'raw data'. As the history of psychophysics illustrates, introspective reports may be highly misleading if the self-reflective concepts used by subjects rely on the wrong implicit assumptions about brain organization and function.

The use of objective evidence to inform scientific accounts also depends on interpretation. Kuhn (1996) eloquently argues that the observations that are used to support theories in physics are always theory-laden. Similarly, objective behavioural data, for example from a perceptual discrimination task, are of little use unless they can be interpreted as representing a genuine attempt by the subject to comply with the task instructions. Only then can percent correct or reaction time

[11] See Laming (1997, pp. 208–209). This needs some clarification. The point is not that, in practice, it should be simple to find objective evidence to corroborate valid hypotheses based on introspective evidence. It is often hard to find evidence for scientific theories. However, there is an onus to establish that any given interpretation can be corroborated in principle by objective evidence. If this cannot be convincingly established then the interpretation must be regarded as questionable.

measures inform hypotheses about the information processing taking place. The critical difference between introspective and objective evidence is that with objective evidence it is possible to go back to the raw data. For instance, a closer examination of subjects' responses may support or invalidate the assumption that they were following the task instructions. This makes it easier to resolve disputes over the interpretation of objective data. In the absence of this safeguard, great care is needed to ensure that interpretations of introspective evidence are well grounded.

How, then, can introspective evidence be used to inform scientific accounts? Essentially, there are two areas for investigation. First, we may attempt to use introspective evidence to examine the self-reflective subsystems – the processes responsible for the 'model' subjects have for understanding their own mental states – and the effect these processes have on thought and behaviour. Second, we may attempt to use introspective evidence to distinguish between mental states – by using the information that is made available to reflective subsystems when subjects introspect.

The processes involved in reflection remain poorly understood. However, an initial attempt to examine the operation of the self-reflective subsystems can be seen in a recent experiment on self-reports of strategy use in children (Siegler & Stern, 1998). In this study, children (around 9 years old) are given arithmetic problems of the form '$Y + X - X =$'. Initially the children solve this problem by first adding 'X' to 'Y', and then subtracting 'X' from the result. However, with experience the children stop performing any arithmetic calculations and simply state the answer as 'Y'. The experimenter can reliably discern the strategy used from the response time. The interesting finding arises from children's reports when they are asked how they solved the problem. Once the strategy is well established, children reliably report its use. However, for the first few trials on which they use the strategy, children report counting just as they had before. In other words, the children appeared to lack awareness of their own discovery and use of the strategy. This study shows that, at least during development, repeated experience of a mental state or process is necessary before the model is updated to allow accurate introspective identification of that state or process. This is surprising, since slow and deliberate processes such as arithmetic calculation and strategy application are usually thought of as directly available for report (Ericsson & Simon, 1993). We take it to support our broad distinction between the information available to introspective processes and the model used to interpret that information.

How can introspective evidence be used to distinguish between mental states? According to the view presented here, introspective processes have access to information concerning limited functional aspects of mental states. However, we do not (usually) interpret this information *as information about our functional states*. Instead, we interpret this information using our own implicit folk-psychological theories. These conceptual frameworks may be developed through consideration of one's own experiential states and attempts to relate this information to observations of behaviour, as well as through conversation with others. Moreover, there will be interpersonal differences in the conceptual frameworks or 'models' used by subjects, even to the extent that subjects may mean different things when they use the same mental state terms (Watson, 1920). In other words, different subjects may use different criteria for response in introspective report tasks (Ericksen, 1960; Kahneman, 1968; Reingold & Merikle, 1990). The resulting 'self-portrait' of the subject's mental state will remain obscure, due to difficulties in understanding their 'palette' of self-reflective concepts.

We propose that the critical process necessary for the productive scientific use of introspective reports is that of *replacing* or *refining* the subject's model for understanding their own mental states. There are two ways of doing this. The first involves providing the *subject* with a well specified model for interpreting their own experience. The second involves *re-interpreting* the subject's reports in terms of a testable functional theory. The first of these methods can be productively used to yield quantifiable empirical data. The second can be used in exploratory studies, which are concerned with the generation and refinement of theoretical accounts.

In some situations, it is a conceptually simple matter to ensure that subjects are using a well specified model for interpreting their reports, provided the relevant states can be reliably elicited. This is well illustrated by the pioneering work of Logothetis and colleagues on bistable percepts, involving the collection of introspective reports from primates (Leopold & Logothetis, 1996; Logothetis & Schall, 1989; Sheinberg & Logothetis, 1997). All that it is necessary to do is to elicit the relevant states, and teach subjects to respond accordingly. This procedure provides subjects with reference points that serve to guide their responses, thus ensuring that subject and experimenter have a common understanding of what is meant by the report response. Thus, Logothetis and colleagues trained primates to respond to visual stimuli presented in isolation. Once the primates could reliably discriminate between the two stimuli, the responses could be used to infer the contents of awareness during binocular rivalry – when the two stimuli were presented simultaneously, one to each eye.

In practice, two sorts of difficulties arise with this procedure. Firstly, anomalous conscious states may arise that are hard to categorize (see Kanwisher). Secondly, the criterion for response may change with time and/or experience. The first of these difficulties can be tackled by employing an initial development phase, in which subjects give free reports of the phenomenology involved in the task. This allows the generation of a range of relevant categories. In binocular rivalry, parts of both images can sometimes be seen simultaneously during an intermediate phase when neither image is dominant. Logothetis and colleagues circumvented this problem by creating images which, when presented alone, were indistinguishable from this experience. The primates were also trained using these images. The solution to the second problem is also illustrated in the experiments of Logothetis and colleagues. During testing, they occasionally used 'catch trials' in which the non-rivalrous images used in training were presented. This allowed them to check that the primates were maintaining the intended criterion for response.

It should be clear from the example above that rigorous methods can be available for specifying the model that subjects use to categorize states of awareness, if two conditions apply. The first is that the relevant conscious states can be reliably elicited by varying the stimulus and/or experimental conditions. The second is that the introspective reports, concerning the subject's state of awareness, are closely related to objective judgements, concerning the world outside the subject. Thus, in the example above, the introspective report concerning the contents of consciousness was effectively equivalent to an objective judgement concerning the stimulus presented. Two other examples of equivalence are given. The introspective judgement of being 'aware of something' is effectively equivalent to an objective judgement of presence or absence. The introspective judgement of being 'aware of a word' is effectively equivalent to an objective judgement of whether a word or a non-word (letter string) was presented.[12] Introspective reports concerning states of perceptual awareness are not generally problematic, since these two conditions can usually be met.

