INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

Cognitive Neuroscience Sector,

SISSA-ISAS, Trieste, Italy.

*

Bridging Access to Consciousness, Cognitive Control and Metacognition : Toward an Application to schizophrenia

Sarah Kouhou,

PhD Thesis,

*

Supervisor:

Tim Shallice, PhD, Professor

*

External Reviewers:

Paolo Brambilla, M.D PhD

Jérôme Sackur, M.D PhD

Acknowledgments

I would like to thank all the people who helped me, knowingly or not, to complete that work:

My supervisor, Tim, for his huge patience and his critical empiricism; my aunt, Marie, for having been there when I needed ; my friends or colleagues in SISSA, especially Georgette and Shima, for their kindness; the Abdu Salam Center of Theoretical Physics (ICTP) in general and overall the library staff, just for providing me with a nice space for work, serious, stimulating and convivial. I also thank Guillaume, for having given me energy and force at the very right moment.

I am also very thankful to Paolo, who showed an interest in my proposal to work with patients and who contributed a lot to accessing the Health Center. Elisa Maso, who was very efficient and available to make me interact with patients, although she was very busy. I am also thankful to the patients themselves, who devoted their time and psychological resources, for free.

I thank also all the subjects who participated to my tiresome experiments, with patience, kindness and seriousness.

Table of contents

*

PART II: Behavioral Evidence for non conscious Priming of Cognitive Control Processes? Which effects on metacognitive performance? Replicating and Exploring

**

PART III: Cognitive control load dependent activations in medial prefrontal cortex by invisible but not by visible primes & an overlap between cognition and metacognition

***

PART IV : Cognitive Control, Access to Consciousness and Metacognition in Psychosis, an Observational Study

****

PART VI: Overall Conclusions, future research

*****

PART I:

Bridging Access to Consciousness, Cognitive Control and Metacognition, A dense overview

From the point of view of the contemporary cognitive neurosciences, cognitive control and consciousness are two clearly distinct notions, each associated with an independent field of investigations and its own conceptual tools. And so it has been through the recent history of the cognitive or psychological sciences: consciousness has long been a concept somewhat monopolized by the psychoanalytic theory, even though that theory has never told anything about consciousness proper, but was only the first one to hypothesize the existence and some properties of the so called Unconscious (Freud, 1896 and later until 1933). Almost in parallel, the notions equivalent to executive control were more developed in the fields of mathematics, computer sciences (Turing, 1936) and what will be become the cybernetics before becoming a paradigm and being introduced in cognitive psychology (Broadbent, 1958) and neuroscience.

Throughout the history of sciences, one generally observes a trend of disciplines and of their concepts to become independent of each other. Thus Physics acquired its modern sense, became independent and completely emancipated from Natural Philosophy with Newton's Principia (1687). But one can also observe the reverse process: two phenomena that used to be considered as independent finally turn out

to be interacting one with each other, and are integrated into a unique model. The best instance is the interaction between electric and magnetic forces. By showing that an electrical current or a temporal variation of electrical current induces a magnetic field (and vice versa), Oersted (1820) demonstrated in the same vein that these physical forces may have a common deeper physical origin. That allowed Maxwell (1864) to propose an unification of the two theories.

Could a unification of models of consciousness and cognitive control occur in the cognitive neurosciences? Such a unification of the models is now possible as, for more than ten years, the phenomenon of Consciousness has been conceptualized within the framework of a specific functional architecture (namely the Global Workspace, see section 2 below). Therefore, that notion can no longer be effectively characterized without referring to this architecture, because (i) the entry of information into the global workspace is equivalent to the access of information into consciousness and, importantly, (ii) the (non) access of information is explained by some properties of that architecture – properties of a central executive system.

On another side, there exists a model of executive functions (Supervisor Attentional System, Norman and Shallice, 1986, see section 3) that, despite the appearances, is NOT equivalent to the Global Workspace although it has some properties in common with it. These (common) properties precisely are those that account for that information accesses (or does not access) to consciousness. In effect, within the SAS framework, the operations that one is held to be conscious of correspond to interactions between the SAS and another downstream system, namely the contention scheduling system (see section 3 for more details). In other words, the SAS framework explains consciousness as a phenomenon emerging from the interactions between two specific subsystems. The space of these interactions would correspond to the global workspace.

In the continuity of that perspective, one can reason that, if consciousness emerges from executive control processes, then manipulating some parameters of cognitive control (which is now possible thanks to the current hierarchical models of cognitive control) should allow one to observe some effects on access to consciousness. Conversely, one can assume that manipulating the accessibility to

consciousness of some task-relevant signals should influence the performance in an executive control task.

Thus, keeping in mind this underlying aim, in this first chapter, I am going to outline models and empirical evidence that suggest a unification of the theories of consciousness and cognitive control. I obviously do not hope to propose such a model, just to produce a small step forward. This first introductory chapter only aims to provide the reader with a minimal knowledge, combined with a necessary conceptual and methodological toolbox, in order to present my perspective on these topics, and so, to understand the underlying purpose of the empirical reports which follow. To resume the analogy with magnetism and electricity, my aim is to find out some experimental situations in which the forces could interact.

My exposé will somewhat follow the recent historical evolution of cognitive psychology and neurosciences, and I will try to show how each notion calls the other one. An important step was made by Antony Marcel (1983), whose pioneer empirical works somewhat 'killed two birds with one stone', since he indeed broke the psychoanalytic monopoly of the notion of unconscious, introduced a new topic in the experimental psychology, and for that had to develop new methods. The cognitive sciences of consciousness have thus begun and grown independently of the psychology of the so called 'superior cognitive functions' –that were born well before.

In effect, Broadbent (1958), Atkinson and Shiffrin (1968), Shiffrin and Schneider (1977) had drawn the distinction between *automatic* versus *controlled processes* and introduced the concept of *selective attention.* Then, attention has been extensively studied by Michael Posner, who emphasized its central role for the cognitive control in a paper explicitly titled *Attention and cognitive control* (1975), and then in his book *Selective Attention and cognitive control* (1986). One will see further how important this notion of selective attention is, as far as it could bridge the most recent models of cognitive control and access to consciousness.

In 1986, Norman and Shallice proposed their model of executive functions (Supervisory Attentional System, or SAS). In 2001, Jack and Shallice carried out an additional step toward a functional link of

cognitive control and consciousness, since within this framework the mental operations corresponding to interactions between the Contention Scheduling system and SAS give are conscious (and therefore reportable).

Independently, Bernard Baars, 1989, carried out a step forward, but in the field of consciousness, with his theory of the Global Workspace (see below, Consciousness as a global workspace), which reminds one of the SAS, since in that framework, the access of information into consciousness is equivalent to its entry into the Global Workspace, or shall we say SAS. A core idea of that model is that accessed information can be broadcast, sent and exchanged with other modules.

Dehaene and Naccache, 2001, took up this central idea of the Global Workspace theory and extensively contributed to the fleshing out of the theory. They even put two new stones to the bridge between consciousness and with cognitive control by considering: (i) the central role of the selective attention, insofar it stands as the fundamental mechanism by which information accesses or does not access the Global Workspace and (ii) the intrinsic *serial modus operandi* of selection, let it be attentional or motor. The interest of such properties is that they allow one to capture, or even to predict, the distinction between (consciously) accessed and non accessed stimuli.

On the side of the cognitive control functions, growing empirical evidence had been accumulated in cognitive psychology, neuropsychology and especially in electrophysiology (including Jacobsen, J. Fuster, P. Goldman-Rakic, E. K. Miller) that allow one to decipher the functional topology of the prefrontal networks and subregions, held to correspond to the SAS. On the basis of these works, of which contents are important to understand what follows, emerged different theories of prefrontal functions (attentional, mnemonic...) and in particular some critical notions: hierarchy, modular segregation, and temporal integration. These notions, now integrated in the contemporary models, constitute the keys to the systematic comprehension of how the cognitive control is implemented within the prefrontal cortex (Koechlin, 2003; Badre, 2007).

It should be noted that most recent models of cognitive control are well established, even quantitative for one, but are unrelated to and developed independently of the Global Workspace theory of

consciousness.

Can they be linked in future research? And how could they be? It will be suggested that the analysis of metacognitive measures can be a way to relate them. Metacognition will be thus the fourth and last section of that introductory chapter, and will hopefully close the loop between cognitive control and access to consciousness.

1. Demonstrating the existence of unconscious processing before all:

The earliest empirical studies of consciousness in psychology were interested in demonstrating the existence of non conscious information processing in the domain of perception, using principally experimental techniques of masking associated with priming (Marcel, 1983; Holender, 2004).

The technique of *masked priming* has indeed become one of the most common and now widespread ways to demonstrate the influence of subliminal stimuli on behavior. The *priming* technique consists in displaying a stimulus (prime) before the target on the basis of which one has to respond, in such a way that the target-based response selection can be facilitated (faster or more accurate) or disturbed (slower or less accurate).

Thus, for example, when one has to decide whether a target letter is a vowel or not by pressing different keys corresponding to yes or no, one's response is faster when the target letter is preceded by the same letter (so called *prime letter*) and slowed down if the letter prime is different –if the target is a vowel and the prime a consonant for instance. Note that the facilitation/perturbation of the target-based decision depends on the relation between the prime and the target property relevant for the task.

The interest of such a technique is that it allows one to tap into the decision making process at different levels of processing, according to the degree of abstraction of the property shared by the target and the prime. If prime and target are physically identical, one will speaks of *sensory* priming ; if they share the same spatial position, of *spatial* priming ; if they share a semantic property, of *semantic* priming and so on… One can even speak of *cross-modal* priming, if the prime is accessed in an auditory format, and the target in a visual one (or the contrary).

Priming effects can be observed even when the prime is below the threshold of awareness – that is to say, in the case of visual stimuli, invisible. The most common way to make a visual stimulus

invisible or subliminal is the so-called *masking* technique, popular since the pioneer works of Anthony Marcel (Marcel, 1983). Basically, a stimulus is visible when briefly displayed (for 33 ms) in the periphery of the visual field, but becomes invisible when it is immediately followed by a second stimulus, namely a *mask*. A parameter critical to determine the degree of visibility of the stimulus is the time interval between the stimulus offset and mask onset (or Inter Stimulus Interval, ISI), or the time interval between stimulus onset and mask onset (or Stimulus Onset Asynchrony, SOA).

At the neuronal level, it seems that the mask selectively interrupts the recurrent interactions between the primary visual cortex and extrastriate areas, preventing the broadcasting of the signal (Lamme et al, 2002), and consequently the "access of information to consciousness" – the expression is put between quotes because it remains to be defined and will be in the next section.

Considerable evidence exists that demonstrates that, masked stimuli can influence behavior and brain activity at different levels of processing, namely sensory (Grill-Spector et al., 2000), attentional (Naccache et al., 2002 ; Woodlan and Luck, 2003; Kiefer and Brendel, 2006; Koch and Tsuchiya, 2007; Bressan and Pizzighello, 2008 ; Kentridge et al., 2008), semantic (Gaillard et al., 2006 ; Van der Bussche et al., 2009) or motor (Dehaene et al, 1998), while subjects report *not having seen them* (Debner & Jacoby, 1994 ; Kouider and Dehaene, 2007).

Is the fact that the subjects report not having seen some stimulus or seeing only a flash/flicker, sufficient to consider that the information is not consciously perceived? What can warrant that subjects are able to rate their own visual perception, and moreover without any bias?

That question seemingly is not trivial since it is still debated, and is at the origin of an operational and rigorous definition of "a consciously perceived stimulus". If consciousness has a functional relevance and reality, then conscious perception should allow some specific behavioral outcome or performance. And consequently, a criterion for consciousness should be possibly defined on the basis of overt behavior. Can one find an objective criterion, related for example to the quality of the response accuracy itself?

The classical toolkit of *d-prime* imported from Signal Detection Theory and applied to psychophysics,

cannot be relevant when applied directly to the discrimination of targets. For the simple reason that one will observe a certain sensitivity, a certain ability of the perceptual system to extract a signal out of a noisy background, (revealed by the d-prime) even when subjects report having not seen the stimulus. This is because the accuracy of the subjects turns out to be far above chance level that one is justified to conclude that an unconscious processing of information takes place (see for instance Jacoby, 1991). BUT one can nevertheless observe another critical behavioral difference susceptible to be formalized by a d-prime: the confidence of the subjects in their own perception (or perception based response) seems to be the main difference.

The fact of being aware of a stimulus (a visual target for instance) that one has to respond to, allows one to self-evaluate one's performance with a high degree of confidence. On the basis of that reasoning stands the idea that conscious perception allows *metacognition*, so that one can model the confidence of the subjects regarding their performance with a second-order d-prime (meta d-prime). Basically, the meta d-prime (Rounis et al, 2010; Rosenthal and Lau, 2010; Maniscalco and Lau, 2011). Maniscalco and Lau, 2011) is a normalized index that reflects how well the confidence (second-order or type 2 decision) of the subjects predicts their objective performance (first-order or type 1 decision). It is related to models of second-order decisions. As outlined by Pleskac and Busemeyer (Pleskac and Busemeyer, 2010), parameters involved in first-order decision seem to be insufficient to account for the outcome second-order decisions, such as confidence: "*More broadly, any hypothesis positing confidence to be a direct function of the diffusion model parameters (δ, θ, z) will have difficulty predicting a difference between correct and incorrect trials, because these parameters are invariant across correct and incorrect trials*." In effect, if one assumes that these two consecutive decisions involve two different systems, then one must consider that *d'≠meta-d'*. Concretely, the sensitivity index *d'* is calculated in a standard fashion[1] when performing a type 1 task (discrimination for example), then a *meta d'* it is applied in a type 2 task where the subjects have to self-evaluate, after each trial, their own accuracy in the type 1 task[2]. One can thus measure how accurate they are in self evaluating their own accuracy. This notion (of

[1] That is to say d'= Z(hit rate) – Z(false alarm rate), where $Z(p)$, $p \in [0,1]$, is the inverse of the cumulative Gaussian distribution.

[2] Note than the analytical formula of the meta d' is analytically different from the *d'*. For more details, see Galvin and al, 2003 and Maniscalco, 2011.

metacognition) will be further developed, but it is important to note that accuracy and meta-accuracy are dissociable: one can be very bad at performing a given task, but accurate in valuating one's performance.

Actually meta-d' is an index that could theoretically be applied to other types of metacognitive task. As an alternative to visibility scaling to rate consciousness, Persaud et al, 2009 proposed a post-decision wagering (PDW) method. However it will not be considered here for several reasons related to the fact the wagering decisions of the subjects do not only reflect their confidence, but also of the contribution of (reward related) processes, which are involved in wagering. It has been demonstrated that these processes could influence response selection even when the incentives are subliminal (Pessiglione et al., 2007). Moreover, the reward mechanisms can selectively influence the betting outcome whereas the (perceptual) confidence remains stable. Finally, the PDW method also rests on the assumption that risk-seeking/reward-seeking biases are homogeneous among subjects or population, which is a very strong and risky assumption easy to show it is false.

Thus, that method can introduce a strong bias and act as confounder, without reflecting the awareness or the confidence of the subjects. As Maniscalco and Lau (2011) point out, this method does not entail any difference in the type 2 ROC curves of the subjects.

The details of these empirical data and theoretical discussions will not be exposed nor discussed because it is outside the scope of the chapter. We will content ourselves with saying/assuming the existence of non conscious information processing is no longer a hypothesis nor something that needs to be demonstrated, so that the topic that becomes more intriguing is Consciousness itself, or rather should we say the nature, including the specificity, of conscious processing.

In effect, within the past decade, while the depth of the non conscious processing was being explored, the center of interest has been slowly moving from the characterization of unconscious processing (Marcel, 1983 ; Holender, 2004) to the issue of the role and specificity of Consciousness in terms of behavioral and neural correlates (Dehaene and Naccache, 2001), and from an information processing point of view, that is to say in terms of cognitive processes entailing/necessitating consciousness, or Type-C processes (Jack & Shallice, 2001).

In this context, although the functional relationship between Cognitive Control functions and Consciousness is very intuitive, it has become the focus of an active line of empirical research in cognitive psychology and neuroscience for some years only (Naccache and Dehaene, 2001 ; Jack and Shallice, 2001).

2. Consciousness as a capacity-limited Global Workspace

The so-called *Global Workspace Theory* (Baars, 1989 ; Dehaene and Naccache, 2001), that assumes a certain architecture of the mind, functional but also neurophysiologically hardwired, has become the most prevalent model of consciousness.

The Global Workspace model is grounded on several assumptions, among which the most important one is the existence of central decision making system, named *Global Workspace,* of which two characteristics are of major importance.

First, it is distinct from other *peripheral* networks, peripheral in the sense that they are specialized for the processing of a particular type of information, that can be more or less abstract (orientation of contrast lines, visual movement, pitches, semantic, face-related for example). Yet, the global workspace is said to be *central*, because it is *amodal* in terms of input and output modalities. Information flow is processed according to its relevance for the pending decision making, independently of its format. Plus, a second important property of the global workspace is its *serial modus operandi*: it is capacity-limited in that sense that only *one* decision (that is to say stimulus based selection) can be made at the same time.

That global workspace can be conceived as a neuronal network, implemented with a specific kind of reciprocally connected neurons. At a given moment that network enters into resonance with another (peripheral) module. This entering in resonance corresponds to the access (see figure A below). The access of information onto consciousness is held to be perfectly equivalent to the entry of information into this global workspace.

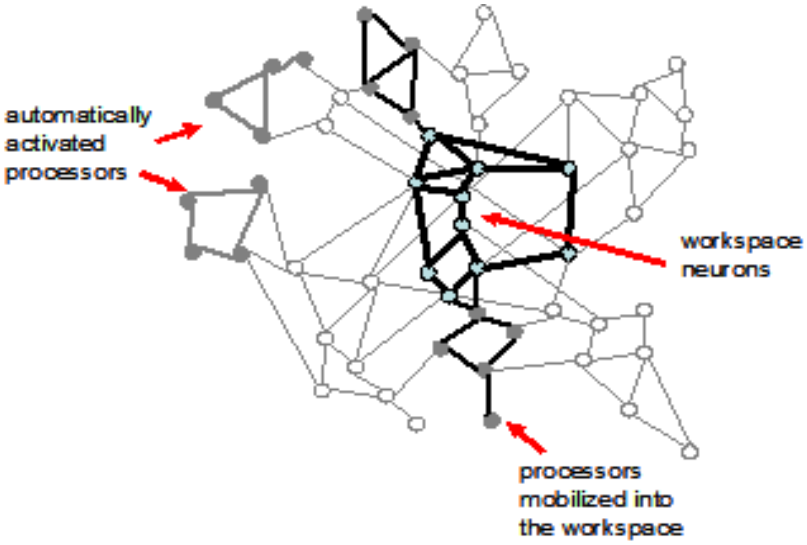


figure 1-A : schematic representation of global neuronal workspace (central blue dots), implemented in long axons neurons distributed in prefrontal, cingulate and parietal cortices, able to be punctually connected with peripheral specialized modules (grey dots). Borrowed from Dehaene & Changeux, Neural Mechanisms of Access to Consciousness, in The Cognitive Neuroscience, Gazzanigga et al., 2003.

To resume the example of the masked stimulus, it indeed triggers a neuronal activity in the primary visual cortex – a module situated outside the global workspace and of which activity is necessary but not sufficient to be associated with a (conscious) visual experience: that activity being very soon interrupted by the mask, the signal does not propagate upstream until reaching the global workspace or "consciousness".

The immediately following question concerns the conditions for information to access into the global workspace, by which mechanisms.

A *"bottom up" mechanism* is of course necessary that would consist in the simple activation of a peripheral module, with a sufficient energy to mobilize the automatic attentional processes. But as said before, this is not a sufficient condition. That module must be anatomically connected to the global workspace and a *top-down mechanism of attentional amplification* has to mobilize (Dehaene and Changeux, in Gazzaniga et al. Ed, 2003) and maintain it.

Although the access is restricted and that the central decision making system works exclusively serially, it is held to carry out a certain kind of manipulation of information which is not possible when information remains subliminal or non accessed (Dehaene & Naccache, 2001). Actually, the specific

nature of the operations made possible by the access of information is not precisely defined within the Global Workspace theory – as said above, the specificity or functional role of consciousness still constitute an open and debated question and even a hot topic in cognitive neuroscience of consciousness.

One can assume, as I do in the content of that thesis, that it allows metacognition, so the possibility of producing recursive knowledge (see section *4. Metacognition*).

Despite that imprecision, the model captures some well-known phenomena that corroborate the existence of a central system, working exclusively serially and capacity-limited : including (i) access to consciousness, (ii) bottleneck effects during two consecutive decisions and (iii) the attentional blink.

Moreover, the advent of neuroimaging has allowed one to individuate a neural network associated with all these aforementioned, including the (non) reportability of the stimulus by the subjects.

Finally, time-resolved techniques (ERP/MEG) provided insights regarding the temporal dynamics of access to the global workspace (Sergent et al., 2005; Del Cul et al., 2007).

2.1 Three pieces of behavioral evidence for a central Global Workspace

In addition of the access of the information into consciousness that we already evoked above, two pieces of behavioral evidence are captured by the model.

The bottleneck of the mind, giving rise to the phenomenon of the *Psychological Refractory Period* and discovered by Welford in 1952, refers to the fact that when one has to make a decision D1 on the basis of a stimulus S1 and *immediately* after, a decision D2 upon a stimulus S2, the response time to the second stimulus is systematically delayed, as if there was an *incompressible time interval* before the end of which the second stimulus cannot be processed. That time interval, named *psychological refractory period*, is thought to correspond to the time necessary for the decision system to reset or disengage/reengage its resources, from one task-set to another one.

In a protocol manipulating the time interval between S1 and S2 (Stimulus Onset Asynchrony, or SOA) and response complexity as well, it has been demonstrated on the basis of fine-grained analysis of the reaction times distributions, that the perceptual and motor stages could be processed in parallel, and

that the delay of the second decision D2 could only be due to a serial central stage of processing (Pashler, 1998; Sigman and Dehaene, 2005; 2006).

The study of the *Attentional Blink* phenomenon has also provided a source of interesting insights (Sergent et al., 2005). First described in 1992 by Raymond, Shapiro and Arnel, the attentional blink is a phenomenon observed in rapid serial visual presentation (RSVP). When a sequence of visual stimuli is displayed in rapid succession at the same spatial location on a screen, and with the subjects having to identify two different visual targets (two letters in a stream of digits for instance) that are inserted into the stream of stimuli, one can fail to detect the second target if one has already detected the first one.

Exactly as for the aforementioned psychological refractory period, or "bottleneck of the mind", what is critical in this protocol is the *time interval between the target onsets* (SOA) : it is precisely when the targets are too close in time (about 100 ms between the *offset* of the first target and the *onset* of the second target) that one more often fails to detect the second target. In other words, if the first target enters the global workspace, the second target cannot *before an incompressible time interval*. In a temporal point of view, the attentional blink is to consciousness what the blind spot is to the retina, and suggests that the conscious (visual) scene is temporally discontinuous despite the fact that we perceive it as a continuous stream.

