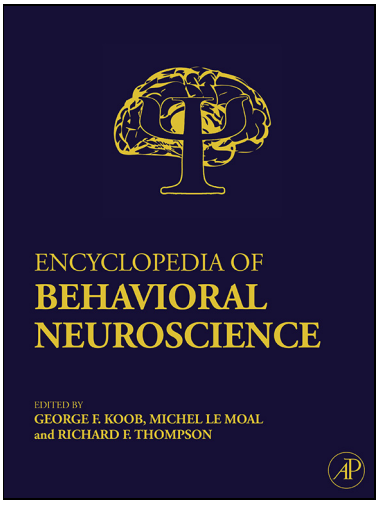

Dienes Z and Seth A K (2010) Conscious and the Unconscious.
In: Koob G.F., Le Moal M. and Thompson R.F. (eds.) *Encyclopedia of Behavioral Neuroscience*, volume 1, pp. 322–327 Oxford: Academic Press.

Conscious and the Unconscious

Z Dienes and A K Seth, University of Sussex, Brighton, UK

Glossary

Continuous flash suppression – A method discovered by Tsuchiya and Koch for making visual stimuli subliminal. When a static stimulus is presented to one eye and a stream of rapidly changing images to the other eye, the static stimulus is consciously suppressed but still processed. Unlike pattern masking, continuous flash suppression allows conscious and unconscious stimuli to be presented for the same period of time (e.g., >1 s).

Event-related cortical potential (ERP) – An electric potential recorded from the scalp by an electroencephalogram in which the trace has been time-locked to the occurrence of a stimulus and noise-reduced by averaging over many trials. Characteristic negative-going and positive-going changes in voltage occur over the several hundred milliseconds following a stimulus onset that can, in principle, indicate the nature of the processing at different times. For example, it has been shown that a positive change in voltage at around 300 ms (the P300), associated with attention and memory processing, can be elicited by stimuli the person is not conscious of.

Implicit learning – Ability to acquire knowledge that is unconscious. For example, we can use the rules of grammar to comprehend and produce grammatical utterances within a fraction of a second, yet we cannot describe more than a few rules of grammar. We can also learn to appreciate certain styles of music, to obey cultural rules, or to gain perceptual motor mastery of a domain without consciously knowing the underlying regularities.

Implicit memory – The influence of a past event, on current behavior and judgments without being able to recollect the prior event. For example, people with anterograde amnesia, due to damage to the hippocampus and temporal lobes, are unable to form new biographical memories yet can be influenced by new events. If shown a list of words, they may not recall seeing the words, yet when given the starting letters of the words they will be more likely to complete the letter stems as the words because of having seen the list of words. Implicit memory is different from implicit learning in that implicit memory need not involve acquiring any unconscious knowledge: completing letter stems as a word involves only knowledge of which the person is conscious (viz., the spelling of the word). In this case, there are no unconscious contents, but implicit memory is often called unconscious memory because it is a form of memory – broadly construed – that the person is not conscious of as memory.

Pattern masking – One of the most common laboratory techniques for producing subliminal perception. The mask consists of jumbled parts of the target stimulus. If the mask occurs just before (forward masking) or just after (backward masking) the target, with the difference in stimulus onsets being on the order of 10–100 ms, an otherwise consciously reportable briefly flashed target is rendered subliminal. The mask is just as effective if presented to a different eye from the target, indicating the mask interrupts processing at a central level.

Signal detection theory (SDT) – It allows the experimenter to distinguish 'discriminability' and 'bias': discriminability is the ability of a subject to discriminate the presence versus the absence of a stimulus, and bias is the tendency of the subject to say 'present' rather than 'absent.' Imagine a light is shown very briefly on half of trials and the subject has to say 'yes' or 'no' in each trial. Subjects may be biased to say 'no' a large or small proportion of the time without their underlying perceptual ability to discriminate changing. In SDT, discriminability depends on the distance apart from the distribution of neural activation for when the stimulus is present and for when it is absent; the bias to say 'yes' depends on whether the criterion amount of neural activation above which 'yes' is said is low or high. With a strict criterion, a person may say 'no' almost all the time when a stimulus is presented and still have a high ability to discriminate.

