Correlating Consciousness: A View from Empirical Science

Axel Cleeremans

Séminaire de Recherche en Sciences Cognitives
Université Libre de Bruxelles

John-Dylan Haynes

Institut für Psychologie und Kognitionforschung
Universität Bremen

Address correspondence to:

Axel Cleeremans
Séminaire de Recherche en Sciences Cognitives
Université Libre de Bruxelles
50 ave. F.-D. Roosevelt CP 122
1050 Bruxelles
BELGIUM

Email: axcleer@ulb.ac.be

Abstract

Research on consciousness is currently enjoying a spectacular revival of interest in the cognitive sciences. From an empirical point of view, the NCC program — the search for the “Neural Correlates of Consciousness” — holds the promise of establishing correlations between physiological and phenomenal states in a way that directly resembles G. T. Fechner´s (1860) so-called “inner psychophysics”. Should the NCC program be entirely successful, we would thus be able to predict phenomenal states based on physiological states. we would be able to predict phenomenal states based on physiological states. In this paper, we explore some of the conceptual and methodological difficulties of this approach. In both neurobiology and psychology, there are serious measurement problems that stand in the way of correlation research, even after the “hard problem” has been set aside. Thus, even if one had identified certain internal functional states as indicators of phenomenal states, the empirical psychologist would still be confronted with fundamental problems, such as determining the absence or presence of these functional states. In this respect, philosophy of science may help and provide a metatheoretical framework for the current interdisciplinary project.

1. Introduction

As if awakening from a bad dream, and newly enlightened by recent progress in brain-imaging[1] techniques, cognitive psychology is again focusing its attention on what has been described as one of the last remaining mysteries: Consciousness. The current interest, however, is by no means new. Indeed, research on consciousness already occupied a central spot during the second half of the nineteenth century, when psychology was gradually emerging from philosophy and physiology as an independent discipline. For example, the main motivation behind G. T. Fechner's (1860) foundation of psychophysics was his interest in the body-mind relationship.

Interestingly, Fechner divided his "exact science of the functional or dependency relationship between body and mind" (translated from Fechner 1860, p. 8) into two parts (see Fig. 1): While the so-called "outer psychophysics" should investigate the relationship between physical stimuli and subjective experience, the "inner psychophysics" was meant to attack the body-mind relationship directly. Fechner's main interest was inner psychophysics, but the limited methodology of his days restricted him to the indirect investigation of that relationship via outer psychophysics. His basic assumption concerning the internal relationship was one of covariance between physiological and phenomenal processes[2]. This was elaborated further in the first three of Müller's famous psychophysical axioms (Müller 1896, pp. 1-2):

> 1. Every state of consciousness is based upon a material process, a so-called psychophysical process, which is a prerequisite of the occurrence of this state of consciousness. (...)

> 2. To an identity, similarity, difference in the constitution of sensations (...) corresponds an identity, similarity, difference of the structure of the psychophysical processes, and vice versa. (...)
>
> 3. If the changes through which a sensation passes have the same direction, or if the differences which exist between series of sensations are of the same direction, then the changes through which the sensation passes, or the differences of the given psychophysical process have the same direction. (...)

Most other prominent early psycho-physiologists, such as Hermann Helmholtz (1861), Wilhelm Wundt (1874) and William James (1890) were similarly interested in the problem of consciousness. This interest quickly waned, however, in the face of immense methodological controversies (Watson, 1913). With the rise of behaviorism, the entire enterprise was soon cast as a "scientifically incorrect" research topic[3].

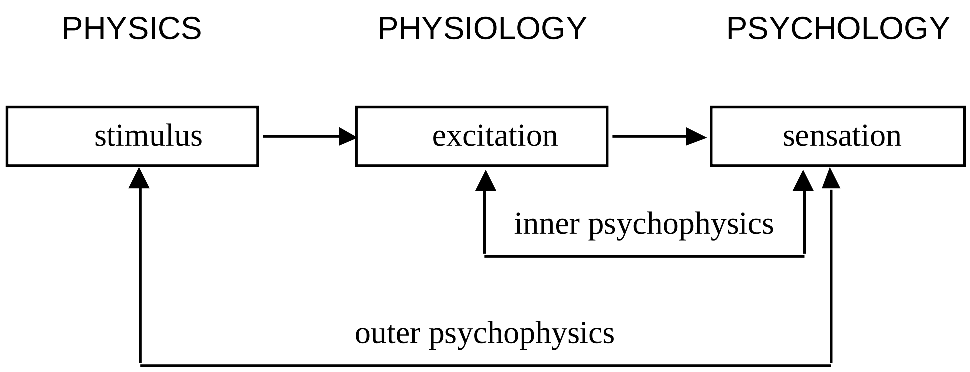

Figure 1: Fechner's (1860) "exact science" of the body-mind relationship. Research methodology in his days was restricted to relationships between physical stimuli and sensation ("outer psychophysics"), which only indirectly allowed inferences about the core problem of the relationship between physiological excitation and sensation. Fechner´s "inner psychophysics" is now embodied in research on the neural correlates of consciousness.

Today, the situation is very different than during the “dark years” of behaviorism. Research on consciousness is highly popular and Fechner’s “inner psychophysics” can be found continued in the search for the “Neural Correlates of Consciousness”, or NCC (Crick & Koch 1990; Chalmers 1996, 1998; Block 1996). In this article, we would like to reflect upon current developments in this field by asking ourselves what justifies the current excitement. In other words, what has changed since Fechner’s days that has so drastically overturned the previously negative outlook of the cognitive science community on consciousness research? Is the current optimism justified by new developments in brain-imaging techniques, or is consciousness research a mere fad that will soon fall back upon itself? In asking these questions, our purpose is to question the assumptions of the NCC program and to delineate its limits. We will suggest that while it is undeniable that the NCC program holds considerable promise in telling us what the relationship might be between brain processes and phenomenal processes, it cannot be taken for granted that the correlational approach is free from the theoretical and conceptual difficulties that have plagued previous attempts. Let us begin by briefly overviewing the NCC program.

2. The NCC program

The NCC program allows the investigation of the laws that describe transitions between conscious states without requiring a prior solution of the mind-body-problem. Indeed, the only assumption such an approach requires is that of a lawful covariance between cerebral and phenomenal processes, as already suggested by Fechner (1860) and Müller (1896). According to the NCC program, it would still be possible to follow a *nomological* research strategy even if the *ontological* problem were unsolvable in principle. In the face of the numerous neurological disorders that are closely related to consciousness (Milner & Rugg, 1992; Young, 1994), it might even be an ethical imperative for an empirical scientist to find out as much as one can about the underlying mechanisms rather than wait for the resolution of the ontological dispute. The search for the neural correlates of consciousness can thus be viewed as the search for bridge laws (Nagel 1961) that map phenomenal onto physiological concepts. If such bridge laws were known, one would be able to predict a subject's experience by a measurement performed on his brain.

Of course, by following this research program, the "explanatory gap" (Levine 1983) — or "hard problem" (Chalmers 1996) — remains unsolved, in the sense that even if we had established a perfect correlation between phenomenal and cerebral processes, we would still fail to know *why it is* that certain brain-processes are accompanied by qualitative experiential states. This problem holds for the same reason that we do not have an answer to the question of *why* the world is such that gravitation decreases as the square of the distance between two masses increases. "That's the way it is" is all a scientist can answer, and several ontologies may be constructed upon the same set of empirical data (Smart 1959). By adopting a nomological approach to consciousness, however, we would consider ourselves quite satisfied by having established the correlation. But even with this restricted goal, there are important problems in our way, including a "functional definition problem", as well as various "measurement problems", some of which may turn out to be of a principle nature.

How would a correlational research strategy on consciousness proceed? The establishment of the correlation seems to be possible via two different, but complementary, paths: Either at the level of models and theories, or by experimental work aimed directly at establishing the correlation empirically. We briefly review these two approaches in the following sections.