It is also clear historically that verbal reports obtained from 'think aloud' protocols can be a valuable source of evidence in other cases (Ericsson & Simon, 1993). This use of introspective evidence appears to be successful for two reasons. Firstly, the instructions for 'think aloud' protocols discourage subjects from providing elaborate interpretations of the mental states they report, thus helping to ameliorate the difficulties associated with rationalization (Gazzaniga, 1985; Nisbett & Wilson, 1977). Secondly, these verbal protocols are used in the development of functional accounts of the processing that the subject is carrying out in the situation. Accounts of this sort may be tested by standard scientific means. For example, in a study of autobiographical memory, Burgess and Shallice (1996a) used a complex retrospective commentary procedure in which subjects produced short descriptions, or even single words, for each experience as they attempted to recall. The tape was then replayed and the subject elaborated on the introspective responses they had produced a minute or two before. The reports were then categorized by the experimenters into 25 types of thought element selected on the basis of pilot studies. A model – a development of that of Norman and Bobrow (1976) – was produced to account for both qualitative and quantitative aspects of the memory retrieval protocols. The model was then applied to a number of findings from objective neuropsychological investigations (Burgess & Shallice, 1996a; Dab, Claes, Morais, & Shallice, 1999; see also the related position of Schacter, Curran, Galluccio, Milberg, & Bates, 1996) and cognitive neuroscience (Fletcher, Shallice, Frith, Frackowiak, & Dolan, 1998; Henson, Shallice, & Dolan, 1999).

In other cases, for example involving neuropsychological and psychiatric disorders, introspective evidence can only be used to inform scientific accounts when the experimenter adopts a different interpretation to the patient. Critically, these cases require the experimenter to do considerable work eliciting reports in order to understand and avoid the erroneous interpretations arrived at by patients. Cytowic illustrates this point very well in his discussion of the work of Heinrich Kluver, who

[12] These examples are chosen because they have been confused in some accounts of perception without awareness. For instance, Holender (1986) interprets evidence of the ability to detect the presence of word stimuli as evidence that subjects are aware of the identity of the words. In doing so he ignores earlier work showing that subjects can base presence/absence responses on perceived duration of the stimulus sequence (Fehrer & Biederman, 1962) and/or apparent motion (Kahneman, 1967).

carried out extensive work attempting to understand the experience of hallucinators. Cytowic (1997) reports that Kluver was initially "frustrated by the vagueness with which subjects described their experience, their eagerness to yield uncritically to cosmic or religious interpretations, to 'interpret' or poetically embroider the experience in lieu of straightforward but concrete description, and their tendency to be overwhelmed and awed by the 'indescribableness' of their visions". Yet, Kluver (1966) eventually identified three classes of visual pathology: (i) 'form' constants, which describe hallucinated patterns, e.g. grating, lattice, honeycomb or chessboard patterns; (ii) alterations in the number, size and shape of perceived objects; and (iii) alterations in spatiotemporal relations between objects. Ffytche and Howard (1999) have further extended this work, illustrating the consistency of these and other pathological reports across a range of clinical conditions, and reviewing neuroscience research that may be relevant to their explanation. For the case of synaesthesia, Cytowic (1997, p. 24) summarizes the attitude that is required in order to reduce introspective reports to scientifically useful descriptions as follows: "Though synaesthetes are often dismissed as being poetic, it is we who must be cautious about unjustifiably interpreting their comments." We regard abstract properties of awareness, derived solely from introspective evidence, as a dangerous base for a science of consciousness. Nonetheless, when introspective evidence is carefully collected and interpreted in specific experimental situations, then it can be of considerable scientific value.

6. A framework for understanding conscious processes

In the information-processing accounts of consciousness developed in the 1970s the unitary nature and control functions of consciousness were explained in terms of the involvement of a limited capacity higher-level processing system (Mandler, 1975; Posner & Klein, 1973). With the diversification of processing systems that cognitive psychology, cognitive neuropsychology and cognitive neuroscience have produced and the realization that processing systems are often informationally encapsulated (Fodor, 1983), it became less plausible to associate the unitary characteristics of consciousness with the operation of any single processing system. In Shallice (1988b) an alternative approach was put forward. It was argued that a number of high-level systems have a set of characteristics in common which distinguish them from the run of cognitive systems which realize routine informationally encapsulated processes. It was held that the contrast between the effective operation of these systems and those realizing informationally encapsulated processes corresponded in phenomenological terms to that between conscious and non-conscious processes.

This approach to consciousness was based on the model of Norman and Shallice (1986) which was originally introduced to explain results from experimental psychological studies on attention and the impairments of patients with prefrontal lesions, the domain to which it has primarily been applied (Della Malva, Stuss, D'Alton, & Willmer, 1993; Shallice, 1988a; Shallice & Burgess, 1996). The

Norman–Shallice model is concerned with action selection. It has three main processing levels. The lowest is that of special-purpose processing subsystems, each specialized for particular types of operation, such as translating from orthographic to phonological representations. Second, there are held to be a large number of action and thought schemas, one for each level of each well learned routine task or subtask. Schemas are selected for operation through a process involving mutually inhibitory competition (contention scheduling). The operation of schemas in a particular situation is dependent on the way their arguments are filled, which is done using representations from other systems, e.g. the perceptual systems. (See Cooper and Shallice (2000) for technical details and simulations of relevant neuropsychological disorders, and Dehaene and Changeux (1997) for a closely related simulation.) Third, to cope with non-routine situations, an additional system – the Supervisory Attentional System (SAS) – provides modulating activating input to schema in contention scheduling.

How does this model relate to consciousness? On the current approach Type-C processes have the following characteristics. (1) They involve the Supervisory System. (2) They lead directly to the selection in contention scheduling of a schema for thought or action, plus its arguments. This selection leads to action and/or to a qualitative change in the operation of lower-level special-purpose processing systems. By contrast, a non-conscious process is one which does not directly involve output from the Supervisory System and where its effects lead to only quantitative changes in the on-line processing systems. On this view, awareness of a particular content will involve either the triggering of a schema, or the modification of a pre-existing schema. However, once a schema is selected, and provided that schema does not conflict with a strongly established schema for action (as in cases requiring inhibition of pre-potent response), then action may proceed without any transfer of information from the SAS.

For some of the Type-C processes discussed previously, the relation to the SAS model is straightforward. An excellent example of the involvement of the Supervisory System is the Jacoby exclusion task (process 5). This case appears to involve the inhibition of a strong pre-potent response, namely the tendency to repeat the word previously presented. In this respect, the task is closely analogous to the Hayling B task (Burgess & Shallice, 1996b), originally used as a neuropsychological test of frontal function. However, the Jacoby task operates in the orthographic lexical domain while the Hayling B task requires the subject to give a completion to a sentence frame that makes no sense in the context of the frame. Tasks involving the inhibition of a strong pre-potent response do not appear to become automated with practice. Therefore, this sort of Type-C process is relatively easy to operationalize in an experimental setting (Jacoby, 1991).