Subliminal perception – The influences on thoughts, feelings, or actions of stimuli that affect neural activity but are not consciously reported. Unlike implicit learning, subliminal perception need not involve changes in the pattern of synaptic connections.

One of the most fundamental problems facing psychology and neuroscience is the nature of consciousness. In this article, we sketch a variety of broad theoretical positions, describe a range of behavioral and neurophysiological measures of consciousness in the context of these theories, and review key findings concerning the existence and nature of unconscious perception and

Encyclopedia of Behavioral Neuroscience, 2010, Vol. 1, 322-327

learning. Finally, we consider possible functions of consciousness.

Theories of Consciousness

Worldly Discrimination Theory

According to worldly discrimination theory (WDT), the content directly expressed in any behavior (e.g., pointing to where a dog is) is the content of a conscious mental state (e.g., consciously knowing "There is the dog!"). Thus, a person shows that s/he is consciously aware of a feature in the world when s/he can discriminate it with choice behavior. This theory often makes use of signal detection theory (SDT), a statistical framework for quantifying the discriminability of a stimulus. SDT itself is mute on the subject of consciousness, and can thus be combined with different theories. The combination of SDT with WDT asserts that continuous information available for discriminations is necessarily the content of conscious mental states. That is, if over many trials a person can reliably indicate that a certain stimulus was present or not (to any degree above chance), then the information allowing that discrimination must be conscious, on WDT. This theory captures one property of conscious knowledge, namely that it allows choice behavior. However, rightly or wrongly, it does not respect other properties associated with consciousness. Consider blindsight patients, who have damage to the primary visual area in the cortex, V1. In the 'blind' part of their visual field they can make some discriminations very accurately; for example, whether an object is moving up or down. According to WDT, blindsight patients see movement consciously because forced-choice discrimination is the result by which we infer they see at all. However, two properties of blindsight suggest intuitively that the seeing is not conscious, and indeed are the reason why their 'sight' is called 'blind.' First, blindsight patients do not spontaneously attempt to use the information practically or inferentially. If they are thirsty, a glass of water in their blind field is not drunk. Second, blindsight patients themselves think they do not see. Indeed, they may insist that the attempt to discriminate shapes, motions, or sizes of objects in their blind field is nonsensical and they can offer only pure guesses. Other theories have tried to locate a divide between conscious and unconscious processes that respect one or both of these intuitions.

Integration Theories

According to integration theories, conscious contents are widely available to many cognitive and/or neural processes. This core idea has been variously expressed as, for example, fame in the brain, broadcast within a global workspace (an idea made popular by Bernard Baars), activity within a 'dynamic core,' or integrated information (as elaborated respectively by Gerald Edelman and Giulio Tononi). The intuition offered by Edelman and Tononi is that, on the one hand, consciousness cannot arise in a system only making a simple discrimination similar to a thermostat: our conscious experience seems to always involve making many discriminations at once. On the other hand, consciousness cannot arise in a system merely making a set of independent discriminations similar to a photographic plate: our conscious experience seems interconnected, integrated, and unified. According to dynamic core and integrated information theories, a mental state is conscious if it provides a sufficiently informative discrimination among a large repertoire of possible states, where successful discrimination requires both differentiation and integration.

Higher-Order Thought Theories

According to higher-order thought (HOT) theories, a mental state is conscious when we are aware of being in that state, a view pursued by David Rosenthal and Peter Carruthers among others. Accordingly, it would be odd to say that a mental state is conscious yet, the person is not aware of being in that state. HOT theories differ from WDT in that it is the ability to tell the mental state one is in, rather than what state the world is in, that determines whether a mental state is conscious. In the context of SDT, HOT theory is associated either with the criterion of standard SDT, or with the second level of discrimination: discriminating not the world (as in WDT), but the accuracy of one's responses (i.e., whether responses reflect knowledge or guessing).

The above theories, excepting some neural integration theories, describe conditions for asserting whether or not a particular mental state is conscious. They do not generally pertain to whether an organism is conscious or unconscious. These are scientifically different problems as, for example, a person who is unconscious (e.g., in a coma or dreaming) can have conscious mental states (as suggested by functional magnetic resonance imaging (fMRI) scans of some coma patients asked to imagine different activities: relevantly different parts of the brain become active). As we will see, measures of consciousness can and do address the conscious status of both mental states and individuals.

