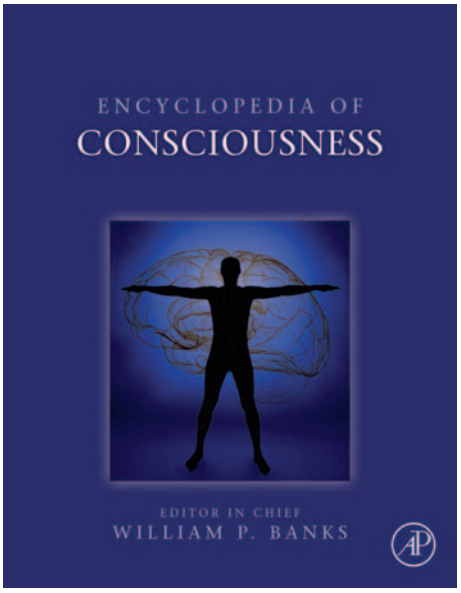

de Gardelle V and Kouider S (2009), Cognitive Theories of Consciousness. In: William P. Banks, (Editor), Encyclopedia of Consciousness. volume 1, pp. 135-146. Oxford: Elsevier.

Cognitive Theories of Consciousness

V de Gardelle and S Kouider, Département d'Etudes Cognitives, CNRS/EHESS/ENS-DEC, Paris, France

Glossary

Connectionism – Connectionism is a framework in cognitive science, according to which all of the processes achieved by the mind can be modeled by parallel and distributed processing among simple operational units. It is mostly based on the development of artificial neural networks, and it has been traditionally opposed to the position that mental processes are based on symbolic computations.

Functionalism – Functionalism is a doctrine in cognitive science, according to which a mental state is defined by its functional role, rather than by its intrinsic structure and its implementation. In other words, a functional model of the mind (or of an operation that is achieved by the mind) involves mental states that are causally related to sensory inputs and other mental states, and behavior.

Homunculus – Literally, 'little man,' in Latin, that is in the context of cognitive theories of consciousness, a conscious observer, which is at the top of the cognitive system.
A Homunculus is a hypothetical construct that operates (i.e., with no further explanation) the very operation that is supposed to be explained. Thus, a theory that relies on a homunculus at some point is incomplete in the same extend.

Introduction

Consciousness is probably the most privileged topic in psychology. The study of consciousness is considered to be at the origin of the separation between psychology and philosophy during the nineteenth century, as psychologists were motivated by the will to tackle this issue in a scientific way. As such, developing a scientific theory of consciousness has been the Holy Grail of psychology since its earliest days. Today, consciousness continues to be a central topic of interest, extending its interest across almost all disciplines of cognitive science.

Studying consciousness, however, has not always been an acceptable question in psychology. Indeed, the issue of consciousness was totally rejected by the dominant behaviorist school during the first half of the twentieth century. The failure of earlier psychologists, who were stuck in unsolvable debates related to introspection, led the behaviorists to reject consciousness as a plausible scientific issue. Instead, behaviorists sought to bring psychology into the scientific domain by restricting it to objective data and reproducible methods. Even the so-called 'cognitive revolution' that transformed psychology during the second half of the twentieth century did not favor a renewal of interest in consciousness. Indeed, the purpose of the cognitive approach was to reintroduce the notion of internal representations or mental states, not the notion of consciousness per se. Furthermore, an important new assumption was that internal representations were largely unavailable to consciousness. Nonetheless, in this new information-processing perspective, cognitive scientists developed several key elements that largely influenced the forthcoming cognitive theories of consciousness. These precursors included new models of attention and working memory, and also new concepts, such as the distinction between modules and central processes, or that between automatic and controlled processes.

In the present article, we will first overview the precursors that allowed the development of cognitive theories of consciousness. Then we will present a selection of influential contemporary accounts of consciousness. These theories will be

Encyclopedia of Consciousness (2009), vol. 1, pp. 135-146

grouped according to three main themes: theories that consider consciousness to result from specific architectural elements within the cognitive system; theories claiming that some features of consciousness are in fact illusory; and theories that focus on the relation between consciousness and learning. We will conclude this article by emphasizing the common challenges that current cognitive models of consciousness have to face: the pressure from the philosophically defined hard problem, on one side, and the pressure from neurobiological evidences, on the other side.

Precursors for a Cognitive Perspective on Consciousness

In this section, we highlight the most influential precursors for a theory of consciousness. Most of these elements emerged during the cognitive revolution. Although they were originally sketched out in light of an information-processing perspective, those elements are now largely linked to the dissociation between conscious and unconscious processing.

