

Bernard J. Baars

How Brain Reveals Mind

Neural Studies Support the Fundamental Role of Conscious Experience

Abstract: *In the last decade, careful studies of the living brain have opened the way for human consciousness to return to the heights it held before the behavioristic coup of 1913. This is illustrated by seven cases: (1) the discovery of widespread brain activation during conscious perception; (2) high levels of regional brain metabolism in the resting state of consciousness, dropping drastically in unconscious states; (3) the brain correlates of inner speech; (4) visual imagery; (5) fringe consciousness; (6) executive functions of the self; and (7) volition. Other papers in this issue expand on many of these points. (Roepstorff; Leopold & Logothetis; Bærentsen; Haggard; Hohwy & Frith).*

In the past, evidence based on subjective reports was often neglected (e.g., Ericsson, this issue). It is still true that brain evidence has greater credibility than subjective reports, no matter how reliable. What is new is increasing convergence between subjective experiences and brain observations. For that reason it is no longer rare to see the word 'consciousness' and 'subjectivity' in major science journals.

No one so far has discovered a gulf dividing mind and brain. On the contrary, the new evidence supports the central role of consciousness as it was regarded over more than two millenia of written thought.

In a sense this was predictable. Nature is full of unexpected convergences — between fruit fly genes and the human body, between the arc of a tennis ball and the orbit of Mars, and between consciousness and the brain. These convergences show once again the remarkable unity of the observable universe.

I: Introduction

Controversy has marked the scientific study of consciousness for more than a century. Some disputes have been about evidence and many more about

Correspondence:
B.J. Baars, 3616 Chestnut St. Apt. 3, Lafayette, Calif. 94549, USA. *Email: baars@nsi.edu*

Journal of Consciousness Studies, **10**, No. ??, 2003, pp. ??–??

philosophy. But scientific controversies are rarely resolved by philosophical reasoning alone.

Many earlier scientific arguments were equally troubling in their time and are now settled. There is general agreement that the earth does go around the sun, that Newton was pretty much right about gravity, and that life is based on carbon chemistry. These questions caused decades of heated debate. Yet somehow consensus emerged. But how? The answer is of course that the path was inductive and empirical, by way of gradually emerging findings and ideas, each one unexpected in its time.

Crucially, the answers that finally emerged were almost always quite different from those that were first envisaged. Early questions are almost always poorly posed, untestable, or based on wrong assumptions.

Consciousness seems to be following a similar path. For 200 years we have collected reliable evidence about sensory consciousness. The decibel scale for sound comes straight from the psychophysics of the 1820s. The color pixels on our screens originated in Newton's experiments in his rooms at Cambridge, showing how a glass prism refracted sunlight into the colors of the rainbow. Both were basic discoveries about conscious perception, because perceived colors and sounds are not in the world. All their perceived properties are in the head. Thus we have a great deal of evidence about consciousness already.

Yet we are still not sure 'what consciousness is' in a deep theoretical sense. This follows the normal history of scientific concepts. Physics in the nineteenth century possessed a wealth of facts about temperature without knowing what temperature really *was*. A workable answer only came with thermodynamic theory, after 1900. In consciousness science our theories also lag behind the evidence, and until a settled theory emerges we will not know what consciousness 'really is'.

The last two decades have shown that conscious brain events are often matched by similar unconscious ones. For example, the meaning of 'this phrase' is unconscious for the reader a fraction of a second before it becomes conscious, even though visual word areas in cortex are already active (Dehaene *et al.*, 2001). After fading from consciousness, the words remain in memory unconsciously for at least ten seconds, as we can tell from sensitive memory tests. The idea that we can often compare conscious and unconscious brain events has profoundly shaped recent science.¹ For the first time we can treat consciousness as an experimental variable, just like any other scientific concept (see Baars, 1997b; Baars, Banks & Newman, 2003). Experimental comparisons make it possible to ask the

[1] Distinctive and widespread brain activity associated with consciousness has actually been known since 1929, when Hans Berger discovered that the entire brain's electrical activity changes visibly when we wake up. Waking activity reveals electrical voltages that are fast, irregular, and low in amplitude. Deep sleep — the least conscious state of the daily cycle — is marked by voltages that are slow, regular, and much higher in amplitude. Other kinds of unconscious states, such as general anaesthesia, epileptic seizures and coma/vegetative states also show massive slow-wave, high-peak activity. Recent studies indicate that frontoparietal regions are markedly lower in metabolism during unconscious states than in waking control conditions (Baars, Ramsoy & Laureys, under review). Thus we already know that conscious and unconscious states involve distinctive brain-wide patterns of activity.

question, ‘What difference does consciousness make in the brain? What does it do for us?’

We can now see much of the living brain at work. Research articles routinely describe brain events reflecting sensory perception, language, mood, mental effort, voluntary control, and much more. In the last decade we have learned an immense amount as a result. (See the classic work of Leopold and Logothetis, Frith and others, this issue).

Obviously the limits of these methods continue to be debated. Yet many scholars now believe that current science is bringing us back to traditional ideas of consciousness, volition, and self that were expelled by behaviorists in the last century. William James would be very much at home with today’s ideas. In that sense, studies of the brain reveal the mind with unprecedented clarity and credibility.

II: The Return of Consciousness, Volition and Self

In the year 2000, 1,400 biomedical articles used the word ‘consciousness.’ In 1950, at the height of behaviourism, there were only five (Baars, 2002a). But in 1950 people were just as conscious as they are today. What has happened is a gradual change in science itself: the slow lifting of the behaviouristic taboo against human subjectivity, which reigned supreme from approximately 1920 to the 1980s.