There is no consensus about which of the above theories best captures intuitions concerning consciousness, which is best supported by evidence, or which should be exclusively adopted in scientific research. Consequently, for current purposes we will treat the measures motivated by each theory as picking out a separate aspect of consciousness or mental life, of interest in its own right. Thus,

the behavioral ability to discriminate a feature in the world can be taken to indicate that a person is, in a very weak sense, conscious of that feature. Saying this much makes a less strong assumption than WDT; one can assume that a person can be (in a very weak sense) conscious of a feature without assuming that they are conscious of it with a conscious mental state. For example, a person with blindsight sees – and is therefore in that sense conscious of – an object moving up, even though the seeing is not conscious. For HOT theory, it is only if the person is aware of seeing that the seeing is a conscious mental state. Some philosophers disagree with this assumption; they prefer to say higher-order thoughts allow introspective or reflective consciousness but are not needed for mental states to be simply conscious. There is no need to quibble over words; clearly, the distinction between those perceptual or learning processes that allow awareness of knowledge and those that do not is interesting, whatever terms one uses (conscious vs. reflectively conscious, etc.). A conscious mental state also may or may not be inseparable from the global availability of the content of the state. However, the availability of information – for example, to figure in intentional control of one's actions – and the fine balance between differentiation and integration, are surely key features of our conscious mental life, regardless of whether one thinks they constitute conscious awareness. The degree of informational differentiation and integration may well be related to the level of consciousness of an individual. In fact, it will help consider a tripartite distinction (adapted from David Rosenthal) of being conscious of the world, being conscious of a mental state, and simply being conscious to some degree.

Behavioral Measures and Findings

Objective Measures

Objective measures refer to the ability to choose accurately under forced choice conditions. If one presupposes WDT, then such ability indicates a conscious mental state; with less strong assumptions, such ability simply indicates that one is conscious of the state of affairs discriminated. Conversely, knowledge is unconscious if a distinction in the world expresses itself only in nonintentional characteristics of behavior (such as its speed), or in physiological characteristics not expressed in behavior at all (e.g., galvanic skin response – how sweaty the person is; or brain activity as revealed by fMRI), or indirectly in other behaviors with related but different contents.

When objective measures indicate that people are not conscious of a masked visual stimulus, such as a word or a number, the impact of that stimulus on behavior and brain is greatly diminished compared to the effects of clearly visible stimuli. Nonetheless, some influences of the masked stimulus persist. Greenwald and his colleagues have shown that when people are not conscious of masked flashed numbers (presented for 33 ms), the size of the number influences the speed of a subsequent judgment of a consciously presented number: semantic characteristics of the subliminal number are partially accessed, even if semantic access is generally diminished when a person is not conscious of a stimulus. Dehaene and colleagues have shown that when people are not conscious of masked words, nonetheless local field potentials, as revealed by implanted electrodes in the amygdala (part of the brain responsible for emotional perception), distinguish threatening from nonthreatening words. Similarly, when people are not conscious of an image of a naked body, nonetheless spatial attention is drawn to the image location. Skeptics worry that these and similar results may be due to the subject being conscious of the stimuli to a small degree: the inevitable noisy estimate of discriminability never allows knowing when discriminability is exactly zero. In any case, one can conclude that stimuli that one is at most minimally conscious of, can influence reaction times, brain activation, and other physiological measures.

Strategic Control

Strategic control determines the conscious status of knowledge by the person's ability to deliberately use or not use the knowledge according to instructions. In Larry Jacoby's process dissociation procedure, a person tries to avoid using the information (exclusion task) or make sure they do use it (inclusion task); any difference in influence of the stimulus between these conditions is taken to indicate conscious knowledge, and any use of it despite intentions in the exclusion condition is taken to indicate unconscious knowledge. This measure can be motivated by integration theories: the ready availability of information to inference and intentions shows its conscious nature on such theories. On higher-order theories, by contrast, there is no incompatibility between control and the relevant mental states being unconscious (e.g., as in hypnosis, where people can engage in strategic behavior without awareness of their strategies).