2.1 Theoretical correlation

What does it mean to correlate consciousness at the level of models and theories? The idea here is to search for and identify specific functional properties of current models of neural processing and to assess the extent to which these properties map onto properties of phenomenal states. This can of course take different forms depending on our assumptions about which functional features are relevant. Many such potential features have been offered over the past few years. For instance, Crick & Koch (1995a) have proposed to characterize visual awareness as an interpretation of a visual scene that is globally available for the control of action. This sort of assumption can help shape further arguments. For instance, if we believe that global availability for control is a functional property of phenomenal states, then we would expect to find direct projections from the NCC to the frontal lobes, which are responsible for action control. Crick and Koch used precisely that reasoning in developing their neuroanatomical argumentation against V1 (primary visual cortex) as a neural correlate of consciousness (Crick & Koch 1995a, 1995b, Pollen 1995): As V1 has no such projections, Crick and Koch doubt it can be an NCC. This type of argument does not require anyone to check if there really is no correlation between visual awareness and neural processes in V1[4]. Instead, the argument purely states a property of phenomenal states (global availability) and rules out one particular NCC-proposal based on observations of the wiring of the brain.

A different example of such a theoretical correlation is also based on neuroanatomy. Global availability suggests that conscious representations are globally broadcasted throughout the brain (Baars 1988). Thus Baars and Newman (e.g. Baars 1988, Baars & Newman 1994) proposed their ERTAS-model (extended reticular-thalamic activating system) to include diffuse thalamo-cortical projections, along with other structures that may support a wide dissemination of information[5].

It is not exclusively neuroanatomy that can give us theoretical hints about the nature of the NCC. For instance, the idea that synchronized gamma-band oscillations (ca. 35-75 Hz) and binding (Gray et al. 1989, Eckhorn & Reitboeck 1990) may constitute potential mechanisms of consciousness (Crick & Koch 1990) are mainly based upon theoretical considerations about the properties of neural processing (Milner 1974, von der Malsburg & Schneider 1986), albeit they have also been established empirically. The important point is that this proposal was a theoretical hypothesis before supporting evidence for it was found in the brain. Further, its continued popularity still appears to be based more on neurocomputational considerations than on actual empirical data.

Theoretical considerations also lie at the heart of many other current proposals. For instance, in a recent paper (O'Brien & Opie, in press), O'Brien and Opie defend the idea that phenomenal experience is caused by stable patterns of neural activity in the brain. In other words, stable patterns of activity in neural tissue are both necessary and sufficient to produce phenomenal experience. Phenomenal consciousness, from this perspective, depends neither on specific mechanisms nor on specific brain regions, but simply emerges as a result of the stability of some representations at some particular point in time. Interestingly, Mathis and Mozer (1996) have independently proposed stability of representation as a condition for conscious awareness to emerge, and have explored the implications of this proposal through connectionist modeling. In one interesting application, the model was found able to account for empirical data such as the differences between conscious and unconscious priming in Marcel's (Marcel, 1980) subliminal perception experiments.

The notion that stability of activation patterns forms the basis of phenomenal consciousness thus differs strongly from the idea that global availability is the relevant functional feature, and it should come as no surprise that such different proposals have important consequences on the conduct of empirical research. It is also important to note that there are now many different such proposals. For instance, Chalmers (1998) has compiled a whole collection of proposed NCCs (his "neural correlate zoo"). The list includes about 20 possibilities, ranging from 40-hertz oscillations (Crick & Koch, 1990) to ERTAS (Baars, 1988).

How are we to decide which proposal is most likely to be correct? Theoretical considerations about which functional properties should correlate with phenomenal experience cannot in and of themselves help advance the debate, but need to be rooted in empirical research. A complementary approach to the NCC problem therefore consists of taking the correlational program by the word, and of developing experimental setups that we can (1) clearly interpret as measuring consciousness, and (2) easily correlate with measurements of brain activity (e.g. Logothetis & Schall 1989). This "online" strategy, however, is confronted with a wide number of methodological problems. In this respect, theoretical considerations, based for instance on features of current models of neural processing, may serve a two-fold function. On the one hand they can show us what certainly cannot be the case (e.g. the pineal gland is certainly not a Cartesian convergence zone for sensory input). On the other hand theoretical considerations can inform empirical research by providing guidance about what to look for and about where to look for it in the brain.

2.2 Online correlation

The "online" logic of correlation research, upon which most of the remainder of this article will be focused, involves three distinct components: (1) measurement of phenomenal states, (2) measurement of physiological states and (3) their correlation. As we will describe in subsequent sections, each of these components raises difficult measurement and analytical issues, but it might be helpful to first describe the "online" strategy at a more general level first.

The online strategy is aimed directly (i.e., by experimentation) at answering the question "What physiological processes are accompanied by what kinds of experiences?" In this context, a process can usefully be thought of as a temporal sequence of states that a system passes through. But how do we describe a system's state? Abstractly, we can assign to it all sorts of properties, so as to describe the state as a point in a high-dimensional property-space. If we add an extra dimension for time, we can then record a process as the trajectory that a system follows in this space. The "stream of consciousness" (James 1892) can thus be defined as a trajectory (sequence of states) a certain system takes through phenomenal space. Likewise, physiological processes can also be described by trajectories within a space of possible physiological states. The goal of the "online" NCC program is then to correlate phenomenal and physiological trajectories[6].

The structure of phenomenal space

What is the exact structure of these spaces? While the basic building blocks of the brain seem to be relatively clear (Kandel et al. 1995), the structure of phenomenal space is to date only vaguely determined. It is a major achievement of analytical philosophy to have provided a clearer categorization of different concepts of consciousness that are in use. For instance, Block´s (1995) distinction between "access consciousness" and "phenomenal consciousness" was extremely helpful in this respect.

Beyond the necessary prerequisite of unspecific wakefulness, phenomenal states can be analytically divided into three classes. First, there are all kinds of *sensory* consciousness that follow the modalities and submodalities. These do not depend exclusively on such input, but can also be triggered internally, as in the case of mental imagery. The relational structure of sensory qualities has been investigated in psychophysics via discrimination spaces (see e.g., Clark 1993). The second class includes affective and emotional states, which can also be given a relational structure. Finally, the third class contains propositional conscious states. These are not necessarily accompanied by sensory or affective qualities, as the famous thought experiments of the Würzburg school of thought psychology early in this century have demonstrated (see Boring 1950, for an overview). Because of its compositional character, the structure of semantic space will be better described in terms of a taxonomy of contents: abstract thoughts about one's cognitive skills (metacognition), about a rule governing the behavior of a system in a learning task, about an abstract problem space, etc. Finally, it is important to note that most phenomenal states are compounds and are thus characterized by several of these aspects. Goals for example are not purely propositional attitudes (“I *want that* I go to the football match”), but also include imagery (the football stadium, the players) and affective qualities (as phenomenal “attractors”). In the following, we will be essentially concerned with phenomenal consciousness. How can we measure phenomenal experience? What are the problems we are likely to face in attempting to do so? We address these issues in the following section.

3. The problem of measurement

Even if the basic structure of neural and phenomenal spaces were clear, both domains bring up serious measurement problems. It should be remembered that every measurement also requires a background theory that allows the interpretation of what has actually been measured. The notion that measurement cannot be dissociated from theory was formulated early in this century for the field of physics by Pierre Duhem: “An experiment in physics is not simply the observation of a phenomenon; it is, besides, the theoretical interpretation of this phenomenon.” (Duhem 1904, p. 144). The gap between reading the display of an oscilloscope and the interpretation that one is recording an action potential in a neuron may be less controversial than the interpretation of data in brain-imaging experiments, which result from all kinds of statistical and filtering processes in order to extract signals. Sometimes, it is not clear what one actually has a measure of. For phenomenal data, this holds even more: While we can at least theoretically imagine methods for recording brain states and processes, the measurement of a system’s phenomenal state is highly problematic. Let us start by focusing on the problems involved with recording neurophysiological trajectories.