Following are seven aspects of subjectivity that are now supported by brain observations. Most were beautifully explored in William James’ great book, more than 100 years ago (1890).

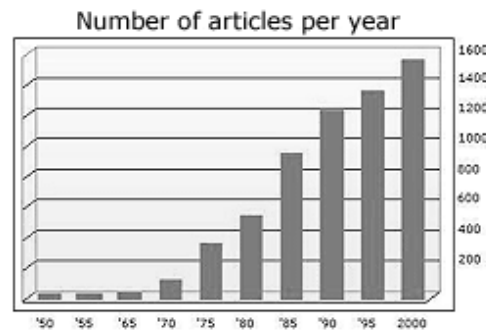


Figure 1. Results of a biomedical database search for the word ‘consciousness’ at five-year intervals from 1950 to 2000. The quantitative results should be interpreted with caution, but they are generally in accord with other sources of information. The behavioristic taboo against the word ‘consciousness’ is clearly no longer dominant. (From Baars, 2002a. Graphics by Thomas Ramsay, www.sci-con.org).

1. Widespread brain activation during sensory consciousness

How important is consciousness in the brain? Could it be largely unrelated to the working of the nervous system? In the nineteenth century T.H. Huxley suggested that consciousness might be a mere by-product of the brain, with no effect upon it, just as ‘the steam-whistle which accompanies the work of a locomotive is without influence upon its machinery’ (quoted in James, 1890).

This question now has relevant evidence. For instance, the conventional view of perception is that it involves the brain’s detection of sensory scenes, objects and events. And so it does. But there is a great difference between the way the

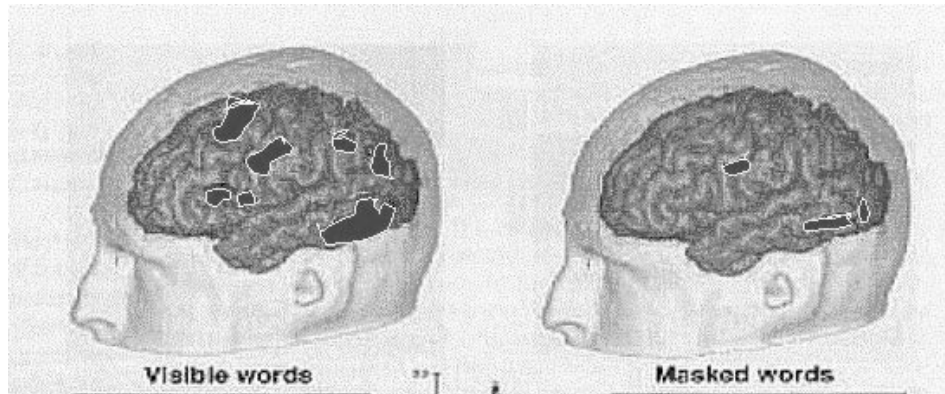


Figure 2. Conscious *versus* masked visual words. Dehaene and co-workers (2001) compared fMRI activation to visible (left) and adjacently-masked (right) words. Red regions indicate peak averaged hemodynamic activity. Words that were reported as conscious show a 12-fold increase in activation in word recognition regions (fusiform gyrus) relative to masked trials, peaking at 200 ms. Parietal and prefrontal regions show activation only in the conscious and not in the unconscious trials. The scans show group activations in the left hemisphere as seen through a translucent three-dimensional reconstruction of the skull and brain of one of the participants. In these transparent views, the deep activation in fusiform, parietal and mesial frontal cortex appear through the overlying lateral cortices. These results are consistent with current theories of conscious sensory perception (see footnote 2). (With permission from the authors).

brain treats conscious and unconscious sensory perception. When a stimulus is presented unconsciously it activates areas in cortex involved in analysing colours, sounds, faces and the like. But when the identical stimulus is shown consciously, it also recruits regions far beyond the sensory cortex.

Here are three examples out of dozens that have been published (reviewed in Baars, 2002b). Dehaene *et al* (2001) compared conscious words on a screen to the same words when they were masked by a pattern presented immediately before and afterwards. Masked words are unconscious, but they are not physically blocked from entering the eyes. They activate retinal receptors with exactly the same energy pattern as conscious words, and evoke neuronal firing in the visual pathway well into cortex. Thus there is a close similarity between the two stimuli, and it makes sense to compare their brain responses.

Figure 2 shows the results. Unconscious words activated vision and word recognition areas, which analyze such things as stimulus location, shape, colour, and word identity. The identical words, when conscious, triggered 12 times more activity in word recognition regions of the visual cortex. In addition, these words evoked a great deal of additional activity in parietal and frontal cortex. It is as if the stimulus activity in the conscious case is widely distributed from visual regions to other areas in the brain (Baars, 1988; 1997a; 2002b). This is in fact predicted by current theory.²

Now consider a very different example, the perception of pain from neurons in the heart region of the body. Rosen *et al* (1996) compared two kinds of neural responses to reduced oxygen supply to the heart. In one group, patients reports intense conscious pain (angina pectoris). The comparison group showed

unconscious cardiac anoxia (silent ischemia). Figure 2 shows the results. Again, conscious pain involves very wide activity spreading from the sensory regions to the rest of cortex. By comparison, unconscious ‘pain’ activity barely reached cortex. As before, similar sensory stimuli have very different effects in the brain, and the difference corresponds to consciousness of the event.

