

# How Can We Construct a Science of Consciousness?

---

**David J. Chalmers**  
**Philosophy Program**  
**Research School of Social Sciences**  
**Australian National University**

## **Abstract**

In recent years there has been an explosion of scientific work on consciousness in cognitive neuroscience, psychology, and other fields. It has become possible to think that we are moving toward a genuine scientific understanding of conscious experience. But what is the science of consciousness all about, and what form should such a science take? This chapter gives an overview of the agenda.

## **1 First-person Data and Third-person Data**

The task of a science of consciousness, as I see it, is to systematically integrate two key classes of data into a scientific framework: *third-person data*, or data about behavior and brain processes, and *first-person data*, or data about subjective experience. When a conscious system is observed from the third-person point of view, a range of specific behavioral and neural phenomena present themselves. When a conscious system is observed from the first-person point of view, a range of specific subjective phenomena present themselves. Both sorts of phenomena have the status of data for a science of consciousness.

Third-person data concern the behavior and the brain processes of conscious systems. These behavioral and neurophysiological data provide the traditional material of interest for cognitive psychology and of cognitive neuroscience. Where the science of consciousness is concerned, some particularly relevant third-person data are those having to do with the following:

- Perceptual discrimination of external stimuli
- The integration of information across sensory modalities
- Automatic and voluntary actions

- Levels of access to internally represented information
- Verbal reportability of internal states
- The differences between sleep and wakefulness

First-person data concern the subjective experiences of conscious systems. It is a datum for each of us that such experiences exist: we can gather information about them both by attending to our own experiences and by monitoring subjective verbal reports about the experiences of others. These phenomenological data provide the distinctive subject for the science of consciousness. Some central sorts of first-person data include those having to do with the following:

- Visual experiences (e.g., the experience of color and depth)
- Other perceptual experiences (e.g., auditory and tactile experience)
- Bodily experiences (e.g., pain and hunger)
- Mental imagery (e.g., recalled visual images)
- Emotional experience (e.g., happiness and anger)
- Occurrent thought (e.g., the experience of reflecting and deciding)

Both third-person data and first-person data need explanation. An example is provided by the case of musical processing. If we observe someone listening to music, relevant third-person data include those concerning the nature of the auditory stimulus, its effects on the ear and the auditory cortex of the subject, various behavioral responses by the subject, and any verbal reports the subject might produce. All of these third-person data need explanation, but they are not all that needs explanation. As anyone who has listened to music knows, there is also a distinctive quality of subjective experience associated with listening to music. A science of music that explained the various third-person data just listed but that did not explain the first-person data of musical experience would be a seriously incomplete science of music. A complete science of musical experience must explain both sorts of phenomena, preferably within an integrated framework.

## **2 Explaining the Data**

The problems of explaining third-person data associated with consciousness are sometimes called the “easy” problems of consciousness. The problem of explaining first-person data associated with consciousness is sometimes called the “hard” problem of consciousness. This is not because the problems associated with third-person data are in any sense trivial, but rather because we have a clear model for how we might go about explaining them.

To explain third-person data, we need to explain the objective functioning of a system. For example, to explain perceptual discrimination, we need to explain how a cognitive process can perform the objective function of distinguishing various different stimuli and produce appropriate responses. To explain an objective function of this sort, we specify a *mechanism* that performs the function. In the sciences of the mind, this is usually a neural or a computational mechanism. For example, in the case of perceptual discrimination, we specify the neural or computational mechanism responsible for distinguishing the relevant stimuli. In many cases we do not yet know exactly what these mechanisms are, but there seems to be no principled obstacle to finding them, and so to explaining the relevant third-person data.

This sort of explanation is common in many different areas of science. For example, in the explanation of genetic phenomena, what needed explaining was the objective function of transmitting hereditary characteristics through reproduction. Watson and Crick isolated a mechanism that could potentially perform this function: the DNA molecule, through replication of strands of the double helix. As we have come to understand how the DNA molecule performs this function, genetic phenomena have gradually come to be explained. The result is a sort of *reductive explanation*: we have explained higher-level phenomena (genetic phenomena) in terms of lower-level processes (molecular biology). One can reasonably hope that the same sort of model will apply in the sciences of the mind, at least for the explanation of the objective functioning of the cognitive system in terms of neurophysiology.

When it comes to first-person data, however, this model breaks down. The reason is that first-person data — the data of subjective experience — are not data about objective functioning. One way to see this is to note that even if one has a complete account of all the objective functions in the vicinity of consciousness — perceptual discrimination, integration, report, and so on — there may still remain a further question: Why is all this functioning associated with subjective experience? And further: Why is this functioning associated with the particular sort of subjective experience that it is in fact associated with? Merely explaining the objective functions does not answer this question.