Attention and the Central Executive

Various influential models developed in the 1960s referred to a central processor, a central executive system, or a supervisory system. Processing within the central system can be considered as analogous to conscious processing, even if the word consciousness was still largely banished in the scientific community. This system is at the top of the hierarchy in the cognitive architecture: it is involved in higher-order computations (decision, monitoring, planning, etc.) and leads to selection and control over lower-level subsystems. As in many contemporary accounts of consciousness, the central system was considered the most integrative element of the cognitive system, granting flexibility and control over behavior.

Another key element was the simple but powerful metaphor of attention as a filtering mechanism that was put forward by Broadbent. In a nutshell, peripheral processors in this theory provide sensory information to the central system dealing with control and decisions. Because multiple sensory channels are continuously acting in parallel, a huge quantity of information becomes available to the rest of the system. However, the central system is very limited in terms of computational resources. Hence, a selection mechanism is needed to prevent overload. As such, attention operates by selecting the most relevant information and by filtering out that which is irrelevant. Then, the most relevant information, which is under the focus of attention, becomes the target of the central system and can thus benefit from deeper and more enriched processing. Once again, although consciousness was not the main concern, one consequence of attentional selection was that it allowed the target information to become conscious. In this perspective, attention and consciousness are two tightly related notions.

The notion of short-term memory put forward by George Miller and later extended to the notion of working memory is also an important precursor. For example, in their model of working memory, Baddeley and Hitch relied on a central executive system, which has top-down control over the distinct specific subsystems, namely the phonological loop and the visuospatial sketchpad. Here, the content of working memory may be roughly equated with the content of consciousness, an aspect that will also be important for future cognitive theories of consciousness.

Norman and Shallice, in turn, proposed a model of action selection implicating a supervisory attentional system. This central system receives sensory evidence and determines the appropriate behavior by selecting instruction schemes for action mechanisms. In addition, the supervisory attentional system can be modulated by the goals of the organism, and it is primarily involved when a new or critical situation appears. Here too, the central part of the model shares some properties that are associated with consciousness, namely flexibility, reactivity regarding unexpected situations, decision, and control over behavior.

In sum, these influential early models depicted the global architecture of the cognitive system by emphasizing the following components: sensory inputs in the periphery that are processed in parallel in multiple channels, attention that performs selection upon these sources of information, a working memory component that keeps tracks of the selected information, and finally a central

system that acts as a supervisor. But one major limitation of this view is that it falls into the homunculus trap, when it comes to the question of consciousness. Indeed, if this central supervisor is governing the whole cognitive system, one may ask who is in turn governing the central supervisor! That is, if we were to rephrase this question by focusing on consciousness, it would be problematic to rely on a hypothetical little man in our head (i.e., a homunculus) that has consciousness, which is the same property we are supposed to explain. This approach unavoidably leads to an infinite regression. Because consciousness was not the main issue for these early models, this crucial issue was left out or even denied during the development of early cognitive models with a central supervisor. As we will see below, current theories of consciousness will overcome this limitation by proposing various cognitive architectures, sometimes including a central system, that take into account the homunculus issue.

Specialized Modules versus Flexible Integration

Closely related to this distinction between central and peripheral processes is the very influential theory of modularity developed by Jerry Fodor. In this framework, modules are fast and efficient devices that process inputs in an automatic and mandatory fashion. They are tuned to a particular kind of computation on a particular kind of information. In other words they are functionally specialized and they constitute the small computational bricks of cognitive architecture. While modules operate in the periphery in this architecture, they have been classically opposed to central processes that can be slow but flexible, and can integrate inputs from different modalities. Closely related to this architectural dissociation is Posner and Snyder's dissociation between, on the one side, automatic processes that are mandatory and fast and, on the other side, controlled processes that are assumed to be strategic and voluntary.

Here again, although this was not explicitly acknowledged in these various works, the fast and automatic processes operated by modules were assumed to reflect unconscious processing, while the control processes involving the central system were assumed to be conscious. These dichotomies between central and peripheral, controlled and automatic, flexible and hard-and-fast processes have provided the ground for the distinction between conscious and unconscious processing, which has been central in the development of current cognitive theories of consciousness. It is of note that the difficulty of studying consciousness did not arise only when researchers decided to face it. Although a few serious attempts have been made to propose functional description of the central system, such as in the Adaptive Control of Though (ACT) theory by John Anderson, this notion was itself often unspecified and often considered as a mysterious but needed component. For instance, Fodor strongly defended the idea that although the program of cognitive science was to understand how modules work, we would surely be in an impasse when trying to address the nature of central processes.

Architectural Accounts of Consciousness

We present in this section three of the most influential cognitive theories of consciousness. For each of them, consciousness is grounded in an information processing system. Baars' global workspace theory uses the metaphor of global broadcasting to describe conscious processing, Jackendoff and Prinz' intermediate level theory emphasizes the need for consciousness to be focused on intermediate representations, and Tononi's information integration theory proposes to relate consciousness with complexity in the cognitive system. The three accounts all share the same will to link consciousness with a particular representational aspect of the cognitive system. These theories differ, though, in many respects and thus provide a diversified sample of what cognitive accounts of consciousness can be.