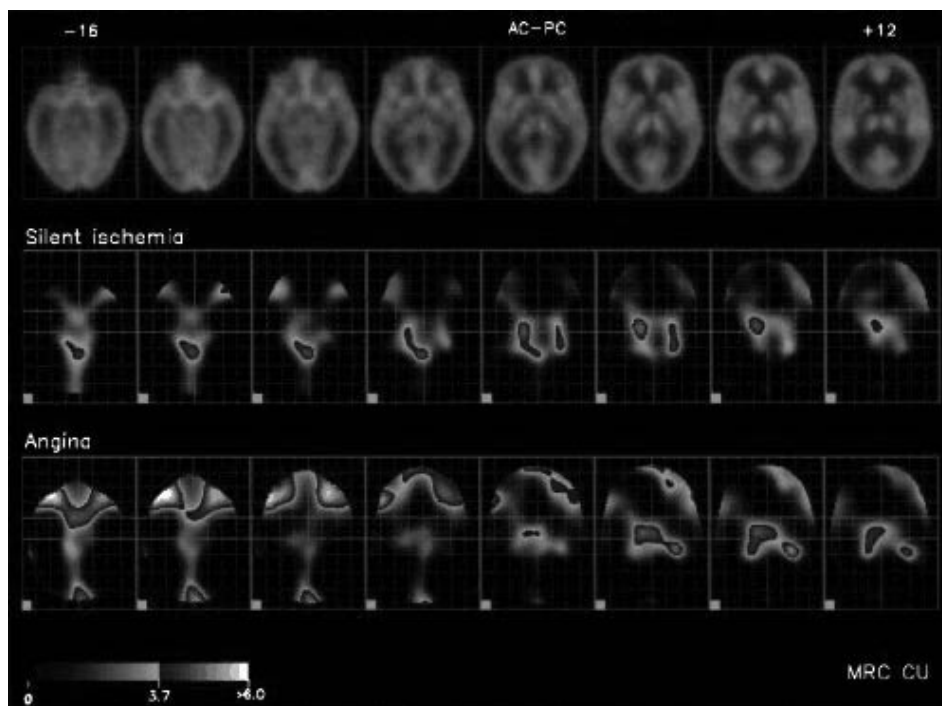


Figure 3. Conscious pain *versus* matched unconscious stimulation. PET was used to measure regional cerebral blood flow changes as an index of neuronal activation during painful and silent myocardial ischemia, induced by intravenous dobutamine. The upper row shows control scans during placebo infusion. The second row shows silent ischemia, with PET activation occurring mostly in bilateral thalami. The bottom row shows significantly higher and more widespread cortical activation during angina (painful ischemia), particularly in the bilateral frontal cortex. Coronal brain slices are viewed from the top, with the eyes looking upward. The leftmost slice is 16 mm below the anterior cingulate (−16). Each succeeding scan to its right is 4 mm higher. (With permission of the authors).

- [2] Global workspace theory suggests that consciousness enables brainwide access between otherwise separate functional networks (Baars, 1988, 1997a, 2002b). It is only one of several ‘globalist’ theories, which interpret consciousness in terms of widespread interactions between many regions in the core brain, regions that are believed to be needed for conscious functions like perceptual continuity, inner speech, imagery, learning and self functions. (Edelman & Tononi, 2000). An increasing number of authors now approach consciousness in terms of a neuronal global workspace capacity. As Daniel C. Dennett recently wrote, ‘Theorists are converging from quite different quarters on a version of the global neuronal workspace model of consciousness’ (2001, p. 42). Global Workspace theory has been implemented in a large-scale computer model by Franklin and colleagues. (Baars & Franklin, 2003). It predicts that the specific pattern of distributed activity associated with conscious events should vary from task to task. What is ‘global’ in such models is the capacity to recruit virtually any set of neuronal resources over a wide range of tasks and conditions.

A recent study by Kreiman et al (2002) adds to these results. Using human patients with electrodes implanted deep in the temporal lobe, to find the source of epileptic seizures, they presented two images to the two eyes. The presentation method, called flash suppression, causes one image to be conscious at a time, though the unconscious image still activates visual cortex. Two-thirds of the deep temporal neurons sampled followed the conscious image; none responded to the unconscious one, even though we know that in visual cortex the unconscious image is represented. Even more important, the temporal regions of the brain do not involve sensory consciousness at all, but rather unconscious aspects of memory and emotion. This shows that conscious events mobilize *unconscious* brain activity outside of the sensory cortex.

Is consciousness a side effect of the brain, like the steam whistle of a locomotive? It seems less and less likely — unless we assume that there is no biological function for the widespread brain activity evoked by the sensory input that is experienced as conscious. The sheer amount of consciousness-related activity suggests a central role for such brain events.

2. High levels of regional brain metabolism in the resting state of consciousness

Our waking hours are taken up with a flow of conscious thoughts, percepts, images, impulses, desires and worries, exertions of will, emotional feelings, inner speech, and intuitions. Yet in spite of sound and reliable self-reported evidence, the stream of thought is rarely studied even today (see Singer, 1993).

