16

Neural Coordination and Human Cognition

Catherine Tallon-Baudry

Abstract

Understanding how thoughts can arise from a hundred billion or so interconnected neu-
rons is the ultimate but still unreached goal of human cognitive neuroscience. The huge
advances in human brain imaging over the last twenty years have led to a parcellation
of the brain into a multitude of functional regions, yet understanding how activity is
coordinated in the activated network remains a great challenge. This chapter summa-
rizes the different modes of neural coordination that have been considered in the human
literature, both theoretically and experimentally. It underlines the distinction between
well-learned behavior, which can take advantage of prespecified neural routes, and dy-
namic neural coordination in flexibly defined neural ensembles, which can generate
new percepts and/or creative behaviors. Discussion is devoted to the role that brain
rhythms play in human cognition, with the underlying assumption that brain rhythms
are a signature of dynamic coordination. A hypothetical but comprehensive schema is
delineated to explain why different frequency bands coexist and interact.

Modes of Coordination

Coordination has distinct meanings. It can refer to the way information is
transmitted from one region to the other. For instance, during attentional ori-
enting, do the frontal areas drive the parietal regions, or is it the other way
round (Buschman and Miller 2007; Grent-'t-Jong and Woldorff 2007; Bressler
et al. 2008; Green and McDonald 2008)? Coordination may also refer to the
way something new is created by the interaction process itself, in agreement
with the aphorism "the whole is larger than the sum of its parts."

Information Flow and Its Limits

Characterizing the information flow between brain regions has been a very
active field over the last fifty years. Roughly, two main streams can be

distinguished. The historically older one is the attempt to organize the ever-increasing number of functional maps of the visual system (Van Essen 1979; Hadjikhani et al. 1998) into a global coherent model (Felleman and Van Essen 1991; Young 1992). This approach is fraught with a number of problems, in particular the choice of the criterion used to place one area on top of the other. Using the pattern of feedforward, feedback, and lateral connections generates too many plausible solutions (Hilgetag et al. 1996), whereas using the sequence of response latency yields a quite different organization of the visual system (Schmolesky et al. 1998).

The other stream is more recent and has been developed more specifically in humans, probably because the noninvasive imaging techniques used in humans (fMRI as well as EEG or MEG) sample the whole brain. Models based on such methods often lead to a rather sequential description of brain regions "lighting up" one after the other. This "boxology" approach has a strong descriptive power and fits well with the models derived from experimental psychology.

Pushing either of these two approaches to its limits raises some difficult issues. The diagrams of brain organization that summarize how information flows from one area to the other are oriented. They usually begin with a sensory input and go through a number of processing stages, which, depending on the authors, end up in supramodal regions or the hippocampal formation (the convergence zones of Damasio 1989), or in the fronto-parietal network as in global workspace theories (Baars 1997; Dehaene et al. 1998). Pushing this line of reasoning to its extreme reveals a number of similarities between convergence zones and grandmother cells. Indeed, most the objections raised against grandmother cells can be transposed to the description level of convergence zones (Singer and Gray 1995). In other words, the general principles of population coding may be relevant no matter whether the code unit is a cell or a functional module. In particular, both convergence zones and grandmother cells integrate across many inputs and, as a result, are faced with similar combinatorial problems. While integration by convergence may take place in dedicated neural circuits for well-learned stimuli (Li et al. 2004), more flexible mechanisms would be needed for new objects and situations.

A particularly crucial issue concerns the endpoint of any process: recognizing an object, a situation, or a mental state for what it is. In other words, when is the readout from the preceding stage sufficient to reach a decision? How does the system know it has reached a state that corresponds to a solution? If one considers a simple stimulus-response association, the process ends with the production of the motor response. However, in the case of mental states that do not end up with a movement, the answer is much less clear. Daniel Dennett is one of the most famous opponents of what he called the "Cartesian Theater" hypothesis (Dennett 1991). In this view, the mind, or the ultimate convergence zone, sits in front of a theater where percepts and thoughts are displayed. The Cartesian Theater hypothesis thus posits the existence of a "person inside" who acts as the ultimate witness (or convergence zone) of