The lesson is that *as data*, first-person data are irreducible to third-person data, and vice versa. That is, third-person data alone provide an incomplete catalogue of the data that need explaining: if we explain only third-person data, we have not explained everything. Likewise, first-person data alone are incomplete. A satisfactory science of consciousness must admit both sorts of data, and must build an explanatory connection between them.

What form might this connection take? An intermediate position holds that although there are two sorts of data, we can explain first-person data wholly in terms of material

provided by third-person data. For example, many think that we might wholly explain the phenomena of subjective experience in terms of processes in the brain. This intermediate position is very attractive, but there are reasons to be skeptical about it. I have discussed this subject at length elsewhere (Chalmers 1996). Here, I will present a simple argument that encapsulates some reasons for doubt:

1. Third-person data are data about the objective structure and dynamics of physical systems.
2. (Low-level) structure and dynamics explain only facts about (high-level) structure and dynamics.
3. Explaining structure and dynamics does not suffice to explain the first-person data.

Therefore

4. First-person data cannot be wholly explained in terms of third-person data.

Here, premise 1 captures something about the character of third-person data: it always concerns the dynamics of certain physical structures. Premise 2 says that explanations in terms of processes of this sort only explain further processes of that sort. There can be big differences between the processes, as when simple low-level structure and dynamics give rise to highly complex high-level structure and dynamics (in complex systems theory, for example), but there is no escaping the structural/dynamical circle. Premise 3 encapsulates the point, discussed earlier, that explaining structure and dynamics is only to explain objective functions, and that to explain objective functions does not suffice to explain first-person data about subjective experience. From these three premises, the conclusion 4 follows.

Of course, it does not follow that first-person data and third-person data have nothing to do with one another; there is obviously a systematic association between them. There is good reason to believe that subjective experiences are systematically correlated with brain processes and with behavior. It remains plausible that whenever a subject has an appropriate sort of brain process, he or she will have an associated sort of subjective experience. We simply need to distinguish correlation from explanation. Even if first-person data cannot be wholly explained in terms of third-person data, the two sorts of data are still strongly correlated.

It follows that a science of consciousness remains entirely possible. It is just that we should expect this science to take a nonreductive form. A science of consciousness will not reduce first-person data to third-person data, but it will articulate the systematic connections

between them. Where there is systematic covariation between two classes of data, we can expect systematic principles to underlie and explain the covariation. In the case of consciousness, we can expect systematic *bridging principles* to underlie and explain the covariation between third-person data and first-person data. A theory of consciousness will ultimately be a theory of these principles.

It should be noted that these foundational issues are controversial, and there are various alternative views. One class of views (e.g., Dennett 1991) holds that the only phenomena that need explaining are those that concern objective functioning. The most extreme version of this view says that there are no first-person data about consciousness at all. A less extreme version of this view says that all first-person data are equivalent to third-person data (e.g., about verbal reports), so that explaining these third-person data explains everything. Another class of views (e.g., Churchland 1997) accepts that first-person data need further explanation, but holds that they might be reductively explained by future neuroscience. One version of this view holds that future neuroscience could go beyond structure and dynamics in ways we cannot currently imagine. Another version holds that if we can find sufficient correlations between brain states and consciousness, that will qualify as a reductive explanation.

I have argued against these views elsewhere (e.g., Chalmers 2002). In what follows, however, I will focus on constructive projects for a science of consciousness. In this discussion I will sometimes presuppose the reasoning sketched above, but much of what I say will have application even to alternative views.

### **3 Projects for a Science of Consciousness**

If what I have said in the previous sections is correct, then a science of consciousness should take first-person data seriously, and should proceed by studying the association between first-person data and third-person data, without attempting a reduction. In fact, this is exactly what one finds in practice. The central work in the science of consciousness has always taken first-person data seriously. For example, much central work in psychophysics and perceptual psychology has been concerned with the first-person data of subjective perceptual experience. In research on unconscious perception, the first-person distinction between the presence and absence of subjective experience is crucial. And in recent years, a growing body of research has focused on the correlations between first-person data about subjective experience and third-person data about brain processes and behavior.

In what follows I will articulate what I see as some of the core projects for a science of consciousness, with illustrations drawn from existing research.

### **Project 1: Explain the Third-Person Data**

One important project for a science of consciousness is that of explaining the third-person data in the vicinity: explaining the difference between functioning in sleep and wakefulness, for example, and explaining the voluntary control of behavior. This sort of project need not engage the difficult issues relating to first-person data, but it may still provide an important component of a final theory.