That may now begin to change. A revealing brain imaging study was published in 2001 by a French group, Mazoyer and colleagues. These scientists did an unusual thing. Most experiments study the brain during very specific cognitive tasks. Mazoyer and his group turned this around. They asked what would happen if people are simply asked to do nothing, compared to nine standard tasks? Sixty-one subjects in nine PET studies were asked to lie down in the darkened apparatus, to ‘keep their eyes closed, relax, refrain from moving, and avoid any structured mental activity such as counting, rehearsing, etc.’ What would brain activity and introspective reports be like?

The biggest surprise came when PET scans consistently showed *more* brain activity in the ‘rest’ condition than in any of nine specific cognitive tasks. Those tasks included imagery, word perception and mental arithmetic. This is quite extraordinary: Whatever subjects were doing while lying in the dark, eyes closed, and trying to relax, requires more metabolic fuel than the standard tasks.

What were they doing? We have two sources of evidence, self-reports and brain activity. Reports from the subjects showed many spontaneous ‘autobiographical reminiscences, recent and ancient, consisting of familiar faces, scenes, dialogs, stories, melodies, etc.’ Four out of five people reported mental imagery, and three out of four inner speech. About half described mild discomfort from the arm catheter that is required for PET.

The second source of evidence was the PET scans themselves. Mazoyer *et al* found, in technical terms, a network of active brain areas during the rest condition, including ‘the bilateral parieto-occipital junction, precuneus, posterior cingulate, and left orbitofrontal cortex’. These areas are involved in:

- (1) immediate memory,
- (2) control of visual imagery and inner speech
- (3) recall of conscious memories
- (4) executive functions, and
- (5) emotions.

It seems that the spontaneous stream of thought does more important things than mental arithmetic. It is more emotionally driven, more self-involved, more apt to dwell on significant past events, and more likely to attend to plans for the future. It also tends to dwell more on interpersonal conflict (Singer, 1993). By comparison, being told to do mental arithmetic seems only remotely related to the subject’s life concerns. Thus the spontaneous stream seems to *do* things for us, even when we aren’t sure what it is doing. That may be why the brain is more active during ‘rest’ than in arbitrary experimental tasks.

Notice how closely the PET scans tracked introspective reports. There is no gulf in this study between objective and subjective evidence. Each is used to interpret the other. Each is revealing in its own way, and together they strengthen our overall understanding.

Does this only prove that subjective reports have been right all along? I believe it does, but it shows more. For one thing, we can now look for unconscious brain activities, the ones people *cannot* report accurately. It is notoriously difficult to introspect on one’s own motives, personality, syntax, interpersonal feelings, details of memory processes and automatic habits. These unconscious influences also drive the stream of thought (Luborsky & Crits-Christoph, 1998). Studies like this may also allow us to explore consciousness in people suffering from brain disorders, in babies, and in other animals.

Most important, perhaps, this study shows no gulf between mind and brain. They emerge as two sides of the same mountain.

3. *Inner speech after a century of dispute*

Human beings talk to themselves every moment of the waking day. Most readers of this sentence are doing it now. It becomes a little clearer with difficult-to-say words, like ‘infundibulum’ or ‘methylparabine’. In fact, we talk to ourselves during dreams, and there is even evidence for inner speech during deep sleep, the most unconscious state we normally encounter (Hobson *et al.*, 2000). Overt speech takes up perhaps a tenth of the waking day; but inner speech goes on all the time. According to careful studies, we devote most of our spontaneous inner speech to ‘current concerns’ (Singer, 1993). Novelists and poets would not find this surprising. But in science, the debate has raged for almost a century whether the introspective evidence was true. Do people really talk to themselves? It is

difficult to convey the endless arguments this simple question generated, years after reliable self-reports showed the answer.

Ten years ago the last justification for skepticism quietly crumbled. Paulesu *et al* (1993) found that mental rehearsal of words triggered high activity near the two classic speech regions of the left hemisphere, Broca's and Wernicke's areas, when compared to no mental rehearsal in the same subjects. In 1861 the French neurologist Paul Broca discovered a patient with a stab wound in the brain just in front of the central fissure. It left him unable to speak, but still understanding speech. Some years later, Carl Wernicke, a German doctor, found the complementary area in another patient, a few centimeters behind Broca's area, where local damage made it impossible to understand speech, though the patient could still articulate. These two regions have been implicated thousands of times in speech input and output, but no one could study them in the living brain. Post-mortems were the sole source of brain evidence.

Today we can see with a simple brain scan how inner speech mirrors outer speech, using the same brain regions. The much-debated gulf between mind and brain may be shrinking a little bit.

4. Are visual images like 'faint copies' of sensations?

Mental imagery has a similar history. The reader can experience imagery simply by looking at a light, closing his or her eyes and noticing an afterimage for a few seconds. Or you can imagine the outer door of your house, and ask yourself on which side the doorknob can be 'seen.' It was Aristotle who first suggested that visual images were 'faint copies' of sensations. The American psychologist C.W. Perky showed this elegantly early in the twentieth century, when she demonstrated that people may confuse faint visual pictures with their own mental images (Perky, 1910). Perky used a back-projected screen with a very faint picture of a banana. People were asked to look at the screen and to imagine an object, a fruit perhaps. As Aristotle might have expected, subjects often thought they were *imagining* a banana that was actually faintly in front of them. If we can confuse our own mental images with real visual objects, imagery and vision would seem to be similar. The Perky Effect has now been replicated a number of times (Segal & Fusella, 1970).

This classic experiment might have settled the question once and for all. But behaviorism was on the upswing in the years after 1910 and consciousness was becoming taboo. Many decades later, psychologists began once again to perform experiments on mental imagery, showing that it was quite reliable. But something very odd happened: Imagery was almost never discussed as *conscious*. (Baars, 1988; 1996). Yet we know that people can report mental images, and accurate report is the standard operational index of consciousness.