Other Type-C processes are more complex. Some examples require a more precise specification of the relation between perceptual awareness and intentional action (as discussed in Section 3). Take the apparently direct perception of an external stimulus. The first case we consider illustrates the need to distinguish what the subject is 'aware of' – for instance, the presence of a stimulus, as opposed to its location or identity. Studies on blindsight show that awareness of the location of the stimulus is not necessary for accurate performance on a simple pointing task when subjects are asked to guess. Nonetheless, in this case, awareness of the presence of the stimulus, which has to be provided by an auditory cue, is necessary for the initiation of the pointing action (Weiskrantz, 1997). On the model the blindsight subject requires input via the SAS to initiate a pre-existing schema for pointing; however, once that schema is initiated, non-conscious information held in special-purpose processing systems can serve to guide the action.

A second case, the experiment by Jack (1998) involving the discrimination of masked letters (process 7 in the list of Type-C processes), illustrates the critical role of prior experience. In that experiment, subjects were initially unable to discriminate a heavily masked stimulus which was not initially specified to be a member of the response set. However, as soon as the stimulus had been presented under lighter masking conditions, and consciously seen, subjects were immediately able to discriminate the stimulus under heavy masking conditions. Awareness of the identity of the stimulus on one occasion allowed subjects to discriminate the stimulus without awareness thereafter. On the model modification of the response set requires top-down change from the SAS, which alters the arguments of the schema controlling discrimination performance. Once the schema and the response set are established then the operation of lower-level processing systems is sufficient for above-chance performance, even for perceptually degraded stimuli (for a related simulation where above-chance forced-choice performance occurs without explicit identification, see Hinton & Shallice, 1991).

Then there are anomalous cases in which an apparently intentional action is initiated in the absence of full awareness, for instance whilst we are in a distracted state or engaged in another task. Examples include reaching for and drinking from a glass whilst talking (Norman & Shallice, 1986), slips of highly routine actions which involve action lapses of the 'capture' error type (Norman, 1981; Reason, 1984), and changing gear or braking whilst driving (this issue is discussed in relation to awareness in Shallice, 1988b). It is not clear that we would wish to speak, in everyday language, of these actions as completely unconscious. Instead, they fit well with the phenomenological distinction between the foreground and background of consciousness (Shallice, 1988b). We explain these anomalous cases by distinguishing between the influence of the stimulus on contention scheduling, and the influence on the SAS. On the model, the selection of well learnt and relatively undemanding schema need not require SAS involvement. For instance, selection may be facilitated because the relevant schema are child-schemas of a larger parent-schema for action. According to this suggestion, the parent-schema for action (i.e. starting the car and beginning to drive) will be selected consciously via the SAS, but the parent-schema may itself include contingencies for the triggering of child-schemas (e.g. braking) without SAS intervention.

Since these anomalous cases do not require focal awareness, and are hypothesized not to involve the SAS, they cannot be classified as involving Type-C processes. However, they are obviously partially analogous to conscious processes, since they do involve schema selection. This suggestion would account for the apparent context sensitivity of anomalous cases (e.g. we don't instantly move our right foot to brake when riding a bicycle). If something like this analysis is correct, then whether or not an action can be classified as involving a Type-C process may depend on the larger task context. We would predict that the action of reaching for a glass and taking a drink would require awareness of the glass in some contexts, for instance if the subject is in an unfamiliar situation and has not already had a chance to look at the glass and mentally rehearse the action of drinking from it. When the action is familiar to the context, awareness may not be necessary to initiate the action. On our approach, objects or thoughts in the 'background of consciousness' would correspond to representations that are rapidly accessible for use by a Type-C process. This would include perceptual stimuli available to be selected by the parietal visuo-spatial attention system discussed by Driver and Vuilleumier.

What account can we offer of conscious reflection (process 1), which we have identified as the prototypically conscious process? For instance, consider the case in which a subject makes a fine discrimination of perceptual quality (e.g. taste or colour) and then responds via a simple two-choice key press. The details of the processes required to accomplish this remain obscure. Nonetheless, the framework offered here would appear to contrast with at least one aspect of that offered by Baars (1988). According to Baars, conscious perceptual information is made 'globally available' for the guidance of response. However, it is not clear that the 'broadcasting' of information encoded in sensory areas to the subsystems controlling movement would be necessary, or indeed possible. In our view, the representational codes would not be compatible. Rather, the critical conscious processes would appear to be as follows: (i) that of modulating the relevant perceptual subsystem, so that it can accomplish the computations necessary to make the appropriate comparison and return information on the result; (ii) the selection of the relevant motor subsystem to make a response contingent on the returned result; (iii) the mediation of the minimal information transfer required between the two subsystems (in this case, a single 'bit' of information).

We will consider just one more example from the list of Type-C processes presented earlier. This is the process underlying encoding of information into episodic memory (process 4). This is of central theoretical importance, for the following reason: the ability to remember a stimulus or thought is the principal criterion for the self-ascription that we were conscious of that stimulus or thought (Allport, 1988). Thus, in our view any putative theory that fails to account for the encoding of conscious information in episodic memory must be considered incomplete.

In our framework, we view episodic memory encoding as a process that results whenever Supervisory System modulation of lower-level processes occurs. This explains why a semantic orienting task is sufficient to give adequate memory encoding even when no instruction to remember is given (Hyde & Jenkins, 1969). It also fits with the computationally based claim that episodic memory encoding processes occur when novel operations are being carried out but not when routine processing is occurring (Sussman, 1975). Thus, episodic encoding may be seen as a by-product of the operation of any Type-C process. It is frequently the case that the only observable consequence of the operation of Type-C processes is the ability of the subject to later recall information about their perceptions and thoughts. How can this fit with our emphasis on understanding the function of consciousness? Schank (1982) argues that a key evolutionary function of episodic memory is that of reminding the subject of relevant autobiographical episodes in order to provide relevant material for strategy development in non-routine situations. On this perspective the subsystems involved in controlling episodic memory retrieval should also be seen as a part of the Supervisory System as their overall function is to assist in coping with non-routine situations.

The aim of this discussion is to indicate the utility of the Shallice (1988a) model as a framework for the description of conscious or Type-C processes, and initial theorizing about those processes. In this discussion, we have attempted to illustrate that the framework can both accommodate various aspects of phenomenology, and coheres with empirical evidence and theorizing in cognitive psychology. Nonetheless, we stress the point that the model only provides a conceptual framework capable of characterizing consciousness in broad information-processing terms. This framework may help to identify some of the computations involved, yet it is a long way from an account of consciousness embedded in neurally plausible computational models of the precise information-processing operations involved. Section 7 aims to illustrate how we may get closer to this goal.