One example of this sort of project is that of explaining binding in terms of neural synchrony (e.g., Crick and Koch 1991). Binding is the phenomenon whereby distinct pieces of information (about the color and shape of an object, for example), represented in different areas of the brain, are brought together for the integrated control of behavior. Some researchers theorize that a key role in this process is played by synchronized neural firing: it might be that information about a single object is represented by neurons in different areas that fire in synchrony with each other, enabling later integration. It is not yet clear whether this hypothesis is correct, but if it is correct, it will provide an important component in explaining the integration of perceptual information, which in turn is closely tied to questions about consciousness. Of course, explaining binding will not on its own explain the first-person data of consciousness, but it may help us to understand the associated processes in the brain.

Research on the hypothesis of a “global workspace” also falls into this class. Baars (1988) postulated such a workspace as a mechanism by which shared information can be made available to many different cognitive processes. More recently, other researchers (e.g., Dehaene and Changeux 2004) have investigated the potential neural basis for this mechanism, postulating a neuronal global workspace. If this hypothesis is correct, it will help to explain third-person data concerning access to information within the cognitive system, as well as data about the information made available to verbal report. Again, explaining these processes will not in itself explain the first-person data of consciousness, but it may well contribute to the project (project 4 below) of finding neural correlates of consciousness.

### **Project 2: Contrast Conscious and Unconscious Processes**

Many cognitive capacities can be exercised both consciously and unconsciously, that is, in the presence or absence of associated subjective experience. For example, the most familiar sort of perceptual processing is conscious, but there is also strong evidence of unconscious

perceptual processing (Merikle and Daneman 2000). One finds a similar contrast in the case of memory, where the now common distinction between explicit and implicit memory (Schacter and Curran 2000) can equally be seen as a distinction between conscious and unconscious memory. Explicit memory is essentially memory associated with a subjective experience of the remembered information; implicit memory is essentially memory in the absence of such a subjective experience. The same goes for the distinction between explicit and implicit learning (Reber 1996), which is in effect a distinction between learning in the presence or in the absence of relevant subjective experience.

Conscious and unconscious processes provide pairs of processes that are similar in some respects from the third-person point of view (e.g., both involve registration of perceptual stimuli) but differ from the first-person point of view (e.g., one involves subjective experience of the stimulus, one does not). Of course, there are also differences from the third-person point of view. For a start, a researcher's evidence for conscious processes usually involves a verbal report of a relevant experience, and evidence for unconscious processes usually involves a verbal report of the absence of a relevant experience. At the same time, there are less obvious differences between the behavioral capacities that go along with conscious and unconscious processes, and between the associated neural processes. These differences make for the beginning of a link between the first-person and third-person domains.

For example, evidence suggests that while unconscious perception of visually presented linguistic stimuli is possible, semantic processing of these stimuli seems limited to the level of the single word rather than complex expressions (see Greenwald 1992). By contrast, conscious perception allows for semantic processing of very complex expressions. Here, experimental results suggest a strong association between the presence or absence of subjective experience and the presence or absence of an associated functional capacity — a systematic link between first-person and third-person data. Many other links of the same sort can be found in the literature on unconscious perception, implicit memory, and implicit learning.

Likewise, there is evidence suggesting distinct neural bases for conscious and unconscious processes in perception. Appealing to an extensive body of research on visuomotor processing, Milner and Goodale (1995; see also Goodale 2004) have hypothesized that the ventral stream of visual processing subserves conscious perception of visual stimuli for the purpose of cognitive identification of stimuli, while the dorsal stream subserves unconscious processes involved in fine-grained motor capacities. If this hypothesis is correct, one can again draw a systematic link between a distinction in first-person data (presence or absence of conscious perception) and a distinction in third-person data (visual processing in

the ventral or dorsal stream). A number of related proposals have been made in research on memory and learning.

### **Project 3: Investigate the Contents of Consciousness**

Consciousness is not simply an on-off switch. Conscious experiences have a complex structure, with complex representational contents. A conscious subject usually has a manifold of perceptual experiences, bodily sensations, emotional experience, and a stream of conscious thought, among other things. Each of these elements may itself be complex. For example, a typical visual experience has an internal structure representing objects with many different colors and shapes, in varying degrees of detail. We can think of all this complexity as comprising the contents of consciousness.

The contents of consciousness have been studied throughout the history of psychology. Weber's and Fechner's pioneering work in psychophysics concentrated on specific aspects of these contents, such as the subjective brightness associated with a visual experience, and correlated it with properties of the associated stimulus. This provided a basic link between first-person data about sensory experience and third-person data about a stimulus. Later work in psychophysics and gestalt psychology took an approach of the same general sort, investigating specific features of perceptual experience and analyzing how these covary with properties of a stimulus.

This tradition continues in a significant body of contemporary research. For example, research on visual illusions often uses subjects' first-person reports (and even scientists' first-person experiences) to characterize the structure of perceptual experiences. Research on attention (Mack and Rock 1998; Treisman 2003) aims to characterize the structure of perceptual experience inside and outside the focus of attention. Other researchers investigate the contents of consciousness in the domains of mental imagery (Baars 1996), emotional experience (Kaszniak 1998), and stream of conscious thought (Pope 1978; Hurlburt 1990).