The Global Workspace Theory of Consciousness

Grounded on the distinction between conscious and unconscious processes, Bernard Baars' global workspace theory is one of the most influential

cognitive theories of consciousness. This theory relies on the metaphor of a theater. In this theater, unconscious specialized processors (equivalent to modules) are assumed to be the actors and the audience. While the audience represents the set of passive processors, actors represents active processors playing on the stage of the theater (i.e., the workspace). These actors are engaged in a competition for being seen by the audience: by broadcasting their information they actually compete for more broadcasting. Active processors with the highest coherent activity can form local coalitions that strengthen them in this competition process. The strongest coalition in this competition dominates the workspace, in a winner-take-all fashion, and corresponds to the content of consciousness. The workspace is equated by Baars to working-memory, in which only the most active content becomes conscious. Additionally, the dominant coalition benefits from global broadcasting, which allows it to recruit new processors from the audience in order to form a global coalition. Here, consciousness allows for the integration of computational resources in a large-scale coordination and for the exchange of information among processors that would otherwise remain separated. In this theory, each processor can operate in the conscious mode if it benefits from global broadcasting through the workspace, or it can operate in the unconscious mode when disconnected from the workspace.

An important feature of the global workspace theory is the presence of contexts as stable coalitions shaping access to the workspace. Contexts are constituted of unconscious processors reflecting, in a hierarchical manner, our expectations, our beliefs, our goals, and ultimately our self. In particular, attention is implemented as a goal context in this theory. It is described as a mechanism that controls access to the workspace, acting as a filter and biasing the competition process toward a particular set of actors. At any given moment, the dominant coalition is under the spotlight of attention, and its informational content becomes the content of conscious experience.

A crucial aspect of Baars' theory is that it avoids the problem of the homunculus by reducing it to an audience of multiple unconscious processors. Here, there is no need for a hypothetical single conscious observer in the system, and thus there is no issue of infinite regression with a homunculus inside another homunculus. Instead, consciousness is considered to reflect the global broadcasting of information to an audience of unconscious processors. As the audience is unconscious, unsupervised, and receptive rather than attending to the information, it does not constitute an internal homunculus.

The Intermediate Level Theory of Consciousness

The intermediate level theory originally proposed by Ray Jackendoff and further defended and specified by Jesse Prinz proposes that within the hierarchy of representations that are used to describe the cognitive system, conscious experience occurs only for specific levels of representation.

The theory is rooted in Jackendoff's analysis of different cognitive systems such as vision, language, and music and the subsequent observation that consciousness does not arise anywhere within these systems. According to Jackendoff, consciousness is not associated with low-level, nor with high-level representations, but rather with those implying intermediate levels of processing. For instance, in the domain of object recognition, it is assumed that the visual system comprises a low level with local computations of visual features, an intermediate level reflecting binding and object recognition, and a higher level computing viewpoint invariance and representing abstract categories. According to Jackendoff and Prinz, conscious experience is not comprised of a disunified picture of visual features, nor is it represented by view-invariant categories. Rather it is composed of bound and specific instances of objects that are assumed to be computed at the intermediate level of representation. In an analogous manner, speech perception can be decomposed into three levels: an acoustic representation of speech sounds at the lower level, a high level involving abstract lexical and syntactic categories, and in between a word recognition level relying on phonological representations. This theory explains why the conscious experience associated with speech perception mostly involves phonological representations, rather than other levels of representations. In Jackendoff and Prinz' theory,

the privileged role of the intermediate level of processing is based on the need for real-time computational efficiency. Indeed, this level of representation is assumed to be the most relevant one regarding ecological and functional needs.

Another important aspect of this theory concerns the central role of attention during conscious experience. Here, attention is defined as a selection process that acts as a gate to working memory mechanisms. It performs the function of selecting the relevant information that is amplified afterward and then becomes conscious. Indeed, Prinz acknowledges that activation of an intermediate-level representation on its own cannot be a sufficient condition for consciousness, given that those representations can be activated during subliminal perception. However, this theory makes the crucial postulate that the amplification of intermediate-level representations by attention is a necessary and sufficient condition for consciousness. In sum, for each domain of processing, the content of consciousness at a particular moment is supported by a representational structure of intermediate level for that domain, which is selected to enter short-term memory, and enriched by attentional processing.