3.1 Recording neurophysiological trajectories

What is our current picture of brain processes? In neural network models of cognition a computation is achieved by the specific arrangement of artificial neurons and their connection weights. This is believed to correspond, at least in a very general sense, to known processing principles of the brain. But how can we find out if our models are correct? In order to validate specific neural processing models it would be necessary to simultaneously record the *individual* activity of connected neurons within large populations. This, obviously, is not possible with current techniques.

Standard methods for in vivo measurement[7] of neural activity vary in terms of their target. Single-cell recording techniques target the activity of individual neurons, while Electroencephalography (EEG) and Magnetoencephalography (MEG) techniques target respectively the collective electric or magnetic activity of large populations of neurons. Finally, techniques such as Positron Emission Tomography (PET) or Functional Magnetic Resonance Imaging (fMRI) provide us with an indirect measurement of neural activity via metabolic processes. The latter two methods entail a trade-off between spatial and temporal resolution: While EEG and MEG allow high temporal but low spatial resolution, PET and fMRI have to trade in their higher spatial resolution for being relatively slow.

Any of these methods can provide measures of neural activity that may covary with performed cognitive processes. They do not, however, in and of themselves offer an answer to the question of how they should be interpreted in terms of the brain processes they are caused by. Event related potentials (ERP), for example, are computed by selective averaging of a number of epochs of EEG signals that are time locked (indirectly via the time of stimulus onset or of a button press) to a repeated cognitive process. These potentials are weak and thus not normally visible in the online EEG. An ERP normally consists of a number of positive and negative potential components, which covary with certain task conditions (Rugg & Coles, 1995). Many different studies have been performed to determine how various experimental conditions influence these components. A specific component structure can thus be viewed as a correlate of a certain class of processes, but the interpretation of these components in terms of brain processes remains very difficult[8]. Thus, while we might identify EEG measures that systematically correlate with some hypothetical underlying process, the question of what functional role they play in the brain is not yet clear (Coles 1995)[9].

Similar issues arise in the case of methods based on the measurement of bloodflow and metabolic processes as *indirect* indicators of neural processes, such as PET or fMRI. The exact relationship between changes in neural activity, metabolism and regional cerebral blood flow (rCBF) is not yet entirely clear (Malonek et al. 1997). For instance, there are time lags in the range of seconds between the onset of neural activity and the subsequent metabolic processes detected by fMRI (Kim & Ugurbil 1997). A promising approach to such measurement issues consists of combining different techniques in such a way as to elaborate a model of which processes different techniques actually measure (e.g., Heinze et al. 1994).

The standard procedure with metabolic measures is to register an increase or decrease in regional cerebral bloodflow (PET) or blood-oxygenation (fMRI) by subtracting the levels observed under experimental conditions from those obtained under one or several control conditions that are supposed to differ from the experimental condition exclusively in the cognitive component of interest[10]. This often leads to a modular picture of the brain that is reminiscent of 19th century phrenology, in that it likewise assigns functions to specific brain areas without considering the brain as an interacting functional system (Luria 1970). In contrast, the currently emerging picture is that a specific state of consciousness can only occur if a wide number of subsystems (each with specific contributions) interact in a coordinated way.

With the current array of relevant methods, a scientist will therefore have to make a priori decisions about which correlates she is looking for, and where they are to be found: Single cell activity (is there a consciousness cardinal cell?), synchronized oscillatory behavior of individual neurons, enhanced 40 Hz activity in the EEG, brain electric microstates (Lehmann et al. 1998), etc. Here it becomes clear again that our current theories about brain processes play an important role in guiding research: We certainly would not look in auditory cortex to find the neural correlate of a visual experience. Computational considerations are similarly important. If we have a certain theory about how the functional unity of perceived objects is realized in the brain (e.g. by synchronized activity), we might look for corresponding neural processes as possible NCC candidates. The choices are often also motivated by pragmatic considerations. For instance, one reason that research on visual awareness currently focuses on binocular rivalry rather than on bistable stimuli such as the Necker cube stimulus is that it is not yet clear where exactly one should look for the neural population that codes Necker cubes (Crick & Koch 1990, Crick 1994). The pitfalls of this selectivity in research were already recognized by Crick (1994, p. 218):

> “(...) although the behavior of the neurons in cortical area MT appears to be correlated with the monkey´s discrimination, and therefore probably with its visual awareness, it does not follow that these particular neurons are the real seat of awareness. They may, by their firing, influence other neurons, perhaps elsewhere in the visual hierarchy, that are the true correlates of awareness.”

In the field of neurophysiology we are confronted with the further problem that we cannot manipulate neural processes directly. Phenomenal states are usually experimentally induced by applying stimuli to the sensory surface of a subject. If we want to know if activity in area V4 is a neural correlate of an experienced color, then we would normally not open the brains of our subjects and stimulate V4 directly[11], but we would, for instance, present colored vs. uncolored stimuli and determine whether we can find a difference in brain activity in the corresponding area (Zeki 1992). The presentation of stimuli causes a chain of neural processes, beginning at the sensory surface, that will also include those that specifically correlate with the conscious state. But which subset of these processes are necessary and sufficient to produce a color experience? One possible solution to control the background noise is to use multistable stimuli (such as the famous Necker cube) that produce changes in conscious percepts while keeping the sensory input constant (e.g. Basar-Eroglu et al. 1993, Logothetis & Schall 1989, Logothetis 1996). A different approach is to artificially time-delay the presentation of the stimulus from the moment of recognition, as can be done with so-called “hidden figures”, where the figure-ground segregation is complicated by a masking background (Landis et al. 1984), or with random-dot autostereograms (“magic-eye” pictures), where the emergence of a 3-D percept occurs some seconds after the onset of the stimulus (Julesz 1978, Revonsuo et al. 1997). Such methods make it possible to separate the primary sensory processing components from the process of conscious recognition.

To summarize, we do not have a “universal brain state meter” that would be able to record neural processes at any desired resolution. But this is not a prerequisite for the conduct of empirical work — for taking this requirement seriously would lead us right down from brains to neurons, proteins, atoms and quarks. In other words, we do not generally believe that we need to know everything about the exact quantum-state of a brain in order to be able to understand its behavior. It is important to ask, however, what the best resolution might be. There is no simple natural answer to this question (see e.g., Cleeremans & French 1996). For instance, it is interesting to note that most connectionist models manage to predict empirical data quite well using principles that are not realized in real neural systems. Thus, it is known, with something approaching certainty, that backpropagation does not exist in the brain, in spite of numerous unsuccessful efforts (Crick, 1989) designed to find processes or structures that could reasonably be said to "backpropagate" error signals to upstream neurons. And yet, no one thinks twice when these ubiquitous feedforward backpropagation networks are used to analyze high-level cognitive phenomena, such as speech production (e.g., Sejnowski & Rosenberg, 1987), sentence parsing (Elman, 1990) or word recognition (e.g., Seidenberg & McClelland, 1989); mid-level phenomena such as implicit learning (e.g., Cleeremans, 1997, Cleeremans, Destrebecqz & Boyer 1998); or even low-level neural phenomena, such as dyslexia (e.g., Plaut & Shallice, 1993). All of these models have been very successful in accounting for empirical data, yet all of them are based on the otherwise completely unsupported notion that learning takes places through back-propagation.

On the other hand, consider neuronal models that hew to the constraints of experimental neurophysiology with unparalleled rigor (e.g., Golomb, Wang, & Rinzel, 1995). Ionic channels, sodium and calcium flows are all modeled to match experimental findings, connection schemes are copied from real neural patterns, neurotransmitter levels are carefully controlled, and the Hodgkin-Huxley equations rigorously respected. From these models, neuronal spiking patterns can be produced and predicted. Higher lever oscillatory firing patterns among groups of "neurons" can be observed, predicted, and modified by changing any of a large number of experimentally observed parameters. Now, a connectionist modeler might think "This is real modeling!" But even these detailed synaptic models are routinely criticized by neurologists as oversimplifications of real neural events.

What are the problems we face when attempting to capture phenomenal experience? We address this issue in the next section.