7. Localization of function: specifying conscious operations

In Section 6 some of the Type-C processes listed – and in particular the operationally more critical experimental Type-C processes (Section 4.2) – relate to individual tasks and therefore come from a very large, if not infinite, set. Can one produce a more basic set of such processes? Secondly, in later versions of the Supervisory System model, the Supervisory System is held to contain a variety of special-purpose subsystems localized in different parts of prefrontal cortex (Shallice & Burgess, 1996). Should the relation between a Type-C process and effective operation of the 'Supervisory System' not then be capable of being specified further?

Work on localization of function in prefrontal cortex can potentially allow us to specify a more basic set of Type-C processes relating to different Supervisory System operations localized in different parts of cortex. There is evidence that functions carried out in prefrontal cortex are compatible with our general view of conscious processing. We claim that supervisory operations are not informationally encapsulated, and thus are not specific to particular modalities of input. This is consistent with one position in a recent debate concerning the lack of material-specificity in operations carried out in regions of left dorsolateral prefrontal cortex (Owen, 1997). We have also characterized conscious processing as an attentive operation (see Section 2). In our view, different conscious processes operating in the same short interval of time must have the same effective input. To judge from psychological refractory period and attentional blink phenomena the short interval of time is of the order of several hundred milliseconds. This view is consistent with a suggestion of Moscovitch (pers. commun.) concerning processes located in prefrontal cortex. He argued that these processes would not be able to be carried out at the same time as a structurally unrelated but demanding task. Support for this position is provided by the functional imaging dual-task study of Shallice et al. (1994), where a demanding but structurally unrelated additional task led to a significant reduction in activation in left dorsolateral prefrontal cortex. On this view, a demanding additional task should lead to the primary task being carried by means of lower-level processing systems alone. It follows from the model that there should not be full awareness of the relevant stimuli, as discussed in Section 6.

It is known experimentally that reading aloud and repeating or writing of a continuous sequence of words can be carried out in parallel to other tasks (Allport, Antonis, & Reynolds, 1972; Shallice, McLeod, & Lewis, 1985). That this is possible fits with work showing that naming, say in word reading or repeating, can be modelled in terms of feed-forward networks (e.g. Plaut, McClelland, Seidenberg, & Patterson, 1996) – suggesting that it can be carried out by lower-level processes alone. More critically, and as the model predicts, in these situations subjects lack full consciousness of the task and stimuli (Spelke, Hirst, & Neisser, 1976). Full awareness of a word would still be necessary for repeating or reading aloud when words are presented alone (i.e. cases involving 'spontaneous recognition', process 2 in the list of Type-C processes). This situation is known to be different, since processing is not properly automated for the first word in a rapidly presented sequence (see Allport & Wylie, 2000; Treisman & Davies, 1973).

What of the specific information-processing operations we have discussed in Section 6? A recent review by Frith (2000) provides excellent evidence that the key operation of modulation of lower-level schemas by the Supervisory System can be localized. Frith (2000) reviews a number of functional imaging tasks, where he argues that "sculpting the response space" is the key process that distinguished experimental and control conditions. The experiments considered involve carrying out a willed action compared with a choice response (Frith, 1992), the generation of a response when there are no strong pre-existing pre-potent responses compared with when such strong tendencies exist, e.g. the Nathaniel-James, Fletcher, and Frith (1997) study using the Hayling task (Burgess & Shallice, 1996b), and random number generation (Jahanshahi et al., see Frith, 2000) which involves the avoidance of responding using routine sets (Baddeley, Emslie, Kolodny, & Duncan, 1998; Jahanshahi & Dirnberger, 1999). The tasks Frith reviews all activate a region of left dorsolateral prefrontal cortex involving the middle and inferior frontal gyri. Further, imaging studies of memory encoding may be interpreted as supporting our claim that episodic encoding results from the operation of this process. Studies of encoding, for instance requiring the active organization of material, activate the same swathe of cortex identified by Frith (Fletcher, Shallice, & Dolan, 1998; Shallice et al., 1994). It has also been shown that carrying out novel operations – requiring schema generation on the model – activates left dorsolateral prefrontal cortex (Dolan & Fletcher, 1997; Tulving, Markowitsch, Kapur, Habib, & Houle, 1994). Now that a plausible anatomical location has been found for this critical executive function, further investigations may serve to give a more precise picture of the information-processing operations involved.

In addition to refining the model as discussed above, evidence from localization of function can extend it. We give two examples. Based on functional imaging evidence, it has been argued that the anterior cingulate cortex plays an essential role in conscious processing (Posner & Rothbart, 1998). The anterior cingulate is a structure activated in many task comparisons but is especially likely to be more activated in more difficult task situations (Paus, Koski, Caramanos, & Westbury, 1998). There is not yet complete agreement in its function. However, it would appear to be the more activated the more concentration is required (Posner & Petersen, 1990) and a meta-analysis has shown that the anterior cingulate tends to be highly active when there is conflict between competing inputs and/or responses (Carter, Botvinick, & Cohen, 1999). Indeed Posner and DiGirolamo (1998) have argued that any Supervisory System operation necessarily involves activation of the anterior cingulate. One possibility which fits with evidence on how its activation is affected by dopaminergic agonists in schizophrenics (Fletcher, Frith, Grasby, Friston, & Dolan, 1996) is that it is involved in top-down supervisory modulation of which processing systems are to be involved in on-line processing. The anterior cingulate therefore appears to complement the left dorsolateral region – involved in the top-down control of content as discussed earlier. On this view, the prefrontal cortex and the anterior cingulate would have complementary roles in conscious processing. A distinction might be made between different aspects of conscious processing with the concentration/mental effort aspects having a separate but linked material basis from those related to conscious content.

Second, imaging investigations inform our view of episodic memory retrieval (listed as pre-experimental process 3 in Section 4.1, but only mentioned in passing in the previous section). It is now well known that memory retrieval tasks activate predominantly right prefrontal cortex (Shallice et al., 1994; Tulving, Kapur, Craik, Moscovitch, & Houle, 1994; but see also Nolde, Johnson, & Raye, 1998). However, a number of different processes and regions appear to be involved (Lepage, Ghaffar, Nyberg, & Tulving, 2000). One process located in right dorsolateral prefrontal cortex appears to be linked to checking the retrieved memory (Henson et al., 1999). Checking retrieved memories may be a special case of a more general process responsible for the monitoring of on-going cognitive operations (Fink et al., 1999; Stuss & Alexander, 1994). Monitoring requires the matching of an overt or covert action with pre-specified criteria. If there is a match then there is no interruption of on-going behaviour. However, if a mismatch occurs a process of correction or more systematic checking takes place. This will involve top-down modulation of schema, somewhat analogous to that occurring with inhibition of a pre-potent response. Since the smooth operation of on-line processing systems is only interrupted in this case, it would follow from the theory given here that there is consciousness of a mismatch but not of a match.