The Information Integration Theory of Consciousness

'The information integration theory of consciousness' has been proposed by Gulio Tononi to explain how consciousness arises from dynamic complex systems. It originates from Tononi's work with Gerald Edelman and their observation that conscious states share two fundamental properties: they are both differentiated and integrated. Conscious states are highly differentiated in the sense that the occurrence of a particular conscious state results from its selection among a huge repertoire of possible conscious states. As such, a conscious state carries an important amount of information, as it reflects a large reduction in uncertainty. At the same time, conscious states are integrated as a unified experience. For instance, one does not have the experience of the color of a particular shape independently from the experience of the shape itself. A given state in a system is considered to be integrated if it results from the interactions of diverse subsets within this system. To account for integration, Edelman and Tononi relied on the notion of neuronal reentry within a thalamocortical dynamic core.

The information integration theory, formulated more recently by Tononi, is more concerned with how any physical system, brain or machine, with both integrated and differentiated information can lead to conscious experience. In this theory, consciousness is a property of a system that can integrate differentiated information: the more one system exhibits integrated and differentiated states, the more it is conscious. Accordingly, Tononi proposed to measure information integration by means of a function labeled Φ, whose value allows one to assess the degree of consciousness within the system. This function Φ takes high values for systems with high complexity, such as small-world architectures where connectivity patterns between units are heterogeneous. Conversely, it has low values for simple and feedforward systems. Importantly, Tononi gives an operational method for the computation of Φ in a given system, based on decomposition of the system into its subsets. As such, he also puts forward the notion of a complex in a system: a complex is mathematically defined as a subset of the system that is not part of a subset of higher Φ value. Importantly, according to the information integration theory, the content of consciousness at a given moment corresponds to the information processed in the complex, which exhibits the highest Φ value, called the main complex of the system. As the system processes information dynamically, interactions between the different parts of the system are continuously changing. Thus, the main complex changes accordingly, and so does the content of consciousness.

One important aspect of this approach is that it considers consciousness to be a quantitative and graduate variable. Furthermore, as consciousness is only determined by the Φ measure, it is only a matter of system complexity in any system. Consequently, animals or mechanical systems exhibiting similar properties can be considered as having a certain degree of consciousness. Still, although the value of Φ can be computed in theoretical situations, with fully specified systems, one obvious difficulty is the measure of Φ in natural systems. The decomposition of the mind into relevant

subunits is still a matter of research, and the assessment of information processed by these subunits has been to date an untargeted issue.

Illusory Features Accounts of Consciousness

Several approaches have claimed that some features associated with conscious experience are in fact illusory. Here, we present the most popular views on this matter. A first perspective is represented by Daniel Dennett's multiple drafts model of consciousness, where the appearance of a unified stream of consciousness reflects an illusion produced during introspection. A second view is held by Daniel Wegner whose theory of apparent mental causation claims that free will and the fact that we consciously determine our actions is illusory. A third account, the sensory-motor theory of consciousness by Kevin O'Regan and Alva Noe, also takes phenomenal experience as a retrospective illusion. However, this theory also associates consciousness with a learning process, and thus so it will be addressed in the next section on 'Learning process accounts of consciousness.'

The Multiple Drafts Model of Consciousness

The quest for a conscious subsystem in the brain has been overtly criticized by Dennett who explicitly related it to the homunculus assumption. Instead, he proposed a multiple drafts model of consciousness in which information does not need to be represented in front of a conscious observer within our heads. In this model, the stream of consciousness is neither unified nor is it produced by a single narrative system. Instead, what makes the stream of consciousness apparently unified is a retrospective reconstruction involving multiple drafts describing the story.

In the multiple drafts model, the cognitive system continuously processes information in parallel in different threads, either in perceptual, conceptual, or motor domains. In fact, threads look like Fodorian modules or specialized processors of the global workspace, and their computations in progress are logged in a temporary draft. As such, multiple drafts are edited in parallel and continuously revised within the system. In addition, these drafts have different fates: some will be read by the rest of the system and will affect subsequent behaviors, while others will simply fade out. In Dennett's model, cerebral celebrity makes a particular draft conscious (or 'fame in the brain'), that is, the extent to which it affects other processes in the system, and eventually subsequent behaviors and responses. In particular, by introspecting ourselves and thus directing our attention to one particular thread, we let the content of this thread affect our behaviors and thus become conscious. Introspection can also have the consequence of modifying the content of the draft itself. For instance, if a thread is probed too late, the associated draft will not be available anymore, or it will be totally reconstructed on purpose. Conversely, if the thread is probed to early, its process is interrupted, and the draft that becomes conscious will not reflect further edition.