This phenomenon is not fully explained yet, but other data suggest it can be or has to be interpreted in terms of central bottleneck of the (conscious) mind, and may be due to the time necessary for the central decision making system to disengage and then reengage the attentional selection processes.

2.2 Neural correlates of a central Global Workspace: a causal role of prefrontal cortex?

The concept of "consciousness of sensory information" has been shown to have a neurophysiological relevance and a signification. Both seen and unseen stimuli have been associated with neural responses, but the corresponding patterns of brain activations differ from a double point of view, namely (i) spatial (neural networks effectively recruited) and (ii) temporal (temporal dynamics of brain responses).

(i) By using functional neuroimaging techniques and determining some minimal contrasts between accessed versus non accessed information, it has been shown that unseen stimuli can nevertheless influence the brain activity at several different levels of processing, including sensory, attentional, semantic or motor (cf. review by Dehaene & Changeux, 2011).

The access of information to consciousness, allowing the subjects to report the stimulus *accurately* and *confidently*, and measurable by a meta-d' (Rounis et al, 2010), has been reported to be associated with the activation of a large network including the *prefrontal, parietal* and *anterior cingulate* cortices, reciprocally connected at large-scale by long-axon neurons (Dehaene, Sergent, Changeux, 2003). As said earlier, that network is *amodal*, it becomes activated independently of the sensory modality, including visual (Dehaene et al. 2001), or tactile (Boly et al., 2007).

Note that neuroimaging studies that have investigated the bottleneck effects reported activations of part of that network, especially the dorsolateral prefrontal cortex (See Marois and Ivanoff, 2005 for a review) – compatible with the *serial modus operandi* of the global workspace.

(ii) In terms of the temporal dynamics, a noteworthy characteristic of non accessed stimuli is that, in event-related potentials, they elicit the same early wave components (until 150 ms, which reminds one of the time interval length involved in the PRP) as the accessed ones –in the occipitotemporal pathway for example. However, on trials when they access the global workspace (that is to say when subjects are able to report the stimulus, *accurately* and *confidently*) one typically observes a late (200-300ms after stimulus onset) and seemingly all-or-none activation of the aforementioned prefronto-parietal network, an amplification of sensory activity concomitant to that activation, and a late global P3b wave component (Melloni et al., 2011).

An EEG study has recently reported a late amplification of broad-band power in the gamma frequencies; an increase of long-distance phase synchronization, particularly in the beta frequencies (Gaillard et al., 2009). Interestingly, the same study reported unidirectional Granger type causality relations from the frontal cortex to the occipital one, and occurring between 200ms and 450ms after stimulus onset, suggesting a significant top-down component when information is accessed and consequently a possible causal role of prefrontal cortex in access.

These studies on the neural implementation of the Global Workspace theory gave rise to new predictions. The hypothesis of a causal role of prefrontal cortex tends to be corroborated by some studies. These include a higher threshold of access to consciousness reported in frontal patients (Del Cul et al., 2009), in healthy subjects after TMS onto dorsolateral prefrontal cortex (Rounis et al., 2010) and in schizophrenic patients (Del Cul et al., 2006), known to present abnormalities in anterior cingulate and dorsolateral prefrontal cortices.

To sum up and close this section about consciousness and the global workspace theory, increasing evidence converge toward the existence of a central decision making system, distributed across a fronto-parietal network. In this framework, the access is conceived as the sudden late activation of that network, conjoint with its entering into resonance with different sensory and associative cortices (depending on the content or properties of the pending stimulus). That putative neural signature of the access to consciousness correlates with the behavioral signature of consciousness, namely the confident and accurate report of the stimulus by the subjects.

Importantly, the access/non access to consciousness could or should be explained by the properties of that Central system. These include in particular *restricted attention selection* and *serial modus operandi* --other properties remaining to be discovered and described at different levels of explanation (from molecules and neurons to behavior). These properties might critically depend on the prefrontal cortex, insofar as recent data suggests it, it might play a causal role in the access – causal being here defined as Granger causality (Granger, 1969).

In the following sections, we are going to keep focusing on the prefrontal cortex, but not in relation with its possible implication in access to consciousness. We will consider it from the point of view of the architecture of cognitive control instead.

3. Executive Control *per se* and its links with consciousness

Among the set of regions associated to the Global Workspace, of particular interest is the prefrontal cortex, both medial and lateral. This is not only because it might play a causal role in the

access to consciousness, but also because it is involved in the cognitive and motivational control of behavior, more commonly referred to as *executive control functions*.

3.1 How to define executive control and how to define its link with consciousness

A major problem for a neophyte wishing to enter into the current research in *executive functions* may be conceptual. According to some 'accessible' definitions that can be found in Wikipedia[3] for instance, executive functions refer to a set of specific processes that manages other cognitive processes, thus allowing one to adapt to novel or unpredicted situations, managing to tasks simultaneously, correcting oneself in case of error, coping with contexts requiring an increase of attention. In the cognitive psychology literature, these skills are held to include attention, error monitoring, flexible behavior, action inhibition, decision making, task setting**.** That list obviously refers more to behavioral features than to brain mechanisms. Probing the underlying neural architecture making these skills possible has been an object of the models of cognitive control. The difficulty may lay in the fact that there is no direct term to term correspondence between the behavioral skills and the underlying neural system proposed by the model. For that reason, I will introduce the key notions by following the "historical" steps of their formulation.

Regarding the issue of the functional links between executive control and consciousness, one must be extremely cautious when reading literature about this. These questions began to be (explicitly) investigated 15 years ago (Eimer and Schlaghecken, 1998), whereas the most important empirical results have been acquired after 2001, namely la bit more than 10 years ago. Things may indeed be very confused insofar one often finds studies that fail to specify which particular information one is (not) conscious of. A typical instance is the confusion between the cognitive operation (that one can be conscious of) and a distractor or any stimulus (of which one is not aware) having a subliminal influence on that same operation. Or vice versa. Dissociating both aspects is fundamental and is not always done, and one even finds some reports of non conscious processing based on the simple d-prime (Lau and Passingham, 2007).

---

[3] On wikipedia one currently finds the following definition "**Executive functions** is an umbrella term for cognitive processes that regulate, control, and manage other cognitive processes, such as planning, working memory, attention, problem solving, verbal reasoning, inhibition, mental flexibility, task switching, and initiation and monitoring of actions. The **executive system** is a theorized cognitive system in psychology that controls and manages other cognitive processes. It is responsible for processes that are sometimes referred to as executive functions, **executive skills**, **supervisory attentional system**, or **cognitive control**."

We are going to present Norman and Shallice model, insofar as it postulates a global architecture and some links with consciousness are sketched. Then we will briefly outline more recent models of cognitive control, presupposed to be based in lateral prefrontal cortex (Badre's model and Koechlin's model).

3.2 Cognitive Control and Global Workspace architectures: Strict equivalence or just partial overlap?

Although in cognitive neuroscience and psychology, Consciousness and Executive control are two independent concepts, phenomena and fields of research, it is worth observing that that even in folk psychology, the notions of consciousness and cognitive control are linked one with the other – presumably because of common underlying neural mechanisms. Although that intuition is very strong, it still remains very vague or controversial from the point of view of the contemporary cognitive neurosciences, because of an intrinsic conceptual difficulty. Two theoretical alternatives indeed are possible and competing one with the other, namely what we4 can call an *ontologist* conception (*access to consciousness is independent of the cognitive control processes*) versus a *reductionist* conception (*access to consciousness is equivalent to, or emerges from, some cognitive control processes*).

That is not only a terminological issue, because different conclusions can be drawn from these positions. It will not be considered though. In my view, this debate no longer makes sense nowadays. By assuming some functional properties of the mind (such as the bottleneck of the mind), the Global Workspace theory, indeed is sufficient to compromise the ontologist position, since it allows one to predict the access or non access of information into consciousness. Therefore, only the reductionist approach will be thus retained and considered.

The SAS model, by Norman and Shallice (1986), is such a reductionist model. It mainly is a description of the executive functions, but also has emphasized/hypothesized the existence of tight functional dependence between some mechanisms and access to consciousness (type C processes, see Jack and Shallice, 2001). That last formulation can sound nonsensical. If the access to consciousness is equivalent to some cognitive control mechanisms, how can these mechanisms depend on consciousness? This is the point.

The model consists in a tripartite architecture (see below for a more detailed description): triggering data base, contention scheduling system, and supervisory attentional system (SAS).

Within this framework, The Contention scheduling is a bottom up automatic system, which can be activated by signals one is not aware of. However, the SAS is recruited according to the complexity, the degree of learning of the task or predictability of the actual (sensory or reward) outcomes, and the quantity of effort entailed by a given goal-directed action or behavior, but cannot manipulate subliminal information.

Importantly, it is not the SAS operations *per se*, but the interactions between the SAS and the Contention Scheduling System that are by definition conscious, reportable, and that correspond to the access of information into consciousness. These top-down interactions make possible other kinds of mental operations –namely metacognitive operations. In that respect, the access to consciousness makes possible metacognition. That point will be resumed later.

For now, simply note that in the common sense, one's actions are usually scaled between two extremes, from involuntary (spontaneous, automatic, even uncontrollable, associated with a seemingly partial awareness of one's decisions) until fully willed (deliberated, planned, controlled and associated with awareness of one's decisions). Of course, the apparent (full or partial) awareness associated with one's phenomenological experience of action reflects the *degree of confidence* regarding one's own (self)perception. This notion of confidence or of certainty about one's own cognitive processes is the hallmark of metacognition, and has been exploited since the first experimental studies of consciousness in order to demonstrate the existence of a non conscious processing (by contrasting a low confidence level in seeing some stimuli with a high discriminatory performance). That point will be considered later, in the section about metacognition.

These differences in the subjective experience of actions presumably lead to suppose the existence of different functional underpinnings. One source might be called *bottom-up*, in the sense that some schemes, including actions or thoughts, are instantaneously or automatically triggered by the context or salient stimuli, such as routine behaviors and procedures (this is the contention scheduling system), whereas the other one might be said *top-down*, meaning that the outputs are selected on the basis of more complex rules, and/or following a slow accumulation of evidence --such as those performed in non-routine, risky or conflicting situations (SAS). One generally attributes the predicate *intentional*

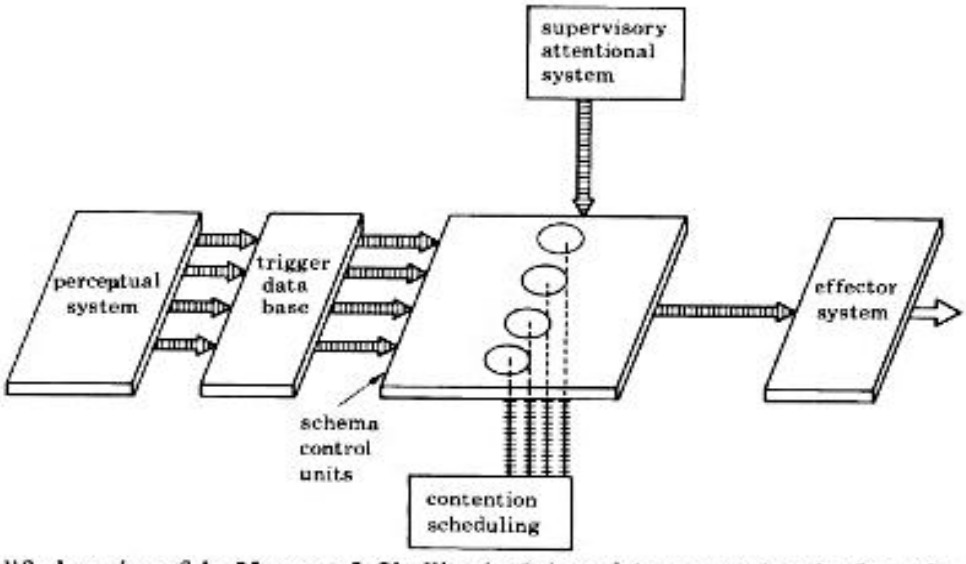
to actions and thoughts originating from the latter source, namely when the Supervisory Attentional System is effectively recruited, in other words when information – including external cues and stimuli, internal signals such as goals, rules, expected rewards, memories, relevant to drive the behavior, seems to be reportable by the agent in a confident way.

Now we are going to present that model in its main lines and concepts. One will see that the tripartite conceptual structure of the model fits almost perfectly a tripartite subcortical - lateral prefrontal network, articulated by the anterior cingulate cortex (medial prefontal) in the midst.

3.3 Cognitive control as a Supervisory Attentional System

3.3.1 Functional properties of the contention scheduling/SAS architecture.

In their seminal paper, Norman and Shallice (1986) proposed a fine-grained [tripartite] architecture implementing the different functional stages by which outputs are generated, from the completely automatic ones (such as walking), until conscious, costly and deliberate ones. That taxonomy rests on the differential recruitment of what the authors named the Supervisory Attentional System (SAS) (cf. Figure B, borrowed from Norman & Shallice, 1980).



A simplified version of the Norman & Shallice (1980) model representing the flow of control information. The lines with arrows represent activating input, the crossed lines represent the primarily mutually inhibitory function of contention scheduling. The term 'effector system' refers to specific purpose-processing units involved in schema operation for both action and thought schemas. In the latter case schema operation involves placing information in short-term stores that can activate the trigger data base.

figure 1-B : SAS architecture, by Norman and Shallice

In this tripartite architecture, designed in order to fit both experimental data and subjective experience of action, the SAS operates by acting on another downstream system, namely the 'contention scheduling system'. It consists in a 'matrix' of schemes of thoughts and actions, some of which being mutually exclusive or overlapped.

Let's note that these schemes are more abstract than purely motor codes, and can by the way be subject to recombination through learning. The activation of a schema depends on two downstream stages of processing: (i) a perceptual stage, since internal and external perceptual cues must *in fine* converge toward a single or several possibly competing schemes, (ii) a 'trigger data base' stage, an intermediary interface presumably linking sensory evidence with restrictive internal variables, such as expected reward, pending resources of the system and the net cost associated with each scheme.

This latter system participates in considerably reducing the number of schemes and sequences upon their relevance/rewarding value, standing as a predictor of optimal action selection by associating schemes with (intended) outcomes.

This architecture is not monolithic but a complex system that needs to be characterized in a more fine-grained functional description – as we will see in the next following sections. However, it is worth mentioning because of the conceptual distinctions it introduced, and the fact that it held some core properties that have been conserved by both models of consciousness and of cognitive control functions.

First, and in particular, due to the inner structure of the system, there is a fundamental constraint consisting in allowing to a single schema to access and monopolize the effectors or motor resources, giving rise to a *bottleneck,* that's to say an intrinsically serial *modus operandi,* of the decision process. The existence of such a bottleneck explains the competition between schemes or action selection.

Secondly, the contention scheduling system can work alone, and even without consciousness. However, when there is noise, entropy, risk or competition, the system requires the SAS intervention in order (i) to elevate or bias the threshold of decision, (ii) to detect possibly co-activated schemes and (iii) to inhibit or maintain them on purpose until a decision is made. In this case, a decision is made as soon as a schema crosses the enhanced threshold and wins the monopoly control of effector resources.

3.3.2 Hardwiring the contention scheduling/SAS architecture.

Triggering data base

It has been showed that ventral striatum, including *nucleus accumbens* and *ventral pallidum*, constitutes an automatic 'low level' functional path implicated in releasing energy in response to incentive cues, even when they are displayed at below a subliminal threshold (Pessiglione et al., 2007). Both the ventral and dorsal striatum receive direct input from dopaminergic neurons of VTA and from substantia nigra, respectively.

The role of dopaminergic signalling in this architecture is noteworthy in that both tonic and phasic dopaminergic signals have been linked to the energization of attentional and motor systems and to the modulation of the threshold of action selection (Niv et al., 2006; Daw & Dayan, 2002).

These structures remind one of the *'triggering data base'* component that drives action selection. It has been defined as an intermediary accumulator which integrates information concerning variables such as expected reward, or opportunity cost – *id est* net attentional and motor costs of the schemes being given the context.

Contention scheduling

The SMA, part of the supplementary motor complex, receives input from the pre-SMA (Nachev et al., 2008). It is not connected to prefrontal cortex but only with spinal cord and primary motor cortex, is strongly recruited during preparation and execution of action. Whereas primary motor cortex is known to encode first order parameters, such as force and direction, the SMA encodes more abstract properties of the movement, especially relative to kinematic parameters, such as speed, order and duration (Tankus et al., 2009). The stimulation of SMA was observed to evoke both movements (consisting of slow postural changes involving several muscle groups, complex motor patterns such as stepping, or even merely the urge to move) and automatic inhibition of motor plan.

In addition, and importantly, SMA and motor cortex can be activated and cause output without any consciousness of the target (Dehaene et al., 1998 ; Sumner et al., 2007), but also give rise to unconscious movements (Desmurget et al., 2009).

The *Contention Scheduling system* might thus include the supplementary motor area (SMA), caudal

dorsal premotor cortex, and the primary motor cortex.

Supervisory Attentional System

Finally, the ***SAS*** is mainly implemented in (lateral and orbital) prefrontal cortex of which the hierarchical architecture has been well captured by more recent models of cognitive control (Koechlin, 2003; Badre, 2007). These models will be described below, in a following section dedicated to the hierarchical models of cognitive control.

One must also include the pre-SMA, the most anterior part of SMA, a tiny network with reciprocal connections with the dorsolateral prefrontal cortex (DLPFC). The DLPFC is recruited proportionally to the charge of contextual executive control and working memory, and typically after an error of commission – or even only risk of error (Swick and Turken, 2002). This posterior DLPFC recruitment correlates with an adjustment of performance and a slowing of time response in the next trial.

Converging evidence suggest that pre-SMA does not implement motor parameters, but rather plays a supervisory role by gating the execution proper, and inhibiting competing responses during conflict as well (Nachev et al., 2007). Lesions of pre-SMA impair inhibition of automatically triggered actions.

This tiny region is (seemingly) recruited when switching from automatic to consciously controlled responding. An electrophysiological study (Isoda & Hikosaka, 2007) recorded the activity of pre-SMA neurons of monkeys which had to perform a saccade overriding task. They designed a protocol made of blocks comprising a varying number of repeated trials (*id est* with the same cue indicating the direction of saccadic movement) in such a way that the response selection gets automatic. However, at an unpredicted moment, the blocks suddenly changed of cue, so that the monkey had to switch from an automatic to a controlled selection. During the 'switching trials', they observed a cost (slower reaction times and higher error percentage) and found that pre-SMA neurons discharged in both correct and incorrect trials, but the time –early or slightly late of neuronal response correlated with the correctness –correct or incorrect, respectively of the behavioral response. They even observed that electrical stimulation of the same pre-SMA neurons replaced incorrect responses by slower correct ones.

In addition, the lateralized readiness potential (LRP) likely originates in pre-SMA, which is notably engaged during free –*id est* not contingent on an external cue conscious initiation of action (Tanji, Mushiake,1996 ; Cunnington et al, 2003, 2007; Colebatch, 2007).

It is noteworthy that if pre-SMA and DLPFC are necessary to switch from automatic to controlled behaviors the aforementioned data only indicate that these networks implement the contextual supervisory control, but do not trigger it. The recruitment of the SAS is modulated by what one can gather under the *'motivationally relevant'* signals and are conveyed by a medial prefrontal structure, namely the Cingulate cortex, on the basis of a action-outcome mapping constantly updated (Behrens, 2007; Alexander ad Brown, 2011).

3.4 Cognitive versus Motivational Control, lateral versus medial prefrontal cortex

Until now, a global vision of the implementation of the supervisory control has been presented, in order to outline different subsystems collaborating/competing in driving the behavioral outputs. Now we are going to focus on the lateral and medial prefrontal cortices, to figure out the mechanisms by which the output actually is (held to be) controlled. *Controlled* here must be understood as *contingent on a cascade of internal (motivational) and external (sensory/contextual) signals hierarchically organized.*

3.4.1 Medial Prefrontal Cortex: motivational control?

Motivational control basically consists in (i) selecting the most rewarding action(s) and (ii) inhibiting possibly punishing or non rewarding action(s) considering the pending sensory and contextual signals. Error monitoring consists in modifying the weights of inhibition/activation of an action set. These weights are updated by reinforcement, and to be optimal, the system must reinforce only the *intended* actions, and not the stochastic noise of the action selection, id est not the actions that are mistakenly selected. Consequently responses to two kinds of errors must be combined when updating the weights of actions, namely the errors of commission and errors of prediction.

Importantly, it follows that these (prediction/commission) error related signals feed the same system, despite having different origins. That point will be important for the following sections, insofar when I will use the expression of 'motivational control', I will refer to a single mechanism that kills two birds with a single stone: it monitors errors and motivates action selection. Evidence exists that suggests that in the brain, such a system is implemented, within the Anterior Cingulate Cortex (ACC).

The Anterior Cingulate Cortex as a cornerstone:

It is useful to discuss the anatomy of the Anterior Cingulate Cortex (ACC) because it will be of importance afterwards. In primates, the ACC comprises BA 24, 25, 32, and 33. It lays in a very strategic anatomical site, receiving input from limbic, motor, executive and memory centers (for a review see Holroyd, in Posner Ed, 2004, and Hayden, 2009).

The ACC is a major target of midbrain dopamine neurons (VTA), known to discharge in response to unexpected reward-related events. It is subdivided into distinct territories, which will be possibly relevant to characterize in functional/computational terms: posterior ACC, connected to the orbitofrontal cortex; dorsal ACC, connected to the lateral PFC and pre-SMA ; CMA (cingulate motor area), the most directly motor of the cingulate areas, connected to the primary motor cortex and SMA. That tripartition within the ACC, drawn on the basis of the connectivity, surprisingly reminds one of the architecture proposed by Norman and Shallice.

The ACC implements several non exclusive functions, which could be encompassed into a single computational one.

(1) First, one of its roles consists in driving the bottom-up action selection, on the basis of an action-outcome mapping, constantly updated.

Some investigations of the motor part of cingulate cortex (Isomura et al., 2003) have the subjects (monkeys) perform a Go/no-Go task when systematically manipulating the contingencies between sensory cues, rewards, and motor responses, in order to test all combinations of parameters (sensory cue, motor response, mapping rule, reward). The responses of ACC neurons correlate with both the motor plan and the reward contingency, but almost never with the attributes (location or color) of the cues. Thus the ACC participates to the selection by relating action sets to their consequences, instantiating a reward/action mapping (Rushworth et al., .2004).

Furthermore, this mapping is dynamically updated based on the error of prediction, itself weighted by a learning rate and depending on the volatility of the environment so that the reward associated with a given action performed at a time i+1, is recursively modified according to the following equation (see

Behrens et al., 2007) :

(*) $r_{i+1} = r_i + \alpha\delta_i$

where : r_i represents the reward associated to the action at the time i
δ_i represents the difference between expected and actual reward of r_i,
α represents the learning rate (between 0 and 1),

(2) Second, the ACC has been considered to be implicated in error detection and correction, because it is responsive to two kinds of errors: (a) *error of commission* which results from an internal feedback, and (b) *error of prediction*, resulting from external feedback.