3.2 Recording phenomenal trajectories

Phenomenal experience is private — hence, as objects of psychological research, phenomenal states have to be approached indirectly, that is, through behavior. Perhaps unsurprisingly, numerous problems may arise when making inferences about phenomenal states based upon behavior. The first problem is to define what kind of *internal* functional states one wants to count as indicators of certain phenomenal states. The second problem is to detect the presence or absence of these functional states based on *external* observable behavior.

The question of adequate measurement procedures is without doubt the most controversial issue in psychological research on consciousness. Indeed, some contemporary areas of psychology (e.g., implicit learning, see Cleeremans 1993; Cleeremans, Destrebecqz & Boyer, 1998) are replete with definitional issues. Mostly, one can trace debates back to differing definitions of measures for consciousness (see e.g., Reingold and Toth 1996). This is very clear, for instance, in the case of research about subliminal perception. The term "subliminal perception" indicates that one is willing to call a representation a "percept" even if the subject is not phenomenally aware of having it. To establish the existence of such representations, the typical research strategy has been to determine the threshold below which a subject reports not to perceive a stimulus (the "prime"), and then to examine whether this stimulus can nevertheless be causally efficacious, that is, whether it can modulate responses such as forced choice guesses or word stem completion. Opponents of this strategy have consistently criticized the threshold determination procedure (e.g. Eriksen 1960; see also Holender 1986), and argue that any remaining discrimination is based on occasional cases where subjects do actually consciously perceive the stimulus, but at such a weak level that they prefer to follow a conservative response strategy and say that they did not do so[12]. In this controversy, two types of percepts playing different functional roles are involved:

• Percept$_1$ is operationally defined as a perceptual state that (a) can selectively influence behavior and (b) that the subject *claims* to have access to.

• Percept$_2$ is operationally defined as a perceptual state that (a) can selectively influence behavior and (b) that the subject *claims not* to have access to.

The dispute is that some scientists (let us call them the “explicitists”) hold the belief that percepts$_1$ *and* percepts$_2$ are accompanied by a phenomenal experience while others (the “implicitists”) believe that percepts$_2$ are not. If one were to perform further tests upon percepts of the second type, and find that they have a variety of potential influences, the explicitists would simply attribute any further findings to conscious percepts and the implicitists to unconscious percepts.

Is the intuition about the mapping of functional roles to phenomenal states thus an a priori which cannot be further scientifically validated and is given into research per definition? In the case of the debates surrounding subliminal perception, Reingold and Toth (1996) have argued that the explicitist’s strategy in fact results in *defining* subliminal perception out of existence, as every remaining discrimination is *interpreted* as a sign of conscious perception. In this context, it might help to have the scientist experience the mapping in his own cognitive system by participating in the experiment himself. An explicitist subject might then have no phenomenal experience with percepts of the second type (i.e. he claims not to have any), only to find himself corrected by a second explicitist (the experimenter) who informs him that he in fact did have a phenomenal experience. Perhaps this situation is the only way towards a phenomenal validation of the measures…

Our intuitions about functional definitions of conscious representations thus seem to be hopelessly entangled with our intuitions about how to measure them. How we choose to map consciousness to functions determines whether we will take a certain behavior to indicate a conscious representation or not. We can call this the "functional definition problem"[13]. If our intuition says that when driving a car in "auto-pilot" while talking to the passenger, one has no phenomenal experience of the driving process, then one might also doubt behavioral measures as indicating conscious states.

Sensitivity and selectivity

Perhaps it would help to look for a more general criterion of consciousness rather than tapping one's intuitions in every single experimental setup. One quite popular belief held about conscious states is that they have the property of being "globally available" (Baars, 1988; Chalmers, 1998) and that they can be flexibly acted upon in various ways. In medical diagnostics there are two criteria that make a symptom a good indicator for a diagnosis (Bishop, Fienberg & Holland 1975). On the one hand, it is supposed to be *sensitive* to all occurrences of the disease it is to indicate. On the other hand, it should be as *selective* as possible and not be present in other diseases. Similar criteria have been proposed in the context of consciousness research. For instance, Shanks & StJohn (1994) have proposed that we respect *information* and *sensitivity* criteria when probing a subject for some knowledge in an implicit learning task. That is, we should make sure that the subject actually needs to know the information we are probing for in order to perform successfully in the task, and we should make sure that the measure we use is sensitive enough to that knowledge. It remains unclear, however, whether the latter criterion can be satisfied at all, because, as Reingold and Merikle (1988) have pointed out (see also Jiménez, 1997), it seems that it would require measures of conscious knowledge that are simultaneously exhaustive (they measure *all* of a subject's conscious knowledge) and exclusive (they measure *only* the relevant conscious knowledge).

In this respect, the popular "global availability" criterion lacks both sensitivity and selectivity as an indicator of consciousness. James's "vague and inarticulate (...) fringe" (1892, 165-166), for example, does not seem to be "globally available" in the same sense as a focally attended conscious percept — precisely for the reason that it cannot be articulated. But if we cut down our criteria for "global availability" so as to make it sensitive to the fringe, for instance by using purely behavioral indicators, we then immediately face the problem of identifying a specific threshold such that our measure does not also spuriously include unconscious stimulus-behavior mappings, as in automatized behavior.

Hence for every available measure, there seems to be a trade-off between sensitivity and selectivity. "Availability" can be quite global for other (unconscious) representations as well — thus decreasing its selectivity as a criterion for consciousness. Imagine the following experiment: A subject is shown photos of 5 people and is given their names. We can then ask him to point to "Tom", or ask him to tell us the name of the person on the left photo, etc. Our intuition says that the subject is not phenomenally aware of the names all the time, but they must be held in short-term memory as they can be retrieved in various ways for the answers. Is this not some kind of global availability for control? It is difficult to find a behavioral criterion that will clearly indicate exclusively availability of conscious representations and not be subject to possible alternative explanations.

Similar considerations would apply for other purported functional properties of conscious states. For instance, stability in time has recently been proposed (O'Brien & Opie, in press) as a functional feature with which to characterize conscious representations. However, "stability", just as global availability, again lacks both sensitivity and selectivity, for it seems plausible that there exist stable patterns of neural activity that one is not directly aware of (i.e., patterns of activity among the light receptors in our retinas).

Sensitivity and selectivity of criteria also play a role depending on whether one wants to demonstrate the presence or absence of phenomenal states. Psychological approaches to consciousness have generally been focused on demonstrating implicit processes (what we can do without consciousness involved), and have therefore faced the difficult task of establishing that certain classes of phenomenal contents were not present during a processing phase. Proving the non-existence of something is known to be more difficult than demonstrating its presence (as all one needs in the latter case is a positive example), however, and also involves different measurement processes. For instance, while most people would agree that verbal reports are valid positive indicators of phenomenal states (in the sense that nobody holds the position that percepts$_1$ are unconscious!), verbal reports are not very good at indicating their absence (Allport 1988). Thus, observing that a subject claims not to have perceived a stimulus, for instance, may be attributed to a failure of recalling the relevant information, to the fact that the relevant knowledge is held with low confidence, to the difficulty of providing verbal descriptions of the details of a complex perceptual state[14], etc.

These difficulties have prompted many authors to dismiss verbal reports as useful indicators of conscious states, and have led them to rely instead on purely behavioral measures, such as button presses. Such measures, however, also suffer from a certain ambiguity, due to the fact that they can be performed automatically and thus may or may not be based on conscious percepts. If one wants to prove that a certain type of perceptual state is simultaneously causally efficacious and *not* available to consciousness, the only uncontroversial criterion seems to be the observation that a subject *cannot avoid* using the stimulus in a task. This ratonale is at the heart of the so-called process-dissociation-procedure, which has been extensively applied in fields such as implicit memory, subliminal perception, and implicit learning (e.g., Debner & Jacoby 1994). But what does it mean if a person *does* have control over a representation? This does not automatically indicate that the subject is aware of it. Given a certain conscious task context it may be that we can intentionally configure our unconscious processes too. All in all, we have no single measure that allows a clear division between conscious and unconscious processes. Verbal report may not capture everything that is available to consciousness, and process-dissociation may not capture everything that is unconscious. This means that we have to decide in advance what we want to be confident about — consciousness or unconsciousness of a representation — and use the appropriate measurement procedure.