In this section we argued that Type-C processes have two general characteristics – they are not informationally encapsulated and they are resource demanding. We have suggested three basic types of Type-C processes: (i) top-down schema modulation, also giving rise to episodic encoding; (ii) retrieval from episodic memory; and (iii) interruption of on-going operations through mismatch detection. Intention realization (Burgess, Quale, & Frith, 2000; Shallice & Burgess, 1991) would be a further possibility. We believe that cognitive neuroscience now has the potential to extend this list further.

8. *Rene

It is now widely accepted in cognitive science that the cognitive subsystems which are concerned with operating on knowledge about ourselves and other minds differ at least in part from the cognitive subsystems concerned with knowledge about physical mechanisms and causation (e.g. Baron-Cohen, Leslie, & Frith, 1985; Brothers, 1995; Frith & Frith, 1999). Consider an artificial intelligence, we shall call '*Rene', whose categories for understanding itself are completely unrelated to those for understanding its external material world, with the same applying to the abstractions it has developed from those categories. By assumption, *Rene's artificial mental states are just functional states. Yet, *Rene would not be able to use its physical or mechanical concepts to categorize its own artificial mental states. Without being told, or conducting its own investigations, *Rene would therefore have no way of knowing what functional state it is in at any particular moment in time.

Nonetheless, if *Rene's self-reflective capacities are to be useful to it – for instance, it could know that a particular type of pain would grow less with time – then its subjective concepts should map, at least broadly, onto functional distinctions between its cognitive states. Thus, another system could use *Rene's introspective reports as a guide to *Rene's functional organization, as well as providing data on the operation of *Rene's self-reflective cognitive subsystems. Similarly, at least some of the subjective concepts we use to differentiate between mental states promise to map directly onto information-processing distinctions between those states (e.g. aware or unaware, intentional or automatic, dream sleep or dreamless sleep). Furthermore, unlike *Rene, the different conceptual systems that humans use to describe themselves are not forced to remain distinct above the level of basic categories. The methods of cognitive science allow us to identify and distinguish between the different functional states involved in different cognitive activities. Through experiencing 'what it is like to' do well specified tasks, we may learn to relate our subjective understanding of our own mental states to such objective specifications of those states.

Would *Rene believe in dualism? There is no determinate answer. Presumably, *Rene could imagine a highly complex mechanism capable of producing the same behaviours as itself. Consequently, *Rene might entertain the possibility that its mental states are just physical states. Yet, crucially, *Rene's understanding of this equivalence could only be highly abstract. *Rene can't simply collapse and simplify his two conceptual systems into a unified whole. Thus, *Rene's understanding of the world (including itself) would remain equally complex, regardless of whether it believed in this equivalence or not. Indeed, in order for *Rene to begin imagining the highly complex mechanism *as experiencing the same mental states as itself*, *Rene would have to make a complete shift of mental set, bringing a wholly new set of concepts into play. *Rene's new train of thought would be so disjointed from the last, it might seem to *Rene that it was thinking about a completely different sort of thing.

*Rene would find no contradiction in imagining an entity that does all the same information processing as itself, yet which lacks its mental states. There would only be a contradiction if there were overlap in the criteria *Rene uses for applying mental and physical concepts. This may explain why we can imagine (in the abstract) an entity that does all the same information processing as us, yet lacks experiential states – the philosopher's zombie (see Dennett). There is no reason to suppose that there could be any actual difference between an entity doing all the same processing as us and a 'conscious' being (as supposed, for example, by Block, 1978; Chalmers, 1996; Kripke, 1972; Searle, 1992). The only difference is the 'attitude', 'stance', or 'mental set' we adopt when we are encouraged to think in different ways about the same thing (Dennett, 1987, 1991, 1996; Papineau, 1998).[13]

9. Summary and conclusion

In this paper we have placed our emphasis on the development of a scientific program for studying consciousness, rather than on a particular account of the neural or computational processes involved. This reflects our belief that the science of consciousness remains in its infancy, and that substantial progress will require a clarification of the deep conceptual and methodological difficulties that surround scientific attempts to understand human experience. In our view, most scientific proposals to date have attempted to bridge the gap between the physical and the experiential too quickly. In his discussion of biological psychiatry and its attempts to account for subjective phenomena (i.e. psychotic symptoms such as hallucinations), Frith (1992, see discussion pp. 25–30) notes that "The history of biological psychiatry is full of 'elephant footprints in the mud' (Lancet, 1978); findings which have made a big impact at the time, but have then faded away." As Frith argues, this has occurred precisely because of a failure to provide an adequate theoretical framework linking physical phenomena to mental phenomena, causing researchers to over-interpret "spurious and irrelevant associations".

In contrast, our approach is to present a theoretical framework to guide further investigation. The principal goal of this approach is the elucidation of the function of consciousness – the question of how conscious information, as opposed to non-conscious information, influences thought and behaviour, and in particular its role in the production of introspective reports. Our strategy for explaining the function of consciousness consists of two distinct components. First, in Section 4 we outline a method for the identification of tasks that provide a handle on relevant psychological

[13] Whilst we credit Dennett with this insight, he does not adopt the same position on philosophical zombies. He claims that we can't properly conceive of zombies, because he doesn't believe that there is a separation between the categories used – Dennett believes that mental state terms only refer to functional concepts or 'dispositions to behave'. We dispute this point (see also Loar, 1996). Ultimately, the issue should be resolved empirically. The view of Papineau (1998) is consistent with ours.

phenomena. This component of the strategy is geared to the production of empirical phenomena that are both suitable for further investigation and in need of explanation, so providing basic data for a theory of consciousness. As the second component, we adopted a particular theoretical framework and used it to understand conscious and non-conscious processes (see Section 6). The basic theoretical elements of this framework, originally put forward to describe the executive functions of prefrontal cortex, are not precisely specified in neural or computational terms. However, they allow an initial grasp of the relevant psychological phenomena using concepts that also link to information processing and/or systems neuroscience descriptions of brain function (see Section 7). Thus, this framework provides a broad structural outline for putative theories of consciousness, and serves to guide experimental work.