(a) In speeded reaction time tasks, or in tasks eliciting a strong conflict of response, one typically observes a frontomedial centered negative deflection in the ERP that peaks about 80-100 ms after subjects make an incorrect response (error related negativity, ERN). This signal does not depend on the modality of output, is thought to originate in BA 24 (Ullsperger 2001; Swick and Turken, 2002; Miltner 03) and is characteristically followed by a more pronounced recruitment of LPFC and pre-SMA (MacDonald et al., 2000), a slowing of reaction time in the next trial, and an adjustment of behavioral performance (Kerns et al., 2004).

Interestingly, it must be noted that the beginning of ERN *precedes* the error (Falkenstein et al, 1990; Dehaene et al., 1994), suggesting not only that a comparator 'knows' the intended response and the currently activated one, but also that the actual selection of response is not be necessary to trigger or alert the mechanisms of supervisory control.

In this respect, it turns out that errors of commission are not necessary to elicit an ERN or activation of ACC. Some studies have shown that an ERN is also generated on trials with a high probability of error, that is to say those eliciting a strong conflict between two or more incompatible competing motor schemes (for instance in the Stroop task).

Last, but not least, awareness of target seems to be required to generate an ACC response to conflict (Kunde et al., 2003; Dehaene et al., 2003), but awareness of error is not necessary to elicit a rERN (Nieuwenhuis et al, 2001). Awareness of error has an effect on the amplitude of the rERN, which is illustrated by a positive correlation between subjective certainty of error and the amplitude of the rERN

(Luu et al., 2000) and is nevertheless required for the subsequent adjustments of supervisory control which led to an overt slowing on the following trial (Endrass, Reuter and Kathmann, 2007).

Conversely, patients with a lesion of the ACC have been reported to show altered generation of rERN, in an Eriksen flanker task (Swick & Turken, 2002; Stemmer et al, 2003). Although the data are not unequivocal, these patients seem to remain able to be conscious of having made an incorrect response (Stemmer et al, 2003). This set of evidence let think that rERN is very likely not the landmark of a conscious but unconscious and automatic mechanism aiming to signal error/ subjective probability of error.

Although the rERN alone is not sufficient to overcome the conflict and/or to strengthen control, the ACC is presumed to be crucial for the recruitment/adjustment of conscious cognitive control, so that an ACC lesion or dysfunction should lead to an impaired ability to strengthen the supervisory control of one's own actions.

(b) As suggested before, as far as it receives input from the VTA, the ACC is also responsive to external negative feedback (absence of expected event/occurrence of unexpected event). The responses to such error of prediction have been named 'fERN' (feedback ERN), as opposed to 'rERN' (response ERN).

This fERN, peaking at ~250-300 ms after feedback onset, is modality--independent, seemingly originating from the same site as the rERN, namely BA24 (Gehring 02, Yeung 04, Bayless 06), but presumably deriving from some mechanisms orthogonal to those implicated in rERN. Whereas the rERN is supposed to result from an efference copy mechanism, fERN results from the error of prediction conveyed by input from midbrain dopaminergic neurons. At least this is a very plausible hypothesis, in an anatomical, functional (Halroyd, 2004), and even temporal point of view. Phasic bursts of VTA neurons are effectively pretty well characterised, in numerous species, by their latency (70-100 ms after feedback onset) and their duration (150-200 ms), so that the offset of their input should terminate at 220-300 ms after feedback onset –a time very compatible with the time of the peak of fERN.

Notably, these two signals (fERN and rERN) intrinsically differ one from each other, and likely play complementary roles. Once selected upon their putative rewarding value, actions are reinforced according to their actual effects – the fERN occurs precisely in a case of negative feedback, reflecting an error of prediction. But in the case of error of commission, *id est* when the intended action is not the actually selected one, the error signal resulting from an efference copy and sent to anterior cingulate cortex –giving rise to an rERN can attenuate the reinforcing effects provided by VTA dopaminergic neurons.

A recent ERP study (Heldmann et al., 2008) directly addressed the question of the functional relationship between rERN and fERN. The authors have the subjects perform an EriksenFlanker task, during which they were told that correct responses had to be accurate and fast, and were given feedback after each trial. Thank to the speed as second criteria of correctness, quite difficult to evaluate for the subject, the authors could manipulate their certainty to have made an incorrect response. Interestingly, they observed a strong rERN and a reduced fERN when the subjects were sure of having made an incorrect response (very late), and low rERN but high fERN when the subjects ignored or were not sure of their performance (hardly late). This pattern of results is consistent with the idea that these two signals play a critical role across learning, since the error of prediction conveyed to ACC by external feedback turns out attenuated by the error of commission.

Under this hypothesis, the absence of rERN might consequently have effects not only on the mobilization of supervisory control, but also on learning, insofar the aforementioned equation (*) can be replaced with the following one:

(**) $r_{i+1} = r_i + \alpha(\delta_i^p - \delta_i^c)$

where
- r_i represents the reward associated to the action at the time i
- δ_i^p represents external error signal (prediction error) of r_i
- δ_i^c represents internal error signal (commission error) of r_i,
- α represents the learning rate (between 0 and 1), according the volatility

Without such an internal signal of accuracy, one can envisage that mistakenly selected responses are

reinforced the same way as the properly selected ones, and considerably slow down the learning of different tasks.

*

To put in a nutshell and close this part, the ACC can be described as comprising two functional loops. The first one is implicated in bottom-up or automatic selection of actions, bridging the *triggering data base* and the *contention scheduling system*. In neural terms, being situated at an intermediary stage between ventral and dorsal striatum (trigger data base) and motor centers comprising SMA and primary motor cortex, it integrates errors of prediction signals conveyed by dopaminergic signaling in order to update reward/actions contingencies in real time and preactivates motor schemas or sequences (via the *CMA-SMA loop*).

The second one is implicated in top-down monitoring of actions, connecting the *Contention Scheduling* and *Supervisory Attentional System*. Sensitive to the magnitude of conflict between competing responses and responsive to discrepancies between intended and actually selected action schemas, it participates to the voluntary inhibition and activation of the outputs implemented in pre-SMA. It is involved also in the strengthening of contextual executive control implemented in lateral PFC (via the *dACC-pre SMA/LPFC loop*) after the occurrence of (response or feedback) ERN.

Note that Alexander ad Brown (2011) proposed a model –the PRO model—that encompasses these apparent different computational functions into a single one. Their model has some difference compared with the classical reinforcement learning algorithms. The value function computed by the network is not a scalar, but a vector of error signals (so that not actions but set of actions are reinforced). Plus, and interestingly, these error signals are general enough to cover a lot of seemingly different experimental conditions, such as error likelihood or response conflict. They can refer to the : (i) occurrence of unexpected reward/punishment, (ii) probability of reward/punishment, (iii) non occurrence of expected reward/ punishment, (iv) probability of non occurrence of expected reward/ punishment.

To conclude, these apparently diverse functions of the ACC, having given rise to different

accounts can be unified under the hypothesis of predictive coding (the neural responses or activity reflect local errors of prediction).

More anterior medial structures:

The hypothesis of the existence of different constants of integration within the medial prefrontal cortex has been developed by Koechlin, Kouneiher and Charron (2009). They stipulate a functional hierarchy for motivational control within the medial PFC, that parallels and energizes the more lateral regions on the basis of reward and error related signals. The segmentation of the lateral region is based upon the relative temporal integration of the network (sensorimotor, contextual, episodic) –see next section for more detailed description of the lateral PFC and cognitive control.

The segmentation of the medial structures is carried out on the basis of the functional connectivity on the medio-lateral axis. Thus, the dACC provides sustained input to mid-lateral PFC, which implements episodic cognitive control on the selection of response, via its input to the posterior lateral PFC. That latter implements the contextual level of cognitive control of the response, by selecting stimulus-response associations (in premotor cortices) on the basis of a contextual cue. The posterior lateral PFC is energized by the preSMA, which conveys motivational signals stemming from current contextual cues.

According to this motivated cascade model of cognitive control, activity in the posterior lateral PFC is modulated by past and current motivational and cognitive cues.

That hierarchy can obviously be discussed. It was proposed very recently by the authors, following the same model they proposed in 2003 for the lateral prefrontal cortex.

3.4.2 Lateral Prefrontal Cortex and cognitive control:

Here we introduce, progressively, the notion of hierarchy. The expression 'hierarchical control of action' is motivated by the fact that in complex behavior, action selection can be contingent on a sensory signal, itself contingent on another signal (named contextual), itself contingent on another signal (named episodic) etc... I used Koechlin's terminology, but that choice is somewhat arbitrary. What must be retained is the idea of a tree-like hierarchy of contingencies, that makes possible the complexity of (Human) behaviors, not only in terms of quantity of signals, but also in terms of the

temporal scale the signals and rewards are integrated at. Human primates are (supposed to be) able to work for a reward extremely remote in time.

Historical steps

The pioneer empirical investigations of lateral prefrontal cortex led to two types of theories, namely *mnemonic* and *attentional* (Petrides, 2000; Lebedev et al, 2004).

Both these trends have contributed to the elaboration of the main key concepts necessary for a global theory: functional segregation, temporal integration of behavioral sequential events, the predictability/unpredictability.

Jacobsen (1936) was the first to show that after a bilateral ablation of prefrontal cortex, monkeys turned out impaired in carrying out a delayed response task, that is to say with some delay introduced between an instruction cue and the occurrence of a 'go' or 'trigger' signal. In 1952, Karl Pribram and collaborators identified the region responsible for such a deficit as the dorsolateral prefrontal cortex.

Nearly two decades later, Fuster and Alexander (1971) carried out a neurophysiological study, and observed a tonic activity during the delay between the instruction cue display and the occurrence of the trigger signal. At that time, that particular temporal pattern of neuronal activity was interpreted as the correlate of a 'working' or 'maintenance memory' by a certain theoretical trend. According that interpretation, the delay period activity corresponded to the active maintenance of perceptual/sensory information lasting until the response is made.

From working memory to modular segregation

Goldman-Rakic (1996), who insisted on the working memory function of the lateral prefrontal cortex, was one of the first to emphasize a modular organization and a functional segregation based on the nature of the task-relevant input. Her model (1995) is based on the idea that working memory is organized topologically, according to the nature of the information currently manipulated. In the continuity of the dorsal/ventral pathways in the parietal and temporal cortices, one finds a similar segregation within the lateral prefrontal cortex.

In an experiment involving a visuo-motor delayed matching to sample, she will describe the “behavior”

of some lateral prefrontal neurons endowed with receptor fields and, preferred direction/location and different location for different neurons.

Sequential context of behaviors and temporal integration

Fuster a priori belongs to the mnemonic trend, but seemingly in a sense extended beyond the single maintenance of perceptual information.

He was interested in the principle of functional organization within the lateral prefrontal networks (Fuster, 1997, 2000, 2001), and stressed the importance of a relative temporal factor that allowed him to distinguish different kinds of memory, or different time-related mechanisms.

He emphasized that the aforementioned *delay period activity* indeed *precedes and lasts until* the selection of the response, but other kinds of temporal pattern of tonic neural activity could be found. Therefore, he attributed to the time or *temporal integration* a key role within the functioning of prefrontal neurons. More than simply being the basis of working memory function, time was conceived as the key for a deep understanding of how sequential overt behaviors are encoded by prefrontal neurons.

As an example, in a neurophysiological study involving monkeys performing a delayed visuomotor response task, Quintana and Fuster (1999) reported different and very specific temporal patterns of discharge by single neurons. At each trial, the animal was displayed the following sequence : (i) a stroboscopic flash (alert), then, after *a delay of 3 seconds*, (ii) one of 4 possible colored cues instructing it about the correct upcoming target to saccade to and the certainty of receiving a reward after a correct response. After a *second delay of 12 seconds* (iii) two colored targets appeared. The animal had to saccade toward the target indicated by the previous cue, and according to different degrees of certainty (75 versus 100%) of being rewarded.

The authors reported that the registered neurons in lateral prefrontal cortex showed a tonic activity that matched the duration period of at least one of the task relevant events (alert signal, short delay, cue display, long delay, targets display, response, post response) without being tuned by any neither sensory, cue-indicated response nor reward-related parameters.

Nevertheless they also reported tonically active neurons (showing the same diversity of period sensitive patterns) but (color) cue sensitive, (cue-indicated) response sensitive, and “reward certainty” sensitive.

Corroborating this idea, Hoshi, Shima and Tanji (1998, 2000) assumed a hierarchical role of the prefrontal cortex in controlling rule-conditioned and goal-directed motor behaviors. They registered movement-related neuronal activity in the dorsolateral prefrontal cortex of monkeys during a task cueing paradigm involving two delayed motor tasks. The first task involved reaching for a *target* that *matched the shape of a cue*. The second task involved reaching a *target* that *matched the location of the cue*. A majority (54%) of 175 movement-related prefrontal neurons seemingly showed a preference for either the *target shape* or the *type of task requirements*. Sixty-four neurons (36%) were selectively active while reaching a *circle* or a *triangle*.

Interestingly, 59 neurons (34%) had an activity that depended on the rule, or stimulus-response mapping (whether the task required matching the *shape* or the *location*). The authors virtually never found such properties in the arm area of the primary motor cortex: only 1 out of 130 movement-related neurons (0.8%) showed task selectivity, and none showed target shape selectivity.

One must also note that Goldman Rakic (1996), although focused on spatial working memory, also underlined the specific temporal pattern of activity, locked on different task related events. She reported the presence of neurons showing a tonic activity during the delay, but she also observed that the same neurons were phasically active when the animal was displayed the target and when the response was initiated.

As a matter of fact, the delay period activities observed in lateral prefrontal neurons seem to instantiate a functional role far more specific than a simple mechanism of perceptual information maintenance, tonically active until the motor response is selected. In effect, patients with damage in the lateral prefrontal cortex still remain able to pass successfully some standard short term memory tests (Petrides, 1989; 2000). Unless redefining that concept, lateral prefrontal cortex cannot be considered as dedicated to that single function.

3.5 Hierarchical Models

There exist two main models of cognitive control. Both share the idea of a rostro-caudal

hierarchical organization of the prefrontal cortex. They are somewhat complementary, insofar one includes the interactions with the medial part (Koechlin), while the other includes the interactions with the striatum (Badre).

3.5.1 the cascade model (Koechlin, 2003)

That model has been proposed by Koechlin in 2003, in a paper published in Science. The main features can be summarized in a few lines:

- rostro-caudal organization according to the relative temporal scales of cognitive control of behavior, including sensory, contextual and episodic.

- The model is quantitative. It indeed uses an information theoretical framework, so that the cognitive control load is a function of the log of the probability of a signal to occur, plus the log of the probability of selecting a response (or a stimulus response association) on the basis of that signal. It follows that the load of cognitive control is cumulative (the load associated of a response is the sum of sensory, contextual, and episodic loads).

- Medial and lateral PFC instantiate different functional roles, namely motivation and cognitive, respectively, but they share the same hierarchical and temporal segmentation.

- the lateral PFC is involved in the tree-like decision per se and receives input from the medial regions corresponding to the same hierarchical level of cognitive control.

Finally, at the apex of that hierarchy stands the frontopolar cortex, with a specific internal structure of branching control, allowing one multitasking (See Sommerfeld, 2007), and thus “to overcome the serial constraint of behavior” (sic).

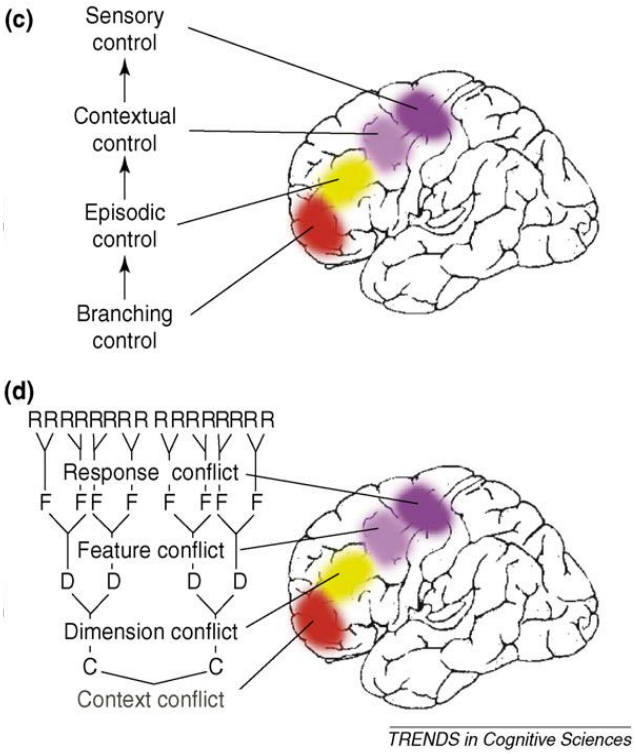


Figure 1-C : different models of hierarchical cognitive control implemented within the lateral prefrontal cortex, (c)Koechlin in (d) Badre and d'Esposito – borrowed from Badre, 2008.

3.5.2 Badre's model

The model of prefrontal lateral cortex proposed by Badre and d'Esposito (Badre, 2008; Badre d'Esposito, 2007; 2009) is hierarchical as well, maybe subtler than Koechlin's one regarding the individuation of the hierarchical levels. In four fMRI experiments, the authors manipulated the cognitive control demands by increasing the number of competing alternatives at four levels of abstraction (competition between responses, competition between Stimulus-Responses associations, competition between sets of Stimulus-Responses associations and so on) In each experiment, the number of alternatives varied between blocks and the subjects were instructed of the number of alternatives on each block. They put in evidence a rostro–caudal frontal gradient of the hierarchy, ranked by the level of abstraction at which representations (of action) compete. Importantly, they observed activations inconsistent with the cascade model proposed by Koechlin. However these models will not be confronted since it is outside the scope of the present purpose.

Note that :

- The model is not explicitly quantitative, but it could be. One could measure the cognitive control load as the quantity of information necessary to select the response (when the response is contingent on the stimulus only, the load is 1 bit ; when there is an additional task-cue, 2bits and so on). An interesting aspect is that they manipulate the complexity of the signal on which the selections are contingent. While Koechlin's model quantifies information serially, that model could also quantify in parallel.
- Some recent works in collaboration with M.J Frank (Badre and Frank, 2011, 2012) integrated neurobiologically plausible mechanisms (dopamine dependent gating) that account for both the selection and the learning of given task. It comes with computational predictions consistent with fMRI data.

- A third aspect (and interest) of such a model is the parallel with the subcortical structures, the striatum (caudate nucleus) in particular. (Desrochers and Badre, 2012). But that is a very recently added feature of the model.

3.6 Decorrelating Cognitive Control Mechanisms and Consciousness

3.6.1 (bottom-up) Influence of non consciously accessed signals on cognitive control mechanisms

A very recent 'new wave' of studies tries to investigate the question of whether subliminal stimuli can influence cognitive control mechanisms. Except in one study, the cognitive control mechanisms that are investigated from that point of view are response inhibition (via a Go/No-Go paradigm), and task setting (via a task selection paradigm and masked priming technique). Before presenting these studies, it is useful to keep in mind the hierarchical frameworks of cognitive control in order to situate these mechanisms, especially from the point of view of the cognitive control load.

The Go/No-go paradigms tap into a stimulus-based response inhibition: the subjects have to select a response, via a stimulus-Response association, and inhibit their response in trials when a (masked) stimulus is displayed. The quantity of information necessary to select or inhibit the response is 1 bit.
In the task setting paradigms, however, assuming that everything is controlled in an adequate way, the

subjects have to select a task-set among at least two possible ones, and then select the response among at least two others. So, the minimal quantity of information necessary to select (or inhibit!) the response is 2 bits. Note that for that kind of experimental issues, the brain imaging or EEG aspect of these studies are fundamental for the demonstrations, insofar one can observe inhibition or priming at different stages occurring before response selection, at perceptual stages for instance. Therefore, an imaging technique is necessary to demonstrate that the priming does NOT occur exclusively at a perceptual stage.

Response inhibition:

An EEG study carried out by Van Gaal et al, 2011, demonstrated that masked No-Go signals could significantly slow down the response selection process, and trigger response inhibition mechanisms that give rise to N2 ERP component –which the landmark of the initiation of inhibitory control.

Task selection

As for the task-set selection (task cueing paradigm), A masked priming paradigm is generally used in order to bias the task selection (Mattler, 2003). Lau and Passingham (2007) reported prefrontal activation by (what they held to be) a non consciously accessed prime. Several points can be discussed in their study, and it has been the starting point of the behavioral study which follows. So I will not present it now, but will do so in the next chapter.

However, a recent work (Wokke et al, 2011) is particularly interesting and elegant, since it reports a context dependent effect of masked Go/No-Go signals. The authors indeed showed that the same masked prime stimulus could exert a substantially different effect on response inhibition and (prefrontal) brain activity according to the context (that was changing on a trial-by-trial basis). It seems to be the first demonstration of a situation whereby an unconscious go or no-go cue processing is contingent on a contextual cue. That flexibility of cognitive control can exist without awareness in some conditions.

3.6.2 Top down influence of cognitive control on subliminal/unconscious information processing

In a very recent review (2012) about the attentional top down modulation of non accessed stimuli, Markus Kiefer wrote: *"if unconscious automatic processing were context-independent, this would result in a tremendous inflexibility of the cognitive system (...): conscious goal-directed information processing would be massively influenced by various unconscious processes. Demands on conscious executive control would be increased, because the intended action could only be ensured by inhibiting numerous interfering response tendencies induced by unconscious information processing"*

Cognitive control (by prefrontal top–down signals) of unconscious cognition (for instance the semantic processing of a masked stimulus) is held to be exerted by modulating the sensitivity of processing pathways for incoming sensory input (Haynes et al., 2007). In that perspective, masked information will only be processed as far as it matches current attentional and task sets.

Several studies suggest that the attention can modulate the processing of subliminal information, depending on the task set (Naccache et al., 2002 ; Woodlan and Luck, 2003; Kiefer and Brendel, 2006; Koch and Tsuchiya, 2007; Bressan and Pizzighello, 2008 ; Kentridge et al., 2008; Kiefer and Martens, 2010 ; Wokke et al, 2011)

3.6.3 Effect of awareness of on cognitive control: the case of error (un) awareness

Awareness of error has an effect on the amplitude of the rERN, what is illustrated by a positive correlation between subjective certainty of error and the amplitude of the rERN (Luu et al., 2000).