Brain imaging may be coming to the rescue again. It has shown routinely that the brain areas needed for visual perception are also used in mental images. This is especially true for vivid images — the ones that are most clearly conscious. (Kosslyn, *et al*, 2001; Ishai & Sagi, 1995). Current brain evidence converges so

clearly with psychological studies that the most radical skeptics may eventually recognize the facts.

5. *Fringe consciousness and the tip-of-the-tongue experience*

William James is the fountainhead of consciousness science in the English language, summarizing a century of remarkable studies that were long denounced as unscientific. Many of these have since been verified (James 1890/1983; Baars, Banks & Newman, 2003). It was James who made a surprising observation about ‘fringe’ or ‘vague’ experiences in the stream of thought. He strongly argued that fringe experiences are much more common than usually thought, and that they are an essential part of human consciousness. In a famous passage he wrote:

Suppose we try to recall a forgotten name. The state of our consciousness is peculiar. There is a gap therein; but no mere gap. It is a gap that is intensely active. A sort of a wraith of the name is in it, beckoning us in a given direction, making us at moments tingle with the sense of our closeness, and then letting us sink back without the longed-for term. If wrong names are proposed to us, this singularly definite gap acts immediately so as to negate them. They do not fit into its mold (James, 1890).

The tip-of-the-tongue (TOT) state is an *intention to recall* the missing word, a mental state that lacks perceived or imagined qualities like color, sound, or taste; it has no clear boundaries in space and time, and no contrast between figure and ground. All expectations and intentions seem to be ‘non-sensory’ in just this sense. To show the power of expectations we need only interrupt some dense flow of predictable experience, for example a printed _____ like this one. Spontaneously we tend to fill in words like ‘sentence’, anything that fits the meaning and grammar of the sentence. We can see the same effect by interrupting a joke just before the punchline, or a musical tune just before returning to the theme. Clever composers continuously play with our expectations. But the sense of seeking a missing word is not a mental image like your visual image of the cover of this magazine, or the inner sound of *these words*.

James thought that ‘fringe’ states like this comprise perhaps a third of our mental life; but some of us now believe that they shape *all* of our experiences without exception. The tip-of-the-tongue state provides a nice case to study, because it draws out a colorless expectation over many seconds.

It is easy enough to induce tip-of-the-tongue states. All we need to do is to ask someone for a difficult word – what do you call the flying dinosaurs? What is the capital of Estonia? Hundreds of studies have been published on fringe experiences. But there is a curious omission in this scientific literature. There is almost no exploration of the *subjective experience* of the fringe. Even though everyone quotes James’ phenomenological description, it has gone virtually untouched.

Consider a very simple hypothesis. Fringe experiences like the tip-of-the-tongue state lack perceptual qualities, like colour, location or timbre. One reasonable guess, therefore, is that it is a non-sensory state. The human cerebral cortex — the source of conscious contents — can be divided into two great portions. The posterior part is largely sensory. But the frontal third has hardly any sensory

areas at all — it deals with motor control, working memory, executive functions, impulse control, aspects of emotion, and the like. Could it be that fringe experiences simply mobilize the front of the cerebrum, and that object-like conscious experiences involve sensory regions in the posterior half?

We now have the first brain imaging study of the tip-of-the-tongue experience (Maril *et al.*, 2001). During TOT states compared to successful word retrieval, Maril and coauthors found, the lateral prefrontal cortex and anterior cingulate are highly activated. This is exactly what we would expect, given the nonsensory ‘fringe’ nature of that state.

Notice how this study gives us a way to think about the phenomenology of the fringe. No longer do we have to avoid this pervasive aspect of reality. We can ask questions about it, test reasonable hypotheses, and hope to learn more. The brain seems to reveal the mind in remarkable detail.

6. *Executive functions of the self—another taboo bites the dust*

For half a century sophisticated scientists and philosophers rejected the idea of ‘self’ as illogical and unscientific. Yet these people were as self-serving as other humans. When criticized they presumably experienced anger and defended themselves. When threatened by professional politics they probably sometimes used the familiar ego defenses we see in the news every day — rationalization, intellectualization of emotion, denial, displacement of blame, the lot. When asked to pass the sugar, they exercised executive control over their skeletal muscles, using prefrontal cortex. The same region of cortex is used to suppress impulsive actions, emotions and appetites, the ones we experience as ‘not the kind of thing “I” would do’. All these standard aspects of self regulate all of our lives. Yet when asked about ‘self,’ philosophers and scientists would denounce it as absurd. Some still do.

Gilbert Ryle is generally cited for his critique of self as a ‘Ghost in the machine’ (1949). Ryle attacked what he claimed was a widespread misconception of self as a homunculus, a little man in the brain. He pointed out that such an observing homunculus could not *explain* self, since it only required another little observer inside itself to make sense of its own experiences, and so on ad infinitum. This critique is still popular.

Most scientific conceptions of self, however, do not involve a homunculus at all (e.g., Baars, 1988; Hilgard, 1977; Dennett, 2001). Ideas of self in neurology, cognitive neuroscience, social psychology, psychodynamics, and personality theory are not homunculi. But Ryle became famous at a time when the whole vocabulary of human psychology was being erased. Self was one more idea for the rubbish heap.