everything that occurs in the conscious mind. Is there, however, another person inside the homunculus, since the sentence, "I am aware that I am aware that I am late delivering my paper," makes sense? Any model of the brain that uses convergence as an intrinsic mechanism confronts, at some point, the "person inside" or homunculus issue. Some posit that activity in the prefrontal cortex is a necessary condition for awareness (Gaillard et al. 2009; cf. Goldberg et al. 2006). Others consider that the neural correlates of perceptual awareness are found in sensory cortices and do not necessarily require a frontal involvement (e.g., Lamme and Roelfsema 2000; Ress and Heeger 2003; Kouider et al. 2007; Wyart and Tallon-Baudry 2009). Recent models such as global workspace theories (Baars 1997; Dehaene et al. 1998) try to circumvent the difficult issue of the homunculus by mapping awareness onto a combination of sensory modules communicating with a large parieto-frontal network.

Emergence

Emergence is a central concept in complex systems analysis and can be roughly summarized as "the whole is larger than the sum of its parts." The idea is that some new information or knowledge is created at the system level, through simple interactions between lower-level components. As a result, some properties that do not exist in any constitutive elements of the system can emerge at the population level. Emergence is typically observed in flocking or herding behavior. A well-known example of such behavior can be found in the field of artificial intelligence (Reynolds 1987): "boids" are moving objects following simple local rules (avoiding bumping into their closest neighbors, moving roughly in the same direction and with the same speed as their closest neighbors, staying close to other boids). These three simple local rules are sufficient to produce a group behavior similar to that of a flock of birds, including the V-shaped flight of ducks. Coherent behavior can thus emerge from local rules, without a need for either an explicit global schema or for a group leader. Let us also consider what happens at the end of a theater performance: clapping usually begins in a loud and disorganized manner, but after several curtain calls applause becomes rhythmic. A temporal structure spontaneously emerges, simply because people tend to listen to each other. This property—emergence of a global coherent behavior without the need of conductor—is particularly interesting when related to the search for the neural correlates of awareness because of the commonly admitted view that there is not a single anatomical module responsible for awareness (Crick and Koch 1990; Engel and Singer 2001; Alkire et al. 2008). In this view, the adequate description level of some cognitive processes would be the whole brain or a substantial part of it (Lashley 1931; Haxby et al. 2001), rather than a functional module, and new cognitive knowledge would be created by between-area interactions, in addition to specific local computations.

Emergence versus Information Transfer?

There are some fundamental differences between the concepts of emergence and of information flow. Let us consider that Mary was in *A* at *t* and in *B* at *t* + 1. If Mary is a commercial traveler and has to meet a client in *B*, then knowing that Mary arrived on time in *B* is sufficient. However, if Mary is a ballet dancer, this description does not help: what matters is whether the other dancers also moved from *A* to *B* or not; in other words, it is the dynamics of the global picture that are relevant. In practice, one or the other strategy may be preferentially used depending on the task to perform. One of the crucial factors is probably the amount of learning and degree of automaticity required, versus novelty and flexibility: a key idea about emergence (or dynamic coordination) is its flexibility, its ability at signaling new relationships.

Feedforward, Feedback, and Recurrent Processing

Emergence and information, as defined here, are concepts rather than experimentally tractable neural mechanisms. How do they relate to the classical typology of neural coordination (feedforward, feedback, and recurrent processing)? The initial wave of feedforward activity has been mainly associated with fast and automatic analysis of the visual scene (Thorpe et al. 1996; Hochstein and Ahissar 2002) and may reflect unconscious processing (Lamme and Roelfsema 2000). Both feedback (Bullier 2001) and recurrent processing (Lamme and Roelfsema 2000) have been related to more sustained levels activity and could be involved in conscious processes. As a first approximation, one could consider that fast feedforward processing is fully automated and, therefore, would not reflect *dynamic coordination* but rather information flowing along a fully prespecified neural route. Conversely, recurrent processing would be more flexible and a signature of emergence. However, there is again another level of approximation, since assessing whether a neural process occurs in a feedforward or feedback manner is always difficult and almost impossible with the noninvasive tools used with humans.

   In the human literature, feedforward processing is often linked to the rapid cascade of early-evoked potentials that is essentially maintained during anesthesia (Alkire et al. 2008). In opposition to the idea that this initial volley of activity is rather automatic, there is growing evidence that top-down expectancies can meet the feedforward stream at very early latencies, before 100 ms (Chaumon et al. 2008; Kelly et al. 2008; Poghosyan and Ioannides 2008; Dambacher et al. 2009). Feedback and recurrent processing are usually associated with longer latencies and with more sustained states. The frequency content of the signal is most often used to describe such sustained states, but late slow waves which can show up in event-related potentials (ERPs), such as the contingent negative variation or the P300, can also be interesting indexes.