Recently, Shalgi and Deouell (2012) investigated in more depth the possible relationships between error, error awareness, confidence and the amplitude of the rERN (using a betting paradigm which allowed them to scale the degrees of confidence). In their paradigm, the participants were displayed three geometrical shapes horizontally, of different colors and size. They were instructed to press the *yes* button if one of the shapes was a designated target shape (Shape target), or if one of the lateral shapes (left or right) was the same shape as the central shape (Matching target), regardless of the size, and to press the *no* button otherwise. In addition, the subjects had to judge their own accuracy after each response (correct or error) and then bet on this judgment using a high, medium, or low amount of money. The average across all subjects regardless of confidence level was consistent with previous work: equal rERN for Aware and Unaware errors which was larger than the correct response negativity

(CRN). However, when they restricted the analysis to high confidence (high bet) trials in confident subjects, a prominent rERN was observed only for Aware errors, while confident Unaware errors (i.e., *error* trials in which subjects made high bets that they were correct) elicited a response that *did not differ* from the CRN elicited by truly correct answers (i.e., *correct* trials in which subjects made high bets that they were correct). However, for low confidence trials in unconfident subjects, an intermediate and equal rERN/CRN was elicited by correct responses, aware and unaware errors. Their results provide substantial evidence that the rERN is related to error awareness, and suggest the amplitude of the rERN/CRN depends on individual differences in error reporting or metacognitive judgment.

These recent data invites one to focus on the ACC, since it could be of particular interest as far as *metacognition* is concerned, and even for cognitive control performance. In effect, some studies suggest that error awareness is required for the subsequent adjustments of supervisory control conjoint with an overt slowing in the following trial (Endrass, Reuter and Kathmann, 2007).

4. Metacognition

4.1 A dependence on access to consciousness?

Metacognition basically refers to cognition about cognition. In Human or even in non Human primates, it refers to the ability to form representations of one's own cognitive processes. Thus, this skill enables us to rate one's visual perception, to attempt to control one's own thoughts or behaviors, to detect one's error without any external feedback, or even to make the distinction between information one does not manage to remember and information one has never received (meta memory).

Actually, it does not seems to be a monolithic system, but on the contrary it seems to follow the same architecture as some first-level cognitive processes (episodic memory, perception, language, motricity…), insofar it can be selectively impaired (Art Shimamura and Larry Square, 1986; Janowsky et al, 1989; Naccache et al., 2005).

The investigation of metacognitive processes is not such a recent topic (Sackur, 2000). However, maybe because of the conceptual but overall experimental and methodological difficulty, metacognition has received little attention until very recently (Smith, 2008 ; Terrace and Son, 2009, Rounis et al., 2010). The reason of the resurgence of this interest for metacognition is very likely related to the tight link existing between metacognition and consciousness, insofar metacognition *presupposes* consciousness. At least, that assertion is suggested by the by the recent history of and the huge progresses made for some decades in cognitive neurosciences and psychology of Consciousness and non conscious processing of information, in patients and normal subjects as well.

Every demonstration of non conscious processing of information indeed rests on the combination of a first-order task and a second-order (or metacognitive) task. It consists in putting in evidence a discrepancy between what subjects report or tell (metacognitive performance, or performance in the second-order task) and what they actually do (cognitive performance, performance in the first-order task).

A simple example can illustrate that point. In an eminently duplicatable experimental situation, subjects are displayed two different kinds of stimuli on the screen of a computer. Some of the stimuli are subliminal, others are not really visible, and others are clearly visible. The first-order task consists in discriminating the two stimuli, by pressing a different key for each stimulus. When the subjects are unable to perceive the stimuli, they are instructed to press a response key 'randomly'. The second-order task consists in rating the visibility of the stimuli after each response. In this context, it is implied that a null or inferior visibility of the stimulus gives rise to a 'random response'.

When one puts in relation the data obtained in each task, one classically observes that, even if subjects report not having seen the stimuli or having seen just a flicker, and thus to have answered randomly, their performance in the first-order (discriminatory) task turns out way above what one should expect a priori from a random process. This has been interpreted as experimental evidence of non conscious information processing – 'information processing' being defined here as the selection of a response contingent on a stimulus, sometimes upon a very abstract property of the stimulus (Dehaene et al., 1998).

As said before, this apparent discrepancy has been solved and captured by the hypothesis of the existence of a central system, so called “Global Workspace” (Baars, 1989; Dehaene and Changeux, 2003). Under certain conditions, the entry of information into this global workspace would correspond to the access of information onto Consciousness, which would afford to manipulate it in different ways that would not be possible if information were unconscious – i.e outside this Global Workspace (cf. section 2, Consciousness as Global Workspace). In short, in this respect at least, an (accurate) metacognitive performance requires awareness of one's cognitive processes.

4.1 Unconscious metacognition : a conceptual problem.

The possibility of unconscious metacognition has been raised (Koriat and Lévy-Sadot, 2000; Charles et al., 2013), so that conceptual clarifications could be needed. One can indeed consider *metaprediction* (see definition below) as a ubiquitous process within the brain, analogous to memory, that is to say, with different levels of explanation and therefore different mechanisms. The most simple and relevant example is the one of error related activity in the anterior cingulate cortex (cf. *section 3.4.1 medial prefrontal cortex: motivational control; the Anterior Cingulate cortex as a cornerstone*).

Charles et al. (2013) given as an example of unconscious metacognition the error related negativity signal observed when subjects were unaware of having made an error. As a reminder, this signal can be observed even in *absence* of overt error. The probability of making a mistake, given the quality and quantity of evidence to be accumulated and actually accumulated, is in itself sufficient to elicit an error related activity. In this respect, the Anterior cingulate cortex behaves as a *metapredictor*: given the activation of a output by a downstream bottom-up mechanism, the Anterior Cingulate cortex predicts/learns to predict the errors of this same mechanism, on the basis of the context.

As already seen above, the main difference between the cases whereby it is associated with error awareness or not, relies on the subsequent events occurring in the neighboring networks (the dorsolateral prefrontal cortex) and the overt performance of the subject in the following trial (Endrass, Reuter and Kathmann, 2007) such as change of strategy, longer reaction times. By contrast, the local error related negativity that we can observe in anterior cingulate cortex when the subject receives negative feedback, when he is coping with a response conflict, or when he makes an error without

being aware of it, has some effects only on the reinforcement weights of the stimulus-responses mappings implemented within the AAC.

The whole conceptual issue of whether there are different levels of explanation of metaprediction, and whether (Bayesian) metaprediction is ubiquitous within the cortex, cannot be considered here and now. Consequently, I will distinguish the concept of *Metaprediction*, namely a signal occurring *within* a neural network without affecting the neighboring networks and the subsequent behavioral outcome, from *Metacognition* which involves a bottom-up transfer of information and can influence the neighboring networks, and the pending behavioral output.

Moreover, and finally, the question of whether metacognition actually entails consciousness (of what?) has been raised, especially with the explosion of Bayesian neural networks to model simple decision making; some might argue that metacognition can be unconscious (for instance Charles et al., 2013). Notwithstanding that a conceptual refinement is needed, this premise is a nonsensical claim : if unconscious metacognition exists, then one no longer has any experimental way to demonstrate the existence of a non conscious processing (since such a demonstration precisely involves a perceptual metacognitive task). Thus, if one were to assume that unconscious metacognition is possible, that would entail throwing away more than twenty years of research on and demonstrations of non conscious processing.

5. Conclusions: which bridges between Metacognition, Conscious Processing and Cognitive control?

The main evidence for a possible link between metacognition and cognitive control comes from neuropsychological and neuroimaging studies, in healthy subjects and diverse populations of patients, which all report or suggest a critical role of prefrontal cortex in metacognition (see Shimamura, 2000; David et al, 2012; Fleming and Dolan, 2012, for a review).However different regions have been reported according to the paradigm and the nature of the metacognitive processes involved (memory, perception…). The regions could be medial (David et al, 2012), but also lateral (Rounis et al., 2010; Fleming et al, 2012a; Fleming et al, 2012b). If one considers two studies having investigated the same metacognitive domain (perceptual metacognition), the performance was linked to more or less anterior

lateral regions, namely dorsolateral (Rounis et al., 2010) or rostrolateral (Fleming et al., 2012). It is unclear which role each of these regions plays in the confidence of metacognitive judgment. Assuming that metacognition involves two different networks, possibly overlapping (one involved in the first-order decision, the other involved in the second-order decision), it follows that inducing a lesion on one of them would alter the metacognitive performance. In the experience reported by Rounis et al (2010), subjects showed a decreased metacognitive performance after a TMS induced lesion onto the dorsolateral prefrontal cortex. What did this lesion do? did it decrease the threshold of perceptual awareness or did it alter the parameters of choice of the second-order decision ? In the study conducted by Fleming et al., the rostrolateral prefrontal cortex showed an activity which correlated with the confidence level. Yet, confidence is also a function of the quality of (sensory) evidence, although indirectly (cf. for example Pleksac and Busemeyer, 2010). Therefore it is difficult to conclude that rostrolateral prefrontal cortex was involved in the second order decision.

We cannot resolve this issue here and now. Nonetheless, that ambiguity related to the diversity of prefrontal sites reported during metacognitive judgment suggests: first, that metacognition is a *relative* process which might depend on the first-order decision domain (memory, visual perception or action selection for instance), but also on the role played by the current information during the first-order behavior. Secondly, that metacognition involves at least two networks, (differentially involved in first-order and second-order decisions) which could be hierarchically organized, since metacognition seems to be associated to a region hierarchically organized.

A way to test these hypotheses could consist in choosing judiciously the first-order decision task, by choosing for example a cognitive control paradigm known to recruit the dorsolateral prefrontal cortex. The aforementioned discrepancy between first-order and second-order tasks could be reproduced, not in order to demonstrate the possible depth of non conscious processing, but on the contrary to investigate the effects on metacognition, when one manipulates the visibility of primes (see above *section 1. Demonstrating the existence of unconscious processing before all* for definition of masked priming), or when one manipulates the cognitive control load (see above *section 3.5 Hierachical models* of cognitive control for a definition of that notion). That will be the object of the first empirical report, in the first

next chapter.

Assuming that visibility or the cognitive control load can influence first-order and/or second-order decisions, visualizing (with fMRI) the networks involved in both first- and second-order decision is necessary step to investigating these hypotheses. That will be the object of the second empirical report, in the second next chapter.

Finally, on the basis of the outcome of the neuroimaging study, one will be able to formulate more precise hypothesis regarding the metacognitive profile of patients with schizophrenia. This psychiatric condition, as one knows, is characterized by important abnormalities, both functional and anatomical, in the Anterior cingulate cortex BA 24 and the prefrontal cortex BA9. That will be the object of the last experimental report.

PART II:

Behavioral Evidence for non conscious Priming of Cognitive Control Processes?

Which effects on metacognitive performance?

Replicating and Exploring

2.1 Introduction :

As said in the Introduction chapter, the hypothesis that, the prefrontal cortex might have a possible causal role in the access to consciousness has been explored for a few years (Schlaweski et al., 2009, Rounis et al., 2010). Parallel to that trend in the cognitive neurosciences of consciousness, the question of whether non consciously accessed stimuli could influence higher level processes, and especially cognitive control processes has been posed (Lau and Passingham, 2007).

These questions are two different aspects of a single one, namely what are the reciprocal influences of cognitive control mechanisms and (access to) consciousness. In the first case, assuming that the prefrontal cortex plays a causal role in the access of information into consciousness entails presupposing that some cognitive control mechanisms located there, hierarchically implemented within the prefrontal cortex, have a causal influence on the access to consciousness (cf *Part I, section 2.2*).

Conversely, assuming that cognitive control operations can be triggered only on the basis of consciously accessed stimuli entails assuming that consciousness plays a specific role in information processing, in the sense that it makes possible a certain kind of operation or influences information processing in such

a way that the control of behavior turns out substantially affected.

2.1.1 Objectives

The first experimental objective was to find a paradigm that would combine these two reciprocal aspects. More schematically, the aim was to design a paradigm that would allow us to observe:

(i) The influences of accessed/non accessed stimuli on specific cognitive control mechanisms, and consequently on the performance in a given task -- this is the first aspect.

(ii) Whether these same influences would then affect the awareness of performance, and consequently the metacognitive performance on the same task – this is the second aspect.

We proceeded in two steps.

The first step was to replicate a quite recent result, that (at that time) was the unique (neuroimaging) study that claimed to demonstrate non conscious priming in a task cueing paradigm, associated with a prefrontal activation (Lau and Passingham, 2007) -- aspect number (i).

The second step -- assuming that it was possible to elicit such a priming at a hierarchical level situated upstream from response selection, and thus to influence the cognitive control mechanisms, consisted in controlling and considering some quantitative cognitive control factors, and adding a metacognitive task, since metacognition is held to entail consciousness[4] of at least an aspect of one's own cognitive processes --aspect number (ii).

2.1.2 Plan

Thus I will present that first empirical part according the following scheme.

For practical reasons, I skip details that progressively lead to the two main experiments.

After a summary of Lau and Passingham's experiment, the results, their interpretation and the problems it rose, I will present the data of preliminary control tests that we carried out regarding the visibility of the primes at each SOA and the priming itself, in a simple discrimination task.

In a second part, I will present the outcome we obtained by trying to replicate Lau and Passingham's

(behavioral) results. Several aspects will be discussed that will justify further choices we made in order to improve the paradigm. That discussion will serve as a transition to introduce the results of what became the definitive version of the paradigm.

Finally, in a third part, I will present it, and the main results it gave rise to: on the basically cognitive aspect (a task-cueing paradigm), and on the metacognitive aspect (self-evaluation). These last data will be discussed.

2.1.3 Lau and Passingham, 2007:

2.1.3.1 Paradigm:

Lau and Passingham (2007) investigated the issue of whether non consciously perceived stimuli could influence the task setting process. They actually carried out a *neuroimaging* study but in this chapter we will focus on the behavioral aspects.

They thus used *a task-cueing paradigm* (figure A) whereby the subjects were displayed one of two geometrical shapes (either a diamond or a square), upon which they had to select one of two possible tasks (phonological, semantic).

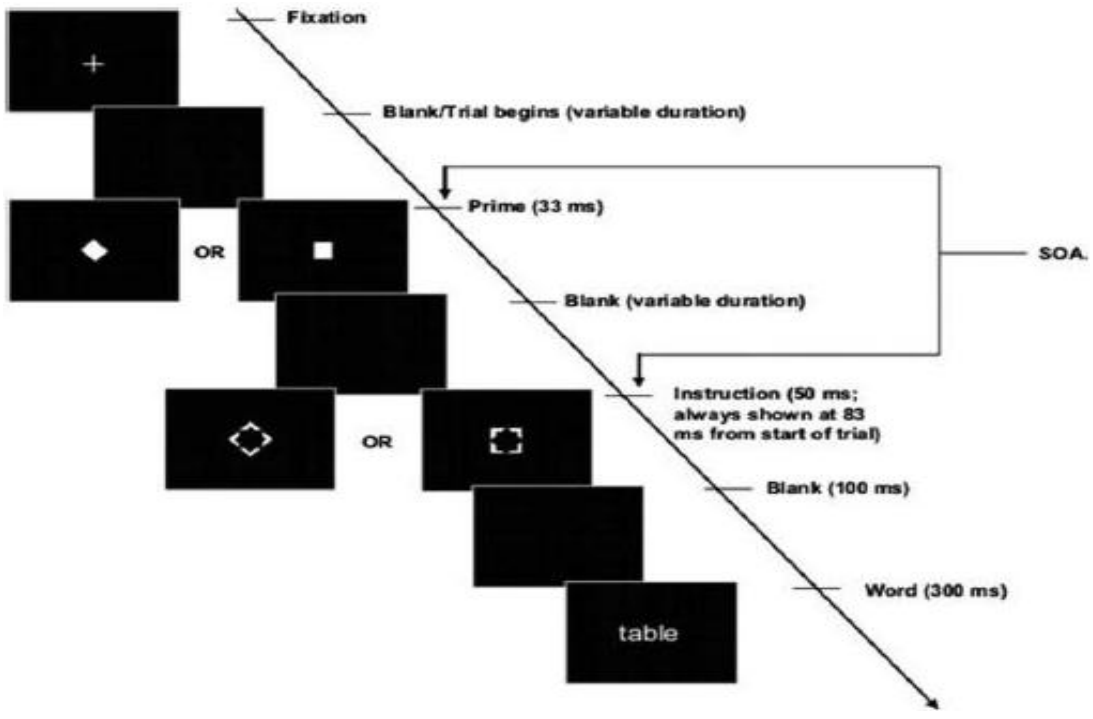


Figure 2-A: from Lau & Passingham, 2007,

These tasks involved binary (motor) responses to an upcoming target letter displayed *100 milliseconds after the task-cue offset (so 150ms after the task-cue onset)*. The phonological task consisted in a yes-no syllabic judgment, and the semantic task consisted in a yes-no concreteness judgment.

In addition, a prime was displayed before the task-cue, which could be either *visible* or *invisible* AND *compatible* (similar) or *incompatible* (dissimilar) with the upcoming task-cue.

The masking paradigm they used is *a metacontrast* paradigm, whereby the primes could be displayed at two different onsets (SOA) respective to the task-cue onset, *namely at 17ms or 84ms.* This type of masking is particular because of its non monotonicity: when plotting visibility against SOA, one can observe an inverted U-shape function of the visibility according to SOA (Breitmeyer, 1984). This effect mainly depends on the luminance of the mask and of the target, that must be carefully controlled – the time display, the retinal position having also an influence. But basically, when the mask/target energy ratio is below 1, a so called *type B* masking is observed (see the book of Breitmayer, 2004, for more details about these masking effects and the corresponding terminology). Using this technique, the authors intended to dissociate the visibility from the strength of the priming, that has been shown to be linearly proportional to SOA length (Vorberg, 2003).

After the experimental session, subjects had to carry on a visibility test, that consisted in discriminating the same masked primes, and displayed at the same SOAs as those used during the experiment proper. They did not report whether the masks were neutral or compatible/incompatible – an aspect that is relevant since the compatibility factor influences performance. Anyway, on the basis of *objective* discrimination performance of the subjects, the authors computed a d-prime and held it as a marker for prime visibility. They reported obtaining a visible prime condition at short SOA and an invisible prime condition at long SOA.

2.1.3.2 Behavioral results:

By splitting the dataset according to the task, a 2-way ANOVA performed on the accuracy revealed a significant *visibility*compatibility* interaction: the difference between incompatible trials and compatible trials was bigger at long SOA and was smaller and non significant at short SOA.

The same 2-way ANOVA performed on reaction times revealed a significant *visibility*compatibility* interaction for the Phonological condition only (the difference between incompatible trials and

compatible trials was bigger at long SOA and was smaller and non significant at short SOA). That was not true for the Semantic conditions because the variances were larger.

2.1.3.3 Problematic aspects:

We considered some aspects of the paradigm to be problematic, that were mainly and generally related to parameters of cognitive control:

(i) Presence of a bottleneck:

A task switching paradigm involves two consecutive decisions, upon two different signals (task selection, then response selection. So the time interval between the respective onsets of the signals (task-cue, target) must be defined on purpose. A 150 milliseconds time interval typically elicit a bottleneck effect, which can be observed in the reaction times (Pashler, Harold, 1994). As a matter of fact, the reaction times obtained in this study are pretty long (means comprised between 1100 and 1200ms).

(ii) The tasks themselves are "noisy",

The task were somewhat inappropriate in the sense that they involved a binary responses (yes-no judgment), to target properties that were either potentially ambiguous (phonological task), or a continuum and therefore less appropriate to yes-no motor response (semantic task, of which the reaction times present an important variance).

(iii) The cognitive load is not controlled (see figure D for that notion and its links with congruency)

A fourth aspect related to the test of visibility for each SOA. The procedure used to test is not reported in full details, so we cannot discuss it. Opting for a conservative attitude, we considered that the use of d-prime (objective performance) as a marker of awareness was inappropriate, especially because the masks were a priori not neutral (but nothing is said about that point).

2.2 General procedure

As said above, the paradigm we intended to replicate involved masked priming. In the upcoming experiments, masked priming will be used in a task-cueing paradigm.

We begin by presenting the outcome of the control of visibility test and simple priming test, despite the fact that they were actually carried out after each experimental session. The aim of this couple of tests was twofold: (i) to confirm that the masking technique gives rise to two clear-cut conditions of visual awareness, regardless the type of masking (A or B) (ii) that significant priming effects were obtained in both conditions, in a simple priming task, so that the failure to obtain priming effects in the task-cueing paradigm would have suggested that the priming had occurred, but nevertheless did not propagate until the response selection stage.

2.2.1 Control of Prime visibility

In order to be sure of the subjects phenomenology associated with each SOA, subjects had to run a visibility test after having completed the experiment. That test consisted in 2*80 trials during which were displayed (i) either a diamond, (ii) or a square (both identical to the primes used during the experiment), or (iii) nothing at all (20% of the trials). These figures were followed by a mask consisting in the two task cues superimposed, that were used during the actual experiment (see figure B).

The subjects had to discriminate the prime (diamond versus square) by pressing a left or a right key. When they did not manage to see them, they had to randomly press one of the keys. After each response, they had to rate the visibility of the prime using the following scale.

Note that they were informed that sometimes there was no prime at all:

- 0 = niente (nothing)
- 0.2 to 0.4 = visto solo un flash (I saw just a flicker)
- 0.6 = visto, ma sono insicuro (I saw a shape, but I am very uncertain)
- 0.8-1 = visto abbastanza o molto bene (quite well or very well seen)

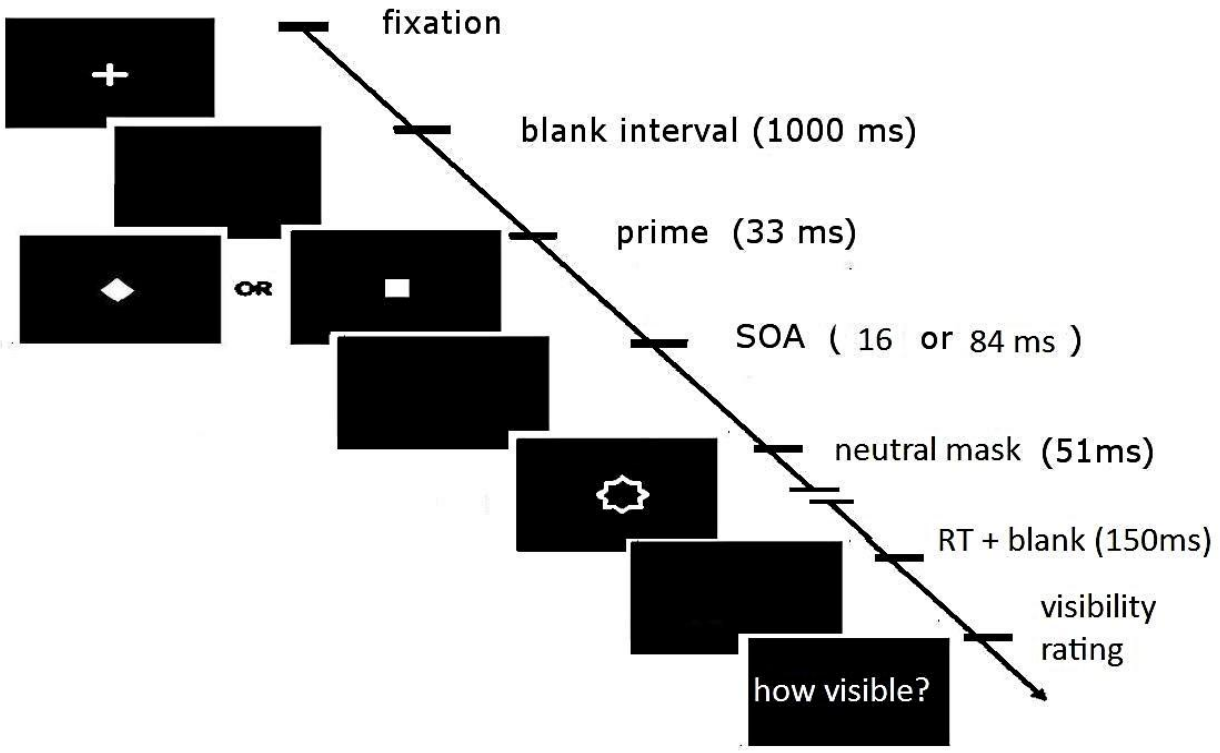


Figure 2-B : trial of the visibility control test

The aim of that test was to be sure that the metacontrast masking paradigm worked correctly (i.e gave rise to 2 clear-cut conditions, namely high visibility and low visibility). Note that Lau and Passingham reported type B masking, with *a high visibility* condition associated with a *short SOA* and a *low visibility* condition with a *long SOA* (note that their long SOA was of 84ms. In our last version, a slightly longer SOA was used to obtain more clear-cut categories of visibility).