The multiple drafts model of consciousness is an early and influential cognitive theory of consciousness, developed with the will to eradicate problematic homunculus assumptions. Indeed, the theory emphasizes that there is not a single observer that would receive all the information and provide a single and unified narrative stream of consciousness. Rather there are multiple on-going processes, from which some drafts have sufficient impact to influence behaviors and lead to consciousness one after the other. According to Dennett, the illusion of a single narrative stream stems from the fact that the story is continuously revised in order to be more plausible. Though less specified than more recent accounts that are similar in principle (e.g., Baars' global workspace theory), this theory remains an interesting instance of a strongly reductive view, which offers to replace the central homunculus system by parallel and distributed processing in a network of threads or daemons. In this approach there is nothing more to consciousness than the causal impact that one particular content has on subsequent processing and behavioral reports. However, one possible criticism linked to this feature is that the explanation provided by Dennett is a theory of report rather than of conscious experience. This latter argument reflects more generally the critical problem of assessing consciousness without relying

on some kind of report (we return to this point in the conclusion of this article).

The Theory of Apparent Mental Causation

In everyday life as well as in scientific accounts, consciousness is usually associated with the determination and control of appropriate behaviors. In his apparent mental causation theory, Daniel Wegner takes a different view in which consciousness and will are actually determined by unconscious causes, and have no real causal role in return. This view is also called epiphenomenalism, as it considers that conscious experience is an epiphenomenon that accompanies unconscious processes, but has no functional role.

In Wegner's theory, our conscious thoughts do not necessarily cause our behaviors. Rather, both conscious thoughts and behaviors are caused by unconscious mechanisms. These underlying unconscious causes of thoughts and the unconscious causes of behaviors are different, though they can influence each other. Because of the mutual influence between these two types of unconscious causes, their effects (i.e., conscious thoughts and conscious behaviors) are correlated as well. Because conscious thoughts happen just before conscious behaviors, they are taken to be the causes of initiated actions. Here, the attribution of a causal role to conscious thoughts is an illusion based on what is apparent to consciousness, not on what really happens. In addition, the theory specifies the condition under which this illusion occurs: thoughts have to appear just before an action (priority), they have to be consistent with the action (consistency), and they have to be the only possible explanation of the action (exclusivity). When these conditions are satisfied, conscious thoughts contain a predictive model of the forthcoming action, and when the action is realized in agreement with the predictions, we grant authorship for it and we experience ourselves as causal agents.

The theory of apparent mental causation does not aim at explaining how consciousness arises in a cognitive system. Rather, it explains how our conscious experience of will is an illusion that stems from our ignorance of actual unconscious causal chains. This approach has the advantage of trying to eliminate a false but still well-established *a priori* about the experience of conscious will. Nonetheless, one might wonder why then would we experience this illusion? Wegner proposes that it may help the subject to maintain his goals through consciousness or to build a better representation of the world, in which his own contributions are marked as such. This theory, however, suffers from an important difficulty, as it is expressed in terms that remind us of the homunculus problem, as pointed out by Dennett. Indeed, there is still in Wegner's account one self: someone who is conscious, someone who attributes causality to conscious thoughts, someone who is experiencing the illusion of conscious will, and who has in fact the properties of a homunculus.

Learning Process Accounts of Consciousness

Here, we present three theories that emphasize the influence of learning on consciousness. In their sensory-motor theory of consciousness, Kevin O'Regan and Alva Noë put forward the notion of learnt sensory-motor contingencies. In both Axel Cleeremans' radical plasticity thesis and Hakwan Lau's higher-order Bayesian decision theory, consciousness results from the ability of the cognitive system to learn about its own internal states.

The Sensory Motor Theory of Consciousness

Most models of vision are based on internal detailed representations that are active when a particular stimulus is present in the visual world. The sensory-motor theory of consciousness proposed by Kevin O'Regan and Alva Noë takes an alternative view in which there is, according to them, no need for detailed representations in the brain, and in which conscious experience is produced by the mastery of sensory-motor contingencies.

In normal situations, the observer knows that he only has to direct his eyes or his attention toward it in order to obtain detailed information. In other words, the world is an external memory, and the information it carries is usually sufficient for action. As a consequence, rather than relying

on internal representations that would be at the origin of conscious experience, this theory considers consciousness to be an active and exploratory process in between the observer and the external environment. In support of this theory, several studies, including some experiments by O'Regan and Noë, have shown that normal observers can suffer from 'change blindness,' a situation predicted by the idea that our memory lies in the outside world. In this situation, observers have the illusion that they are conscious of the whole visual scene while, actually, they fail to notice important modifications in the scene. Importantly, these changes are noticed when participants direct their eyes or their attention to the critical location. These findings show that observers have an illusory and overconfident estimation of their visual capacities. The sensory-motor theory of consciousness also proposed to explain some features of conscious experience on the basis of the characteristics of the sensory-motor contingencies, i.e., the principles that link exploration acts to sensorial consequences. In vision, for instance, a saccade to the left will shift the visual input on the retina accordingly, but even if the position of the object in front of you has changed on your retina, you would still feel that this object has not moved: this principle is embodied in your sensory-motor contingencies. Besides, the different sensorial modalities are different means for exploring the environment, and among these modalities, the differences in the sensory-motor contingencies (e.g., optical laws differ from acoustical laws) are the basis for the differences in the structure of conscious experience. Importantly, these contingencies apply at different levels of abstraction: some relate to the physical apparatus of the stimuli in a given modality, while others relate to features or categorical attributes. When we look at a particular object from a changing viewpoint, the visual image changes but the category of the object remains constant.