If we accept the simplifying assumption that early ERPs reflect mainly feed-forward processing and that oscillations reflect mainly dynamic coordination, a crucial issue is how the two interact. This is only beginning to be addressed experimentally. For instance, several groups showed that the amplitude of the response evoked by the stimulus depends on the phase and/or amplitude of ongoing alpha (Jansen and Brandt 1991; Becker et al. 2008) or gamma rhythms (Fries, Neuenschwander et al. 2001). The mechanisms responsible for these interactions could be a general phase resetting of ongoing oscillations by stimulus onset (Makeig et al. 2002), a controversial proposal (Shah et al. 2004; Mazaheri and Jensen 2006; Risner et al. 2009), or an asymmetry in ongoing alpha oscillations (Nikulin et al. 2007; Mazaheri and Jensen 2008). Last but not least, it seems that the nature and strength of the interaction could depend on the resting state frequency characteristics of each subject (Koch et al. 2008).

Distinct Roles for Different Frequency Bands

In this section, discussion is devoted to brain rhythms, with the underlying assumption that brain rhythms are one of the signatures of dynamic coordination. More precisely, I will focus on how brain rhythms are related to cognitive processing.

Mapping Cognitive Functions onto Frequency Bands

Historically, each frequency band has been associated preferentially to a given type of cognitive or physiological process. Delta waves were associated with sleep, theta activity with memory, and alpha rhythms with vigilance fluctuations, whereas the beta and gamma ranges were associated initially with active awake stages and more recently to feature binding, attention, and memory. It is beyond the scope of this chapter to review all of the literature on all frequency bands (for recent reviews, see Jensen et al. 2007; Klimesch et al. 2007; Palva and Palva 2007; Schroeder and Lakatos 2009; Tallon-Baudry 2009), but a few examples are sufficient to demonstrate the absence of a strict correspondence between a frequency band and a cognitive process.

There is a large and converging body of evidence that grouping features into a coherent percept is accompanied by changes in the gamma range (reviewed by Jensen et al. 2007; Tallon-Baudry 2009). However, the formation of coherent percepts can also be accompanied by modulations of oscillatory synchrony in the alpha range (Mima et al. 2001; Freunberger et al. 2008). The alerting, orienting, and executive attentional networks engaged in many attentional tasks affect oscillatory synchrony in different frequency ranges, from theta to gamma frequencies (Thut et al. 2006; Fan et al. 2007; Siegel et al. 2008). Episodic memory encoding and retrieval typically affects both theta and gamma oscillatory synchrony (Sederberg et al. 2003; Osipova et al. 2006),

but some also report modifications in the alpha band (Klimesch, Doppelmayr, Schimke et al. 1997; Klimesch et al. 1999; Sauseng et al. 2002). Visual short-term memory retention is associated with sustained gamma and beta oscillations that originate from distinct areas (Tallon-Baudry et al. 1998, 2001; Tallon-Baudry 2004), but also with modulations in the alpha range (Jensen et al. 2002; Jokisch and Jensen 2007; Grimault et al. 2009). The historical associations between a frequency band and a cognitive process should therefore be reconsidered: cognitive functions do not map directly onto frequency bands.

It is also important to underline a fact that sounds trivial: the functional role of oscillatory synchrony in distinct frequency bands may simply depend on the functional specialization of the area that generates these oscillations (Tallon-Baudry et al. 2005), much as the functional significance of ERPs depends on the areas that generate them. This might seem like a statement of the obvious, but this simple statement had surprisingly disappeared from human literature on brain rhythms. For instance, local gamma oscillations are observed in a wide range of areas in human intracranial recordings, from visual (Lachaux et al. 2000, 2005; Tallon-Baudry et al. 2005), to frontal (Howard et al. 2003; Mainy et al. 2007), and medial temporal lobe structures (Tanji et al. 2005; Sederberg et al. 2007). It would indeed seem quite unlikely that those gamma oscillations should all reflect the same cognitive function, given the variety of their anatomical location. Intracranial data in humans remain scarce, but fortunately increasingly more MEG/EEG studies include a source reconstruction approach that provides quite a precise localization of increase and decrease of oscillatory synchrony (Grimault et al. 2009; Hillebrand et al. 2005; Hoogenboom et al. 2006; Medendorp et al. 2007; Siegel et al. 2008; Gross et al. 2004; Wyart and Tallon-Baudry 2009).