Is there a "functional state meter"?

Even if we could pragmatically agree upon a property such as "availability for global control" as an indicator of conscious representations, we would need to determine in each individual case whether a specific representation of this type was present at a certain time. This is not as easy as it may seem, because we usually have certain degrees of freedom in interpreting a certain experimental behavior by the internal functional processes it may have been produced by. Not only do we lack a "consciousness meter" (Chalmers 1998), we also lack a "functional state meter". This is what we could dub the "functional state measurement problem".

Thus, both the implicitist and the explicitist could share the "global availability" criterion. The explicitist could simply give a different *functional* interpretation and say that percepts$_2$ *are* principally available for global control, but the subject chooses not to exert that control due to a conservative response bias. How could we test such a hypothesis? All we have access to is the external behavior of a cognitive system. Its internal functional processes have to be inferred indirectly. The main difficulty here is that a single observable behavior may arise from a variety of different internal causes. To take a technical example, one's computer may fail to operate for a number of different reasons. The CPU may be blown, a cable may be defect, the hard disk may be broken, etc. A skilled technician could certainly open up the computer and precisely identify the cause of the failure, but she can only do so because she has a detailed theory of how computers work. Psychologists, unlike computer technicians, do not yet possess similarly detailed theories about the workings of the mind.

"Global availability" is a *dispositional* property of an internal representation. In order to know whether a representation is globally available, we would have to test if it really can be used in different ways. This cannot be done in a single experimental trial because we can only test for one type of availability at once due to the complex measurement process involved. The problem that such measurement processes induce changes in the system also holds for measurements of most neurophysiological processes, but in these cases the interaction is kept relatively small (e.g. radioactive injections, application of magnetic fields), so that they are not likely to influence neural or cognitive processes in a fundamental and systematic way. But there seem to be only two possibilities to measure what a person experiences: (1) One can instruct the person in advance to monitor his experiences and to act upon them (verbally or behaviorally), or (2) one can ask the person retrospectively about prior experiences, without giving her instructions before the phenomenal event is poised to occur.

In the first case, we are then confronted with a "superposition problem", because we might somehow disturb the phenomenal trajectory by the instruction[15]. An advance instruction to a subject to report his momentary phenomenal state necessarily triggers a superimposed cognitive process that performs the measurement. As phenomenal states are likely to depend on transient cognitive processes, this may lead to a major interference of the state to be measured[16]. The only states that can be recorded in this manner are those that remain invariant under the measurement process (eigenstates under measurement M). These eigenstates are basically the "clear cases", about which a consensus that they remain unchanged by the measurement procedure can be developed. Such cases can be systematically induced in experimental settings. For instance, if someone is instructed to focus on a red square, we would agree that asking the person about his experience would not change it substantially. On the other hand, if a subject is instructed to monitor whether a particular task involves conscious processing of certain stimuli, it might be that a positive report is only due to an attentional shift that would not have been present in the undisturbed case[17.]

Given these difficulties, should we use retrospective measures instead? Memory dependent measures are used in many settings. In the experiments on subliminal perception mentioned above, a visual stimulus is presented to the subject for a brief time, and is then followed by a second stimulus appearing at the same location. This results in a so-called "backwards masking effect", which prevents the first stimulus from becoming conscious. Or does it just erase our memory of a transient experience of the first stimulus? Representation in short-term memory is generally taken as the demarcation line between unattended and non-perceived stimuli: If there is memory immediately after the presentation, there was awareness, if there is no memory, there was no awareness (Rock & Gutman 1981). This criterion differs from "global availability" because it extends access in time: The representation in short-term memory is available for control some time after being perceived.

An interesting retrospective procedure is the so-called "thought-sampling", introduced by Klinger et al. (1978), which uses a random interval beeper to instruct subjects to report "What went through your mind just before a signal occurred" each time the beeper sounds. This is a short-term-memory dependent measurement procedure, which (due to its unpredictability) has the advantage of recording phenomenal trajectories in a rather undisturbed way. Lehmann et al. (1998) recorded subjects' EEG while applying the thought-sampling technique. The reported experiences were categorized into "abstract" vs. "imagery" and then correlated with pre-beep EEG-microstates (phases of "similar" brain electric activity). The microstates about 120 ms before the beep were significantly different for the two classes of experience, which indicates a participation of functionally different neural systems. This was interpreted by the authors as evidence for the existence of “atoms of thought” with a duration corresponding to that of the microstates.

But are memory-dependent measures a good solution? While short-term memory is a good positive measure, as Allport (1988) has pointed out, a negative result, by contrast, is not such a good indicator of the absence of awareness, because the relevant information could have simply been forgotten. Another issue is that the reconstructive nature of memory makes it difficult to decide about the truth of retrospective statements. Research on dreaming shows this clearly: Online recording of phenomenal states is impossible, so subjects are asked about their dreams after they have woken up. It cannot be decided by this method if the memory relates to an actual conscious process or is a mere retrospective illusion. The constructive nature of memory processes (Bartlett 1932) is widely accepted within cognitive psychology, and causes serious potential difficulties for the validity and reliability of this type of measurement. A final problem related to memory is the restriction that memory imposes on the number of concepts we can form of phenomenal states. This is what Raffman (1995) has called the “memory constraint”: subjects are able to discriminate many more qualitative differences when stimuli are presented synchronously than they are able to re-identify.

Hence, while we are in advantage over Fechner, Müller, and Köhler in terms of our methods for the measurement of physiological processes, it seems that there has been no major advance in the field of phenomenology. To recapitulate, we are confronted with the following problems:

- mapping of phenomenal to functional states
- selectivity and sensitivity trade-off of measures

- awareness or unawareness as target
- measurement of functional states
- superposition of measurements
- memory constraints
- communicability constraints

This list is far from complete — complex statistical measurement problems have been left out completely for example (e.g., Greenwald & Draine 1997). In short, the lack of an adequately elaborated phenomenology is the major unsolved problem in consciousness studies in general, and thus of course also stands in the way of the NCC research program.

4. Establishing the correlation

Once suitable parallel recordings of phenomenal and neurophysiological trajectories have been established, one is still confronted with the problem of correlating them. A particularly clear example of how the correlation problem can be thorny is provided by cases where one is targeting phenomenal *events* rather than static states. For instance, in the famous experiments of Libet (Libet et al. 1983, Libet 1985), the issue is what happens in the EEG just before a person consciously decides to perform an action. This volitional act is an event, and in order to correlate it, it is necessary to localize it as precisely as possible in the time domain. The same problem arises in the case of perceptual transitions, when a subject is presented with multistable stimuli (Basar-Eroglu et al. 1993, Logothetis & Schall 1989, Lumer et al. 1998). In this case also, phenomenal transitions, thus events, are to be identified and temporally localized. If one uses button presses to indicate these events, this adds a neural motor component that again needs to be separated from the brain processes that are responsible for the phenomenal change (Basar-Eroglu et al. 1996).

Even if we had singled out only conscious states, we would still have a selectivity problem. How many different phenomenal states does a certain neural substrate correlate with? While subcortical activity of the reticular formation is a necessary condition for a red experience, it is at the same time a necessary condition of a tone experience, because it is part of the more general condition of wakefulness. On the other hand, cortical activity seems to correlate more specifically with differences in content (Baars 1995). This leads to a problem already stated by Mill (1843) in the field of causality[18]:

> "It is seldom, if ever, between a consequent and a single antecedent, that this invariable sequence subsists. It is usually between a consequent and the sum of several antecedents; the occurrence of all of them being requisite to produce, that is to be certain of being followed by, the consequent. In such cases it is very common to single out one only of the antecedents under the denomination of Cause, calling all the others merely Conditions." (Mill, 1843, 327).