People flashed words can be asked to complete word stems either (1) with the flashed word (inclusion instructions) or (2) making sure it is not the flashed word (exclusion instructions). If the word is flashed quickly enough (about 50 ms), then people will complete the stem with the word at above baseline levels even when told not to (i.e., while under exclusion instructions). Although people use the word under inclusion instructions (so they are conscious of the word in a weak sense), they cannot use the information that the word was present in their intentions, as revealed by their inability to exclude the word (thus, there is poor integration of the information with mental activities generally). Similarly,

when words are presented to people under general anesthesia, in the next few days people have a tendency to reproduce the words to word stems even when told to avoid any words they think they heard under anesthesia.

People can gradually acquire knowledge of the structure of an environment (by implicit learning) in such a way that the knowledge does not allow strategic control. For example, in the serial reaction time (SRT) task, used by Axel Cleeremans and others, people have to press one of four buttons depending on whether an 'X' appears in one of four corresponding locations. Unbeknownst to people, there is a predictable sequence of locations. Consequently, people become faster and faster at pressing the buttons. People are subsequently told about the existence of the sequence, and asked to produce a sequence that was not the one they had just been exposed to (exclusion instructions). If the training sequence is not highly predictable and is presented quickly, people cannot help but tend to produce the training sequence in exclusion, that is, even when trying to avoid it.

Subjective Measures

Subjective measures test whether people are aware of the mental states they are in. When a person successfully makes a series of discriminations about the world, to test for awareness of knowing, confidence ratings can be given on each trial. If, on all the trials, when the person says 'guess' nonetheless the discrimination performance is above baseline, then there is evidence that the person has knowledge about the world that they do, not know have: unconscious knowledge according to higher order theories (the guessing criterion). Further, if a person is aware of when they know and when they are just guessing, there should be a relation between confidence and accuracy. Thus, no relation between confidence and accuracy is another indication of unconscious knowledge according to higher-order theories (the zero-correlation criterion). Sometimes methodologically weaker forms of subjective measures are used, for example, only after a block of trials or a whole experiment are people asked to report if they saw anything or knew about a relation.

Right from the very beginning of research into subliminal perception in the nineteenth century, subjective measures were used. The fact that people can be substantially above chance in perceptual judgments (e.g., about what words were displayed) when believing they are just guessing is a robust phenomenon. Such stimuli can produce widespread brain activations and strong effects on choice behavior. Further, flashing unusual shapes that people are not aware of seeing can make people like those shapes more in the future, a phenomenon called 'the mere exposure effect.' Flashing brand names of a thirst-quenching drink can dispose people to choose that brand if they are thirsty. When people are not aware of seeing stimuli, the stimuli can have further intriguing effects. For example, flashing images of very intelligent people can reduce self-esteem; flashing images of clowns or unattractive people can increase self-esteem. Complicated masking procedures are not necessary to interfere with awareness of seeing; unattended words can influence people who think they did not see any words (so-called inattentional blindness). In contrast to these positive effects, there is no evidence that commercially available subliminal self-help tapes have any effect beyond placebo on study effectiveness, memory ability, or self-esteem.

Various types of brain damage interfere with awareness of seeing. People with blindsight can discriminate shape, size, orientation, and even the emotional expression of faces without being aware of seeing, though processing the meaning of words is limited. People with visual neglect (typically with damage to the right inferior parietal lobe) are often not aware of seeing stimuli in the left side of space. Nonetheless, even the meaning of such stimuli can facilitate processing stimuli on the right-hand side of space for these people.

People can gradually acquire knowledge of the structure of a domain (implicit learning) without being aware of their knowledge. Arthur Reber asked people to memorize strings of letters made by an artificial grammar. When people were subsequently informed about the existence of a grammar, they could classify new strings as obeying the grammar or not without being able to say what the rules were. Indeed, people can classify above chance when they believe that they are literally guessing or using nothing but intuition.