An important line of research investigates the contents of consciousness in abnormal subjects. For example, subjects with synesthesia have unusually rich sensory experiences. In a common version, letters and numbers trigger reports of extra color experiences, in addition to the standard perceived color of the stimulus. Recent research strongly suggests that these reports reflect the perceptual experiences of the subjects, and not just cognitive associations. For example, Ramachandran and Hubbard (2001) find that certain visual patterns produce a perceptual "pop-out" effect in synesthetic subjects that is not present in normal subjects. When first-person data about the experiences of abnormal subjects are combined with third-

person data about brain abnormalities in those subjects, this yields a new source of information about the association between brain and conscious experience.

#### **Project 4: Find the Neural Correlates of Consciousness**

This leads us to what is perhaps the core project of current scientific research on consciousness: the search for neural correlates of consciousness (Metzinger 2000; Crick and Koch 2004). A neural correlate of consciousness (NCC) can be characterized as a minimal neural system that is directly associated with states of consciousness. Presumably the brain as a whole is a neural system associated with states of consciousness, but not every part of the brain is associated equally with consciousness. The NCC project aims to isolate relatively limited parts of the brain (or relatively specific features of neural processing) that correlate directly with subjective experience.

It may be that there will be many different NCCs, for different aspects of conscious experience. For example, it might be that there is one neural system that is associated with being conscious as opposed to being unconscious (perhaps in the thalamus or brainstem; see e.g., Schiff 2004), another neural system associated with the specific contents of visual consciousness (perhaps in some part of the visual cortex), and other systems associated with the contents of consciousness in different modalities. But any such proposal can be seen as articulating a link between third-person data about brain processes and first-person data about subjective experience.

In recent years, by far the greatest progress has been made on the study of NCCs for visual consciousness. Milner and Goodale's work on the ventral stream provides an example of this sort of research. Another example is the research of Nikos Logothetis and colleagues on binocular rivalry in monkeys (e.g., Logothetis 1998; Leopold, Maier and Logothetis 2003). When difference stimuli are presented to the left and right eyes, subjects usually undergo alternating subjective experiences. Logothetis trained monkeys to signal such changes in their visual experience, and correlated these changes with changes in underlying neural processes. The results indicated that changes in visual experience are only weakly correlated with changes in patterns of neural firing in primary visual cortex: in this area, neural firing was more strongly correlated with the stimulus than with the experience. But changes in visual experience were strongly correlated with changes in patterns of neural firing in later visual areas, such as inferior temporal cortex. These results tend to suggest that inferior temporal cortex is a better candidate than primary visual cortex as an NCC for visual consciousness.

Of course, no single experimental result can provide conclusive evidence concerning the location of an NCC, but a large amount of evidence concerning the location of NCCs for

vision has accumulated in the last few years (Koch 2004), and one can expect much more to come. If successful, this project will provide some highly specific connections between brain processes and conscious experiences.

### **Project 5: Systematize the Connection**

To date, links between first-person data and third-person data have been studied in a somewhat piecemeal fashion. Researchers isolate correlations between specific aspects of subjective experience and certain specific brain processes or behavioral capacities, in a relatively unsystematic way. This is to be expected at the current stage of development. But we can hope that as the science develops, more systematic links will be forthcoming. In particular, we can hope to develop principles of increasing generality that link a wide range of first-person data with a correspondingly wide range of third-person data. For example, there might eventually be an account of the neural correlates of visual consciousness that will not only tell us which neural systems are associated with visual consciousness but will also yield systematic principles telling us how the specific content of a visual experience covaries with the character of neural processes in these systems.

A few principles of this sort have been suggested to date in limited domains. For example, Hobson (1997) has suggested a general principle linking certain levels of neurochemical activity with different states of consciousness in wakefulness, sleep, and dreaming. It is likely that any such proposals will be heavily revised as new evidence comes in, but one can expect that in coming decades there will be increasingly well-supported principles of this sort. The possibility of such principles holds out the tantalizing prospect that eventually, we might use them to predict features of an organism's subjective experience based on knowledge of its neurophysiology.

### **Project 6: Infer Fundamental Principles**

If the previous project succeeds, then we will have general principles connecting third-person data and first-person data. But these general principles will not yet be fundamental principles. The principles might still be quite complex, limited to specific aspects of consciousness, and limited to specific species. A science of consciousness consisting of wholly different principles for different aspects of consciousness and different species would not be entirely satisfactory. It is reasonable to hope that eventually, some unity could be discovered behind this diversity. We should at least aim to maximize the generality and the simplicity of the relevant principles wherever possible. In the ideal situation, we might hope for principles that are maximally general in their scope, applying to any conscious system

whatsoever, and applying to all aspects of conscious experience. And we might hope for principles that are relatively simple in their form, in the way that the basic laws of physics appear to be simple.

