THE EXPLANATORY GAP IN NEUROSCIENCE

Lucia Melloni1† & Wolf Singer1,2,3

Introduction

We all know what it is like to be conscious, to be aware of something, to be aware of ourselves. However, even though there is this consensus and even though most would agree that consciousness is intimately related to brain functions, a universally accepted definition is still lacking. It is one of the goals of cognitive neuroscience to arrive at a better characterization of consciousness through a better understanding of the underlying neuronal mechanisms — and there have been remarkable advances in the study of the Neuronal Correlates of Consciousness (NCC). Since the seminal paper of Crick and Koch (1990), several theoretical proposals (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Lamme, 2006; Singer, 2000) as well as important empirical contributions regarding NCC have been made. However, important questions remain unresolved. In the following we propose that the search for a neuronal correlate of consciousness is equivalent with the search for neuronal mechanisms that have to account for several constitutive properties of consciousness. Therefore, we shall first summarize these characteristics and then explore putative mechanisms. Based on theoretical considerations and empirical evidence we propose that neuronal synchrony between distant neuronal assemblies might be a key mechanism for perceptual awareness. Finally, some pressing questions in the field of consciousness research will be discussed.

Some basic definitions

Rather than aiming at a comprehensive definition of consciousness or awareness we prefer an operational definition of these terms. We consider a
cognitive process as conscious, if the subject is aware of it and can report about it. If a subject can report the presence (or absence) of a stimulus (detection) or its identity (discrimination), we assume that the subject was conscious of the respective cognitive content. Failure to detect or to identify will be taken as a sign of unawareness. Even though this is not an exhaustive definition, it allows for an objective evaluation of subjective data, a scientific program called ‘heterophenomenology’ (Dennett, 1992). Because of the brain’s complexity it is difficult to induce specific and selective modifications of neuronal activation patterns. This often precludes establishment of causal relations between neuronal and cognitive processes and restricts research to the collection of correlative evidence. Therefore, consciousness research has focused on the search for neuronal correlates that are defined as the minimal set of neuronal events which are necessary and/or sufficient for perceptual awareness (Chalmers, 2000). This definition, however, confounds the search for a mechanism with the identification of mere correlates. For example, if a response in a certain brain region is associated with a consciously perceived stimulus, activation of this area cannot per se be taken as an NCC, because the observed activity could be either the prerequisite for or the consequence of conscious processing. What is needed instead is a model that causally relates certain neuronal mechanisms to consciousness. This in turn requires induction of predicted changes of consciousness by interfering with the putative mechanisms. In the following we define several requirements that need to be met by a mechanism that supports consciousness. Subsequently we shall discuss whether neuronal synchronization fulfills these premises.

Properties of consciousness constraining neuronal implementations

One central property of perceptual awareness is its unified character. However, the architecture of the brain is distributed, supporting multiple, parallel operations in numerous specialized but highly interconnected processing areas. This raises the question of how the unity of conscious experience can arise from the distributed processing of the brain.

Another characteristic of conscious processing is that the contents of our experience constantly change over time but, at any given moment, are discrete and distinct. Thus, the neuronal implementation of consciousness has to meet the requirement to support a seamless flow of ever changing contents that are experienced as coherent in space and time.

A particularly striking feature of consciousness is its limited capacity. At any moment in time, only a small fraction of the ongoing processes in the brain gains access to consciousness. Thus, being conscious of something al-
ways implies prevention of other contents from reaching consciousness at the same time, suggesting the action of a powerful selection process that gates access to consciousness. This in turn raises the question as to what distinguishes processes that enter consciousness from those that do not. As most results of sensory and executive processing have in principle access to consciousness – with the exception of certain vegetative signals – the distinguishing signature cannot be content specific but should be assignable in a dynamic way to all activity patterns that can, in principle, give rise to conscious experience.

Subjectively it appears to us that our actions and decisions depend mainly on those processes that we are conscious of. Experimental evidence indicates, however, that stimuli and processes which have no access to consciousness also have an impact on behavior. They can either trigger or inhibit a particular behavior. These unconscious determinants of behavior are not restricted to low level sensory processes but include the results of deep semantic decoding and the programming of complex motor responses (Dehaene et al., 1998; van Gaal, Ridderinkhof, Fahrenfort, Scholte, & Lamme, 2008). Thus, theories about the neuronal correlates of consciousness have to address the question of how signals are selected for access to awareness and which signatures distinguish conscious from unconscious information processing.