Daniel C. Dennett, who studied with Ryle, has answered the homunculus critique by pointing out that ‘an analysis of the Subject (is) a necessary component of any serious theory of consciousness.’ (2001, p. 221) As long as the self is accounted for in terms that are not just other selves, Dennett sees no problem. In that sense he is in agreement with scientists who have studied ego functions for the last century.

Brain imaging has also come to the rescue. Neurologists have long known that damage to the front edge of the cortex can result in profound alterations of personality. Most famous is the classic case of Phineas Gage, the nineteenth-century American railroad foreman who was injured by a tamping spike driven explosively through the orbit of his left eye and frontal brain. As Damasio *et al.* (1994) point out, Gage's doctors were surprised by his robust physical health. But his personality changed. From a model worker Gage turned into an angry drifter, unable to plan for his future or to control impulsive feelings. Such personality changes are common in frontal lobe damage.

Today we can see brain images of prefrontal damage. Even better, we can watch frontal executive regions doing their job in normal, healthy people. In the last few years the literature on the topic has grown so quickly that a search for 'executive prefrontal' shows 289 empirical articles. They include such 'self'-related topics as voluntary decision-making, emotion and motivation, criminal behavior, schizophrenia, and much more.

Self functions engage more than just frontal cortex. But for humans, that part of the brain is involved in personality traits such as persistence and self-control, postponing immediate gratification, moral commitment, and even conscious emotional feelings.

Today the question of the self is back in full force.

7. *The rediscovered problem of volition — not free will*

In the nineteenth century psychology focused on three topics: cognition, volition, and emotion (Hilgard, 1977). To volition William James devoted a remarkable, fact-filled chapter. James discussed such things as people with 'inhibited will' (now called dysexecutive syndrome), 'explosive will' (now called impulse-control disorder), and the like. He also provided a subtle and beautiful theory of voluntary control, called the ideomotor theory, which is still highly plausible (James, 1890/1983; Baars, 1988; 1993; Franklin, 2002). James' ideas were of course erased during the decades of behavioristic dominance.

In the late twentieth century the word 'cognition' and even 'emotion' returned to science, but 'volition' is still a bit lost, in spite of massive evidence (Baars, 1988; 1992; 1993). Indeed, many scientists still claim — voluntarily, of course — not to believe in volition at all (e.g. Wegner, 2001).

The free will debate continues to muddy these waters. Voluntary actions are usually accompanied by a sense of free will. People value that sense of freedom, and may even kill or die for it. It is of fundamental importance. But the *metaphysical* question of free will is a very different matter. It is hard even to state coherently. To conflate volition with free will is to make it impossible to study. For scientists, a useful working assumption is that the feeling of freedom is a deeply held human experience, but that our actions are determined by some causal network like any other. That allows us to explore it empirically.

In sum, the conscious *sense* of free will is real and important, but metaphysical free will is a recipe for endless, useless debate.

III: Brain Studies of Volition

Volition can be studied as an empirical question, as fundamental as consciousness itself. The evidence for both is immense. Major regions of the nervous system are dedicated to voluntary control, especially in prefrontal cortex. Other divisions of the nervous system are independent of voluntary control, such as the autonomic nervous system (so called because it is normally autonomous — free from voluntary control). Billions of neurons control nonvoluntary aspects of action in the cerebellum and basal ganglia. Clinical neurologists, who deal with brain damaged patients every day, have never erased the term ‘voluntary’ from their vocabulary. The evidence is just too obvious.

Recent brain studies show robust effects of volition. For example, one can show a striking difference in the brain between voluntary swallowing and spontaneous swallowing in humans. Like other vital functions (breathing, for example), swallowing has long been believed to have a dual brain control system. We can do it either voluntarily or spontaneously. If we are trying to swallow a big pill, voluntary effort may be needed. That difference is now seen in brain scans. (Kern, Jaradeh *et al.*, 2001).

There are many other ‘double dissociations’ between voluntary and involuntary brain events, showing that they do not have the same neuronal basis. For instance, Iwase *et al.* (2002) showed that voluntary smiles show quite different cortical activity compared to spontaneous (unintended) smiles. Yet the behaviour of smiling is similar in both cases.

The famous knee jerk reflex is experienced as involuntary. One can try to imitate it voluntarily with the same spring-loaded dynamic acceleration. But it feels different: People may be surprised by their own reflexes; we are rarely surprised by our voluntary decisions and actions.

These facts mean that volition can be treated as a variable, just like consciousness. (Baars, 1988; 1992; 1997a; 2002; Baars, Banks & Newman, 2003). That is, we can easily compare similar voluntary and involuntary actions. Scientifically this is the only way to study any topic.

Volition is not one problem, but a set of empirical questions. Recent brain studies show progress in understanding the voluntary sense of effort, for example. This is now believed to be the common factor in fluid intelligence tests, the ability to deal flexibly with difficult problems. In a remarkable meta-analysis of many different brain studies, Duncan & Owen (2000) have shown that the same part of the brain is activated with mental effort, regardless of the specific task involved. Since mental effort is a fringe experience (without sensory consciousness), it is not surprising that it activates the dorsolateral prefrontal cortex.

It is impossible to review the relevant studies here. But the sheer number of brain experiments on volition tells the story. Like consciousness and self, the question of voluntary control is back on the frontier.