It is unreasonable to expect that we will discover principles of this sort anytime soon, and it is an open question whether we will be able to discover them at all. Currently, we have little idea what form such principles might take (Chalmers 1996 speculates that they might involve the notion of information). But if we can discover them, principles of this sort would be candidates to be fundamental principles: the building blocks of a fundamental theory of consciousness. If what I said earlier is correct, then something about the connection between first-person data and third-person data must be taken as primitive, just as we take fundamental principles in physical theories as primitive. But we can at least hope that the primitive element in our theories will be as simple and as general as possible. If, eventually, we can formulate simple and general principles of this sort, based on an inference from accumulated first-person and third-person data, I think we could be said to have an adequate scientific theory of consciousness.

What would this entail about the relationship between physical processes and consciousness? The existence of such principles is compatible with different philosophical views. One might regard the principles as laws connecting two fundamentally different domains (Descartes 1641/1996; Popper and Eccles 1977). One might regard them as laws connecting two aspects of the same thing (Lockwood 1989; Chalmers 1996). Or one might regard them as grounding an identification between properties of consciousness and physical properties (Smart 1959; Papineau 2002). Such principles could also be combined with different views of the causal relation between physical processes and consciousness (see Chalmers 2002). But for many purposes, the science of consciousness can remain neutral on these philosophical questions. One can simply regard the principles as principles of correlation, while staying neutral on their underlying causal and ontological status. This makes it possible to have a robust science of consciousness even without having a widely accepted solution to the philosophical mind/body problem.

#### **4 Obstacles to a Science of Consciousness**

The development of a science of consciousness I have presented thus far may sound remarkably straightforward. We simultaneously gather first-person data about subjective experience and third-person data about behavior and brain processes, isolate specific correlations between these data, formulate general principles governing these correlations,

and infer the underlying fundamental laws. But of course, it is not as simple as this in practice. There are a number of serious obstacles to this research agenda. The most serious obstacles concern the availability of the relevant data, in both the third-person and first-person domains. In what follows I will discuss a few of these obstacles.

### **Obstacles involving Third-Person Data**

The third-person data relevant to the science of consciousness include both behavioral and neural data. The availability of behavioral data is reasonably straightforward: here, one is constrained only by the ingenuity of the experimenter and by the limitations of experimental contexts. And in practice, researchers have accumulated a rich body of behavioral data relevant to consciousness. But the availability of neural data is much more constrained by technological limitations, and the body of neural data that has been accumulated to date is correspondingly much more limited.

In practice, the most relevant neurophysiological data come from two or three sources: brain imaging via functional magnetic resonance imaging (fMRI) and Positron emission tomography (PET) technology, single-cell recording through insertion of electrodes, and surface recording through electroencephalographs (EEG) and magnetoencephalography (MEG). Each of these technologies is useful, but each has serious limitations for the science of consciousness. EEG and MEG have well-known limitations in spatial localization. Brain imaging through fMRI and PET is better in this regard, but these methods are still spatially quite coarse-grained. Single-cell recording is spatially fine-grained but is largely limited to experimentation on nonhuman animals.

These limitations apply to all areas of cognitive neuroscience, but they are particularly pressing for the science of consciousness, because the science of consciousness relies on gathering third-person data and first-person data simultaneously. By far the most straightforward method for gathering first-person data is verbal report; but verbal report is limited to human subjects. By far the most useful third-person data are data at the level of single neurons, where one can monitor representational content that correlates with the content of consciousness (as when one monitors a neuron with a specific receptive field), but these experiments are largely limited to nonhuman subjects. As a result, it is extremely difficult to discover strong associations between first-person data and corresponding neural data with current techniques.

There have been numerous ingenious attempts to circumvent these limitations. The most well-known include Logothetis's experiments on monkeys, in which they are trained extensively to provide a substitute for a verbal report of visual consciousness by pressing a

bar. Research on blindsight in monkeys by Cowey and Stoerig (1995) has done something similar. But the very fact that researchers have to go to such great lengths in order to gather relevant neural data simply illustrates the problem. Others have performed neuron-level measurements on human surgical patients (e.g., Kreiman, Fried and Koch 2002), but there are obvious practical limitations here. Many others (e.g., Rees 2004) have tried to get as much relevant information as they can from the limited resources of brain imaging and surface recording; nevertheless, fewer deep correlations have emerged from this sort of work than from neuron-level studies.