Despite a certain variability among subjects, the masking technique indeed worked well. The subjects were able to discriminate the absence versus the presence of a stimulus; they generally reported to see the stimulus quite well in one condition, and to see only a flicker in another one (figure 1). A pairwise Wilcoxon test confirmed a significant difference between the subjective ratings of visibility ($p<0.002$) associated with the SOAs. At the short SOA, the mean subjective rating suggest that the subjects were not aware of the figure that was displayed. At the long SOA, the mean subjective rating suggest they were.

This supports the splitting of data by SOA to examine the properties of the conscious and unconscious processes separately.

However, the masking effect we obtained was 'classical', or type A masking, being given that a low visibility condition was associated with the shortest SOA. Therefore, we did not manage to replicate the metacontrast effect obtained by Lau and Passingham in their study, but that could possibly be explained by difference in the mask/target energy ratio -- an LCD screen was used, contrary to Lau and Passingham, and we did not control the luminance.

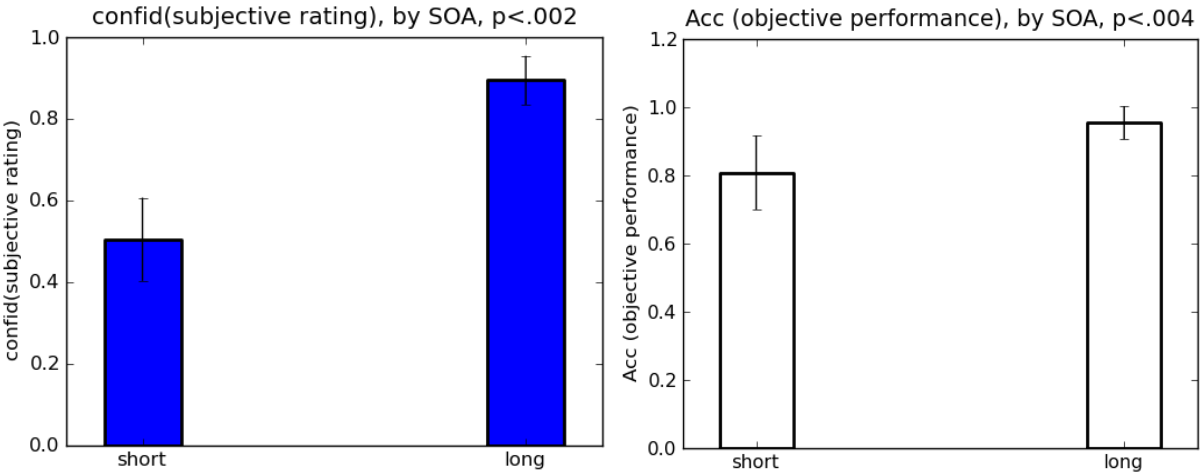


Figure 2-1: *Left,* subjective report :
mean confidence level by SOA,
Short = 16ms ; long = 84ms (n= 16 for this version)
Visibility Scale : **0** = nothing ; **0.2-0.4** = flicker ;
0.6 = very uncertain, but saw a shape; **0.8-1** = more or less well seen
Only trials corresponding to a correct discrimination are reported.
***Right*, objective performance:**
mean accuracy by SOA, in discriminating the targets
(bars represent Standard error)

For each upcoming experiment, the two subjects who showed a different pattern of visibility were removed. One of them was able to perceive well at both SOAs (visibility rating of 0.92 for short SOA, and 0.94 for the long SOA). The second one seemingly perceived very poorly at both SOAs (visibility rating of 0.17 for short SOA, and 0.29 for the long SOA).

2.2.2 Control of Priming in a simple discrimination task

We made a second test in addition to the visibility test. After the experiment proper, the subjects ran a short experiment during they had to discriminate the task-cue (diamond versus square) by pressing a left or a right key. The figures to discriminate were preceded by some primes, exactly as

during the experiment, that is to say *visible* or *invisible*, *compatible* or *incompatible*. The aim of that test was to see whether we could observe a priming effect in each visibility condition. In the contrary case, we would have changed the masking or priming technique.

Accuracy:

Non normality of raw and arcsine transformed accuracy was confirmed by a Shapiro-Wilk test ($p<.001$), so we carried out pairwise Wilcoxon tests to test the compatibility factor in each SOA condition.

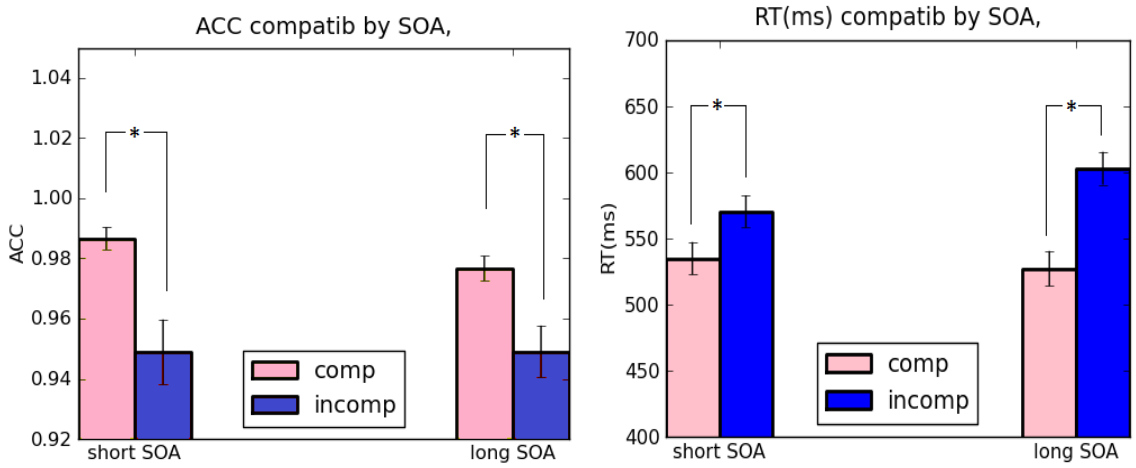


Figure 2-2 : priming (difference between compatible and incompatible) in different visibility conditions (low versus high visibility, for short versus long SOA) Accuracy on the left, Reaction Times on the right. The bars represent standard errors.

We observed significant differences between the incompatible and compatible conditions at both the short SOA ($p<.022$) and long SOA ($p<.02$).

Reaction Times:

Normality of raw reaction times was indicated by a Shapiro-Wilk test ($p>.32$), so we did not carry out any log transformation.

Reaction Times were then entered in two (within-subject) one-way ANOVA with compatibility as single factor, one for each SOA.

Short SOA:

In short SOA trials (see figure 2), the one-way ANOVA with compatibility as the main factor carried out on Reaction Times revealed a significant effect of prime compatibility ($F=4.46$, $p<.042$)—longer RTs in incompatible condition.

Long SOA:

In long SOA trials (see figure 2), the same one-way ANOVA (compatibility as main factor) carried on Reaction Times revealed a significant effect of prime compatibility ($F=8.21$, $p<.007$))—longer RTs in incompatible condition.

2.2.3 Preliminary Conclusions

To sum up, this preliminary methodological part aimed to control the visibility and the efficiency of the priming technique by metacontrast. These controls were actually performed after the experiment proper, although they are reported here before –for simplicity. They were carried out mainly because of the visual phenomenology corresponding to the different SOAs reported by Lau and Passingham.

The authors reported obtaining a high visibility condition associated with the 16ms SOA and a low visibility condition to the 84ms SOA –which is not a classical masking effect. It is certainly not easy to replicate. We indeed failed to replicate that masking effect, and instead obtained a classical masking effect –that we judged satisfactory since it gave rise to two clear cut conditions of visibility.
Regarding the priming itself, we observed significant priming effects at both SOAs.

2.3 First Pilot : replicating Lau and Passingham, 2007

For purposes explained in the Introduction chapter, we wanted to replicate the results reported by Lau and Passingham in 2007. The paradigm was formally equivalent to a task-switching paradigm, where at each trial subjects have to switch from a simple task to another one on the basis of a cue

displayed before an upcoming target. The specificity of their paradigm consisted in the use of display of a prime, *visible* or *invisible* AND *compatible* or *incompatible* with the task-cue (that is to say identical or different), in order to facilitate or elicit a conflict at the task selection stage. The expectations regarding the behavioral data could thus be as follows.

Following the additive factors method developed by Sternberg (1969), the (strongest) hypothesis regarding the reaction times would have been that, if the factors manipulated (visibility and compatibility of the primes) actually act at the same processing stage, in the present case, at the task selection stage, then the reaction times should reveal a *visibility*compatibility* interaction. A less ambitious expectation would have consisted in obtaining a compatibility effect in the condition of invisibility –that would have demonstrated that the nature of the prime influenced the response selection, even if the prime was not visible.

As for accuracy, which might be more relevant than reaction times because an unseen stimulus should a priori not involve an additional serial (conscious) operation but should nevertheless influence the bottom up mechanisms of selection, one was expecting the same pattern as for the reaction times, namely: either *a visibility*compatibility* interaction, or at least a compatibility effect in the low visibility condition. One of these possibilities is sufficient to demonstrate that an unseen stimulus (the prime) influences the selection of the response and thus the selection of the task set.

But things may be more complex, because after examining in detail the fine timing of their paradigm (see figure below) we suspected that the very short time interval between the task-cue offset and the target onset used by Lau and Passingham was too short (100 ms), in that it was very likely that a non identified bottleneck effect (see Introduction for the definition of that concept) could have interacted with the factors of interest or with the task-cueing itself, which may recruit a similar frontal network (for a review of the neuroimaging data of the bottleneck effects, see Marois and Ivanoff, 2005) . Consequently, we introduced a third factor, named *bottleneck*, by manipulating the time interval between task-cue offset and target onset (100ms versus 500ms).

The design of our paradigm thus became a 2x2x2 one, with SOA(2), compatibility(2) and bottleneck(2) as main factors. In the same vein, this revisited design allowed one to estimate how 'pure' the results they reported were. In particular, if the bottleneck factor went significant or interacted with the visibility or the compatibility of the prime, then it would be necessary to modify and adapt the paradigm.

2.3.1 Paradigm

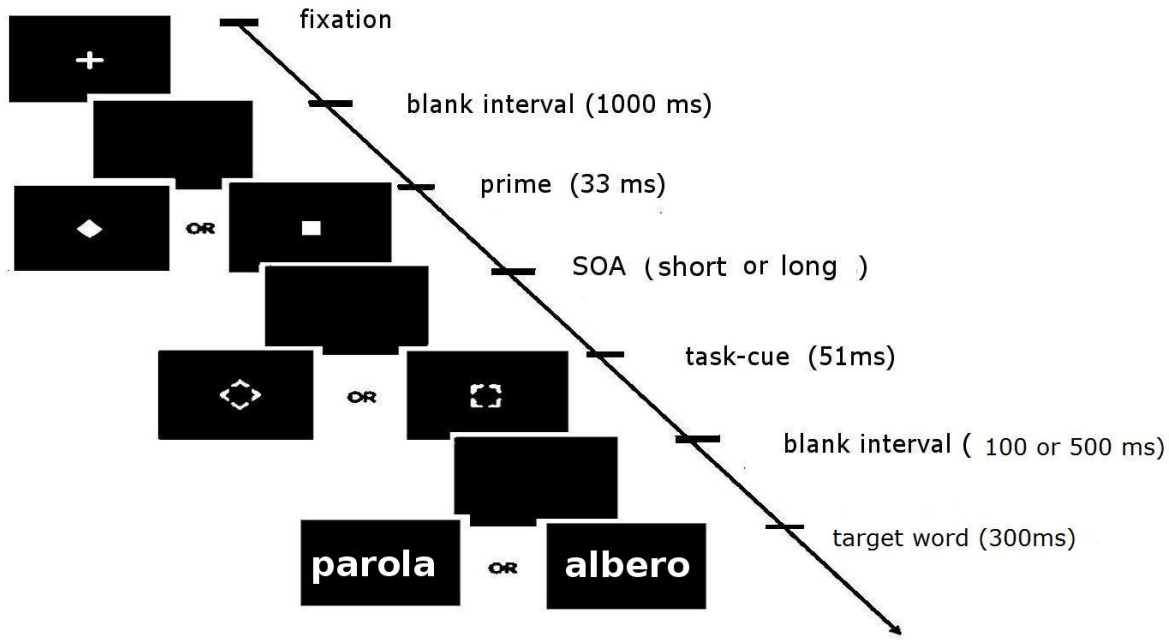


Figure2- C : trial procedure :
At each trial, subjects are displayed the following sequence:
a prime (little square or little diamond) – then a task cue (big square or big diamond) which also plays the role of mask, and finally a target word.
The priming can be *compatible* (prime and cue identical) or *incompatible* (prime and cue different), *invisible* (SOA = 16.7 ms) or *visible* (SOA = 100.2ms).
The task cue indicates which one of the two tasks to perform regarding the upcoming word. The subject must answer yes or no as fast as possible, by pressing a left or right key.

2.3.2 Subjects, Material and Methods

Subjects

15 right-handed subjects (mean age 25.6 ± 4.8 years) participated in the experiment. All were healthy, had no psychiatric history running in family and did not follow any pharmacological treatment. All gave informed consent.

Procedure

The whole experiment comprised 4 blocks comprising 72 trials separated by an interval of 1 second. The key/response mapping was counterbalanced across subjects, thus splitting them into 2 groups.

The subjects were instructed to ignore the prime whenever they were able to see it, and to pay attention to the mask/cue and to the word. They had to answer a yes/no question about the upcoming target word, according to the shape of the cue. The *square* always indicated the question '*has it a concrete content?*' whereas the diamond always indicated the question *'is it bisyllabic?'*.

The subjects answered **yes** or **no** by pressing one of two keys as fast as possible (the mapping key/answer was constant during all the experiment, but differed across groups).

Material

The experiment was run on an Asus laptop (frame rate was 60 Hz, 17 inches) and programmed in Python. The words (1° * 1° by letter, Arial font) were displayed in lowercase, in the **center** of a black screen, whereas the shapes 1,5° * 1,5° for the cue, and 1° * 1° for the prime) were displayed 3° above or below the center (to obtain a more reliable masking effect).

Targets:

The possible target words (72 by block) were chosen among a set comprising about the same proportion of bisyllabic, trisyllabic, concrete and abstract words. Before being used for the test, that

word list was displayed to several Italian subjects who gave a consensus about the properties of the words. Any ambiguous or infrequent word was replaced.

2.3.3 Results

The reaction times superior to 2000 ms or inferior to 300 ms were eliminated, and only those corresponding to a correct response were analyzed.

The data were first submitted to a normality test (Shapiro-Wilk, **alpha= 0.01**). If necessary they were transformed (logarithmic or arcsine for RT and accuracy, respectively), and tested a second time. After this second normality test, they we were entered either in a pairwise Wilcoxon test, or in a within-subject 3-way ANOVA with prime compatibility (2), soa (2) and bottleneck (2) as main factors.

Finally, a total of four subjects were removed from the sample because their mean accuracy, or reaction times or scores in visibility were 2 standard deviations outside the mean of the sample.

Accuracy:

A Shapiro normality test revealed that both raw accuracy ($p<.001$) and arcsine transformed accuracy ($p<.001$) did not follow a normal distribution. Therefore we did not perform any analysis of variance, but used a non parametric test instead.

Short SOA:

In short SOA trials, the analysis did not reveal any significant effect of compatibility ($V=61$; $p>.61$) nor bottleneck ($V=64$; $p>.85$).

Long SOA:

In long SOA trials, the analysis revealed no significant effect of compatibility ($V=81$; $p>.24$) but a trend toward a bottleneck effect ($V=92$; $p<.07$) –see figure 3 above.

Reaction Times:

A Shapiro test of normality confirmed that the raw reaction times did not follow a normal distribution ($p<.001$), and confirmed that log-transformed reaction times did ($p>.54$).

Short SOA:

In short SOA trials, the 2-way ANOVA with compatibility and bottleneck as main factors revealed no significant effect of compatibility (p>.61) but a significant effect of bottleneck (F=10.94, p<.002) –cf figure 3 below.

Long SOA:

In long SOA trials, the 2-way ANOVA with compatibility and bottleneck as main factors revealed no significant effect of compatibility (p>.81) but a significant effect of bottleneck (F=10.75, p<.002) –cf figure 3 below.

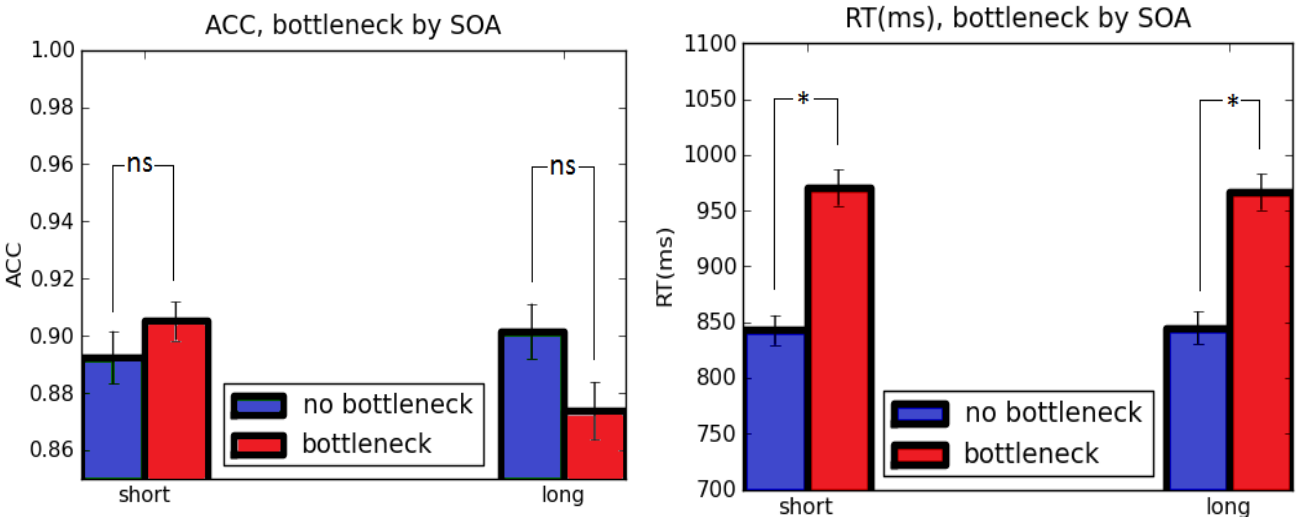


Figure 2-3: Accuracy (left) and Reaction Times (right), bottleneck factor by SOA, The bars represent Standard errors.

2.3.4 Preliminary remarks and conclusions

On the basis of the result (consisting in a strong bottleneck effect for reaction times at both SOA and a trend of bottleneck effect for accuracy, at long SOA) obtained on that first pilot, we decided to opt for a revisited version of the paradigm because of diverse problems exposed below.

We retained the metacontrast masking procedure since it gave rise to satisfying visibility results in the

subjects, notwithstanding that their reports were discrepant compared with those obtained by Lau and Passingham. Lau and Passingham had reported ***low visibility associated with the long SOA***; their interpretation was made so on the basis on the d' values of the subjects in a discrimination task (hence *not on the basis of subjective rating of visibility*), and in addition they removed two subjects from their sample to reach significance.

The masked priming tested in a simple discrimination task seemingly was efficient as well, in each visibility condition. For some reason we were not able to determine, the priming did not propagate downstream, namely onto response selection stage. Possible reasons are given below, that justify why we modified the paradigm.

Main bottleneck effect:

The reaction times revealed a main bottleneck effect. That was linked to the time interval between the task cue and the target, that affected the reaction times. In Lau and Passingham's paradigm, the duration of blank interval between the task-cue and target was 100ms. The reason why such a bottleneck can be critical is that such a phenomenon is linked to central decision making-processes and thus affects response selection, although indirectly. Moreover, bottleneck effects, such as the 'psychological refractory period' or the attention blink, recruit the (dorsolateral) prefrontal region implicated during task-cueing protocols (cf. Marois & Ivanoff, 2005, for a review) – which could constitute a problem for the future neuroimaging study we intended to carry out.

Tasks and term-to-term correspondence between target properties and responses:

A second problem one met concerned the tasks themselves.

The phonological judgment consisted in deciding whether an upcoming word contained 2 syllables. The volunteers had to press one key for '*yes*' and another one for *'no'.* The semantic judgment consisted in deciding whether an upcoming word refers to a

concrete thing. The volunteers had to press a key for 'yes' and another one for 'no'.

The main problem arose with the semantic judgment task. The concreteness/abstractness property of a word actually is a continuum, and consequently is less appropriate for yes-no answers. Subjects indeed

asked for clarifications and hesitated a lot, which increased both the reaction times and their variability. For that reason we decided to change both the task and the targets for our upcoming future paradigms, in order to have binary properties of targets giving rise to binary responses, and thus discrete probability of response (which will make easier the eventual quantification of the cost associated with response selection).

Since those tasks used by Koechlin (2009), that we have already described in the previous Introduction chapter (PART I, section 1.2), had been proven to work well, we chose letters as targets, and for the tasks, a case judgment as the first task, and a consonant judgment as the second task. Because of the previous point regarding the bottleneck effect, we opted for a 300 ms time interval between task-cue offset and target onset.

Cognitive control load: how many signals is response selection contingent on?

The term-to-term correspondence between binary properties of targets and binary responses was not only necessary to reduce the variability in reaction times, but also to quantify the cognitive control load.

The congruence of the targets allows one to manipulate the (discrete) quantity of information necessary and sufficient to select the response. If we consider for instance the tasks used in Lau and Passingham's paradigm, some target words give rise to the same response whatever the task cue and are said to be *congruent* for that reason, whereas others give rise to different responses according to the task-cue, and are said to be *incongruent.* That distinction is represented below (figure D).