More generally, the presentation (and, to a much greater extent, the generation) of this approach has required us to consider some fundamental conceptual and methodological issues relating to consciousness. The term 'Introspective Physicalism' reflects the conclusions we have reached in three ways. Our first step is to adopt, and defend, a form of physicalism[14] (see Section 8). The goal of a theory of consciousness cannot be to tell us 'what it is like to be' in a mental state (as supposed by Jackson, 1995; McGinn, 1989; Nagel, 1974). Nor should we naively suppose that every subjective concept, however 'self-evident', accurately describes some aspect of the mind. Subjective concepts can only be acquired through consideration of our own experience (Lewis, 1990). Inevitably, some of these concepts will 'carve nature at its joints', whilst others will simply serve to mislead. Misleading concepts will not map onto functional distinctions between mental states. However, science may still study them from a distance by investigating the self-reflective processes that give rise to them. The closest that science can come to accounting for subjectivity is through elucidating the mechanisms that allow us to understand ourselves from our own point of view. Thus, our second step is to argue that a theory of consciousness must account for the processes underlying introspection.

Our third step is to emphasize the role of introspective evidence in the formulation of scientific accounts. As physicalists, we reject meta-physical dualism. Yet, we support methodological dualism, and attempt to address the specific methodological issues that arise concerning the use of introspective evidence. Although frequently overlooked, the history of psychology provides important lessons about the subtle complexities and difficulties associated with introspective evidence. Ultimately, it should be possible to account for all 'phenomenal appearances'. However, in so far as introspective observations are taken to reflect properties of the mental states under consideration, it is not yet clear which observations will ultimately be considered veridical, which will need to be re-described in order to cohere with a scientific understanding of the mind, and which will be explained as outright illusions. Thus,

[14] More specifically we are 'token' physicalists and 'type' functionalists (see Davidson, 1980). Every instance, or instantiation, of a mental state is identical with a physical state (physicalism). However, the type of mental state is determined solely by the causal relationships that the token mental state has with other mental states (functionalism).

we argue that obtaining valid introspective evidence is a complex craft. In our discussion, we have stressed the need to take a sceptical approach to observations based solely on introspective evidence, pending the collection of objective evidence that can validate the interpretation placed on that evidence. Nonetheless, introspective evidence can and should play both a major and an explicit role in the development of information-processing theories. Introspective evidence is an essential component of our research proposal for the identification of the processes necessary and sufficient for awareness, 'Type-C' processes. It also informs the theoretical framework we propose for understanding those processes.

From a philosophical perspective, this view of the use of introspective evidence in cognitive psychology relies on an inversion of the argument of Nagel (1974). Nagel argues that it is our knowledge of 'what it is like to be' in certain mental states that presents a barrier to the science of the mind. The argument here is the converse: it is precisely because we know what it is like to be in certain mental states that we are able to bring this evidence to bear on functional theories in general, and on theories of consciousness in particular. Scientific theories that are informed by introspective evidence in this way can justifiably claim to provide an account that links the mental and the physical.

Acknowledgements

The preparation of this paper was supported by a research grant from the Welcome Trust (053288/Z/98/Z/JRS/JP/JAT). We would like to thank Ned Block, David Papineau, Stanislas Dehaene, Michael Martin, Patrick Haggard, Vinod Goel and two anonymous reviewers for comments on previous versions of this work.

References

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

Allport, A., & Wylie, G. (2000). Selection-for-action in competing (Stroop) tasks: 'task-switching', stimulus-response bindings, and negative priming. In S. Monsell, & J. S. Driver (Eds.), *Control of cognitive processes: attention and performance*, Vol. XVIII. Cambridge, MA: MIT Press.

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: a disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, *24* (2), 225–235.

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

Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology A*, *51* (4), 819–852.

Badgaiyan, R. D. (2000). Executive control, willed actions, and nonconscious processing. *Human Brain Mapping*, *9* (1), 38–41.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21* (1), 37–46.

Bisiach, E. (1988). The (haunted) brain and consciousness. In A. J. Marcel, & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 101–120). Oxford: Clarendon Press/Oxford University Press.

Block, N. (1978). Troubles with functionalism. In C. W. Savage (Ed.), *Perception and cognition: issues in the foundation of psychology*. Minneapolis, MN: University of Minnesota Press.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*, 227–287.

Bornstein, R. F., & D'Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology*, *63* (4), 545–552.

Bornstein, R. F., & D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Social Cognition*, *12* (2), 103–128.

Bowers, K. S. (1984). On being unconsciously influenced and informed. In K. S. Bowers, & D. Meichenbaum (Eds.), *The unconscious reconsidered* (pp. 227–272). New York: Wiley.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.

Brothers, L. (1995). Neurophysiology of the perception of intentions by primates. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1107–1115). Cambridge, MA: MIT Press.

Burgess, P. W., Quale, A., & Frith, C. D. (2000). Brain regions involved in prospective memory according to positron emission tomography. Manuscript submitted for publication.

Burgess, P. W., & Shallice, T. (1996a). Confabulation and the control of recollection. *Memory*, *4* (4), 359–411.

Burgess, P. W., & Shallice, T. (1996b). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, *34* (4), 263–272.

Campion, J., Latto, R., & Smith, Y. M. (1983). Is blindsight an effect of scattered light, spared cortex, and near-threshold vision? *Behavioural and Brain Sciences*, *6*, 423–448.

Carter, C. S., Botvinick, M. M., & Cohen, J. D. (1999). The contribution of the anterior cingulate cortex to executive processes in cognition. *Annual Review of Neuroscience*, *10* (1), 49–57.

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

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology*, *40* (4), 343–367.

Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*, 297–338.

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

Crick, F. H. C. (1994). *The astonishing hypothesis: the scientific search for the soul*. New York: Charles Scribner's Sons.

Crick, F., & Koch, C. (1990). Some reflections on visual awareness. *Cold Spring Harbor Symposia on Quantitative Biology*, *55*, 953–962.

Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, *375* (6527), 121–123.

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

Cytowic, R. E. (1997). Synaesthesia: phenomenology and neuropsychology. In S. Baron-Cohen, & J. E. Harrison (Eds.), *Synaesthesia* (pp. 17–42). Oxford: Blackwell.

Dab, S., Claes, T., Morais, J., & Shallice, T. (1999). Confabulation with a selective descriptor process impairment. *Cognitive Neuropsychology*, *16*, 215–242.

Davidson, D. (1980). *Essays on actions and events*. Oxford: Oxford University Press.

Dehaene, S., & Changeux, J. P. (1997). A hierarchical neuronal network for planning behavior. *Proceedings of the National Academy of Sciences USA*, *94* (24), 13293–13298.

Dehaene, S., Naccache, L., Le Clec, H. G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395* (6702), 597–600.

Della Malva, C. L., Stuss, D. T., D'Alton, J., & Willmer, J. (1993). Capture errors and sequencing after frontal brain lesions. *Neuropsychologia*, *31*, 363–372.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press/A Bradford Book.

Dennett, D. C. (1988). Quining qualia. In A. J. Marcel, & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 42–77). Oxford: Clarendon Press/Oxford University Press.