One can hope that this is a temporary limitation imposed by current technology. If a technology is eventually developed that allows for noninvasive monitoring of neuron-level processes in human subjects, we might expect a golden age for the science of consciousness to follow.

### **Obstacles involving First-Person Data**

Where the availability of first-person data is concerned, there are a number of related obstacles that run quite deep. I will discuss three of these obstacles here.

#### **1 Privacy**

The most obvious obstacle to the gathering of first-person data concerns the privacy of such data. In most areas of science, data are intersubjectively available: they are equally accessible to a wide range of observers. But in the case of consciousness, first-person data concerning subjective experiences are directly available only to the subject having those experiences. To others, these first-person data are only indirectly available, mediated by observation of the subject's behavior or brain processes. Things would be straightforward if there were a "consciousness meter" that could be pointed at a subject, revealing his or her subjective experiences to all. But in the absence of a theory of consciousness, no such consciousness meter is available. This imposes a deep limitation on the science of consciousness, but it is not a paralyzing limitation. For a start, any subject has direct access to first-person data concerning his or her own conscious experiences. We could imagine that Robinson Crusoe on a desert island (equipped with the latest brain imaging technology) could make considerable progress toward a science of consciousness by first-person observation. More practically, each of us has indirect access to first-person data concerning others' experiences, by relying on behavioral indicators of these data.

In practice, by far the most common way of gathering data about the conscious experiences of other subjects is to rely on their verbal reports. Here, one does not treat the verbal reports just as third-person data (as a behaviorist might, limiting the datum to the fact

that a subject made a certain noise). Rather, one treats the report as a report of first-person data that are available to the subject. Just as a scientist can accumulate third-person data by accepting reports of third-person data gathered by others (rather than simply treating those reports as noises), a scientist can also gather first-person data by accepting reports of first-person data gathered by others. This is the typical attitude that researchers adopt toward experimental subjects. If there is positive reason to believe that a subject's report might be unreliable, then a researcher will suspend judgment about it. But in the absence of any such reason, researchers will take a subject's report of a conscious experience as good reason to believe that the subject is having a conscious experience of the sort that the subject is reporting.

In this way, researchers have access to a rich trove of first-person data that is made intersubjectively available. Of course, our access to this data depends on our making certain assumptions: in particular, the assumption that other subjects really are having conscious experiences, and that by and large their verbal reports reflect these conscious experiences. We cannot directly test this assumption; instead, it serves as a sort of background assumption for research in the field. But this situation is present throughout other areas of science. When physicists use perception to gather information about the external world, for example, they rely on the assumption that the external world exists, and that perception reflects the state of the external world. They cannot directly test this assumption; instead, it serves as a sort of background assumption for the whole field. Still, it seems a reasonable assumption to make, and it makes the science of physics possible. The same goes for our assumptions about the conscious experiences and verbal reports of others. These seem to be reasonable assumptions to make, and they make the science of consciousness possible.

Of course, verbal reports have some limits. Some aspects of conscious experience (e.g., the experience of music or of emotion) are very difficult to describe; in these cases we may need to develop a more refined language. And verbal reports cannot be used at all in subjects without language, such as infants and nonhuman animals. In these cases, one needs to rely on other behavioral indicators, as when Logothetis relies on a monkey's bar-pressing. These indicators require further assumptions. For example, Logothetis's work requires the assumption that monkeys are conscious, and the assumption that visual stimuli that the monkey can exploit in the voluntary control of behavior are consciously perceived.

These assumptions appear reasonable to most people, but they go beyond those required in the case of verbal report. The farther we move away from the human case, the more questionable the necessary assumptions become. For example, it would be very difficult to draw conclusions about consciousness from experiments on insects. But in any case, verbal

reports in humans, combined with behavioral indicators in primates, give researchers enough access to first-person data to enable a serious body of ongoing research.

## **2 Methodologies**

A second obstacle is posed by the fact that our methods for gathering first-person data are quite primitive compared with our methods for gathering third-person data. The latter have been refined by years of scientific practice, but the former have not received nearly as much attention. Where simple first-person data are concerned, this problem is not too pressing: there is usually no great difficulty in determining whether one is having an experience of a certain color in the center of one's visual field, for example. But where more subtle aspects of subjective experience are concerned, the issue arises quickly.

Even with a phenomenon as tangible as visual experience, the issue arises in a number of ways. Visual experiences as a whole usually have a rich and detailed structure, for example, but how can subjects investigate and characterize that detail? Most subjects have great difficulty in introspecting and reporting this detail more than superficially. Particular difficulties arise in investigating the character of consciousness outside attention. To introspect and report this structure would seem to require attending to the relevant sort of experience, which may well change the character of the experience.