**Neuronal synchrony as a key correlate of perceptual awareness**

We shall argue that neuronal synchrony possesses most of the features that we have defined above as constitutive for consciousness or for processes leading to conscious awareness.

In order to create a unified conscious experience of multimodal contents, the modular architecture of the brain has to be overcome. One solution would be convergence of all information in a singular center. This option is considered theoretically implausible (Dennett, 1992), and so far no such center has been identified. Furthermore, behavioral and brain imaging studies have shown that unconscious processing engages very much the same cerebral areas as conscious processing, including frontal and prefrontal cortex (Lau & Passingham, 2007; van Gaal et al., 2008). Thus, there is no compelling evidence for specific areas supporting conscious processing. An alternative to convergence is that the distributed processes are bound together dynamically by rendering the signals coherent in time (Hebb, 1949), e.g. by synchronization (Singer & Gray, 1995). In this framework, the presence or absence of a particular feature is signaled through the activity of feature-selective neurons, while the grouping of those elementary
features into coherent representations of cognitive contents is achieved via synchronization of the respective neurons, forming a distributed representation of a perceptual object. Neuronal synchronization is a self-organizing process that allows rapid formation and dissolution of neuronal assemblies defined by coherence. Such dynamic binding is an economical and flexible strategy to cope with the representation of the virtually unlimited variety of feature constellation characterizing perceptual objects. Taking the unified nature of conscious experience and the diversity of possible contents into account, coherence (synchrony) offers itself as a mechanism allowing representation of ever changing constellations of content in a unifying format. In this way the assemblies that represent the different but unified contents of the continuous flow of consciously processed items could be bound together in ever changing constellations into a coherent, but permanently changing whole (meta-assembly). The updating of this meta-assembly would then be achieved by continuous cooption and exclusion of sub-assemblies. In this framework the rate-limiting factor for the formation of a new meta-assembly corresponds to the time needed to establish stable coherence. In case coherence is expressed by synchrony, this would be the time needed to establish stable phase relations. Stable states would then be reached once the relative phase between local oscillations ceases to change (Tognoli & Kelso, 2009).

Synchronization is also ideally suited to contribute to the selection of contents for access to consciousness. Synchronization enhances the saliency of signals and thereby facilitates their propagation in sparsely connected networks such as the cerebral cortex (Abeles, 1991; Fries, 2009; Jensen, Kaiser, & Lachaux, 2007). Gamma band synchronization, in particular, assures coincidence among distributed inputs with millisecond precision. Furthermore, when neuronal responses engage in synchronized oscillations, frequency and phase adjustments can be exploited for the selective routing of activity and the dynamic gating of interactions between interconnected neurons. At the level of individual neurons, oscillations are associated with periodic alternations of phases with high and low excitability, the latter resulting from the barrage of synchronized IPSPs that have both a shunting and a hyperpolarizing effect (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008).Excitatory inputs that arrive at the depolarizing slope of an oscillation cycle generate large responses, whereas inputs arriving at the falling slope and trough are shunted and ineffective. Hence, neuronal oscillations define temporal windows for effective communication between neurons, providing a mechanism to selectively and flexibly bias the communication between neuronal groups (Fries, 2009). When two groups of neurons open their windows of susceptibility
their excitatory phases) at the same time, they are more likely to interact with each other, to increase their synchrony and, as a consequence, to also have enhanced influence on other groups of neurons. By adjusting oscillation frequency and phase, groups of neurons can either be linked together into tightly coupled assemblies or be segregated into functionally isolated sub-groups. This mechanism can act both within and across areas, and can, in principle, account for selective and flexible routing of information within networks with fixed anatomical architectures. Taken together, the oscillatory patterning of activity and the option to adjust frequency and phase of the oscillations could serve three complementary functions: gain control, selective and flexible routing of information between neuronal groups, and formation of coherent representations. Furthermore, if depth of processing is determined by the extent of propagation of information in the brain, this can also account for the observation that conscious perception is associated with deeper processing than unconscious perception.