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

Dennett, D. C. (1996). Facing backwards on the problem of consciousness. *Journal of Consciousness Studies*, *3* (1), 4–6.

Dixon, N. F. (1971). *Subliminal perception: the nature of a controversy*. London: McGraw-Hill.

Dixon, N. F. (1981). *Preconscious processing*. Chichester: Wiley.

Dolan, R. J., & Fletcher, P. C. (1997). Dissociating prefrontal and hippocampal function in episodic memory encoding. *Nature*, *388* (6642), 582–585.

Driver, J., & Mattingley, J. B. (1998). Parietal neglect and visual awareness. *Nature Neuroscience*, *1* (1), 17–22.

Ericksen, C. W. (1960). Discrimination and learning without awareness: a methodological survey and evaluation. *Psychological Review*, *67*, 279–300.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: verbal reports as data (Rev. ed.)*. Cambridge, MA: MIT Press.

Fehrer, E., & Biederman, I. (1962). A comparison of reaction and verbal report in the detection of masked stimuli. *Journal of Experimental Psychology*, *64*, 126–130.

Ffytche, D. H., & Howard, R. J. (1999). The perceptual consequences of visual loss: 'positive' pathologies of vision. *Brain*, *122* (Pt. 7), 1247–1260.

Fink, G. R., Marshall, J. C., Halligan, P. W., Frith, C. D., Driver, J., Frackowiak, R. S., & Dolan, R. J. (1999). The neural consequences of conflict between intention and the senses. *Brain*, *122* (Pt. 3), 497–512.

Fletcher, P. C., Frith, C. D., Grasby, P. M., Friston, K. J., & Dolan, R. J. (1996). Local and distributed effects of apomorphine on fronto-temporal function in acute unmedicated schizophrenia. *Journal of Neuroscience*, *16* (21), 7055–7062.

Fletcher, P. C., Shallice, T., & Dolan, R. J. (1998). The functional roles of prefrontal cortex in episodic memory. I. Encoding. *Brain*, *121*, 1239–1248.

Fletcher, P. C., Shallice, T., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1998). The functional roles of prefrontal cortex in episodic memory. II. Retrieval. *Brain*, *121*, 1249–1256.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove: Lawrence Erlbaum Associates.

Frith, C. D. (2000). The role of dorsolateral prefrontal cortex in the selection of action, as revealed by functional imaging. In S. Monsell, & J. Driver (Eds.), *Control of cognitive processes: attention and performance*, Vol. XVIII. Cambridge, MA: MIT Press.

Frith, C. D., & Frith, U. (1999). Interacting minds – a biological basis. *Science*, *286* (5445), 1692–1695.

Frith, C., Perry, R., & Lumer, E. (1999). The neural correlates of conscious experience: an experimental framework. *Trends in Cognitive Sciences*, *3* (3), 105–114.

Fuhrer, M. J., & Ericksen, C. W. (1960). The unconscious perception of the meaning of verbal stimuli. *Journal of Abnormal and Social Psychology*, *62*, 432–439.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition*, *7*, 1–26.

Gazzaniga, M. S. (1985). *The social brain*. New York: Basic Books.

Goel, V., Grafman, J., Tajik, J., Gana, S., & Danto, D. (1997). A study of the performance of patients with frontal lobe lesions in a financial planning task. *Brain*, *120* (Pt. 10), 1805–1822.

Hameroff, S., & Penrose, R. (1996). Conscious events as orchestrated space-time selections. *Journal of Consciousness Studies*, *3* (1), 36–53.

Henson, R. N., Shallice, T., & Dolan, R. J. (1999). Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. *Brain*, *122* (Pt. 7), 1367–1381.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: investigations of acquired dyslexia. *Psychological Review*, *98* (1), 74–95.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: a survey and appraisal. *Behavioral and Brain Sciences*, *9* (1), 1–66.

Humphrey, G. (1951). *Thinking: an introduction to its experimental psychology*. London: Methuen.

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, *82*, 472–481.

Jack, A. I. (1998). *Perceptual awareness in visual masking*. Unpublished doctoral dissertation, University College London.

Jackson, F. (1995). What Mary didn't know. In P. K. Moser, & J. D. Trout (Eds.), *Contemporary materialism*. London: Routledge.

Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30* (5), 513–541.

Jacoby, L. L. (1998). Invariance in automatic influences of memory: toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24* (1), 3–26.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: false recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, *118* (2), 126–135.

Jahanshahi, M., & Dirnberger, G. (1999). The left dorsolateral prefrontal cortex and random generation of responses: studies with transcranial magnetic stimulation. *Neuropsychologia*, *37* (2), 181–190.

James, W. (1890). *The principles of psychology*. New York: Holt.

Kahneman, D. (1967). An onset-onset law for one case of apparent motion and metacontrast. *Perception and Psychophysics*, *2*, 577–584.

Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin*, *70* (6), 404–425.

Kluver, H. (1966). *Mescal and mechanisms of hallucinations*. Chicago, IL: University of Chicago Press.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103* (3), 490–517.

Kripke, S. A. (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: University of Chicago Press.

Laming, D. R. J. (1997). *The measurement of sensation*, Oxford: Oxford University Press.

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, *379*, 549–553.

Lepage, M., Ghaffar, O., Nyberg, L., & Tulving, E. (2000). Prefrontal cortex and episodic memory retrieval mode. *Proceedings of the National Academy of Sciences USA*, *97* (1), 506–511.

Lewis, D. (1990). What experience teaches. In W. C. Lycan (Ed.), *Mind and cognition: a reader* (pp. 499–519). Oxford: Blackwell.

Loar, B. (1996). Phenomenal states. In N. Block, O. Flanagan, & G. Guzeldere (Eds.), *The nature of consciousness*. Cambridge, MA: MIT Press.

Logothetis, N., & Schall, J. (1989). Neuronal correlates of subjective visual perception. *Science*, *245*, 761–763.

Mandler, G. (1975). *Mind and emotion*. New York: Wiley.

Mandler, G. (1980). Recognising: the judgment of previous occurrence. *Psychological Review*, *87*, 252–271.

Mandler, G., Nakamura, Y., & Van Zandt, B. J. (1987). Nonspecific effects of exposure on stimuli that cannot be recognized. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13* (4), 646–648.

Marcel, A. J. (1983). Conscious and unconscious perception: experiments on visual masking and word recognition. *Cognitive Psychology*, *15* (2), 197–237.

McGinn, C. (1989). Can we solve the mind-body problem? *Mind*, *98*, 349–366.

Merikle, P. M. (1984). Toward a definition of awareness. *Bulletin of the Psychonomic Society*, *22* (5), 449–450.

Merikle, P. M. (1992). Perception without awareness: critical issues. *American Psychologist*, *47* (6), 792–795.