Here we can expect that at least some progress can be made by developing better methods for the gathering of first-person data. It may be reasonable to pay attention to traditions where the detailed study of experience has been explored in detail. These traditions include those of western phenomenology, introspectionist psychology, and even Eastern meditative traditions. Even if one is skeptical of the theories put forward by proponents of these traditions, one might still benefit from attending to their data-gathering methods. This research strategy has been pursued most notably in the "neurophenomenology" of Francisco Varela and colleagues (Varela 1995; Lutz et al 2002), in which neurophysiological investigation is combined with phenomenological investigation in the tradition of Husserl. A number of other attempts at refining first-person methods are discussed in the papers collected in Varela and Shear (2001).

Of course, any method has limitations. Subjects' judgments about their subjective experiences are not infallible, and although training may help with this, it also introduces the danger that observations may be corrupted by theory. The introspectionist program of experimental psychology in the nineteenth century famously fell apart when different schools could not agree on the introspective data (Boring 1929). Still, our ambitions need not be as grand as those of the introspectionists. For now, we are not aiming for a perfect characterization of the structure of consciousness, but simply for a better characterization.

Furthermore, we are now in a position where we can use third-person data as a check on first-person investigation. Experimental investigation has helped us to distinguish circumstances in which first-person reports are unreliable from those in which they are more reliable (Schooler and Fiore 1997), and there is room for much more investigation of this sort in the future. So it is reasonable to hope that there can be at least a modest refinement of our methods for the reliable investigation of first-person data.

### **3 Formalisms**

A final obstacle is posed by the absence of general formalisms with which first-person data can be expressed. Formalisms are important for two purposes. First, they are needed for data gathering: it is not enough to simply know what one is experiencing, one has to write it down. Second, they are needed for theory construction: to formulate principles that connect first-person data with third-person data, we need to represent the data in a way that such principles can exploit.

The main existing formalisms for representing first-person data are quite primitive. Researchers typically rely either on simple qualitative characterizations of data (as in “an experience of red in the center of the visual field”) or on simple parameterization of data (as when color experiences are parametrized by hue, saturation, and brightness). These simple formalisms suffice for some purposes, but they are unlikely to suffice for the formulation of systematic theories.

It is not at all clear what form a proper formalism for the expression of first-person data about consciousness should take. The candidates include (1) parametric formalisms, in which various specific features of conscious experience are isolated and parametrized (as in the case of color experience above); (2) geometric and topological formalisms, in which the overall structure of an experience (such as a visual experience) is formalized in geometric or topological terms; (3) informational formalisms, in which one characterizes the informational structure of an experience, specifying it as a sort of bit-by-bit state that falls into a larger space of informational states; and (4) representational formalisms, in which one characterizes an experience by using language for the states of the world that the experience represents (one might characterize an experience as an experience as of a yellow cup, for example). All of these formalisms may have limitations, a detailed study of various alternative formalisms is likely to have significant benefits.

## 5 Conclusion

Overall, the prospects for a science of consciousness are reasonably bright. There are numerous clear projects for a science of consciousness that take first-person data seriously. One can recognize the distinctive problems that consciousness poses and still do science. Of course, there are many obstacles, and it is an open question how far we can progress. But the last 10 years have seen many advances, and the next 50 years will see many more. For now, it is reasonable to hope that we may eventually have a theory of the fundamental principles connecting physical processes to conscious experience.

## Bibliography

- Baars, B.J. 1988. *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.
- Baars, B.J. (ed.) 1996. Mental Imagery [Special issue]. *Consciousness and Cognition* 5(3).
- Boring, E.G. 1929. *A History of Experimental Psychology*. New York: Century.
- Chalmers, D.J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford, England: Oxford University Press.
- Chalmers, D.J. 2002. Consciousness and its place in nature. In *Philosophy of Mind: Classical and Contemporary Readings*. Oxford, England: Oxford University Press.
- Churchland, P.S. 1997. The hornswoggle problem. In J. Shear, ed, *Explaining Consciousness: The Hard Problem*. Cambridge, Mass.: MIT Press.
- Cowey A. and P Stoerig. 1995. Blindsight in monkeys. *Nature* 373:247-249.
- Crick, F. and C. Koch. 1990. Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences* 2:263-275.
- Crick, F. and C. Koch. 2004. A Framework for Consciousness. In Michael Gazzaniga ed., *The Cognitive Neurosciences III*, 3rd edition. Cambridge, Mass.: MIT Press.
- Dennett, D.C. 1991. *Consciousness Explained*. Boston: Little-Brown.
- Dehaene S. and J Changeux. 2004. Neural mechanisms for access to consciousness. In Michael Gazzaniga ed., *The Cognitive Neurosciences III*, 3rd edition. Cambridge, Mass.: MIT Press.
- Descartes, R. 1641/1996. *Meditations on First Philosophy*. J. Cottingham, trans.-ed. Cambridge, England: Cambridge University Press.
- Gazzaniga, M.S. ed. 2004. *The Cognitive Neurosciences III*. Cambridge, Mass.: MIT Press.
- Goodale, M.A. 2004. Perceiving the World and Grasping It: Dissociations Between Conscious and Unconscious Visual Processing. In Michael Gazzaniga ed., *The Cognitive Neurosciences III*, 3rd edition. Cambridge, Mass.: MIT Press.
- Greenwald, A.G. 1992. New Look 3: Reclaiming unconscious cognition. *American Psychologist* 47: 766-779.