Subdivisions within Frequency Ranges

Another source of complexity stems from the existence of functional subdivisions within a given frequency range. It has long been known, for instance, that the upper and lower alpha ranges could display distinct functional variations (Klimesch, Doppelmayr, Pachinger et al. 1997; Petsche et al. 1997). More recently, we showed that distinct cognitive processes can elicit gamma oscillations in different locations: grouping and selective attention simultaneously affect gamma-band oscillations, but in distinct subfrequency bands and at distinct locations (Vidal et al. 2006). Similarly, learning and conscious perception are associated with oscillations in the gamma range, but in different subbands and in distinct areas (Chaumon et al. 2008). Finally, varying attention and awareness simultaneously revealed that distinct frequency bands within the gamma range varied separately with visual awareness and spatial attention (Wyart and Tallon-Baudry 2009). It has also been suggested that the detailed frequency content of gamma-band oscillations could encode specific physical features, such as spatial frequency (Hadjipapas et al. 2007) or sound

lateralization (Kaiser et al. 2009). There is no doubt that gamma-band oscillations are influenced by stimulus low-level features in sensory regions (Hall et al. 2005; Adjamian et al. 2008), but whether this still holds true for higher-level areas remains an open issue. In any case, oscillatory synchrony in a given frequency band should not be considered as a single phenomenon, functionally and anatomically homogenous (Tallon-Baudry 2009).

What Are the Relevant Criteria of Frequency-band Selection?

The absence of a direct correspondence between a frequency range and a cognitive function raises a fundamental issue: what determines the preferential use of a given frequency? There appears to be a large flexibility: the same fronto-parietal network (Buschman and Miller 2007) or the same visual (Wyart and Tallon-Baudry 2009), olfactory (Cenier et al. 2009), or audio-visual (Chandrasekaran and Ghazanfar 2009) region can engage into oscillatory synchrony at distinct frequencies, involved in distinct cognitive functions. This flexibility may be subtended by distinct local networks and cellular types: *in vitro* experiments reveal that there are distinct frequencies (20–30 Hz vs. 30–70 Hz) in the infra- and supra-granular layers, respectively (Cunningham et al. 2004; Roopun et al. 2006). It is tempting to suggest that each frequency band corresponds to a specific microcircuitry; for instance, gamma-band oscillations would critically depend on $GABA_A$ interneurons in upper layers. This basic cellular equipment, present in each and every cortical area but in varying proportions, would define natural frequency domains corresponding to the typical frequency ranges, from theta to alpha and gamma. In a given cognitive task, the use of a frequency band would depend on two sets of factors.

The first group of factors relates to the task's physiological requirements. It was initially suggested that frequency depends on the network's size and geometry (Kopell et al. 2000; von Stein and Sarnthein 2000): because conduction delays increase in large network, synchronization takes place at lower frequencies (Buzsáki and Draguhn 2004). Another potentially important factor is the time constant of the biological mechanisms involved (Koch et al. 1996) and the coding precision required (Desbordes et al. 2008). If, for instance, time-dependent synaptic plasticity is required, then a precision of 10–20 ms is necessary (Bi and Rubin 2005; Markram et al. 1997) and the whole network might shift to the gamma frequency range. Finally, the metabolic costs of establishing sustained oscillations may vary between frequency bands. It has been suggested, in particular, that there may be a stronger relationship between the BOLD response and gamma-band activity (Mukamel et al. 2005; Niessing et al. 2005), although the relationship at rest may be more complex (Mantini et al. 2007; Nir et al. 2008). The preferential use of a frequency band might therefore also be influenced by metabolic demands.