Merikle, P. M., Joordens, S., & Stolz, J. A. (1995). Measuring the relative magnitude of unconscious influences. *Consciousness and Cognition: an International Journal*, *4* (4), 422–439.

Merikle, P. M., & Reingold, E. M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17* (2), 224–233.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63* (2), 81–97.

Miller, J. (1987). Priming is not necessary for selective-attention failures: semantic effects of unattended, unprimed letters. *Perception and Psychophysics*, *41* (5), 419–434.

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

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*, 435–450.
Nathaniel-James, D. A., Fletcher, P., & Frith, C. D. (1997). The functional anatomy of verbal initiation and suppression using the Hayling test. *Neuropsychologia*, *35* (4), 559–566.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, *75*, 522–536.
Nolde, S. F., Johnson, M. K., & Raye, C. L. (1998). The role of prefrontal cortex during tests of episodic memory. *Trends in Cognitive Sciences*, *2* (10), 399–406.
Norman, D. A. (1981). Categorisation of action slips. *Psychological Review*, *88*, 1–15.
Norman, D. A., & Bobrow, D. G. (1976). On the role of active memory processes in perception and cognition. In C. N. Cofer (Ed.), *The structure of human memory* (pp. 114–132). San Francisco, CA: Freeman.
Norman, D. A., & Shallice, T. (1986). Attention to action: willed and automatic control of behavior (Rev. ed.). In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4). New York: Plenum Press.
Owen, A. M. (1997). The functional organization of working memory processes within human lateral frontal cortex: the contribution of functional neuroimaging. *European Journal of Neuroscience*, *9* (7), 1329–1339.
Papineau, D. (1998). Mind the gap. In J. Tomberlin (Ed.), *Language, mind and ontology. Philosophical perspectives* (Vol. 12, pp. 373–388). Oxford: Blackwell.
Paus, T., Koski, L., Caramanos, Z., & Westbury, C. (1998). Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: a review of 107 PET activation studies. *NeuroReport*, *9* (9), 37–47.
Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.
Posner, M. I., & DiGirolamo, G. J. (1998). Executive attention: conflict, target detection, and cognitive control. In R. Parasuraman (Ed.), *The attentive brain*. Cambridge, MA: MIT Press.
Posner, M. I., & Klein, R. M. (1973). On the functions of consciousness. In S. Kornblum (Ed.), *Attention and performance* (Vol. IV, pp. 21–35). New York: Academic Press.
Posner, M. I., & Petersen, S. E. (1990). The attentional system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.
Posner, M. I., & Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *353* (1377), 1915–1927.
Reason, J. T. (1984). Lapses of attention. In R. Parasuraman, R. Davies, & J. Beatty (Eds.), *Varieties of attention*. Orlando, FL: Academic Press.
Reber, A. S. (1997). Implicit ruminations. *Psychonomic Bulletin and Review*, *4* (1), 49–55.
Reingold, E. M., & Merikle, P. M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind and Language*, *5*, 9–28.
Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, *94* (3), 329–359.
Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
Schacter, D. L. (1989). On the relation between memory and consciousness: dissociable interactions and conscious experience. In H. L. Roediger, & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: essays in honour of Endel Tulving* (pp. 355–389). Hillsdale, NJ: Erlbaum.
Schacter, D. L., Curran, T., Galluccio, L., Milberg, W. P., & Bates, J. F. (1996). False recognition and the right frontal lobe: a case study. *Neuropsychologia*, *34* (8), 793–808.
Schank, R. C. (1982). *Dynamic memory*. Cambridge: Cambridge University Press.
Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
Shallice, T. (1972). Dual functions of consciousness. *Psychological Review*, *79*, 383–393.
Shallice, T. (1988a). *From neuropsychology to mental structure*. New York: Cambridge University Press.
Shallice, T. (1988b). Information-processing models of consciousness: possibilities and problems. In A. J. Marcel, & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 305–333). Oxford: Clarendon Press/Oxford University Press.
Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114* (Pt 2), 727–741.

Shallice, T., & Burgess, P. (1996). The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society of London, Series B*, *351* (1346), 1405–1412.

Shallice, T., Fletcher, P., Frith, C. D., Grasby, P., Frackowiak, R. S., & Dolan, R. J. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature*, *368* (6472), 633–635.

Shallice, T., McLeod, P., & Lewis, K. (1985). Isolating cognitive modules with the dual-task paradigm: are speech perception and production separate processes? *Quarterly Journal of Experimental Psychology A*, *37* (4), 507–532.

Shanks, D. R., & Johnstone, T. (1997). Implicit knowledge in sequential learning tasks. In M. A. Stadler, & P. A. Frensch (Eds.), *Handbook of implicit learning*. London: Sage.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17* (3), 367–447.

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organisation. *Proceedings of the National Academy of Sciences USA*, *94*, 3408–3413.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: a microgenetic analysis. *Journal of Experimental Psychology: General*, *127* (4), 377–397.

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition*, *4* (3), 215–230.

St. John, M. F., & Shanks, D. R. (1997). Implicit learning from an information processing standpoint. In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 162–194). Oxford: Oxford University Press.

Stuss, D. T., & Alexander, M. P. (1994). Functional and anatomical specificity of frontal lobe functions. In L. S. Cermak (Ed.), *Neuropsychological explorations of memory and cognition: essays in honor of Nelson Butters* (pp. 191–200). New York: Plenum Press.

Sussman, G. J. (1975). *A computational model of skill acquisition*. New York: American Elsevier.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science*, *282* (5395), 1846–1851.

Treisman, A. M., & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), *Attention and performance* (Vol. IV). London: Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press.

Tulving, E., Kapur, S., Craik, F. I., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings. *Proceedings of the National Academy of Sciences USA*, *91* (6), 2016–2020.

Tulving, E., Markowitsch, H. J., Kapur, S., Habib, R., & Houle, S. (1994). Novelty encoding networks in the human brain: positron emission tomography data. *NeuroReport*, *5* (18), 2525–2528.

Van Gulick, R. (1994). Deficit studies and the function of phenomenal consciousness. In G. Graham, & L. Stephens (Eds.), *Philosophical psychology* (pp. 25–50). Cambridge, MA: MIT Press.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*, 158–177.

Watson, J. B. (1920). Is thinking merely the action of language mechanisms? *British Journal of Psychology*, *11*, 87–104.

Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, *171* (4356), 737–738.

Weiskrantz, L. (1986). *Blindsight*. New York: Oxford University Press.

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

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19* (6), 1235–1253.

Whittlesea, B. W., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, *29* (6), 716–732.

Wilkes, K. V. (1988). —, yishi, duh, um, and consciousness. In A. J. Marcel, & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 16–41). Oxford: Clarendon Press/Oxford University Press.