These "conditions" can be related to the numerous unspecific (e.g. subcortical) correlates that are necessary for a conscious state to occur. Mill's argument raises the interesting question of whether the same phenomenal state can be accompanied by several *different* sufficient sets of physiological conditions, as is possible with some other cognitive functions, which may be achieved in different ways.

The gap between psychology and neurobiology is thus still very large. Imagine a person who is asked to think either of the abstract concept "teleological functionalism" or "type identity theory", and to press on one of two corresponding buttons five minutes later. It would not be problematic to simply ask the subject what she is thinking of and to use this information to predict her behavior. But if our current theories of semantic representation are correct, and if meaning is therefore distributed over large populations of neurons in the specific structure of their synaptic connectivity, then, given the currently available neurobiological techniques (see above), it will remain impossible for quite some time to measure the state of this interconnected neural representation system. We might be able to localize the part of the brain in which the entire "abstract semantic population" resides, but this will not allow us to predict human behavior over even short time scales, when it is influenced by semantic properties.

5. Conclusion

This article may appear to be overly pessimistic about the possibility of correlation research on consciousness. However, this need not be the case — a clearly structured problem space is one of the most important steps towards a solution. Figure 2 gives a sketch of our view of this problem space. Along with many empirical scientists we believe that a scientific approach to consciousness is possible without the prior solution of the body-mind problem — that is, in the form of a purely nomologically oriented research program. It is important to remember, however, that this promising research strategy is confronted with many fundamental problems, the most important of which are outlined in Figure 2. In this article, we have attempted to delineate a few of these problems: The functional definition problem, the functional measurement problem, and physiological measurement problem. Thus, not only are our current methods for brain research still rather limited, but there is furthermore no empirical approach to the measurement of phenomenal data that has so far achieved to go beyond the very "clear cases". As a case in point, it can be surprising to find assumptions (e.g., "is unconscious representation possible?") that are hotly debated in one context (e.g., implicit learning or subliminal perception) go wholly unquestioned in another. As we described earlier, there is no universally accepted definition of consciousness that incorporates sufficiently clear functional markers. Our intuitions about when we want to ascribe phenomenal states to subjects are dependent on attentional and memory processes, and thus we seem to be perpetually confronted with the problem of determining whether we are really investigating consciousness, or a functional compound about which we cannot ascertain how (or if at all) it maps to phenomenal states.

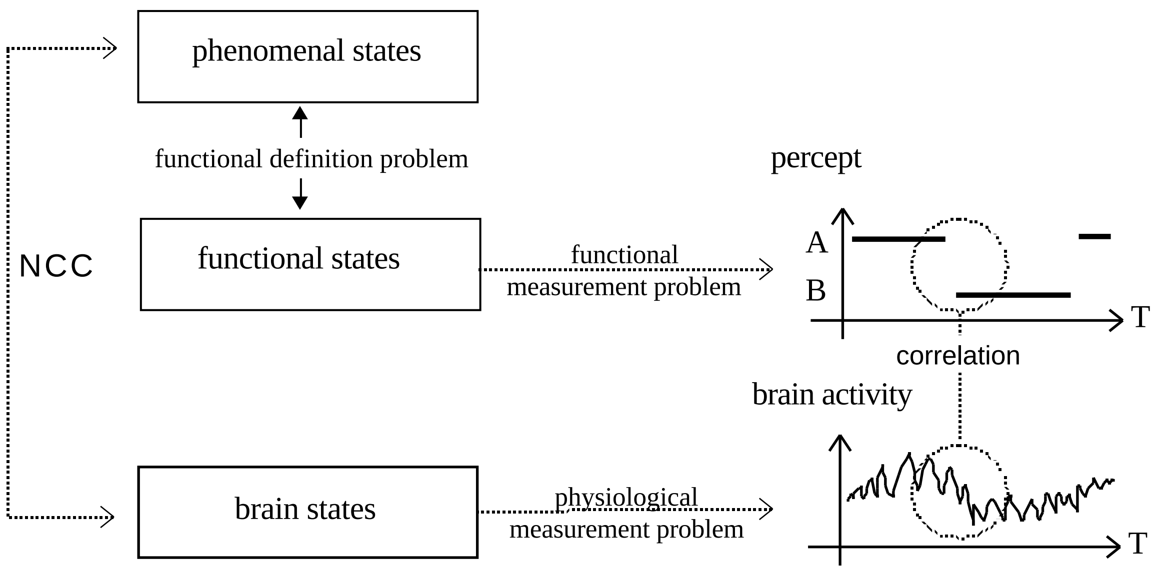

Fig. 2. The complex definition and measurement problems encountered in research on the neural correlates of consciousness (see text for further details).

The most important problem is that we seem to have to bring in the subjectivity of the scientist in order to make statements about the mapping of functional to phenomenal properties. As psychologists we are in a similar situation to Jackson's (1986) neuroscientist Mary. We can know everything about the cognitive or functional properties of a person (which is far beyond where we are at the present time), but be blind in terms of their mapping to phenomenal properties. We therefore have to use "first person evidence" (Chalmers 1998) — but one person's first person evidence is another person's third person evidence. Further, even this evidence is not purely introspective, but based upon complex inferences: I do not *see* that my conscious representations have the functional property of being globally available, but I *infer* it.

Perhaps we should follow a path sketched above: Like in Mary's excursion to the colored world, we could simply have the *experimenter* experience his own experiment, and thus produce the functional and neural processes in his own cognitive system. This might make him more careful in attributing rudimentary discriminations to conservative response biases. But of course, there are also situations where we would even deny the experimenter access to his phenomenal states. It seems that we are caught in a vicious circle between first person and third person evidence. All that we can do is wait and hope for “converging evidence” (an expression which nowadays can be found in virtually every other article in cognitive neuroscience). We will have to somehow integrate all sorts of theories, models and data from different approaches: psychology, neuropsychology, neurobiology, computational modeling and introspection. But how do the pieces of this mosaic fit together?

Perhaps something can be learned from philosophy. The NCC research strategy brings about a change in the focus of the philosophical problem. Because of the complex problems encountered in empirical research, the scientist is more likely to find help in resorting to literature in the field of philosophy of science rather than in philosophy of mind. There is a great need for a metatheoretical reflection on the research process because of several distinct reasons. First, the NCC project is inherently interdisciplinary, and hence involves radically different levels and methods of approach[19]. Second, the goal of the project is a *nomological* reduction of phenomenal processes to physiological processes. This involves a reflection on the principles of reduction and (nomological) explanation that are not normally considered in any single discipline’s methodology. Third, the problem of measurement has received considerable attention in philosophy of science, and is obviously of great importance to the NCC research program. On the one hand, measurement presupposes a background model of what is actually being measured (Duhem 1904) — a problem that appears in the neurobiological as well as psychological aspects of NCC research. On the other hand, the special intersubjectivity problem of phenomenal data will need to be dealt with as a fundamental epistemological problem (cf. Goldman 1997). Finally, it is likely that we will have to rely on "converging evidence". A critical reflection on this term within the framework(s) provided by philosophy of science may be highly useful.

Acknowledgments

Axel Cleeremans is a research associate of the National Fund for Scientific Research (Belgium). John-Dylan Haynes is a research assistant at the Institute for Psychology and Cognition Research, University of Bremen (Germany). The authors wish to thank Thomas Metzinger, Sven Schütt, Daniel Strüber, Michael A. Stadler and Günter Vetter for valuable discussions on the current paper. Parts of the chapter on phenomenal measurement are based upon an unpublished manuscript by John-Dylan Haynes and Sven Schütt.

References

Allport, A. (1988). What concept of consciousness? In Marcel, A. J. & Bisiach, E. (Eds.), Consciousness in Contemporary Science. Oxford: Clarendon Press.

Anderson, J.R. (1978). Arguments concerning representations for mental imagery. Psychological Review, 85, 249-277.

Baars, B.J. (1988). A Cognitive Theory of Consciousness. Cambridge University Press.