- Hobson, J.A. 1997. Consciousness as a state-dependent phenomenon. In J. Cohen and J. Schooler, eds, *Scientific Approaches to Consciousness*. Mahwah, N.J.: Erlbaum.
- Hurlburt, R.T. 1990. *Sampling Normal and Schizophrenic Inner Experience*. New York: Plenum Press.
- Kaszniak, A. (ed) 1998. *Emotions, Qualia, and Consciousness*. Singapore: World Scientific.
- Koch, C. 2004. *The Quest for Consciousness: A Neuroscientific Approach*. Englewood, Colorado: Roberts.
- Kreiman, G., I. Fried, and C. Koch. 2002. Single neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings of the National Academy of Sciences USA*, 99:8378-8383.
- Leopold, D.A., A. Maier, and N.K. Logothetis. 2003. Measuring subjective visual perception in the nonhuman primate. *Journal of Consciousness Studies*. 10 (9-10): 115-130
- Lockwood, M. 1989. *Mind, Brain, and the Quantum*. London: Blackwell.
- Logothetis, N.K. 1998. Single units and conscious vision. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 353:1801-1818.
- Lutz, A., J. Lachaux, J. Matrinerie, and F. & Varela. 2002. Guiding the study of brain dynamics by using first-person data: Synchrony patterns correlate with ongoing conscious states during a simple visual task. *Proceedings of the National Academy of Science USA* 99:1586-91.
- Mack, A. and I. Rock,. 1998. *Inattentional Blindness*. Cambridge, Mass.: MIT Press.
- Merikle, P.M. and M. Daneman. 2000. Conscious vs. unconscious perception. In (M.S. Gazzaniga, ed.) *The New Cognitive Neurosciences*. Cambridge, Mass.: MIT Press.
- Metzinger, T. 2000. *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, Mass.: MIT Press.
- Milner D.A. and M.A. Goodale. 1995. *The Visual Brain in Action*. Oxford, England: Oxford University Press.
- Papineau, D. 2002. *Thinking about Consciousness*. Oxford University Press.
- Pope, K.S. 1978. *The Stream of Consciousness: Scientific Explorations into the Flow of Human Experience*. New York: Plenum Press.
- Popper, K. and J. Eccles. 1977. *The Self and Its Brain: An Argument for Interactionism*. New York: Springer-Verlag.
- Ramachandran, V.S. and E.M. Hubbard. 2001. Psychophysical investigations in to the neural basis of synaesthesia. *Proceedings of the Royal Society London* 268: 979-983.
- Reber, A. 1996. *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford, England: Oxford University Press.
- Rees G. 2004. Neural Correlates of Visual Consciousness in Humans. In Michael Gazzaniga ed., *The Cognitive Neurosciences III*, 3rd edition. Cambridge, Mass.: MIT Press.
- Schacter, D.L. and T. Curran. 2000. Memory without remembering and remembering without memory: Implicit and false memories. In M.S. Gazzaniga, ed., *The New Cognitive Neurosciences*. Cambridge, Mass.: MIT Press.
- Schiff, N.D. 2004. The Neurology of Impaired Consciousness: Challenges for Cognitive Neuroscience. In Michael Gazzaniga ed., *The Cognitive Neurosciences III*, 3rd edition. Cambridge, Mass.: MIT Press.

- Schooler, J.W. and S.M. Fiore. 1997. Consciousness and the limits of language. In J. Cohen and J. Schooler, eds., *Scientific Approaches to Consciousness*. Hillsdale, N.J.: Erlbaum.
- Smart, J.J.C. 1959. Sensations and brain processes. *Philosophical Review* 68:141-56.
- Treisman, A. 2003. Consciousness and perceptual binding. In A. Cleeremans, ed., *The Unity of Consciousness: Binding, Integration, Dissociation*. Oxford England: Oxford University Press.
- Varela, F. 1997. Neurophenomenology: A methodological remedy for the “hard problem”. In J. Shear, ed., *Explaining Consciousness: The Hard Problem*. Cambridge, Mass.: MIT Press.
- Varela, F. and J. Shear J. 2001. *The View from Within: First-Person Methodologies*. Exeter, England: Imprint Academic.