A possible weakness of subjective measures arises when the terms on the scale used are not well defined to people. The term 'guessing' can mean simply "I am not very sure" in everyday life: having knowledge without being absolutely certain does not entail the person is unaware of the knowledge. Nonetheless, in the case of artificial grammar learning at least, people can classify above baseline even when they would bet on a random event such as a coin toss rather than on their own classification decision. In sum, people can have knowledge even when they sincerely believe they do not.

Thus far, the measures have dealt with whether or not a person is conscious of either an event or a mental state (conscious contents). Next we consider some measures that can also bear on the conscious level of an individual.

Brain Measures

EEG Measures

In 1929, Hans Berger discovered that waking consciousness is associated with low-amplitude, irregular electroencephelographic (EEG) activity in the 20–70 Hz range and it

is now known that unconscious conditions such as non-dreaming deep sleep, coma, general anesthesia, and epileptic absence seizures show predominantly low-frequency, regular, and high-amplitude oscillations. Event-related cortical potentials (ERPs) have been used to assess whether a stimulus is consciously perceived or not, though there is dispute about whether early components are most indicative of conscious awareness (i.e., amplitude changes ~100 ms vs. ~300 ms after stimulus onset). Such ERPs are also associated with other functions beyond consciousness *per se*, for example, in novelty detection, so are unlikely to serve as pure consciousness indices. The proprietary 'bispectral index' combines various aspects of the EEG signal to estimate anesthetic depth (conscious level) and hence, probability of accidental waking during surgery. EEG measures of consciousness float free of theory, only gaining purchase through attempted reliable correlations with one or more of the behavioral measures above.

Neural Correlates

A useful way of addressing experimentally the distinction between conscious and unconscious processing is via the influential notion of a 'neural correlate,' which refers to patterns of activity in brain regions or groups of neurons that have privileged status in the generation of conscious experience. For example, Stanislas Dehaene and colleagues have shown using fMRI that consciously perceived words evoke widespread cortical activity as compared to equivalent masked stimuli the subject is not conscious of, which evoke only local activity. This evidence is congruent with integration theories such as global workspace theory, though other evidence suggests that being conscious versus unconscious of a mental state may involve a difference in activation only in a small area: the mid-dorsolateral prefrontal cortex. Another ongoing controversy is whether synchronized neural activity, for example, in the so-called 'gamma' band (~20–50 Hz) and especially around 40 Hz, constitutes a dynamical correlate of consciousness; the intuition here is that neural synchrony may account for the 'binding' of contents within integrated conscious scenes.

In addition to investigating the neural correlates of conscious contents (of the world or of mental states), it is also possible to examine the neural correlates underlying the conscious level of an individual. Here, ample evidence indicates that normal consciousness is specifically associated with thalamus and cortex. Regions such as the hippocampal system and cerebellum can be damaged without a loss of consciousness *per se*; indeed, in cases such as Rasmussen encephalitis, an entire cortical hemisphere can be surgically removed without loss of consciousness (though not without other effects). Whereas highly localized cortical damage can delete specific conscious contents such as color, shape, and motion, damage to certain thalamic nuclei or to brainstem regions such as the reticular formation can abolish consciousness permanently. Finally, brain imaging evidence (pioneered by Steven Laureys) has shown that even in apparently unconscious conditions such as the 'persistent vegetative state,' stimuli can evoke significant though not widespread cortical responses, as compared to healthy controls. Such imaging studies provide a much needed fine-grained means of diagnosing disturbances of conscious level and establishing recovery potential of brain trauma patients.

Neural Complexity

Several recent measures build on the observation that conscious scenes are distinguished by being simultaneously integrated (each conscious scene is experienced 'all of a piece') and differentiated (each conscious scene is composed of many distinguishable components and is therefore different from every other conscious scene). The 'dynamic core hypothesis' (DCH) of Gerald Edelman and Giulio Tononi proposes that consciousness arises from neural dynamics in the thalamocortical system with just these features, as measured by the quantity 'neural complexity' (C_N). C_N is an information-theoretic measure that is high if each subset of a neural system can take on many different states and if these states make a difference to the rest of the system. The 'information integration theory of consciousness' (IITC; proposed by Tononi a few years after the DCH) shares with the DCH the idea that conscious experiences provide informative discriminations among a vast repertoire of possible experiences. In the IITC, the quantity Φ is defined as the 'information integrated' across the informational 'weakest link' of a system. Importantly, Φ is a measure of the capacity of a neural system to integrate information, whereas C_N is a measure of the actual dynamics of the system. A third measure, 'causal density,' measures the fraction of causal interactions among elements of a system that are causally significant; it is low both for highly integrated systems and for collections of independent elements. While complexity measures such as these cannot yet be applied to interactions between all sets of neurons in a whole brain, they can be applied to the summaries of neural activity at multiple sites as provided by, for example, EEG or fMRI. Work employing and developing these measures has only just begun.