As previously mentioned, any neuronal correlate of consciousness should exhibit signatures that differ between conscious and unconscious forms of information processing. Regarding neuronal synchrony, this prerequisite is fulfilled with respect to its spatial scale: Following the distinction between local and global scale integration by Varela et al. (2001), processing carried out unconsciously (which is automatic and modular) should be based mainly on local integration in divergent-convergent feed-forward architectures, whereas conscious processing should involve large-scale integration via the extended networks of re-entry loops that couple neurons both within and across the various levels of the cortical processing hierarchies (for a similar proposal see Dehaene et al., 2006). Anatomical and physiological studies suggest that neuronal, in particular cortical, architectures share features of ‘small world networks’ (reviewed in Bassett & Bullmore, 2006). These allow for the coexistence of both local modular and global, distributed processes. Important properties of this architecture are minimization of path length between any pairs of nodes, optimization of the number of connections and the possibility of coexistence of multiple local processes and globally ordered states. Moreover, such networks can operate in critical states, allowing for fast reconfigurations of network dynamics (Bassett & Bullmore, 2006; Sporns & Zwi, 2004; Yu, Huang, Singer, & Nikolic, 2008). Cortico-cortical connections can be subdivided in two major subgroups. Local, intra-cortical connections that run tangentially to the layers and link neurons that share similar response properties and are separated by only a few hundred micrometers, and long-distance connections that often but not always run through the underlying white matter and link neurons in different cortical areas. The latter serve exchange of
information between distinct cortical sites and can establish globally coordinated activation patterns (Varela et al., 2001).

We started this chapter describing four conditions that any theory of NCC should satisfy. We then discussed how neuronal synchronization could fulfill each of them, namely that neuronal synchronization could account for (i) the unity of experience as well as for (ii) its temporal dynamics, (iii) for the selection of signals entering conscious processing and for (iv) the distinction between conscious and unconscious processing related to the spatial scale of the synchronized activity. We shall now review evidence relating to neuronal synchronization to consciousness.

Evidence relating to long-range synchronization and consciousness

Masking offers an interesting possibility to distinguish between conscious and subconscious processing, since the same physical stimuli can be either perceived or not perceived depending on the temporal and spatial sequence of stimuli that surround them. In our studies, we capitalized on this phenomenon and presented words that could be perceived in some trials and not in others (by adjusting the luminance of the mask) and simultaneously performed electroencephalographic (EEG) recordings (Melloni et al., 2007). Several measures were analyzed: time-resolved power changes of local signals, the precision of phase synchronization across recording sites over a wide frequency range, and event-related potentials (ERPs). A brief burst of long-distance synchronization in the gamma frequency range between occipital, parietal and frontal sensors was the first event that distinguished seen from unseen words, while local synchronization was similar between conditions. Interestingly, after this transient period of synchronization, several other measures differed between seen and unseen words: we observed an increase in amplitude of the P300 ERP for visible words which most likely corresponds to the transfer of information to working memory. In addition, during the interval period in which visible words had to be maintained in memory, we observed increases in frontal theta oscillations. Theta oscillations have been related to maintenance of items in short-term memory (Jensen & Tesche, 2002).

To test whether the increase in long-distance synchronization relates to awareness or depth of processing, we further manipulated the depth of processing of invisible words. It has previously been shown that invisible words can be processed up to the semantic and motor level (Dehaene et al., 1998). In a subliminal semantic priming experiment we briefly presented words (invisible) that could either be semantically related or not related to a second visible word on which subjects had to carry out a semantic classification
task. Invisible words were processed up to semantic levels as revealed by modulation of the reaction times depending on the congruency between invisible and visible words: congruent pairs exhibited shorter reaction times than incongruent ones. We observed increases in power in the gamma frequency range for unseen but processed words. For visible words we additionally observed increases in long-distance synchronization in the gamma frequency range (Melloni & Rodriguez, 2007). Thus, local processing of stimuli is reflected in increases in gamma power, whereas long-distance synchronization seems to be related to awareness of the stimuli. This suggests that conscious processing requires a particular dynamical state of the cortical network that is characterized by the dynamic configuration of widely distributed networks through transient synchronization. The large-scale synchronization that we observed in our study could reflect the transfer of contents into awareness and/or their maintenance. We favor the first possibility given the transient nature of the effect and argue that the subsequent theta oscillations might support maintenance. It is conceivable that short periods of long-distance synchronization in the gamma band reflect the update of new contents, while the slower pace of theta oscillations might relate to the sustained integration and maintenance of local results. The interplay between these two frequency bands might underlie the phenomenon of continuous but ever changing conscious experience (see below).