Baars, B.J. (1995). Surprisingly small subcortical structures are needed for the state of waking consciousness, while cortical projection areas seem to provide perceptual contents of consciousness. Consciousness and Cognition, 4, 159-162.

Baars, B. J. & Newman, J. (1994). A neurobiological interpretation of Global Workspace Theory. In Revonsuo, A. & Kamppinen, M. (Eds.), Consciousness in Philosophy and Cognitive Neuroscience. Hillsdale, NJ: Lawrence Earlbaum Associates.

Bartlett, F.C. (1932). Remembering: A Study in Experimental and Social Psychology. Cambridge: Cambridge University Press.

Basar-Eroglu, C., Strüber, D., Stadler, M. & Kruse, P. (1993). Multistable perception induces a slow positive EEG wave. International Journal of Neuroscience, 73, 139-151.

Basar-Eroglu, C., Strüber, D., Kruse, P., Basar, E. & Stadler, M. (1996). Frontal gamma-band enhancement during multistable visual perception. International Journal of Psychophysiology, 24, 113-125.

Bishop, Y., Fienberg, S. & Holland, P. (1975). Discrete Multivariate Analysis: Theory and Practice. Cambridge, MA: MIT Press.

Block, N. (1995). On a confusion about the function of consciousness. Behavioral and Brain Sciences, 18, 227-297.

Block, N. (1996). How can we find the neural correlates of consciousness? Trends in Neuroscience, 19, 456-459.

Boring, E.G. (1950). A History of Experimental Psychology. New Jersey: Prentice-Hall.

Bornstein, R.F: & Pittman, T.S. (1992)(Eds.). Perception without Awareness: Cognitive, Clinical, and Social Perspectives. New York: Guilford.

Chalmers, D. (1996). The Conscious Mind. Oxford: Oxford University Press.

Chalmers, D. (1998). On the search for the neural correlates of consciousness. In Hameroff, S., Kaszniak, A., & Scott, A. (Eds.). Toward a Science of Consciousness II. Cambridge, MA: MIT Press.

Clark, A. (1993). Sensory Qualities. Oxford: Clarendon.

Cleeremans, A. (1997). Principles for implicit learning. In D. Berry (Ed.), How implicit is implicit learning? Oxford: Oxford University Press.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. Trends in Cognitive Sciences, 2, 406-416.

Cleeremans, A., & French, R.M. (1996). From chicken squawking to cognition: Levels of description and the computational approach in psychology. Psychologica Belgica, 36, 5-29.

Coles, M.G.H. (1995). Event-related brain potentials: an introduction. In Rugg, M.D. & Coles, M.G.H. (Eds.), Electrophysiology of Mind: Event-Related Brain Potentials and Cognition. Oxford: Oxford University Press.

Crick, F. (1989). The recent excitement about neural networks. Nature, 337, 129–132.

Crick, F.C.H. (1994). The Astonishing Hypothesis: The Scientific Search for the Soul. New York: Scribners.

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

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

Crick, F.C.H. & Koch, C. (1995b). Cortical areas in visual awareness - reply. Nature, 377, 294-295.

Debner, J.A. & Jacoby, L.L. (1994). Unconscious perception: attention, awareness and control. Journal of Experimental Psychology: Learning, Memory and Cognition, 20, 304-317.

Duhem, P. (1904). The Aim and Structure of Physical Theory. Translated by Philip P. Wiener, 1974. New York: Atheneum.

Eckhorn, R. & Reitboeck, H. J. (1990). Stimulus-specific synchronizations in cat visual cortex and their possible role in visual pattern recognition. In Haken, H. & Stadler, M. (Eds.), Synergetics of Cognition. Berlin: Springer.

Elman, J.L. (1990). Finding structure in time. Cognitive Science, 14, 179–211.

Eriksen, C.W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. Psychological Review, 67, 279-300.

Fechner, G.T. (1860). Elemente der Psychophysik. Leipzig: Breitkopf und Ertel.

Goldman, A.I. (1997). Science, publicity and consciousness. Philosophy of Science, 64, 525-545.

Golomb, D., Wang, X. & Rinzel, J. (1994). Synchronization Properties of Spindle Oscillations in a Thalamic Reticular Nucleus Model. Journal of Neurophysiology, 72, 1109-1126.

Gray, C. M., König, P., Engel, A. K. & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which exhibits global stimulus properties. Nature, 338, 334-337.

Greenwald, A.G. & Draine, S.C. (1997). Do subliminal stimuli enter the mind unnoticed? Tests with a new method. In Cohen, J. D. & Schooler, J. W. (Eds.), Scientific Approaches to Consciousness. NJ: Lawrence Erlbaum Associates.

Hardcastle, V.G. (1995). Locating Consciousness. Amsterdam: John Benjamins.

Heinze, H.J., Mangun, G.R., Burchert, W., Hinrichs, H., Scholz, M., Münte, T. F., Gös, A., Scherg, M., Johannes, S., Hundeshagen, H., Gazzaniga, M.S. & Hillyard, S.A. (1994). Combined spatial and temporal imaging of brain activity during selective visual attention in humans. Nature, 372, 543-546.

Helmholtz, H. (1861). Handbuch der Physiologischen Optik. Leipzig: Voss.

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, 1-66.

Jackson, F. (1986). What Mary didn´t know. Journal of Philosophy, 83, 291-295.

James, W. (1890). Principles of Psychology. Reprinted 1981. Cambridge, MA: Harvard University Press.

James, W. (1892). Psychology. Reprinted 1948. Cleveland: World Publishing Company.

Jiménez, L. (1997). Implicit learning: Conceptual and methodological issues. Psychologica Belgica, 37, 9-28.

Julesz, B. (1978). Global stereopsis: cooperative phenomena in stereoscopic depth perception. In Held, R., Leibowitz, H. W. & Teuber, H. L. (Eds.), Handbook of Sensory Physiology Vol. VIII: Perception. Berlin: Springer.

Kandel, E.R., Schwartz, J. H. & Jessell, T. M. (1995). Essentials of Neural Science and Behavior. Appleton & Lange.

Kim, S.G. & Ugurbil, K. (1997). Functional magnetic resonance imaging of the human brain. Journal of Neuroscience Methods, 74, 229-243.

Klinger, E. (1978). Modes of normal conscious flow. In Pope, K. S. & Singer, J. L. (Eds.), The Stream of Consciousness: Scientific Investigations into the Flow of Human Experience. New York: Plenum.

Köhler, W. (1923). Zur Theorie der stroboskopischen Bewegung. Psychologische Forschung, 3, 397-406.

Kosslyn, S.M. & Pomerantz, J.R. (1977). Imagery, propositions, and the form of internal representations. Cognitive Psychology, 9, 52-76.

Landis, T., Lehmann, D., Mita, T. & Skrandies, W. (1984). Evoked potential correlates of figure and ground. International Journal of Psychophysiology, 1, 345-348.

Lehmann, D., Strik, W.K., Henggeler, B., König, T. & Koukkou, M. (1998). Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. visual imagery and abstract thoughts. International Journal of Psychophysiology, 29, 1-11.

Levine, J. (1983). Materialism and qualia: the explanatory gap. Pacific Philosophical Quarterly, 64, 354-361.

Lewis, C.I. (1929). Mind and the World Order. New York: C. Scribner´s Sons.

Libet, B., Gleason, C. A., Wright, E. W. & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity. Brain, 106, 623-642.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. Behavioral and Brain Sciences, 8, 529-566.

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

Logothetis, N. (1996). What is rivaling during binocular rivalry? Nature, 380, 621-624.

Lumer, E.D., Friston, K. J. & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. Science, 280, 1930-1934.

Luria, A.R. (1970). The functional organization of the brain. Scientific American, 222, 2-9.

Marcel, T. (1980). Conscious and preconscious recognition of polysemous words: Locating the selective effects of prior verbal context. In Attention and Performance (Nickerson, R.S., Ed.), pp. 435-457. Lawrence Erlbaum.