The most original idea expressed by the sensory-motor theory is that external stimuli do not have to be represented in detail in the brain. Importantly, however, O'Regan and Noë are NOT against any form of representation, or any storage of information in the brain. Following their own terms, they grant at least that the knowledge of the laws of sensory-motor contingencies have to be represented. What is rejected is the notion of continuous detailed representations of the outside world, and the fact that having these representations could be sufficient to create consciousness, without making use of it, in the sense of exploring it through sensory-motor contingencies. Regarding this question, one interesting argument put against this theory was the issue of dreams or mental imagery. Since those phenomena provide compelling intuitive support for the existence of such detailed internal representations, how does the sensory-motor theory deal with that? The answer provided by O'Regan and Noë is that there are still some differences between normal visual experience and dreaming or visual imagery situations, which correspond to the fact that in the latter cases the subject cannot make use of all the sensory-motor contingencies. Additionally, they deny that dream-experiences are stable in the details, as they miss the stability of the world as a memory.

The Radical Plasticity Thesis of Consciousness

While many cognitive models use symbolic and discrete representations, connectionist models rely on sub-symbolic and distributed units in artificial neural networks. In these models, representations are patterns of activation over processing units. Following this perspective, Axel Cleeremans proposed a conceptual framework termed the Radical Plasticity Thesis that put a strong emphasis on the link between conscious awareness and learning. This theory is based on three main principles.

The first main idea states that learning is a mandatory consequence of information processing, leading the cognitive system to develop representations of higher quality. Here, the quality of a representation is assessed by the stability and strength of activation in the dynamic network, and distinctiveness, which is equivalent to differentiation in the theory put forward by Tononi (see 'The information integration theory of consciousness' above). The second important idea is that consciousness reflects the quality of representations within the cognitive system. In this theory, the more representations achieve high

quality (i.e., high strength, stability, and distinctiveness), the more they participate in conscious experience. Hence, in this theory, conscious experience is a graded and continuous variable. Given those two principles, learning is associated with higher quality representations, which are in turn more likely to be conscious. Finally, the third principle highlights the implication of metarepresentations for self-consciousness. Cleeremans proposes that high-quality representations are efficient detectors of a particular content, and that they can be the target of metarepresentations. These metarepresentations capture the associations between first-order representations, which are developed through learning and past experience. The theory also proposes that a metarepresentation helps the first-order representations on which it is focused to achieve higher quality. In other words, one system can support consciousness insofar as it is able to learn about its environment and create internal representations, and also be able to learn about its own representations and increase their quality. Here, the more the system knows about its own rules, the more it is assumed to be conscious.

Cleeremans further distinguishes between different aspects of conscious experience and describes how these aspects correlate with the increase, through learning, in the quality of representations. The formation through learning of internal representations is depicted in three stages. The first one relates to implicit cognition: a poor-quality representation can influence behavior, but it is not strong enough to let the subject know about these influences or to have much control over them. Through exposure and learning, the representation achieves higher quality and becomes explicit. In this second step, the availability to control and the potential impact on the cognitive system also increase dramatically and reach a maximum. When the representation is sufficiently learned, it becomes automatic. According to Cleeremans, this third stage is associated with high-quality representations readily available to conscious awareness, though the subject has less control over their influences as they operate in a mandatory way. Thus, in this final idea, Cleeremans takes a view that can be contrasted with the classical assumption that automatic processes are unconscious.

The Higher-Order Bayesian Decision Theory of Consciousness

Signal detection theory and Bayesian frameworks have recently undergone a great renewal of interest among cognitive scientists. These conceptual tools bring useful insights in the description of behavioral performance, such as discrimination, detection, and decision. In a nutshell, signal detection theory proposes that discrimination between target and noise relies, on the one hand, on the objective distance (discriminability) between their two signal distributions on a psychophysical continuum and, on the other hand, on the particular setting of a decision threshold (criterion or bias) on that continuum. Bayesian decision theory, in turn, proposes a way to optimize the setting of the decision threshold, through prior learning over time of the probability distributions of the noise and target signals.