Recently, Gaillard et al. (2009) revisited the question of processing of visible and invisible words. In intracranial recordings in epileptic patients they observed that invisible words elicited activity in multiple cortical areas which quickly vanished after 300 ms. In contrast, visible words elicited sustained voltage changes, increases in power in the gamma band, as well as long-distance synchronization in the beta band that showed bidirectional Granger causality. In contrast to our study, Gaillard et al. observed a rather late (300-500 ms) rise of long-distance synchronization. However, it is important to note that in the study of Gaillard et al., phase-synchrony was analyzed mostly over electrodes within a given cortical area or at most between hemispheres, and thus this analysis raises an important methodological issue. Earlier synchronization events could have passed undetected because of lack of electrode coverage. Since with intracranial recordings the electrode placement is based on medical and not experimental considerations, analyses are restricted to the available electrodes and locations. Despite of these restrictions, this study provides one of the most compelling pieces of evidence for a relation between long-distance synchronization and consciousness.

Another commonly used paradigm in studies of consciousness is binocular rivalry. When two images that cannot be fused are presented simultaneously
to each of the two eyes, perception fluctuates between the two images. Thus, despite constant stimulation conditions, perception alternates. This suggests a selection mechanism that gates access to consciousness. Several studies using different stimulus materials as well as recording methods (single cell analysis and local field potential recordings in animals and electroencephalographic and magnetoencephalographic registration in human subjects) have shown increased synchronization and phase locking of oscillatory responses to the stimulus that was consciously perceived and controlled the pursuit eye movements (Cosmelli et al., 2004; Fries, Roelfsema, Engel, König, & Singer, 1997; Srinivasan, Russell, Edelman, & Tononi, 1999). Cosmelli et al. (2004) extended the findings obtained in human subjects by performing source reconstruction and analyzing phase-synchrony in source space. These authors observed that perceptual dominance was accompanied by coactivation of occipital and frontal regions, including anterior cingulate and medial frontal areas. Recently, Doesburg et al. (2009) provided evidence for a relation between perceptual switches in binocular rivalry and theta and gamma band synchronization. Perceptual switches were related to increments in long-distance synchronization in the gamma band between several cortical areas (frontal and parietal) that repeated at the rate of theta oscillations. The authors suggested that transient gamma-band synchronization supports discrete moments of perceptual experience while theta oscillations structure their succession in time, pacing the formation and dissolution of distributed neuronal assemblies. Thus, long-range gamma synchronization locked to ongoing theta oscillations could serve to structure the flow of conscious experience allowing for changes in content every few hundred millisecond. Further research is required to clarify the exact relation between the two frequency bands and their respective role on the generation of percepts and the pacing of changes in perception.

Another paradigm in consciousness research exploits the attentional blink phenomenon. When two stimuli are presented at short intervals among a set of distractors, subjects usually detect the first (S1) but miss the second (S2) when the two stimuli are separated by 200-500 ms. Increases in long-range neuronal synchrony in the beta and gamma frequency ranges have been observed when S2 is successfully detected (Gross et al., 2004; Nakatani, Ito, Nikolaev, Gong, & van Leeuwen, 2005). Furthermore, Gross et al. (2004) observed that successful detection of both S1 and S2 was related to increased long-distance synchronization in the beta range to both stimuli, and this enhanced synchrony was accompanied by higher desynchronization in the inter-stimulus-interval. Thus, desynchronization might have facilitated the segregation of the two targets, allowing for identification of the second stimulus (also see Rodriguez et al., 1999). Source analysis revealed, as in the
case of binocular rivalry, dynamical coordination between frontal, parietal, and temporal regions for detected targets (Gross et al., 2004).

In summary, studies of masking, binocular rivalry, and the attentional blink support the involvement of long-range synchronization in conscious perception. Recent investigations have suggested further that a nesting of different frequencies, in particular of theta and gamma oscillations, could play a role in pacing the flow of consciousness. Furthermore, the study of Gross et al. (2004), suggests that desynchronization could serve to segregate representations when stimuli follow at short intervals. These results are encouraging and should motivate further search for relations between oscillatory activity in different frequency bands and consciousness, whereby attention should be focused not only on the formation of dynamically configured networks but also on their dissolution.