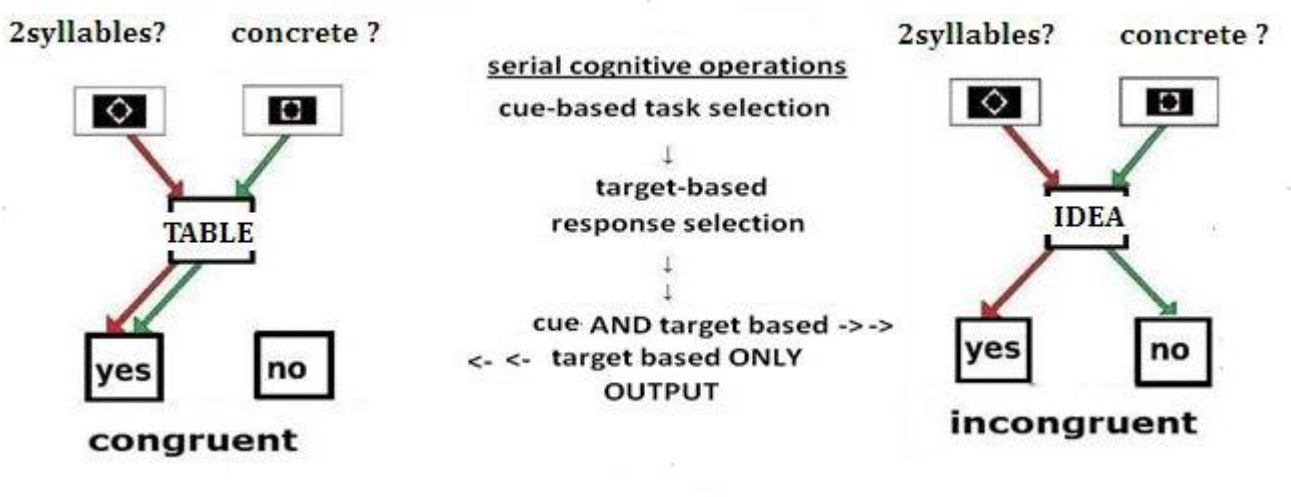


Figure 2-D : congruent versus incongruent targets
The targets differ regarding the (discrete) quantity of information necessary to select the response. The congruent ones need one signal, the incongruent ones need two signals.

Such a distinction among trials might turn out critical if one considers the paradigm from the point of view of cognitive neurosciences and not only in terms of cognitive psychology of cognitive control. Effectively, on the basis of previous neuroimaging studies and according to most prevalent current models (Badre & d'Esposito 2007, Koechlin 2003), that are based on a hierarchical organization of prefrontal cortex, incongruent trials should at least imply a more "costly" response selection so that the reaction times and/or accuracy should be affected by that factor. If SOA and prime compatibility indeed affect response selection after having biased task selection, then the effects of SOA and/or prime compatibility could likely depend on or interact with congruence.

The (speculative) basis for such a behavioral hypothesis is that dorsolateral and/or anterior cingulate activity should necessarily be modulated by such a difference in the "cost" of response selection at the time of response selection itself (cf *part I, section 3.4 Cognitive versus motivational control, lateral versus medial prefrontal cortex*). But that hypothesis should be further investigated with EEG and neuroimaging techniques.

2.4 Second Pilot :

The expectations were nearly the same as for the previous version of the experiment, except that (i) we considered a third factor (named target *congruence,* figure E), that was supposed to reflect

the cognitive control load that we thought it would influence performance, and that (ii) we also added a 'metacognitive task', in order to observe the possible effects of the prime visibility, prime compatibility and even congruence onto the awareness of performance.

Regarding the basic task (i.e the performance on the task-cueing paradigm):

Following the additive factors method developed by Sternberg (1969) regarding reaction times, it was reasonable to expect an effect of compatibility in the high visibility condition, or for a compatibility*congruence interaction in the high visibility condition only. That was a priori motivated by the theoretical framework of the Global Workspace theory, that suggests that consciously accessed stimuli (and their putative collateral effects) must be carried out serially (Dehaene and Naccache, 2001). In other words, an invisible prime might be processed and its putative distracting effects inhibited by parallel bottom-up mechanisms, but the effects of a visible one should be processed or inhibited in a different way, namely serially, because of the intrinsic properties of the decision system within the global workspace (cf. *chapter I, section 2.1 three pieces of behavioral evidence for a central Global Workspace*). Consequently, a visible incompatible prime should slow down the decision processes, eventually more when the target is incongruent (i.e the cognitive control load is higher).

As for accuracy, as said before, it might be more relevant than the reaction times. An unseen stimulus should not a priori involve an additional serial operation but might nevertheless influence some (parallel) mechanisms of selection. That being said, we were expecting for a global additive or multiplicative effects of congruence and compatibility effects, or at least a compatibility effect in the low visibility condition.

Actually we were in a phase of exploration, so that to retain the paradigm it was necessary to get at least a compatibility effect in the low visibility condition, since it was necessary and sufficient to demonstrate that an unseen stimulus (the prime) could influence the selection of the response and thus the selection of the task set.

Regarding the metacognitive task (i.e the awareness of the one's performance in the task-cueing paradigm):

For this variable we recorded only the accuracy (named meta-accuracy to avoid confusion). That was also an exploratory test, but since we made the hypothesis that the prefrontal network(s) involved in the task selection involved similar structures to the 'metacognitive task' (dorsolateral prefrontal cortex and SMA), and since we considered that network to be capacity-limited (locus of bottleneck effects) we were expecting for the congruence or visibility to impair error detection.

We had no specific hypothesis, neither regarding the false alarms (reporting an error when performance is correct) nor regarding the misses (reporting a correct response when performance is actually incorrect).

2.4.1 Paradigm

The paradigm remained generally structurally unchanged, but we changed the tasks, and removed the bottleneck effect – we opted for a blank interval of 300 ms (instead of 500 ms) between the task cue and the target, in order to minimize the total length of the experiment and to maximize the number of trials.

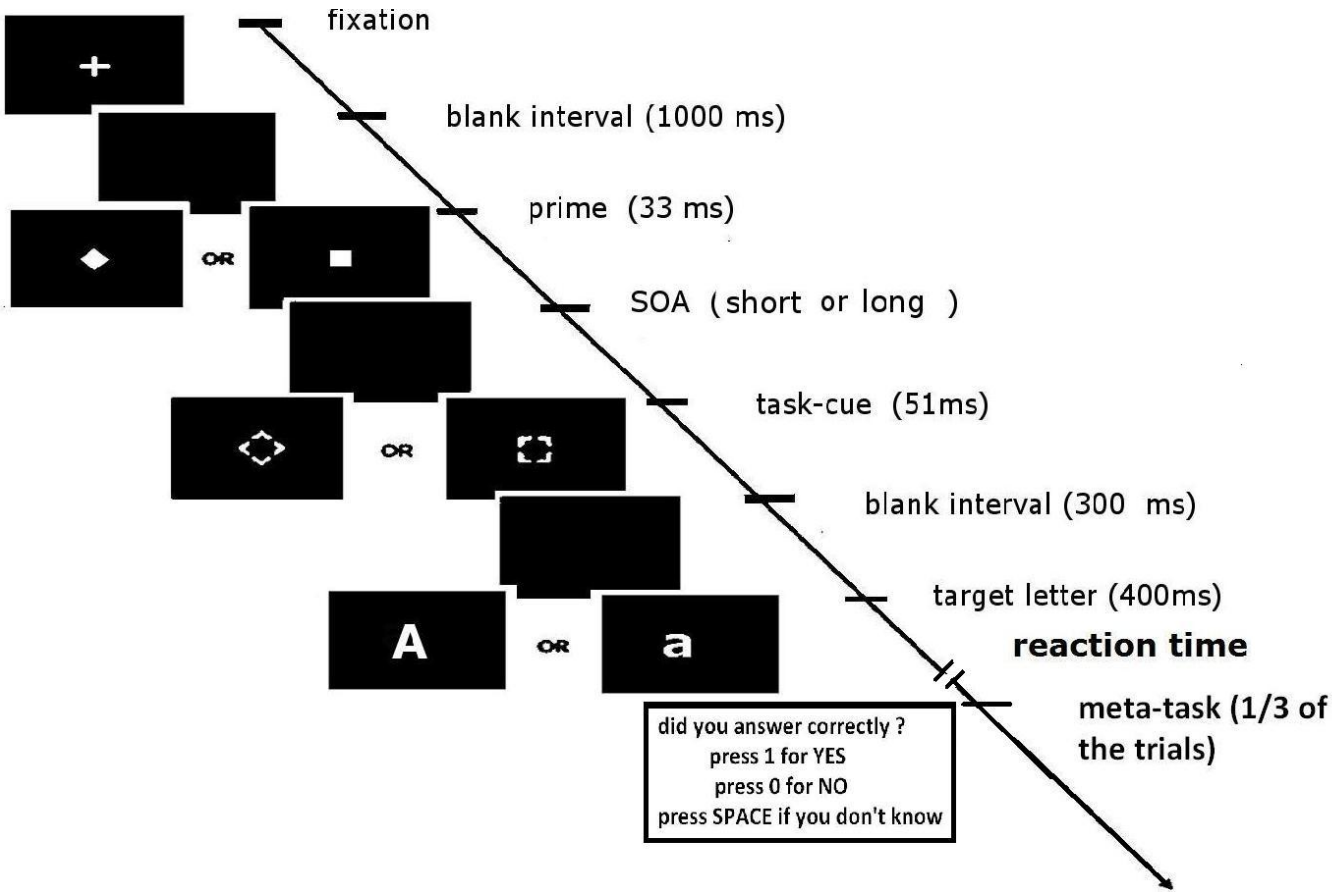


Figure 2- E : trial procedure

task cueing : At each trial, subjects are displayed the following sequence :
a prime (little square or little diamond) – then a task cue (big square or big diamond) which also plays the role of mask, and finally a target letter.
The priming can be *compatible* (prime and cue identical) or *incompatible* (prime and cue different), *invisible* (SOA = 16.7 ms) or *visible* (SOA = 100.2ms).
The task cue indicates which one of the two tasks to perform regarding the upcoming letter. The subject must answer yes or no as fast as possible, by pressing a left or right key.

Metacognitive task: In **one third** of the trials, immediately after having answered, the subjects were asked about the correctness of their response. They could answer *'yes', 'no', 'don't know/not sure'*. Only the confident responses were analyzed.

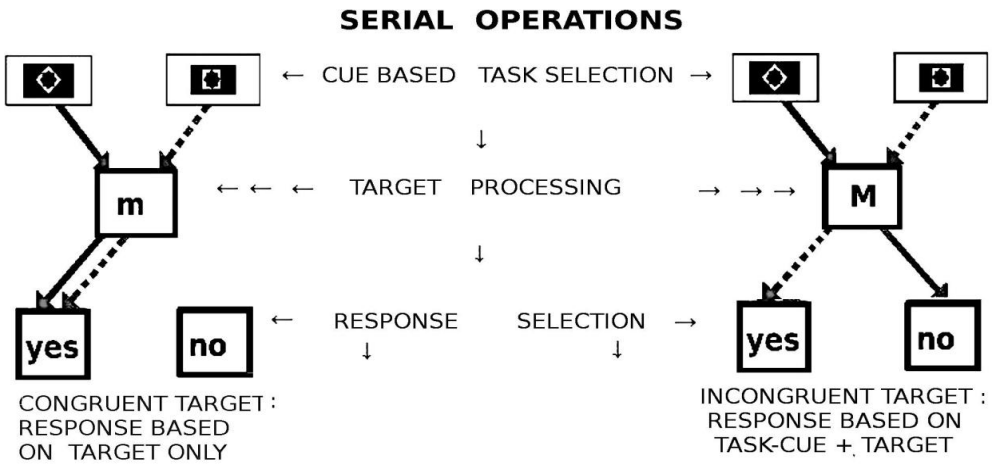


Figure 2-F : manipulating cognitive control load with target congruence
according to the task to perform, some targets gave rise to a single (left, congruent) or two (right, incongruent) possible responses, so that in the first case, only one information is necessary to select the response, whereas in the other case, two information are needed.
In a theoretical point of view, a critical difference between these trials consists in that only the incongruent trials involve a cue-based task selection (the response selection is conditioned by task cue *and* the target), whereas the congruent trials are equivalent to a simple target-based response selection (the response selection is conditioned by the target *only*).
The prefrontal activity is a priori more important in incongruent trials

2.4.2 Subjects, Material and Methods

Subjects

40 right-handed subjects (mean age 24,2 ± 3.8 years) participated in this experiment. All were healthy and gave informed consent.

Procedure

The whole experiment comprised 4 blocks each comprising 128 trials, each trials being separated by an interval of 1 second. Since the blocks involved different sets of target letters, the order of the blocks across subjects was varied according to a 4*4 latin square design. The key/response mapping was counterbalanced across subjects, thus splitting them into 2 groups.

The subjects were instructed to ignore the prime whenever they saw it, and to pay attention to the mask/cue and to the letter. They had to answer a yes/no question about the letter according to the shape of the cue. The *square* always indicated the question '*is it a consonant?*' whereas the diamond

always indicated 'is *it in lowercase?*'.

The subjects answered **yes** or **no** by pressing one of two keys as fast as possible (the mapping key/answer was constant during all the experiment, but differed across groups).

Training

Subjects had an intensive training phase with feedback until they reached excellent performance, namely superior to 90% correct and faster than 1000ms. Each training sequence comprised 90 trials. At the end, the program informed the subjects of their mean speed and accuracy. The subjects needed a mean of 2,5 training sessions (about 225 trials, SD unknown) before running the experiment proper.

Finally, during the experiment proper, a meta-task was introduced on one third of the trials. After having responded, an instruction might appear on the screen asking subjects to indicate whether they answered correctly or not. For that question, they could say *YES* or *NO,* by pressing two other corresponding keys, or *I don't know*, by pressing the space bar. They were encouraged to answer precisely, the speed not being important, and to answer yes or no only when they were confident.

Material

The experiment was run on an Asus laptop (frame rate was 60 Hz, 17 inches) and programmed in Python. The letters (1° * 1°) were displayed in the **center** of a black screen, whereas the shapes 1,5° * 1,5° for the cue, and 1° * 1° for the prime) were displayed 3° above or below the center (to obtain a more reliable masking effect).

Targets:

The possible targets (8 by block) were chosen among this set of letters ['M', 'R', 'B', 'T', 'm', 'r', 'b', 't', 'A', 'E', 'U', 'I', 'a', 'e', 'u', 'i'] and differed among blocks in order to avoid habituation. Consonant/vowel and upper/lowercase proportions were equally distributed across the different blocks. Furthermore, in each block, the congruent target letters giving rise to identical responses (yes or no) were equally distributed according to the response they gave rise to.

2.4.3 Results

Since we were interested in the pattern of effects in *invisible prime* versus *visible prime* conditions, the dataset was split according to SOA in order to be entered in two different within subject 2-way ANOVA, with *compatibility(2) * congruence (2)* as factors.

Note that the behavioral variables of interest were of two types.

The first type comprised the variables giving information about performance in the *task-cueing paradigm* itself: *reaction times* (in milliseconds), *accuracy* (in percentage of correct responses).

The second category comprised the metacognitive performance, that is to say the ability of the subjects to self-evaluate with a reasonable certainty the correctness of their response on a given trial. We named the global accuracy in that (meta) task *meta-accuracy.* Furthermore, the meta-accuracy was then split into *false alarms* (FA), and *hits* (HIT). False Alarms consist in reporting an incorrect response when the response actually was correct. Hits consist in reporting an incorrect response while the response indeed is incorrect. Ideally we intended to develop more refined signal detection related analysis, but we went unable to do so, because of the unbalanced structure of the HITs (at least one condition missing in about two thirds of the subjects). This made impossible to have a term-to-term correspondence between HIT and FA in each of the 8 conditions.

Reaction times superior to 2000 ms or inferior to 300 ms were eliminated, and only those corresponding to a correct response were analyzed.

Finally, we removed three subjects from the sample because their mean accuracy, meta-accuracy or reaction times were 2 standard deviations outside the mean of the sample.

2.3.3.1 Basic task: task cueing

Accuracy:

Preliminary Shapiro-Wilk tests of normality revealed that both raw accuracy and arcsine transformed accuracy distributions differed significantly from normality ($p<.001$ for both). We therefore used non parametric statistics (pairwise Wilcoxon test).

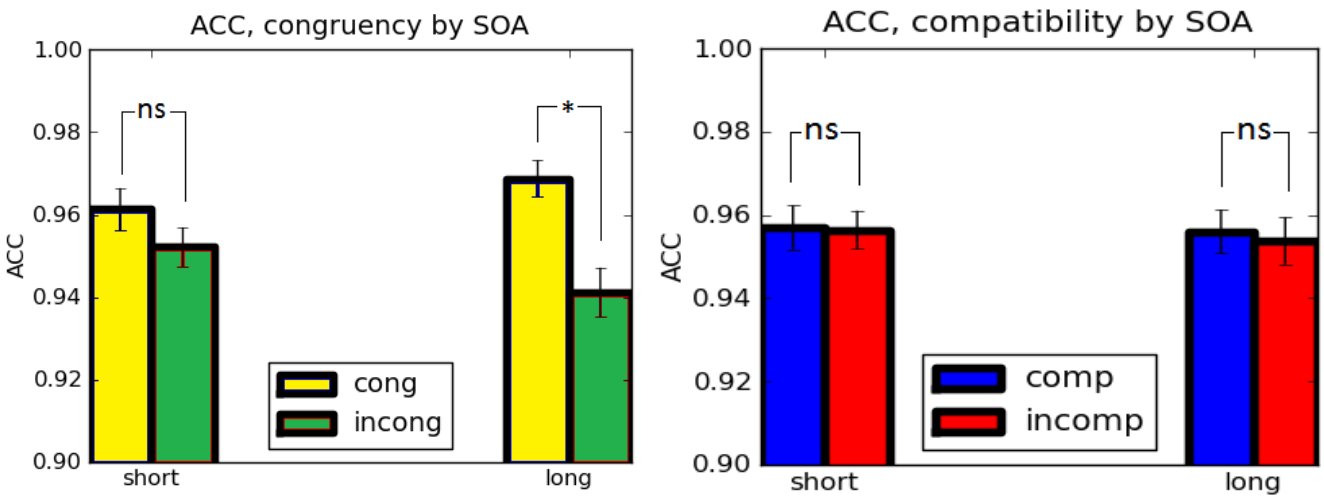


Figure 2-6: mean ACC by main factor.
The bars represent Standard errors.

Short SOA trials:

In short SOA trials, the congruence factor did not reach significance (V=250.5; p > .083), nor did the prime compatibility factor (V=299; p>.30).

Long SOA trials:

In long SOA trials (when the prime was *visible*), there was a significant effect of congruence (V=153.5; p<.0028), but not of compatibility (V=307; p>.50).

That pattern of results suggests that the mean accuracy was impaired by a higher cognitive load when the subjects were displayed a *visible prime* before selecting the task set, let it be compatible or not.

Reaction Times:

A Shapiro-Wilk (alpha= 0.01) normality test indicated that the raw reaction times did not follow a normal distribution (p<.0001), and that log-transformed reaction times did (p>.037). We thus carried

out an analysis of variance on log-transformed reaction times.

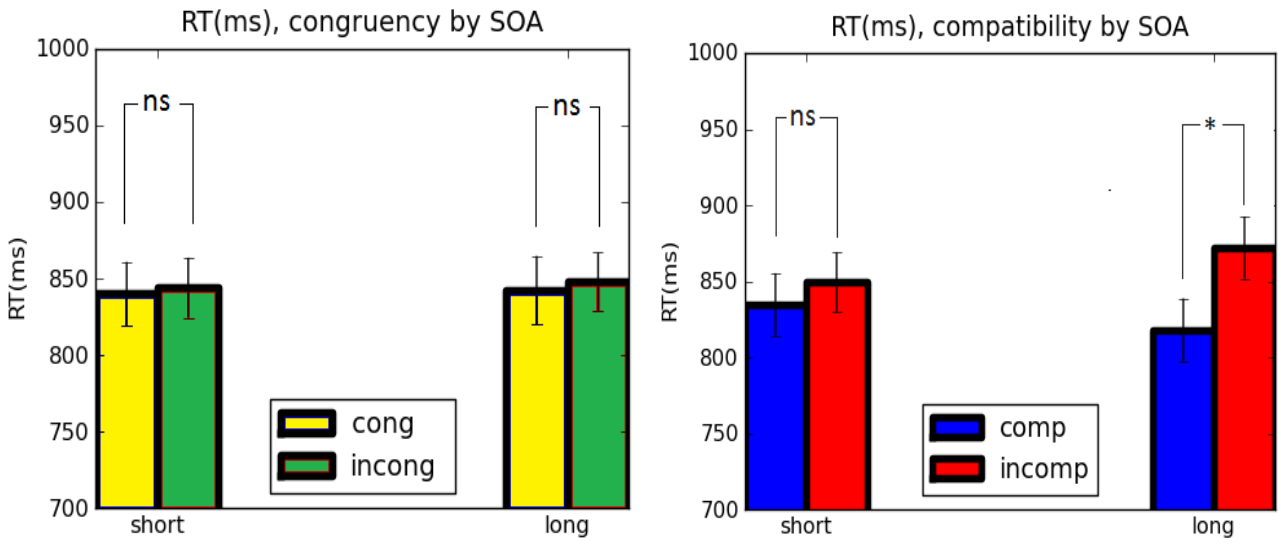


figure 2-5 : mean RTs by main factor, for each SOA.
The bars represent Standard errors.

Short SOA trials:

The 2-way ANOVA *(compatibility (2) * congruence (2))* performed on short SOA trials did not reveal any significant effect (F=0.30, p>.57 for the compatibility factor, F=0.06, p>.79 for the congruence factor).

Long SOA trials:

The 2-way ANOVA *(compatibility (2) * congruence (2))* on long SOA trials revealed that the effect of prime compatibility was significant (F=5,31; p<.023). Prime compatibility thus influenced Reaction Times only when the prime was a priori visible (figure 5).

We did not obtain any other significant result.

Summary of the results in the basic task:

The subjects were significantly slower when task selection was preceded by a conflict elicited by a visible incompatible prime.

The error rates were influenced by the *congruence*, but that influence was restricted to the condition where a *visible prime* was displayed before the task cue, seemingly independently of the compatibility of the prime.

Note that, contrary to our expectations, we did not observe any evidence of a non conscious priming (effect of compatibility in the low visibility condition) in that task-cueing paradigm. The preliminary control tests carried out on *masked priming* had shown that a priming process occurred in each visibility condition. That suggests that these priming effects nevertheless do NOT propagate downstream, onto response selection, and thus do not influence the overt performance. For the moment, only factors affecting the internal state of the decision system (access/non access to consciousness, load of cognitive control) seemed to affect the overt performance.

2.3.4.2 Metacognitive task: confident self-evaluation

Meta-accuracy:

Shapiro-Wilk tests of normality revealed that both raw meta-accuracy and arcsine meta-accuracy did not follow a normal distribution ($p<.001$ for both), so that we used non parametric statistics (pairwise Wilcoxon test).