IV: The Myth of Introspectionism

For many years, the study of conscious experience was derided as ‘introspectionism’. Psychology students were taught that introspection had been tried in the nineteenth century and had failed. In fact, that critique was part of the behaviouristic myth of origins propagated in the early twentieth century (Blumenthal, 1979; Danziger, 1979; Hilgard, 1977; Baars, 1986; 2003; in press). Major figures in earlier psychology, like Wilhelm Wundt, vehemently criticized introspective methods as empirically unreliable. Yet Wundt was labeled the ultimate introspectionist, a cruel irony. The supposed failure of consciousness science was meant to justify behaviouristic prohibitions against subjective experience, even when tested and validated objectively. The result was poor science, wrong history, and dehumanized education.

In fact, the nineteenth century discovered the majority of mental phenomena we know today. Human consciousness has been studied scientifically with great success at least since the 1820s. Yet this misleading history continues to be taught.

Reports of reliable conscious experiences are much like other empirical indices in the history of science. They have pros and cons, and are continuously being improved methodologically. New brain-based measures also are emerging. Such a standard scientific effort needs no defense, as long as experiential reports are carefully collected under optimal conditions, and verified in a variety of ways. They are extraordinarily useful. The evidence they provide is often accurate, cumulative, empirical, theoretically elegant, and humanly significant.

V: Where Are We Going? Toward a Neo-Jamesian Science

Science and its technological offspring are unpredictable. But if the evidence arrayed in this paper is roughly true, we may be on the verge of a historic understanding of our own minds, both the conscious and unconscious sides.

Human beings, it is said, cannot stand too much reality. We are our own worst observers. Can we tolerate an accurate understanding of our own minds? It is not at all obvious, and we can imagine bad as well as good ripple effects. Hot debates will no doubt arise about privacy and personal control in a world where our consciousness is no longer walled off from public knowledge. Nightmare scenarios can be found in any science fiction library.

I prefer to hope for a more positive future, a neo-Jamesian science in which scientific knowledge reveals our own lives and makes us more deeply humane. Predictions of disaster do not have a strong track record. The end of privacy was forecast when computers started to become powerful, but in fact, so far the personal computer has vastly expanded the access of average people. Genetic breakthroughs were greeted with dire predictions that have yet to come true. Pandora’s curse may be seen in the uncontrolled spread of nuclear science; but on average, the ripple effects of fundamental new findings seems positive.

The truth is that any group merciless enough to use torture already has access to the private thoughts of its victims. Perhaps half the governments in the world resort to such methods. The twentieth century was simply filled with horrific examples of whole populations for whom thinking was not private and not free. Arguably the behaviouristic conception of human beings bereft of consciousness and volition contributed to a mindset in which people became nothing but puppets, to be manipulated at will.

In a neo-Jamesian universe ethical questions become obvious. If we can literally see the pain we inflict upon animals, on babies and perhaps fetuses and each other, the dilemmas can no longer be rationalized and evaded. That may not make life easier, but it makes it more honest. Ultimately, consciousness is a piece of reality, and by and large, we are better off trying to understand it than to evade it.³

References

- Baars, B.J. (1986), *The Cognitive Revolution In Psychology* (New York: Guilford Press).
- Baars, B.J. (1988), *A Cognitive Theory of Consciousness* (Cambridge: Cambridge University Press).
- Baars, B.J. (1992), *Experimental Slips and Human Error: Exploring the Architecture of Volition* (New York: Plenum Press).
- Baars, B.J. (1993), 'Why volition is a foundation issue for psychology', *Consciousness & Cognition*, **2** (4), pp. 281–309.
- Baars, B.J. (1996), 'When are images conscious? The curious disconnection between imagery and consciousness in the scientific literature', *Conscious Cogn.*, **5** (3), pp. 261–4.
- Baars, B.J. (1997a), *In the Theater of Consciousness: The Workspace of the Mind* (New York: Oxford University Press).
- Baars, B.J. (1997b), 'Treating consciousness as an empirical variable: The contrastive analysis approach', in *The Nature of Consciousness: Philosophical Controversies*, ed. N. Block, O. Flanagan, G. Guzeldere (Cambridge, MA: Bradford/ MIT Press).
- Baars, B.J. (2002a), 'Recovering consciousness: A timeline', *Science and Consciousness Review* (www.sci-con.org).
- Baars, B.J. (2002b), 'The conscious access hypothesis: Origins and recent evidence', *Trends in Cognitive Science*, **6** (1), pp. 47–52. <http://www.nsi.edu/users/baars/BaarsTICS2002.pdf>
- Baars, B.J. (2003), 'The double life of B.F. Skinner', *Journal of Consciousness Studies*, **10** (1), pp. 5–25.
- Baars, B.J. (in press), 'I.P. Pavlov and the freedom reflex', *Journal of Consciousness Studies*.
- Baars, B.J. & Franklin, S. (2003), 'How consciousness interacts with working memory', *Trends in Cognitive Sciences*, **7** (4), pp. 166–71. <http://www.nsi.edu/users/baars/>
- Baars, B.J., Ramsay, T.Z. & Laureys, S. (under review), 'Globalist theories of (un)consciousness: Recent evidence'.
- Baars, B.J., Banks, W.P. & Newman, J. (2003), *Essential Sources In the Scientific Study of Consciousness* (Cambridge, MA: MIT Press; <http://mitpress.mit.edu/0262523027>).
- Blumenthal, A.L. (1979), 'Wilhelm Wundt, the founding father we never knew', *Contemporary Psychology*, **24**, pp. 547–50.
- Damasio, H., Grabowski, T., Frank, R., Galaburda, A.M., Damasio, A.R. (1994), 'The return of Phineas Gage: Clues about the brain from the skull of a famous patient', *Science*, **264** (5162), pp. 1102–5.
- Danziger, K. (1979), 'The positivist repudiation of Wundt', *Journal of the History of the Behavioral Sciences*, **15**, pp. 205–26.
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D.L., Mangin, J.F., Poline, J.B., Riviere, D. (2001), 'Cerebral mechanisms of word masking and unconscious repetition priming', *Nat Neurosci*, **4** (7), pp. 752–8.