Pressing question in the field of consciousness research

One influential view on the function of consciousness posits that consciousness allows for exchange of contents between several independent processing modules, which in turn makes this information available to a variety of processes including perceptual categorization, unification in short- and long-term memory, evaluation, etc. (Baars, 1997). If consciousness is a prerequisite of these cognitive processes, how can the requirements for and the consequences of consciousness be distinguished? If conscious and unconscious processing did not only differ with respect to their respective qualitative properties, but also with respect to their consequences, such a distinction might indeed be impossible: If consciousness, as proposed, enhances the depth of processing then a method that contrasts differences in perceptual states of physically identical stimuli will reveal not only activity related to awareness per se but also activity related to its consequences, i.e. deeper processing, episodic memory formation, etc. (for a similar argument see Lau, 2008). This might explain why different research groups find signatures of consciousness at different moments in time. Neurons in the medial temporal lobe (MTL) show an all or none response profile depending on whether stimuli are processed consciously. When briefly presented images are made unperceivable through backward masking, neurons in MTL do not respond. In contrast, when the same images are recognized, MTL neurons exhibit clear responses with a latency of about 300 ms (Quiroga, Mukamel, Isham, Malach, & Fried, 2008). Following the logic of the contrastive approach, this result can be taken as a correlate of consciousness. The question is: is this a correlate of consciousness or a correlate of the consequences of consciousness? Given the intimate
relation between MTL and hippocampus and hence with networks responsible for the management of episodic memory, it is also conceivable that the rather late MTL responses have to do with memory formation following conscious perception rather than with the access to consciousness per se. This view is supported by the fact that complete resection of the hippocampus and adjacent temporal cortex does not lead to deficits in conscious perception per se (Postle, 2009). However, this does also not imply that MTL does not contribute to conscious perception. It is important to distinguish between neuronal signatures which correlate with the content of our moment-to-moment conscious awareness, and those that contribute to the feeling of a stream of consciousness (continuous present). Thus, MTL activity might not contribute to the former, but be essential for the latter. In order to experience a continuous present, memories from events that have just occurred should be linked to those occurring at the present time. Thus, the delayed MTL activity could serve to assure the continuity of the flow of consciousness.

The conundrum of cause and effect

How else could one solve the conundrum to distinguish between cause and effect? There is no panacea. Obviously, the most straightforward approach is to obtain comprehensive data on the sequence of effects distinguishing conscious from unconscious processing, assuming that causes precede effects. This could be complemented with attempts to interfere with these sequential processes using transcranial magnetic stimulation. For instance, if blocking the late activity in MTL does not abolish conscious perception but its continuity we would be one step further. Following this strategy those processes could be discarded one by one that figure as consequences of consciousness, leaving us with those more closely related to consciousness per se. Secondly, one could evaluate introspective reports as a direct measure of the quality of the experience. In this case one would not contrast conscious with unconscious processing, but arrive at a parametric estimate of ‘consciousness’. For instance, in visual experiments, we could ask subjects about how clear their content of awareness is, and then correlate increases in subjective clarity with measures of neuronal activity.

Conclusions

It is evident that the study of consciousness has greatly profited from the search for neuronal correlates. However, simply showing that the brain makes a difference between conscious and unconscious processing is not sufficient.
We propose here to go one step further and to develop mechanistic explanations that establish plausible and, hopefully, at later stages even causal relations between brain processes and consciousness. Oscillatory synchrony is one candidate mechanism, and it has the advantage that it can be measured relatively directly in humans who are able to give detailed descriptions about their conscious experience. However, oscillations and synchrony seem to be mechanisms that are as intimately and inseparably related to neuronal processing as the modulation of neuronal discharge rates. Thus, without further specification these phenomena cannot be addressed as NCC apart from the triviality that consciousness does not exist without them. We and others (Varela et al., 2001) propose that the spatial scale and perhaps also the precision and stability of neuronal synchrony might be taken as more specific indicators of whether the communication of information in the brain is accompanied by conscious experience or not. In this framework, conscious experience arises only if information that is widely distributed within or across subsystems is not only processed and passed on to executive structures but in addition bound together into a coherent, all-encompassing, non-local but distributed meta-representation. This interpretation is compatible with views considering consciousness as the result of the dynamic interplay of brain subsystems; but it poses the challenging question related to the ‘hard problem of consciousness’ research: is there something on top of this distributed meta-assembly that makes us experience? From a truly dynamicist point of view, the answer is ‘probably not’. However, this should not discourage us from trying to separate the contribution of each subsystem and then go on to characterize how they interact. This would be a real alternative to the ever-lurking homunculus.

Acknowledgements
This work was supported by the Max Planck Society. We are indebted to Caspar M. Schwiedrzik for insightful comments on this manuscript.

References
Chalmers, D.J. (2000). What is a neural correlate of consciousness? In T. Metzinger (ed.), Neural correlates of consciousness: em-
practical and conceptual questions (pp. 17–40). Cambridge, MA: MIT Press.


of conscious recognition. Proc Natl Acad Sci USA, 105(9), 3599-3604.