In many empirical studies on consciousness, participants' awareness of a given stimulus is equated with their performance on discrimination tasks (i.e., discriminability). Conversely, chance level performance on a discrimination task is often assumed to imply that the participant is completely unaware of the feature targeted by the task. Hakwan Lau's higher-order Bayesian decision theory of consciousness uses empirical dissociations between performance and awareness to support the idea that consciousness may not always be associated with an increase in discriminability. Rather, the hypothesis defended here is that it is related to the setting and the maintaining of the criterion threshold used for the perceptual decision. This theory associates some features of the higher-order thought theory with the Bayesian decision framework. More precisely, it proposes that while the lower-order system implements discriminability, the higher-order system, in turn, implements the decision threshold. In this view, the lower-order system performs a certain number of discriminations upon external signals, and the higher-order system learns about the distribution of states of the lower-order system, so as to interpret the signal, and to be able to set the threshold in an optimal manner. While Lau's theory is to date clearly not developed as far as other proposals, this work provides a new idea to the current theoretical landscape. It addresses a theoretical issue related to signal detection theory, which

is a methodological tool of increasing importance in the field of consciousness.

Conclusion

In this article we have presented an overview of the most representative cognitive accounts of consciousness. Most of these theories radically differ in their conception of what consciousness is. While some consider that it reflects the activation of attended intermediate level representations (Jackendoff, Prinz), or the involvement of complexes in a system (Tononi), others would equate consciousness with global broadcasting (e.g., Baars, Dennett), and still others would associate it with learning upon ones' own representations (e.g., Cleeremans, Lau) or upon sensory-motor contingencies (O'Regan and Noë). As such, it is obvious that consciousness is not yet a well-defined notion. In addition, cognitive accounts are now facing two epistemological constraints that impose important pressure on their development. The first one has been put forward by philosophers and corresponds to the need to focus on the 'hard problem' rather than the 'easy problem' of consciousness. The second one is related to the increasing amount of empirical evidence resulting from the study of the brain. We conclude this article by focusing on these two constraints.

The philosopher David Chalmers termed a dichotomy between the easy problem and the hard problem of consciousness in order to delineate the two major features related to consciousness. On the one hand, consciousness offers a processing advantage, as it allows for the information in working memory to be processed in a long-run by multiple devices. On the other hand, consciousness carries the qualitative property of subjective experience. Chalmers states that the first issue is in fact an interesting though only a computational problem; hence it is easy to study scientifically. The second one, however, is much more mysterious. How can the subjective quality of experience arise from squishy organic matter is a question that seems to go far beyond our possible understanding. The distinction between the hard problem and the easy problem can be mapped onto the dichotomy between access consciousness and phenomenal consciousness proposed by the philosopher Ned Block. Access consciousness relates to the global use of conscious information, and the possibility through consciousness to trigger complex and integrated processes such as reasoning, control of actions, decisions, and verbal reports. In contrast, phenomenal consciousness refers to the mere subjective experience, the 'what it is like' question expressed also by Thomas Nagel. Both Chalmers and Block defend phenomenal consciousness and the hard problem, claiming that conscious experience of a stimulus is not reducible to its information processing and its causal influences in the system.

It turns out that most cognitive models are expressed in information processing terms, and as such they are bound to take a reductive approach when trying to explain phenomenal consciousness. Intrinsically, they favor functionalist perspectives whereby information processing is all there is to conscious experience. On a more general perspective, science deals with measurements and measurements are by definition targeted to a piece of information that is measured. Hence, in most cognitive accounts, the hard problem is either reduced to the easy one or even completely denied. For instance, some will stand that it might be in fact necessary to revise our definition of what consciousness is, in order to eradicate any reference to some mysterious 'phenomenal' properties of the mind (e.g., Dennett). Indeed, one should not overlook the possibility that phenomenal and access consciousness are two notions that have to be dissociated only conceptually. They are not easy to dissociate experimentally since any measure of phenomenal consciousness can hardly be dissociated from the involvement of access consciousness. Indeed, measuring phenomenal consciousness in an experimental perspective must be based on some form of report, hence on access consciousness.

The other limit of purely functional accounts of consciousness comes from the brain. Indeed, as they remain distant from the biological implementation, purely functional or philosophical perspectives on consciousness are now likely to miss this crucial dimension, and source of evidence. Consequently, they will lack the same amount of explanatory power. Indeed, consciousness has become for many scientists a biological problem whose answers will be found by studying the brain. In

fact, some authors go a step further in arguing that the operational definition for the scientific study of consciousness should be expressed in neural terms. In that perspective, the psychological tools that we use may all be discarded in favor of a more physiologically grounded approach. Even if we do not want to go that far, it is a matter of fact that today basic observations of brain processes might help, by providing new concepts that would help directing research, and new critical test that would help discarding unfitting theories. Memory is a good example. Functional accounts of memory had to go back to the drawing board when the neurology of memory began to be understood in a more precise way. On the other hand, functional and neurological accounts have sometimes worked together productively, as for instance Baars' global workspace theory, which has been extended at the neurobiological level by Stanislas Dehaene and colleagues. Of course, without theoretical knowledge of the functions that are to be explained, a purely biological theory of cognition would be impossible to construct. In other terms, the union between psychological and neurobiological perspectives makes both approaches stronger, and future models of consciousness will be bound to include brain evidences and hence to be transformed into neurocognitive rather than purely cognitive accounts of consciousness.