The Function of Consciousness

The different theories of consciousness imply different functions. According to WDT, consciousness allows behavioral discrimination in the world; yet, WDT begs the question of why consciousness is required: intuitively

most people accept that automata without consciousness can make discriminations. Integration theory offers clear functions, depending on the theory, to do with flexibility or dealing with novelty. Higher-order theories need to locate the function of consciousness not in the powers of any mental state about the world, but in what can be gained by being aware of such mental states.

Functions for consciousness are easy to suggest though hard to establish. It has been suggested that consciousness allows flexibility, intentional action, or ability to deal with novelty. Yet hypnotizable people can engage in novel intentional actions without being aware of their intentions. For example, they can 'forget' the number 4, counting '1, 2, 3, 5, . . .', flexibly overcoming habit, without being aware of any intention or mental state containing the content 'four.' It has been suggested that consciousness allows rationality, yet at least sometimes its reasoning function is only *post hoc*, as revealed by 'choice blindness': people are asked to select the most attractive of two faces and then to describe why they made each selection. On some trials the pictures were switched by sleight of hand immediately after a choice was made; subjects often failed to detect the switch but nevertheless offered a plausible account why they chose a particular face. Indeed, it has been shown that sometimes deliberating on a complex choice leads to worse decisions than being distracted by another task. It has been suggested that consciousness of our own mental states allows us to be conscious of others' mental states, allowing more sophisticated social interactions. Yet we can be affected by the emotion of a face without being conscious of it. Finally, it has been suggested that our rich consciousness of our mental lives arose from sexual selection: it arose precisely because, similar the peacock's tail, it is costly yet pointless.

See also: Amnesia; Declarative Memory; Hallucinations in Neuropsychiatry and Drug Abuse: From Phenomenology to Pathophysiology; Implicit Learning and Memory: Psychological and Neural Aspects; Sleep: Learning and Memory; Subjective Experience and the Expression of Emotion in Man.

Further Reading

Baars BJ, Banks WP, and Newman J (ed.) (2003) *Essential Sources in the Scientific Study of Consciousness*. Cambridge, MA: MIT Press.

Banerjee R and Chakrabarti B (eds.) (2008) Models of brain and mind: physical, computational and psychological approaches. *Progress in Brain Research* 168: 1–266.

Banks W (ed.) (2009) *Encyclopaedia of Consciousness*. Amsterdam: Elsevier.

Cleeremans A (ed.) (2003) *The Unity of Consciousness*. Oxford: Oxford University Press.

Hassin RR, Uleman JS, and Bargh JA (eds.) (2005) *The New Unconscious*. Oxford: Oxford University Press.

Jamieson G (ed.) (2007) *Toward a Cognitive Neuroscience of Hypnosis and Conscious States*. Oxford: Oxford University Press.

Laureys S (ed.) (2006) *The Boundaries of Consciousness: Neurobiology and Neuropathology*. Amsterdam: Elsevier.

Metzinger T (ed.) (2000) *Neural Correlates of Consciousness*. Cambridge, MA: MIT Press.

Velmans M and Schneider S (eds.) (2007) *The Blackwell Companion to Consciousness*. Chichester: Blackwell.

Wilken P, Bayne T, and Cleeremans A (eds.) (2009) *Oxford Companion to Consciousness*. Oxford: Oxford University Press.

Zelazo PD, Moscovitch M, and Thompson E (eds.) (2007) *The Cambridge Handbook of Consciousness*. Cambridge: Cambridge University Press.