[3] The author gratefully acknowledges support from The Neurosciences Institute and The Neurosciences Research Foundation. 10640 John Jay Hopkins Drive, San Diego, CA 94549. www.nsi.edu

- Dennett, D.C. (2001), 'Are we explaining consciousness yet?', *Cognition*, **79**, pp. 221–37.
- Duncan, J. & Owen, A.M. (2000), 'Common regions of the human frontal lobe recruited by diverse cognitive demands', *Trends Neurosci.*, **23** (10), pp. 475–83.
- Edelman, G.M. and Tononi G. (2000), *A Universe of Consciousness: How Matter Becomes Imagination* (New York: Basic Books).
- Hilgard, E.R. (1997), *Divided Consciousness: Multiple Controls In Human Thought and Action* (New York: Wiley).
- Hobson, J.A., Pace-Schott, E.F., Stickgold, R.. (2000), 'Dreaming and the brain: Toward a cognitive neuroscience of conscious states', *Behav Brain Sci.*, **23** (6), pp. 793–842;
- Ishai, A. & Sagi, D. (1995), 'Common mechanisms of visual imagery and perception', *Science*, **268**, pp. 1772–4.
- Iwase, M., Ouchi, Y., Okada, H., Yokoyama, C., Nobezawa, S., Yoshikawa, E., Tsukada, H., Takeda, M., Yamashita, K., Takeda, M., Yamaguti, K., Kuratsune, H., Shimizu, A., Watanabe, Y. (2002), 'Neural substrates of human facial expression of pleasant emotion induced by comic films: A PET Study', *Neuroimage*, **17** (2), pp. 758–68.
- James, W. (1890), *The Principles of Psychology* (New York: Holt).
- Kern, M.K., Jaradeh, S., Arndorfer, R.C., Shaker, R. (2001), 'Cerebral cortical representation of reflexive and volitional swallowing in humans', *Am J Physiol Gastrointest Liver Physiol*, **280** (3), pp. G354–60.
- Kosslyn, S.M., Ganis, G. & Thompson, W.L. (2001), 'Neural foundations of imagery', *Nat Rev Neurosci.* **2** (9), pp. 635–42.
- Kreiman G, Fried I, Koch C. (2002), 'Single-neuron correlates of subjective vision in the human medial temporal lobe', *Proc Natl Acad Sci U S A*. Jun 11, **99** (12), pp. 8378–83.
- Logothetis, N.K. and D.A. Leopold (1998), 'Single-neuron activity and visual perception', in *Toward a Science of Consciousness II*, ed. S. Hameroff, A. Kaszniak, A. Scott (Cambridge, MA: MIT Press).
- Luborsky, L. and Crits-Christoph, P. (1998), *Understanding Transference: The Core Conflictual Relationship Theme Method*, 2nd Edition: Contents (Washington, DC: APA Books).
- Maril, A., Wagner, A.D. & Schacter, D.L. (2001), 'On the tip of the tongue: An event-related fMRI study of retrieval failure and cognitive conflict', *Neuron*, **31**, pp. 653–60.
- Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houde, O., Crivello, F., Joliot, M., Petit, L., Tzourio-Mazoyer, N. (2001), 'Cortical networks for working memory and executive functions sustain the conscious resting state in man', *Brain Research Bulletin*, **54** (3), pp. 287–98.
- Milner, A.D. & Goodale, M.A. (1995), *The Visual Brain in Action* (Oxford: Oxford University Press).
- Paulesu, E., Frith, C.D. & Frackowiak, R.S. (1993) The neural correlates of the verbal component of working memory. *Nature*, **362** (6418):342-5.
- Perky, C.W. (1910), 'An experimental study of imagination', *American Journal of Psychology*, **21**, pp. 422–52.
- Rees, G., Kreiman, G., Koch, C. (2002), 'Neural correlates of consciousness in humans', *Nat Rev Neurosci.*, **3** (4), pp. 261–70.
- Rosen, S.D., Paulesu, E., Nihoyannopoulos, P., Tousoulis, D., Frackowiak, R.S., Frith, C.D., Jones, T., Camici, P.G. (1996), 'Silent ischemia as a central problem: Regional brain activation compared in silent and painful myocardial ischemia', *Ann Intern Med*, **124** (11), pp. 939–49.
- Ryle, G. (1949), *The Concept of Mind* (London: Hutchinson).
- Segal, S.J. & Fusella, V. (1970), 'Influence of imaged pictures and sounds on detection of visual and auditory signals', *Journal of Experimental Psychology*, **83** (3), pp. 458–64.
- Singer, J.L. (1993), 'Experimental studies of ongoing conscious experience', in *Experimental and Theoretical Studies of Consciousness*. Ciba Foundation Symposium 174, rd. G.R. Bock & J. Marsh (London: Wiley Interscience).
- Wegner, D. (2002), *The Illusion of Conscious Will* (Cambridge, MA: MIT Press).