See also: History of Philosophical Theories of Consciousness; Neurobiological Theories of Consciousness.

Suggested Readings

Anderson JR (1983) *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Baars BJ (1988) *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.

Baddeley AD and Hitch GJ (1974) Working memory. In: Bower GA (ed.) *The Psychology of Learning and Motivation*, pp. 47–89. Academic Press.

Block N (1995) On a confusion about a function of consciousness. *Behavioural and Brain Sciences* 18: 227–247.

Block N, Flanagan O, and Güzeldere G (eds.) (1997) *The Nature of Consciousness: Philosophical Debates.* Cambridge, MA: MIT Press.

Broadbent DE (1958) *Perception and Communication*. Pergamon Press.

Chalmers D (1995) Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2(3): 200–219.

Chalmers DJ (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Cleeremans A (2005) Computational correlates of consciousness. *Progression in Brain Research* 150: 81–98.

Cleeremans A (2008) Consciousness: The radical plasticity thesis. *Progression in Brain Research* 168: 19–33.

Dehaene S and Naccache L (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79(1–2): 1–37.

Dennett DC (1991) *Consciousness Explained*. Little, Brown & Company.

Edelman GM and Tononi G (2000) *A Universe of Consciousness: How Matter Becomes Imagination*. New York: Basic Books.

Fodor J (1983) *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press.

Jackendoff R (1987) *Consciousness and the Computational Mind*. Cambridge: MIT Press.

Lamme VA (2006) Towards a true neural stance on consciousness. *Trends in Cognitive Sciences* 10: 494–501.

Lau HC (2008) A higher-order Bayesian decision theory of perceptual consciousness. *Progression in Brain Research* 168: 35–48.

Mack A and Rock I (1998) *Inattentional Blindness*. Cambridge, MA: MIT Press.

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

Nagel T (1974) What is like to be a bat? *Philosophical Review* 83: 434–450.

Norman DA and Shallice T (1986) Attention to action: Willed and automatic control of behavior. In: Davidson R, Schwartz G, and Shapiro D (eds.) *Consciousness and Self Regu lation: Advances in Research and Theory,* 2nd edn., vol. 4, pp. 1–18. New York: Plenum.

Online Papers on Consciousness. Compiled by David Chalmers.http://consc.net/online/

Posner MI (1994) Attention: The mechanism of consciousness. *Proceedings of the National Academy of Sciences of the USA* 91(16): 7398–7402.

Posner MI and Snyder CRR (1975) Attention and cognitive control. In: Solso RL (ed.) *Information.*

Prinz J (2005) A neurofunctional theory of consciousness. In: Brook A and Akins K (eds.) *Cognition and the Brain: Philosophy and Neuroscience Movement*, pp. 381–396. Cambridge: Cambridge University Press.

Rumelhart DE and McClelland JL (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1*: *Foundations*. Cambridge, MA: MIT Press.

O'Regan JK and Noë A (2001) A sensorimotor account of vision and visual consciousness. *Behavioural and Brain Sciences* 24(5): 883–917.

Schneider W and Shiffrin RW (1977) Controlled and automatic human information processing: Vol. I. Detection, search, and attention. *Psychological Review* 84: 1–66.

Tononi G (2004) An information integration theory of consciousness. *BMC Neuroscience* 5: 42.

Tononi G and Edelman GM (1998) Neuroscience – Consciousness and complexity. *Science* 282(5395): 1846–1851.
Velmans M and Schneider S (eds.) (2007) *The Blackwell Companion to Consciousness*, pp. 225–235. Oxford, UK: Blackwell Publishing.
Wegner DM (2002) *The Illusion of Conscious Will*. MIT Press.
Wegner DM (2003) The mind's best trick: How we experience conscious will. *Trends in Cognitive Sciences* 7: 65–69.

Biographical Sketch

Vincent de Gardelle is a doctoral student supervised by Sid Kouider at the Ecole Normale Supérieure (Paris, France). His research addresses the question of biased conscious perception, by examining the perceptual illusions that might result from a subject's expectations under poor sensory evidence.

Sid Kouider is a cognitive neuroscientist working at the Ecole Normale Supérieure (Paris, France) on the neurobiological and psychological foundations of consciousness. His work focuses on contrasting conscious and unconscious processes both at the psychological and at the neural level, using various behavioral and brain imaging methods. Recently, he extended this line of research to study the neural correlates of consciousness in prelinguistic babies.