Malonek, D., Dirnagl, U., Lindauer, U., Yamada, K., Kanno, I. & Grinvald, A. (1997). Vascular imprints of neuronal activity: relationships between the dynamics of cortical blood flow, oxygenation, and volume changes following sensory stimulation. Proceedings of the National Academy of Sciences USA, 94, 14826-14831.

Mathis, W.D., & Mozer, M.C. (1996). Conscious and unconscious perception: A computational theory. In Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society, pp. 324-328. Lawrence Erlbaum.

Merikle, P.M. (1998). Psychological investigations of unconscious perception. Journal of Consciousness Studies, 5, 5-18.

Mill, J. S. (1843). A System of Logic. Books I - II. Collected Works of John Stuart Mill Vol. VII. Reprinted 1974. Toronto: Routledge and Kegan Paul.

Milner, P. M. (1974). A model for visual shape recognition. Psychological Review, 81, 521-535.

Milner, A.D. & Rugg, M.D. (1992)(Eds.). The Neuropsychology of Consciousness. London: Academic Press.

Müller, G.E. (1896). Zur Psychophysik der Gesichtsempfindungen. Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 10, 1-82.

Nagel, E. (1961). The Structure of Science. New York: Harcourt, Brace and World.

O'Brien, G., & Opie, J. (in press). A connectionist theory of phenomenal experience. Behavioral and Brain Sciences.

Paivio, A. (1976). Images, propositions, and knowledge. In Nicholas, J. M. (Ed.), Images, Perception, and Knowledge. Dordrecht: Reidel.

Penfield, W. (1958). The Excitable Cortex in Conscious Man. Liverpool: Liverpool University Press.

Plaut, D.C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. Cognitive Neuropsychology, 10, 377–500.

Pollen, D.A. (1995). Cortical areas in visual awareness. Nature, 377, 293-294.

Raffman, D. (1995). On the persistence of phenomenology. In Metzinger, T. (1995)(Ed.). Conscious Experience. Thorverton, UK: Imprint.

Reber, A. S. (1993). Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious. Oxford: Oxford University Press.

Reingold, E.M., & Merikle, P.M. (1988). Using direct and indirect measures to study perception without awareness. Perception and Psychophysics, 44, 563-575.

Reingold, E.M. & Toth, J.P. (1996). Process dissociations versus task dissociations: a controversy in progress. In Underwood, G. (Ed.), Implicit Cognition. Oxford: Oxford University Press.

Revonsuo, A., Wilenius-Emet, M., Kuusela, J. & Lehto, M. (1997). The neural generation of a unified illusion in human vision. Neuroreport, 8, 3867-3840.

Rock, I. & Gutman, D. (1981). The effect of inattention on form perception. Journal of Experimental Psychology: Human Perception and Performance, 7, 275-285.

Rugg, M.D. & Coles, M.G.H. (1995). Electrophysiology of Mind: Event-Related Brain Potentials and Cognition. Oxford: Oxford University Press.

Salzman, C.D., Murasugi, C.M., Britten, K.H. & Newsome, W.T. (1992). Microstimulation in visual area MT: effects on direction discrimination performance. Journal of Neuroscience, 12, 2331-2355.

Seidenberg, M., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, 96, 523-568.

Sejnowski, T. and Rosenberg, C. (1987). Parallel networks that learn to pronounce english text. Complex Systems, 1, 145–168.

Smart, J. J. C. (1959). Sensations and brain processes. Philosophical Review, 68, 141-156.

Smythies, J. (1997). The functional neuroanatomy of awareness: With a focus on the role of various anatomical systems in the control of intermodal attention. Consciousness and Cognition, 6, 455-481.

Shanks, D.R. & St.John, M.F. (1994). Characteristics of disociable learning systems. Behavioral and Brain Sciences, 17, 367-395.

von der Malsburg, C. & Schneider, W. (1986). A neural cocktail-party processor. Biological Cybernetics, 54, 29-40.

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

Wundt, W. (1874). Grundzüge der Physiologischen Psychologie. Leipzig: Engelmann.

Young, A.W. (1994). Neuropsychology of awareness. In Revonsuo, A. & Kamppinen, M. (Eds.), Consciousness in Philosophy and Cognitive Neuroscience. Hillsdale, NJ: Lawrence Earlbaum Associates.

Zeki, S. (1992). The visual image in mind and brain. Scientific American, 267, 68-77.

Footnotes

1 The term “brain-imaging” is preferred to “neuro-imaging” because several methods (PET, fMRI) register neural activity only indirectly and rely upon complex inferences. This point is elaborated further in subsequent sections of the text.

2 Fechner can be seen as one of the early propagators of double-aspect theory. He believed the body-mind mapping to be described by the Weber-Fechner law: $S = k \log R$.

3 Only gestalt psychologists continued research on bridging phenomenal and neural processes up to the present day. Köhler (1923) for example coined the term “psychophysical level” for those parts of the brain that have the potential to mediate conscious experience.

4 Of course there would be an empirical correlation between V1 activity and visual awareness even if it were not part of the NCC, because V1 is a relay station to the other visual areas. The problem of separating out the *necessary* conditions is elaborated below.

5 For a discussion of the relationship between reticular formation, thalamic nuclei, cortex and consciousness see Smythies, 1997.

6 Currently, the recording of continuous trajectories rather than single states is only an ideal goal. While it can be approximately achieved in physiology, it may remain impossible to go beyond discretely sampled states in phenomenology.

7 If we want to correlate consciousness, we will of course use living subjects.

8 This holds mainly for the late components which reflect endogenous cognitive processes. The interpretation of the early (exogenous) components is quite clear — they reflect purely stimulus-dependent activity in the sensory pathways.

9 EEG can of course deliver many interesting data about the spatial and temporal localization of processes, as indicated by Hardcastle (1995) for the field of consciousness.

10 This is, in a sense, the inverse of the process-purity assumption, which has been heavily criticized by a number of psychologists, especially in the field of implicit learning (see for instance, Cleeremans, 1997). In that case, the problem is to ascertain whether two measures actually probe two *completely separate* processes. In this context, the question is whether two *similar* processes differ exclusively in one aspect.

11 Direct stimulation of the brain via electrodes (e.g. Penfield 1958, Salzman et al. 1992), transcranial magnetic stimulation and neuropharmacological methods are rare exceptions.

12 This is only a brief sketch of one frontier in the controversy on subliminal perception. The so-called process-dissociation procedure (PDP) applied by Debner & Jacoby (1994) provides somewhat less controversial results by making more conservative assumptions about the relationship between the different measures involved. For more extensive reviews see e.g. Bornstein & Pittman (1992), Holender (1986) and Merikle (1998).

13 Cf. Block´s (1995) access consciousness vs. phenomenal consciousness problem.

14 Even focally attended percepts have limits to their communicability. Visual experiences, for example, are most likely of a pictorial nature (i.e. they have the format of a spatial array, Anderson 1978, Kosslyn & Pomerantz 1977, Paivio 1976). This leads to serious problems because pictorial representations need to be transformed into a propositional code in order to be communicated. Theoretically a very long serial string of symbols would be needed in order to code the information provided by a single visual image.

15 For an overview of the early controversy on the superposition vs. memory issue see James (1890, 186 ff.).

16 This is also a reason why it seems impossible to find cognitive *correlates* of consciousness. Indeed, doing so would require one to have a cognitive process running whilst recording the phenomenal trajectory, which would of course severely disturb the process. This does not mean that there can be no cognitive theories of consciousness, but only that the online correlation strategy presented here can not be applied to search for cognitive correlates of consciousness.

17 An example for the problem of a superimposed measurement is the statement: “Of course we are always conscious of the fringe, just pay attention to it!”.

18 In this paper we only explore correlations and thus make no further statements on causal relationships.

19 Even within psychology there is little interaction between neighboring research fields. Reber (1993, 109) points out this Balkanization: “(...) the two research programs [of implicit learning and implicit memory] have, unfortunately, traveled parallel courses with precious little interaction. (...) [I]f one were to construct a Venn diagram of the literature citations in these two related domains, the intersection would be very nearly the empty set (...)”.