A second group of factors can be found in cognitive constraints. First, the time constant of the task is likely to influence the pace of the system: if there

are only 500 ms to complete a visual search, for instance, frequencies below 5–10 Hz are unlikely to be relevant, whereas if there is no time constraint to perform a task, then one might shift to lower frequencies. Similarly, if there is any regularity in the temporal structure of the task, subjects are likely to form windows of temporal expectancies (Tallon-Baudry 2004; Praamstra and Pope 2007; Schroeder and Lakatos 2009). Second, oscillations could be used to define chunks of processing, in which data will be grouped and isolated from those of the preceding and following period. Sensory or cognitive chunks can potentially be created at many timescales; for instance, arbitrary associations can be learned through a wide range of time intervals (Balsam and Gallistel 2009). Examples of chunking can be found in vision and olfaction and have been related to beta-range oscillations (Uchida et al. 2006; VanRullen et al. 2006). Along the same line of reasoning, it would be tempting, although premature, to relate very slow (< 0.3 Hz) oscillations (Monto et al. 2008; Nir et al. 2008) to the "psychological present," the few seconds during which successive events form a perceptual unity and can be apprehended without voluntary recall. Third, another interesting potential constraint is the number of cognitive processes to be multiplexed. Searching for someone in a crowd typically involves retrieval of information about the person from long-term memory, attentional suppression of nonmatching faces, and bottom-up feature-binding processes. One possibility to coordinate these three cognitive processes would be to use distinct frequencies, as detailed in the next section. In this view, the frequency tuning of each process is likely to depend on the total number of processes required by a task.

Integrating between Frequencies

Multiplexing, Integration/Segregation

Analyzing a situation and reacting in an appropriate manner requires the coordination of a number of cognitive processes. As shown above, searching for someone in a crowd, for example, involves a sensory analysis of the visual scene, the recall of a face template from memory, and the attentional scanning of all the faces potentially matching that template. All these cognitive processes have to be integrated into the general search task, but because they reflect distinct operations they should nevertheless remain segregated. The neural correlates of these distinct cognitive processes are likely to show up in different frequency bands, and one way to coordinate them without fusing them is to coordinate activity between frequency bands. Coordination of one frequency band with another would be an elegant solution to multiplex information while keeping a reasonable trade-off between integration and segregation. Besides, because the number of simultaneous frequencies as well as the number of multiplexing patterns is limited, the number of concurrent processes is naturally

restricted to a finite number, in line with the idea that simultaneous tasks might tap into shared and limited resources.

Different Types of Coupling Can Be Considered

Coupling between frequency and/or areas can appear in many different ways. The first candidate that was considered was between-area phase coupling in the same frequency band (Lachaux et al. 1999; Varela et al. 2001). In this view, two or more distant areas oscillate in the same frequency range with relatively constant phase relationships, and there is indeed experimental evidence that such phase coupling between distant sites can occur and play a cognitive role (see, e.g., Tallon-Baudry 2004; Tallon-Baudry et al. 2001; Uhlhaas, Linden et al. 2006; Melloni et al. 2007; Doesburg et al. 2008). Between-area coupling can also appear as amplitude covariation of the signal at the same frequency (Bruns and Eckhorn 2004). Both phase- and amplitude-coupling can also occur between frequency bands, recorded at the same site or at different sites (Bullock et al. 1997; Palva et al. 2005; Meltzer et al. 2008). Finally, a mixed version of phase- and amplitude-coupling occurs when high frequency oscillations occur preferentially during peaks (or troughs) of a lower frequency rhythm. Such coupling patterns have been observed, for example, in rat entorhinal (Chrobak and Buzsáki 1998a) and neocortex (Sirota et al. 2008), as well as in cat visual cortex (Grenier et al. 2001). So-called nested oscillations attracted a great deal of interest because of an influential model of memory storage that would account for the limits of human memory capacity by an interplay between theta and gamma oscillations (Lisman and Idiart 1995), and there is growing evidence in humans for such theta/gamma relationships (Canolty et al. 2006; Sauseng et al. 2008).

Conclusions and Prospects

A picture emerges in which activity in the brain can be coordinated at multiple spatial and temporal scales using different frequency bands. This overall picture is much more complex than the initial attempts at matching a cognitive process with a type of neural coordination, but much more flexible and deemed to have a bright future. Indeed, it suggests that cross frequency coupling could be used to integrate as well as segregate information over behavioral time and cognitive space, a feature that is necessary to obtain a high level of flexibility. In addition, because the biologically constrained circuits that generate oscillations in different frequency bands are limited in numbers, the numbers of different frequencies, and therefore the number of cross frequency coupling schemata, is limited. This inbuilt constraint of the brain would be reflected in the well-known cognitive capacity limitations.