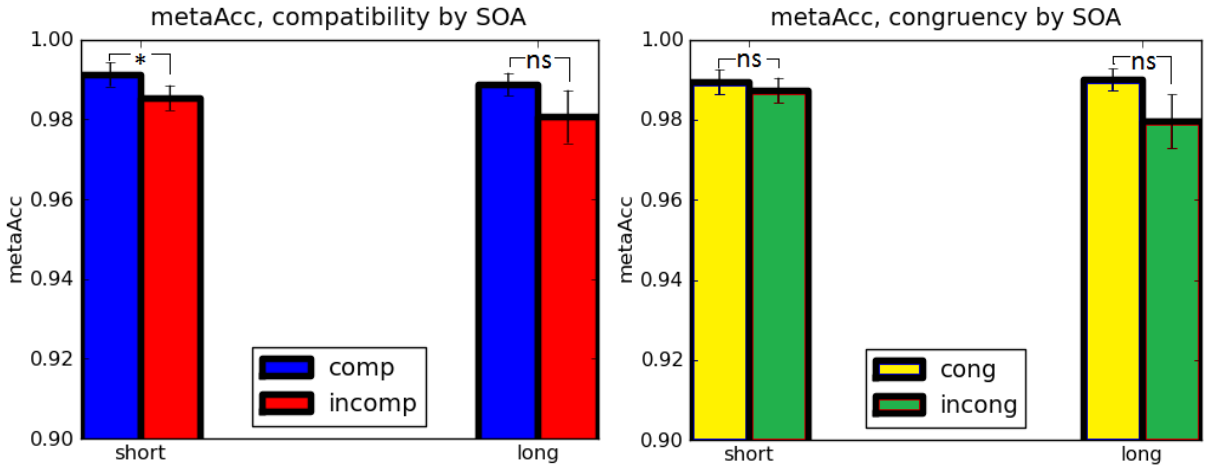


Figure 2-7: mean meta-ACC by main factor, for each SOA.
The bars represent Standard errors.

Short SOA trials:

In short SOA trials, we observed a significant effect of compatibility (V=43.5p<.04) (figure 7), but not of congruence (V=131; p>.15).

Long SOA trials:

In long SOA trials, the compatibility effect was not significant (V=98; p>.14), nor was the congruence factor (V=191; p>.21)

The analysis of global meta-accuracy suggests that the metacognitive performance of the subjects was less good when they had *to cope with a non conscious distractor* (compatibility effect only in short SOA condition), but that does not give any information about the way the compatibility influences the self-evaluation processes (overestimation versus underestimation of one's own performance). Exploring the FA and HIT rates will allow a more refined insight of that phenomenon. Since the congruence factor was significant for the accuracy in the LONG SOA condition, we also split the data according to the congruence of the trials, using a Bonferroni correction of the threshold (threshold set at 0.025 after correction).

Error reported after a correct first-order response {False Alarms, FA}:

The dataset corresponding to FA is rich of insight insofar it is supposed to provide information about the factors that bias the subjective certainty of having made a mistake. Given the quite high accuracy level, that dataset is large and all the conditions are reasonably equally represented.

Shapiro tests of normality carried out on raw false alarms and arcsine false alarms (p<.001 for both) did not allow to carry out an analysis of variance, so we carried out non parametric statistics (pairwise Wilcoxon tests).

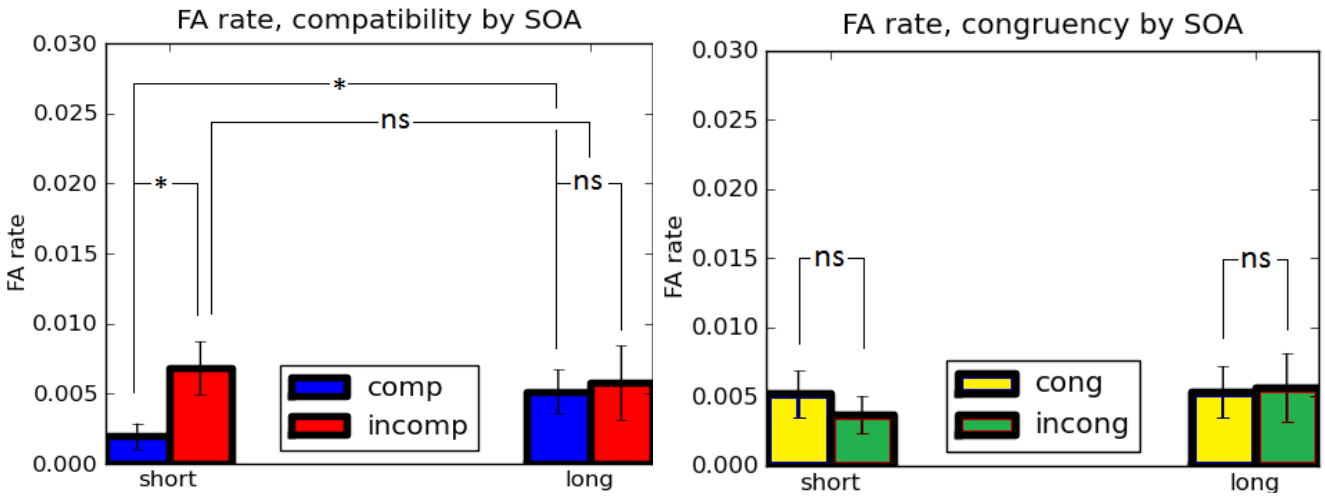


Figure 2-8: FA rates by factor.
The bars represent Standard error.

Since we wanted to figure out whether the compatibility effect depended on the visibility of the prime, we looked at the compatibility factor in short SOA versus long SOA trials, and within each visibility condition, in congruent versus incongruent trials as said above.

Short SOA trials :

In short SOA trials, we found that the compatibility factor was significant ($V=80$; $p<0.017$), but the congruency was not ($V=34$; $p>.44$) .

Congruent trials :

In short SOA and congruent trials, we found no significant effect of compatibility (Wilcoxon test, $p>.29$)

- Incongruent trials :

In short SOA and incongruent trials, we found a significant effect of compatibility($V=42$, $p<.024$) –figure below.

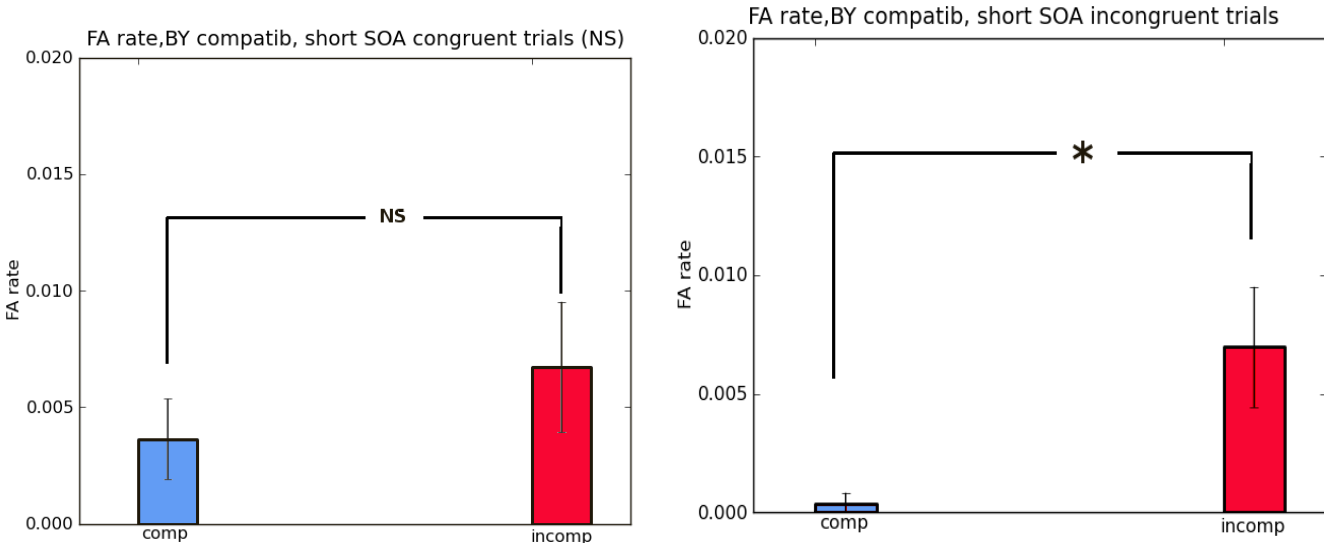


Figure 2-9: FA rates by compatibility, in SHORT SOA trials ;
Left: congruent trials; *right:* incongruent trials.
The bars represent Standard error.

Long SOA trials :

In long SOA trials, there was no significant difference between compatible and incompatible conditions (V=24 ; p>0.25)

- Congruent trials :

In long SOA and congruent trials, we found no significant effect of compatibility (V=10 ; p>.55)

- Incongruent trials :

In long SOA and incongruent trials, we found no significant effect of compatibility (V=24, p>.8)

Looking for a possible interaction between SOA and compatibility, we compared the FA rates in the long versus short SOA within the incompatible trials and we observed no difference between short and long SOA conditions (pairwise Wilcoxon, p>0.17).

We even observed no significant difference of FA rates between *visible compatible* condition and *invisible incompatible* condition (pairwise Wilcoxon, p>0.12).

In short, we observed that a *visible* prime elicited a constant rate of FA, let it be compatible or not.

However, the nature of the prime (compatible or incompatible) had a different impact on FA rates only when it was *NOT* visible.

Errors successfully detected, i.e errors reported after incorrect first-order responses {HIT}:

The dataset complementary to the FA is potentially very interesting, since it is supposed to reveal the factors that elicit a blindness to their own cognitive processes (task selection and or response selection in particular).

That part of the analysis should be carefully considered. Given the quite high general level of accuracy, that dataset is very small (about 5% of the total trials), so that about two thirds of the subjects had 'empty conditions', and the conditions themselves comprise very few trials. In other words, the conditions are not equally represented within and between subjects and the data points are very dispersed.

Furthermore, the mean hit rates seem to be very low (error detection was quite low, about 60% –the chance level being 33%), but that might be due to the small size of the dataset.

Shapiro-Wilk tests of normality made on raw hits and arcsine transformed hits revealed that both did not follow a normal distribution ($p<.001$ for both). We therefore performed pairwise Wilcoxon tests to compare the means in each condition.

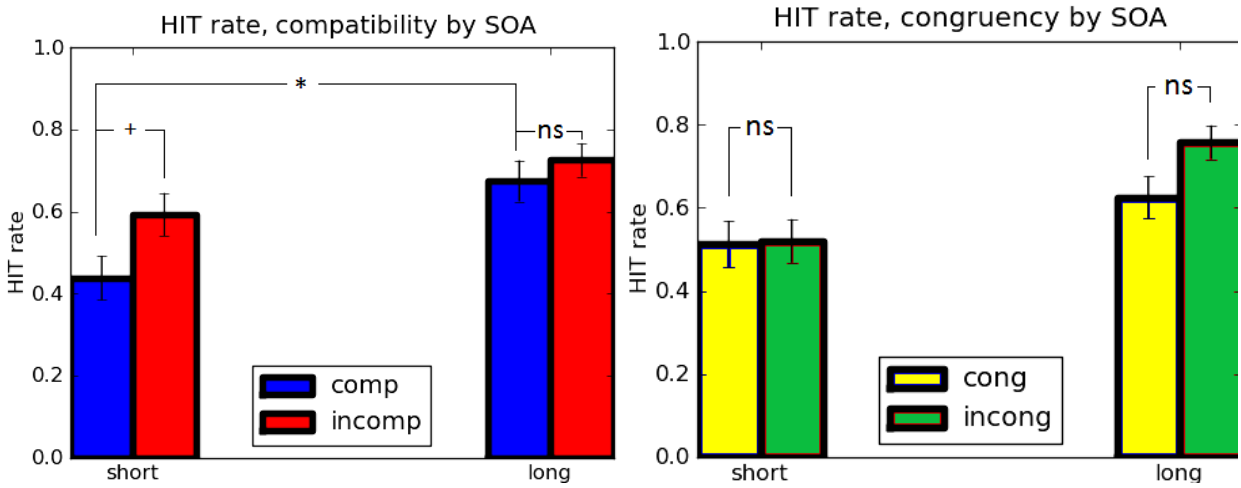


Figure 2-10 : HIT rates by main factors, for each SOA.
The bars represent standard errors.

Short SOA trials:

In short SOA trials, we observed only a trend of the compatibility factor (p=0.09).

- Congruent trials :

  In short SOA congruent trials, we found no significant effect of compatibility (Wilcoxon test, p>.88)

- Incongruent trials :

  In short SOA incongruent trials, we found a significant effect of compatibility (Wilcoxon test, p<.018) –figure 11 below.

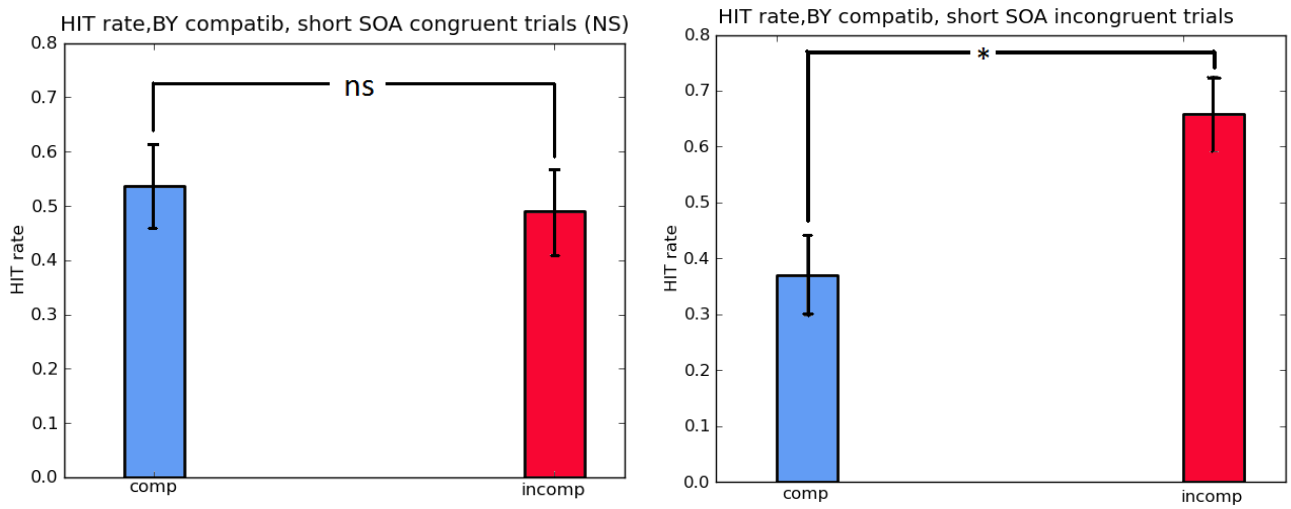


Figure 2-11: HIT rate by compatibility, in short SOA trials.
Left: congruent trials; right : incongruent trials.
The bars represent standard errors.

Long SOA trials:

In long SOA trials, nor the congruence ($p>.63$), neither the compatibility ($p>.66$) were significant.

- Congruent trials :

  In long SOA congruent trials, we found no significant effect of compatibility (Wilcoxon test, $p>.19$)

- Incongruent trials :

  In long SOA incongruent trials, we did not find any significant effect of compatibility (Wilcoxon test, $p<.49$).

Since the pattern of results of the HIT rates (figures 10 and 11) looked quite similar to the one of the FA rates (figure 8 and 9), with nearly significant effect of compatibility in short SOA trials only, and significant compatibility effect in short SOA incongruent trials), we compared all the columns two-by-two as we did for FA rates. We observed nearly the same result as for the FA rates -- the only difference was that, for the HITs, the compatibility factor in short SOA condition failed to reach significance.

We also compared long versus short SOA conditions within the incompatible trials, and observed no difference (pairwise Wilcoxon, $p>0.40$).

We even observed no significant difference between the *visible compatible* condition and the *invisible incompatible* condition (pairwise Wilcoxon, $p>0.73$).

In short, we observed that a *visible* prime elicited a constant rate of HIT responses, let it be compatible or not. The nature of the prime (compatible or incompatible) *tended* to have a different impact on HIT rates only when it was *NOT* visible (figure 10). That latter effect was restricted to incongruent trials.

Summary of the results in the metacognitive task:

The analysis of meta-accuracy, considered globally independently of the correct or incorrect trials in the basic task, revealed a main effect of prime compatibility. That effect was strong when the prime was generally not visible (in short SOA trials), and absent when the prime was generally visible (in long SOA trials).

The decomposition of meta-accuracy into FAs and HITs revealed rather similar patterns of results. We obtained a significant (for false alarms) or nearly significant (for hits) effect of compatibility in short SOA trials only, that is to say when the prime was not visible.

Furthermore, we observed a significant compatibility effect for both the HITs and FAs in short SOA incongruent trials *only*. In short SOA incongruent trials, the occurrence of a compatible prime was associated to less frequent false alarms, but also to less frequent hits.

Said more briefly, the occurrence of a generally invisible compatible prime was associated to less frequent reports of error. That effect seemed to be restricted to incongruent trials (for HITs and FAs, we observed a significant compatibility effect in short SOA incongruent trials, but not in short SOA congruent trials), that is to say when the cognitive control load was higher.

We were unfortunately unable to compute a meta-d' because of the lack of data (for the hits, at least one condition out of eight was empty for almost all subjects).

2.4.4 Preliminary Conclusions

Basic task:

Regarding the performance in the ***basic task***, a first question of the experiment was whether unconscious priming effects would have been obtained at the so-called task-selection stage. The subjects were indeed slower when they were presented with an incompatible prime, but that effect occurred only when the prime was generally visible (cf. Reaction Times, effect of compatibility in long SOA trials only). That result is not consistent with those reported by Lau and Passingham (Lau and Passingham, 2007), or more particularly with their interpretation of their own results (the claim of high visibility associated with short SOA). However, if we ignore their phenomenological interpretation and consider the physical variable, namely the SOA, our results are similar.

Note that this result is compatible with the Global Workspace framework, which stipulates that information entering into the global workspace (that is to say consciously accessed) must be processed serially, because of the intrinsic structure of the global workspace itself. In such a scenario, a visible incompatible prime would activate the switching to the wrong task set. After the display of the task-cue (signaling the right task set to select) the presence of a wrong task-selection would be signaled by a mechanism of error signal-based control, itself triggering the activation of a second mechanism of inhibition/correction. Only some time after task-cue onset, would the task relevant decision processes be resumed.

An hypothesis compatible with the slowing down would be based on a drift diffusion model (Ratcliff, 1985, 2002 ; Ratcliff and MacKoon, 2008 for a review), assuming a serial processing *modus operandi*, but also an incompressible period of evidence accumulation necessary for the task-set to be selected. In that scenario, the incompatible prime would trigger the accumulation of evidence for the wrong task-set, and the display of the task-cue, some milliseconds later, would triggers the accumulation of evidence for the right task-set. The wrong task-set would therefore be inhibited by a lateral inhibition mechanism. The additional (first scenario) or delayed (second operation) covert operation would manifest itself in overt behavior through longer reaction times.

The analysis of accuracy (cf. Accuracy, pairwise Wilcoxon in short SOA trials, no compatibility effect) supports the idea that the nature (compatible or incompatible) of invisible primes had no significant

effect on the accuracy of the subjects, since there was no prime compatibility effect, whether the prime be visible or invisible. Instead, one observed a lower performance when the response selection was 'more expensive' in terms of cognitive control (incongruent trials). This effect was restricted to when subjects were displayed a *visible prime before the task cue, let it be compatible or not* (cf. Accuracy, pairwise Wilcoxon, congruence effect in long SOA trials).

In a nutshell, no influence of unconscious or subliminal stimuli is obtained the present results. On the contrary, our data show that subjects were significantly slowed down by *incompatible primes, but only when they were visible.* In addition, they were less accurate when the 'cost' of the response selection was higher (incongruent trials) AND when they were displayed a visible prime before the task-cue.

Meta-task:

We obtained unexpected effects regarding the meta-accuracy (idem, i.e prime compatibility effect in the short SOA condition). In other words, although incompatible primes did not affect the performance of the subjects in the task itself, they influenced the 'awareness' of their performance—and consequently the metacognitive mechanisms. It must be first reminded that we considered exclusively the (a priori) 'confident' self-evaluations of the subjects. These effects are as follows:

The global meta-accuracy, including hits and false alarms, was mainly influenced by prime compatibility, and we found that this effect was restricted to the generally non visible primes (short SOA).

Moreover, the false alarms and hits, considered according to prime visibility (short/long SOA) and compatibility, tend to show a similar pattern : (i) they tend to be/are affected by invisible primes (significantly for false alarms, a trend for hits) (ii) they both show a significant compatibility effect in short SOA and incongruent trials, but not in congruent trials and (iii) they 'behave' as if the non conscious occurrence of a compatible prime was associated to less frequent reports of error (decreased frequency of reported errors when the prime was invisible and compatible).

2.5 More general Conclusions, Discussions and further investigations

2.5.1 Interpretations of these results: Locus of the effects?

One of the aspects of that study consisted in investigating to which extent it is true that cognitive control mechanisms are by definition conscious and are made on the basis of conscious

information – a definition including awareness of one's performance as a part of the performance.

A first aspect of the question was whether non consciously perceived stimuli could influence the overt performance of subjects in a cognitive control task..

On the basis of the performance of the subject in the basic task (task-cueing paradigm), we did not observe that the invisible primes significantly affected behavioral performance, at least in this paradigm. Subjects were slowed down or speeded up only by incompatible or compatible generally *visible* primes, respectively, and were less accurate because of factors unrelated to the compatibility of the primes, but by endogenous factors such as the cognitive control required for response selection (congruence effect in long SOA trials). In summary, we did not observe any effects of generally invisible stimuli on the performance (no compatibility effect in short SOA trials).

The second aspect of the question was whether non consciously perceived stimuli could influence the *awareness of that same overt performance* in a task that we named the meta-task, providing a third variable named meta-accuracy.

On the basis of the performance of the subjects in the meta-task (consisting in error detection), although the compatibility and visibility factors we manipulated had very little influence on the performance in the basic task, they considerably influenced the awareness of that performance (effect of compatibility in short SOA trials).

Interestingly, FA (confident report of an error while the response actually is correct) and HIT (confident report of a correct response while the response indeed is correct) rates showed a rather similar pattern (no effect in long SOA trials, significant or nearly significant of compatibility in short SOA trials, significant compatibility effect in short SOA incongruent trials). They both decreased when an invisible compatible prime was displayed beforehand.

This apparently similar behavior of FA and HIT rates suggests that FA and HIT are influenced by the same mechanism. Basically, this putative mechanism would be sensitive to the presence of a conflict – at least a conflict occurring non consciously. Therefore, the presence of a compatible prime (equivalent to an absence of conflict), could exert a bias that manifests itself through a less frequent 'reports of error', giving less FAs in some cases or less HITs (accurate report of error) in other cases.

Here, 'less frequent' must be understood relatively to a sensitivity baseline remaining to define. The fact that one observed a significant compatibility effect in *incongruent short SOA trials* but nothing significant in *long SOA or congruent trials*, for both the FAs and HITs can suggest that the detection of error on the basis of the absence/presence of conflict might depend on a threshold, and that this threshold might be modulated by the cognitive control load.

Transposed in the terminology of the Signal Detection Theory, it is likely that the sensitivity, indexed by the meta-d', of that mechanism remains constant across conditions (see figure 2-G below for a simplified representation), but the threshold of decision would be switched up when the cognitive control is higher.

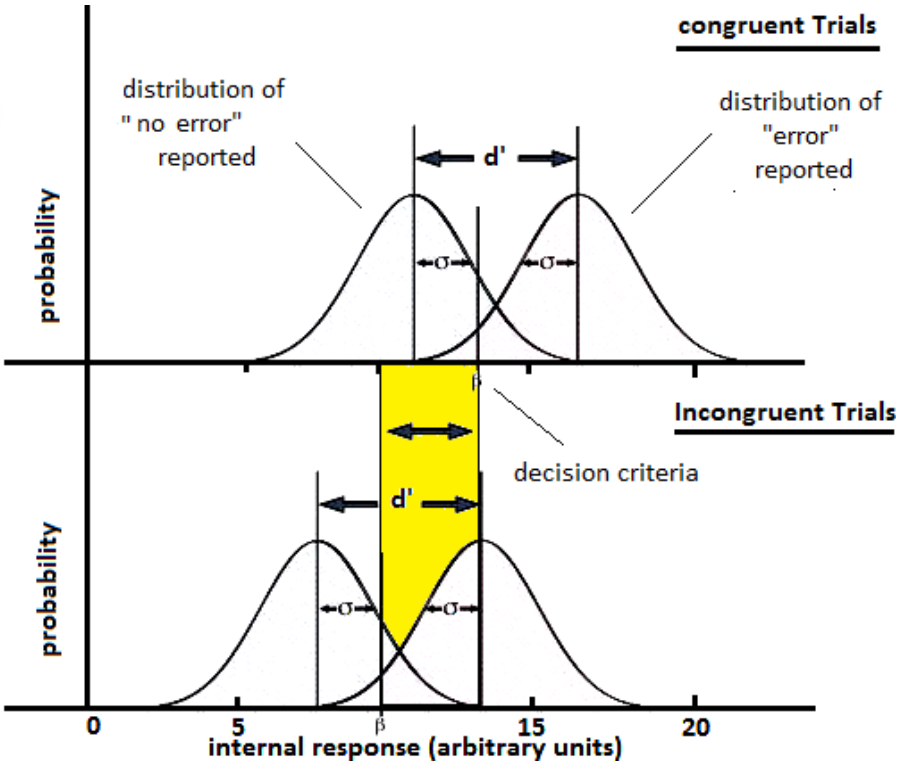


Figure 2-G: constant sensitivity (d-prime) but possible change in decision criteria (beta) according to cognitive control load, when displayed an invisible prime (compatible or not).
Top: congruent trials ; Bottom : Incongruent Trials.

The schematic diagram illustrates the idea that (1) Prime compatibility would only affect the internal response within the action selection processes, on which a decision is made
(2) the cognitive control load (lower or higher, corresponding to congruent versus incongruent trials, respectively) might have an influence not on the *sensitivity* per se (indexed by the d-prime) but only the *criteria* for error detection (indexed by the beta).
The yellow area represent the difference of the beta in congruent versus incongruent trials.

However, the decision criteria (beta), corresponding to the threshold of decision, would be higher when the cognitive control demand is relatively unimportant (congruent trials) or when the primes are visible (long SOA trials), and lower when the prime is invisible (short SOA trials) and the cognitive control load relatively more important (incongruent trials).

*

Different questions follow this pattern of results, concerning the (brain) mechanisms recruited by basic task itself, and those involved in the metacognitive task.

Regarding the basic task, we did not find any compatibility effects in the short SOA condition for accuracy, but we did for the meta-accuracy, so we can suppose despite invisible primes having no effect on the mechanisms recruited, they nevertheless activate them and left a trace on brain activity. A trace that would be (retrospectively) exploited in an error detection task.

Since the FA and HIT rates suggest an effect of compatibility in short SOA and incongruent trails, one could expect for observing a conflict-related during task selection (after the task-cue onset) or response selection (after the target onset).

Regarding the metacognitive task, it would be possible to make the hypothesis that subjects were aware of being slower and thus 'inferred' a probability of error. In the present case this possibility seems unlikely for two reasons.

First, the reaction times were affected by the compatibility in the LONG SOA condition, whereas the false alarms and hits were affected by compatibility in the SHORT SOA+INCONGRUENT condition.

Secondly, even by splitting the dataset according to the SOAs, no correlation was found between reaction times and FA or hits – one should have found some correlation if that scenario was plausible.

It seems that things are simpler, suggesting the existence of a simple 'associative' and bottom up mechanism that, above some degree of perturbation of response selection process, signals the potential for an error (which would explain why when there is no conflict, that is to say when the prime is compatible, both FA and HIT rates significantly decrease).

A recent study (Wenke, Fleming, Haggard, 2010) is consistent with our results or at least with the possible existence of such a simple mechanism based on the detection of conflict during action selection processes.

In their paradigm, the participants had to freely choose between two visual targets displayed on the screen, by pressing a left or right key. Before the target was presented, they were displayed a subliminal arrow, in order to influence the choice of one key and thus exert a bias on the selection of action. The response of the subjects caused the display of colors, which depended on whether the action of the subject was compatible with the prime or not. After having responded, the participants had to rate how much control they thought to have over the different colors. Note that the sense of control, or authorship, entails being aware of one's own action selection processes.

The authors observed that the priming also had effects on the subjective sense of control of the subjects over the effects of their action: the subjects reported more control experienced when the prime was compatible than when it was incompatible. They also demonstrated that the action-effect contingencies were not sufficient to explain the subjective rating of authorship.

This paradigm is comparable to ours in the sense that they both studies attempted to manipulate action (or task) selection processes with supraliminal primes. Moreover they are formally equivalent, involving a *basic task* (free choice of a target, task-cueing), a *meta-task* requiring to be aware of his/her response selection processes (reporting the sense of control over action effects, self-evaluation), and the use of *masked primes.*

The results or an aspect of their results is similar to ours since we observed an effect of the prime compatibility (only for invisible primes) on the error detection capacity of the subject, with fewer errors reported when the prime is compatible (less FAs and less HITs). The idea developed by Wenke et al, that a 'smooth and uncontested' action selection gives rise to a higher sense of control, fits perfectly the idea of a mechanism detecting the 'conflict' or the noise in action selection processes.

2.5.2 Further investigations:

Assuming there were some covert effects of the invisible primes, the compatibility effect does not allow one to situate these covert effects a specific level or stage of processing. There are various

possibilities (Kiesel, Kunde, Hoffman, 2007):

(i) At a perceptual presemantic stage :

Some findings support the existence of priming at perceptual stages, and demonstrate that primes identical to targets give rise to faster responses than only compatible primes (Bodner and Dypvik, 2005). In our case, the primes and task cues were physically rather different, so that a priming at this stage is less likely but not impossible.

(ii) At a semantic stage of processing :

The prime and the task-cue share some semantic properties. Semantic priming is thus likely.

(iii) At a task selection level.

Some studies having investigated the neural signature of effects of priming reported, in the compatible condition, a decrease of activity in both the regions coding for the prime/targets and the prefrontal cortex. A greater functional connectivity between the two regions was also reported (Ghuman et al., 2008).

It is therefore possible that the priming effects observed at perceptual levels could also involve some 'central' processes, those that carry out the selection of task sets or responses on the basis of perceptual input.

In that case, one should be able to observe task-cue locked effects of compatibility, eventually lasting after the target onset or until the response is selected.

(iv) At response selection

Even if priming effects occur at a task selection level, it does not necessarily affect the following stage, namely response selection. That question can be disambiguated only with an imaging technique. To demonstrate (with neuroimaging) that the priming occurs at a task selection, and propagates to the response selection level, brain activity must be considered from the *target onset*, namely after the prime and task cue displays. Yet, in their neuroimaging experiment, Lau and Passingham (2007) convolved the signal from the fixation onset, namely from the beginning of the trial. The prefrontal activation that

they report is not sufficient to demonstrate that priming occurs at a task selection stage and affect response selection.

The next (neuroimaging) investigation will first consist in determining whether response selection is influenced by the non visible primes. A second question will concern the network involved in the basic task, versus in the metacognitive task – the aim being to find a possible overlap as suggested by the two-stage model proposed Pleskac & Busemeyer (2010).

That experiment has been carried out and will be described in the next chapter.

Appendix:

Visibility tests

SOA	Mean confid	Stdev confid	Mean ACC	stdev ACC
16	0,535	0,125	0,823	0,097
84	0,898	0,067	0,964	0,049
Total	0,716	0,211	0,894	0,104

Table 1.1: subjective rating of visibility and objective performance reported for each SOA, 16ms and 84ms)

Basic task (or first-order task):

SOA	CONG	COMP	mean RT	StDev RT	mean ACC	StDev ACC
short	**cong**	**comp**	813.309	170.810	0.961	0.049
		incomp	825.253	162.232	0.962	0.049
	cong Total		819.281	165.571	0.962	0.049
	incong	**comp**	815.732	154.095	0.961	0.043
		incomp	831.689	160.673	0.951	0.040
	incong Total		823.710	156.572	0.956	0.042
ShortSOA_Total			821.496	160.615	0.959	0.045
long	**cong**	**comp**	789.219	170.949	0.962	0.052
		incomp	852.024	174.557	0.967	0.035
	cong Total		820.622	174.494	0.965	0.044
	incong	**comp**	800.736	150.884	0.950	0.047
		incomp	853.754	150.176	0.940	0.070
	incong Total		827.245	151.886	0.945	0.060
longSOA_Total			823.933	163.072	0.955	0.053
Total			822.714	161.586	0.957	0.049

Table 1.2: behavioral performance, first order task

Metacognitive task (or second-order task):

SOA	CONG	COMP	mean meta_ACC	StDev meta_ACC
short	**cong**	**Comp**	0.992	0.018
		Incomp	0.979	0.030
	cong Total		0.986	0.025
	incong	**Comp**	0.989	0.025
		Incomp	0.978	0.025
	incong Total		0.983	0.025
ShortSOA_Total			0.985	0.025
long	**cong**	**Comp**	0.986	0.024
		Incomp	0.990	0.020
	cong Total		0.988	0.022
	incong	**Comp**	0.988	0.019
		Incomp	0.971	0.057
	incong Total		0.979	0.043
longSOA_Total			0.984	0.034
Total			0.984	0.030

Table 1.3: behavioral performance, second order task

SOA	CONG	COMP	% "unknown"
short	cong	comp	1.15
		incomp	3.89
	cong Total		2.52
	incong	comp	2.43
		incomp	3.39
	incongTotal		2.91
ShortSOA_Total			2.67
long	cong	comp	1.52
		incomp	2.81
	congTotal		2.29
	incong	comp	2.42
		incomp	2.12
	incongTotal		2.24
longSOA_Total			2.27
Total			2.42

Table 1.4: percentage of "unknown" responses, second-order task

SOA	CONG	COMP	mean hit	StDev hit
short	cong	comp	0.382	0.470
		incomp	0.614	0.437
	cong Total			0.463
	incong	comp	0.519	0.500
		incomp	0.558	0.497
	incongTotal			0.494
ShortSOA_Total			0.522	0.475
long	cong	comp	0.672	0.441
		incomp	0.640	0.401
	congTotal			0.419
	incong	comp	0.552	0.482
		incomp	0.730	0.408
	incongTotal			0.452
longSOA_Total			0.649	0.431
Total			0.584	0.458

Table 1.5: mean hits (errors detected), second-order task

SOA	CONG	COMP	mean FA	StDev FA
short	cong	comp	0.004	0.012
		incomp	0.010	0.023
	cong Total		0.007	0.018
	incong	comp	0.001	0.005
		incomp	0.009	0.021
	incongTotal		0.005	0.015
ShortSOA_Total			0.006	0.017
long	cong	comp	0.005	0.016
		incomp	0.003	0.013
	congTotal		0.004	0.014
	incong	comp	0.003	0.007
		incomp	0.006	0.027
	incongTotal		0.005	0.020
longSOA_Total			0.004	0.017
Total			0.005	0.017

Table 1.6: mean FA (false alarms, error reported despite a correct response), second-order task

PART III:

Cognitive control load dependent activations in medial prefrontal cortex by invisible but not by visible primes, and an overlap between cognition and metacognition

3.1 Scope and description of the study

In the last chapter we reported a behavioral experiment whereby subjects had to perform a paradigm very similar to the one used by Lau and Passingham in their neuroimaging study (2007), namely a task-cueing paradigm conjoint with masked visual priming similar to the one used by Mattler (2003). The participants were thus displayed *masked* and *unmasked* primes before the task cues (See figure 2-1, chapter II, for the trial procedure). In order to bias the task selection, the primes could be either *compatible* with the task cue (inducing a facilitation) or *incompatible* (eliciting a conflict). In addition, in one third of trials, the subjects were asked to evaluate the accuracy of their response.

As discussed in Part II, we obtained unexpected results. Several hypotheses were thus set up on the basis of this pattern of results that could be explored in a neuroimaging study (cf. *Part II, section 2.5.2, Further investigations*):

- The neural network involved in response selection effectively is influenced by the non visible primes,
- That influence on brain activity should be associated with an effect of prime compatibility

on the accuracy or on meta-accuracy, in the short SOA condition.

- An overlap exists between the networks involved in the basic task (task-cueing), and those involved in the corresponding metacognitive task.

Here we report an fMRI study of which main purpose was to determine whether response selection could be influenced by the non visible primes, in other words, whether unseen primes could influence brain activity at a level hierarchically superior to response selection (in the sense defined by Badre or Koechlin, cf Part I, *Hierarchical models of cognitive control*) and propagate down to response selection.

3.1.1 Notions keys and factors:

The behavioral paradigm we used in that study was similar to the one reported in the previous chapter, but was adapted for a neuroimaging study (jitters, additional baseline conditions allowing one to make contrasts, reduced number of trials).

As in the purely behavioral version of that study, in addition to the prime visibility and compatibility, we considered a third factor, quantitative and with two levels only, namely the cognitive control load associated with response selection. We did so by manipulating the *congruency* of the targets. That notion, and its link with the cognitive control load, have already been defined and developed in the previous chapters, but is re-explained and schematized in Figure 3-2. The trials could then be split into two types, namely those involving a higher cognitive control load (incongruent targets) and those involving a lower cognitive control load (congruent targets).

The interest of that factor consists in setting two levels of cognitive control load. A higher cognitive control load a priori is associated with relatively more activity within the lateral prefrontal cortex. A gradient of the cognitive control load has been demonstrated to parallel a gradient of activations within the lateral prefrontal cortex, and to be associated with longer reaction times as the load is increasing (Koechlin et al., 2003).

Furthermore, diverse electrophysiological studies have reported several interesting properties of the (lateral) prefrontal neurons. These studies suggest that the tonic activity of prefrontal neurons influence

the current response selection: (i) the tonic neuronal activity begins at task-cue display and lasts until the onset of the response (Quintana and Fuster, 1999), (ii) the tonic neuronal activity is motor task dependent (Hoshi et al, 1998) even though the rule, on which the decision is contingent, is very abstract (Wallis and Miller, 2003).

Importantly, despite the fact the congruency and compatibility factors both make the response more 'demanding', and can both give rise to longer reaction times by adding an additional mental operation, they intrinsically differ regarding the nature of the mechanisms of the decision making processes they tap in. (cf. Part I section 3.4, *Motivational versus Cognitive Control, medial versus lateral prefrontal cortex*)

The ***congruency*** factor is held to depend on the *hierarchical* and *serial* top-down control of response selection, and affects the quantity of information necessary to select the correct response (cf. Part I for the definition of this notion, section 3.4, *Motivational versus Cognitive Control, medial versus lateral prefrontal cortex*). The number of serial decisions is proportional to that quantity – this is why incongruent targets give rise to longer reaction times – as we observed in our previous behavioral study (see Part II). Insofar as it only refers to the number of serial branches of the tree-like decision process, that factor purely pertains to *cognitive control*.

The ***compatibility*** factor, however, depends on *parallel* bottom up processes of response- or task-selection. It is supposed to affect a single decision step by eliciting a conflict/facilitation of the selection. The main current issue is whether this conflict/facilitation can influence the following steps –in the present case, if the conflict at a task selection level can affect the upcoming selection of the response. Assuming it does, overcoming a response conflict or being facilitated in response selection involves a variation of the energy release required to select the response. In that respect, the compatibility stands as a *motivational control* factor (cf. Part I section 3.4, *Cognitive versus motivational control, lateral versus medial prefrontal cortex*). The subjects knew they could be presented with a prime, but the compatible/incompatible nature of the prime remained unpredicted and unpredictable. According to the frameworks that have been presented beforehand, the display of the prime should elicit a (medial

prefrontal) neuronal response due to a variation of the probability of selecting the incorrect response, (cf. Part I section 3.4, *Cognitive versus motivational control, lateral versus medial prefrontal cortex*).

These notions of *cognitive* versus *motivational* control have been defined previously (cf. Part I section 3.4), and associated with different loci of activation within the prefrontal cortex, namely lateral and medial, respectively. The functional segregation between lateral and medial activations has been underlying our expectations regarding the patterns of activations in different conditions.

In effect, a plethora of studies converge toward the global view that the medial prefrontal structures energize the lateral ones upon conflict-, error- or reward-related signals (i.e motivational factors), independently of the output modality and domain processing (Barch et al, 2001, Koechlin et al., 2009). More specifically related to response conflict, previous studies (Botvinick et al., 2001; Miller and Cohen, 2001 ; Kerns et al., 2004 ; Johnston et al., 2007) provided evidence that the anterior cingulate cortex (ACC) contributes to action selection by tracking conflict between competing responses, thus reflecting the (motivational) demands of the current response selection (Yeung and Nieuwenhuis, 2009). Increased anterior cingulate cortex activity is generally expected in demanding situations such as when overriding habitual actions or coping with a response conflict –giving rise to an error or not.

In the present case, if the conflict at the task setting stage influences upcoming response selection, then some activations could be expected in the medial regions typically involved in response selection and/or response conflict monitoring (including anterior cingulate cortex), and possibly more anterior prefrontal regions since the response is conditioned by a rule, (see below for more details regarding our expectations).

3.1.2 Some important aspects

Finally, note that, since we were expecting prefrontal activations, we tried to neutralize the possible confounds that could interact with the factors we manipulated, namely:

(i) The practice level, that should be as perfect as possible insofar an insufficient practice can increase prefrontal activations (Fletcher, Shallice, Dolan, 2000);

(ii) The choice of the task sets (since the paradigm involved two consecutive decisions, we opted for

binary questions giving rise to binary properties of the cue and targets, in order to minimize the variance of the resulting reaction times) ;

(iii) The temporal structure within each trial (to prevent bottleneck effects).

Moreover, we were looking for prefrontal activation related to response selection; we therefore "locked" the analysis on the *target*, and did not consider the hemodynamic activity occurring before target onset.

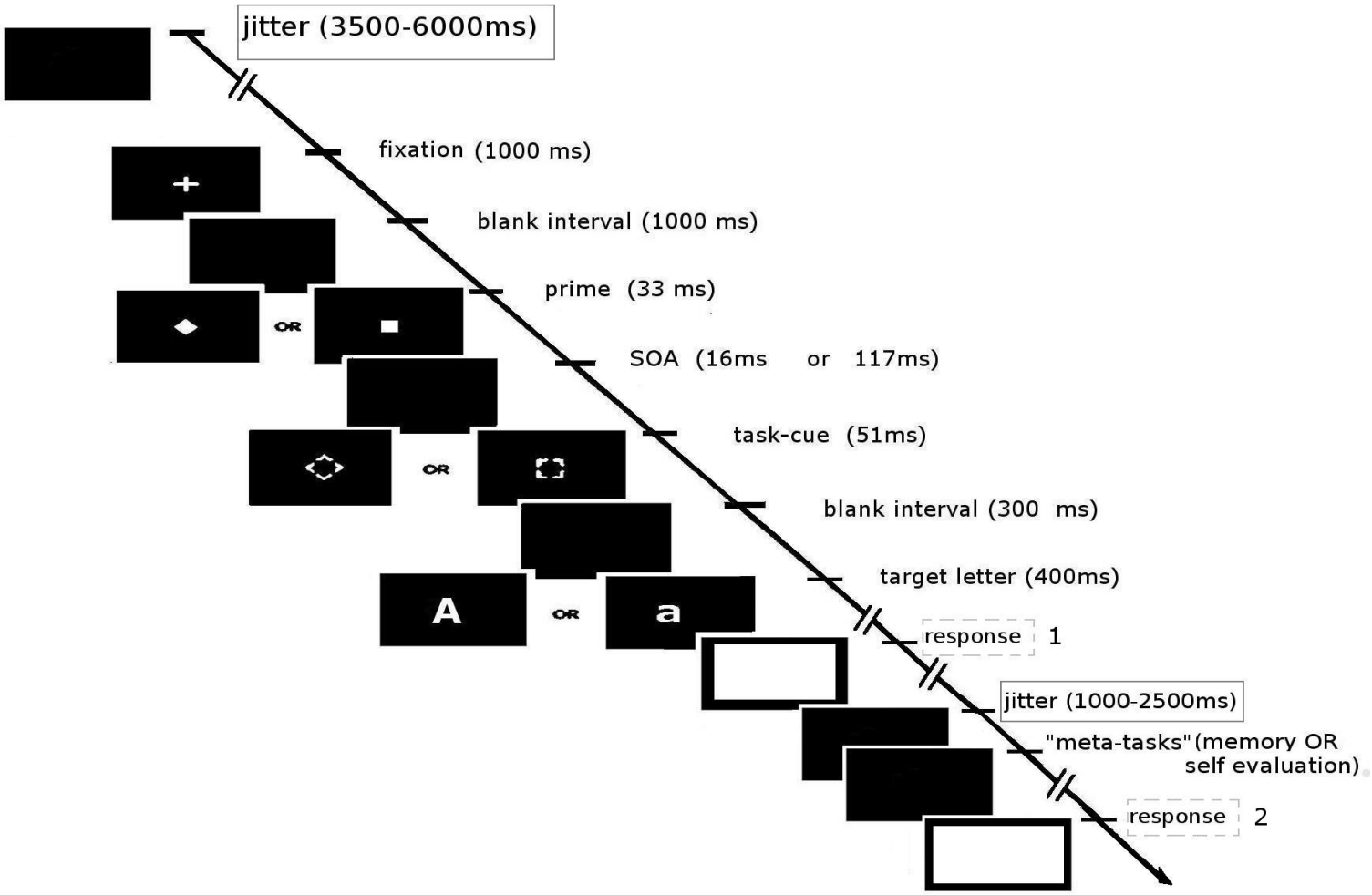
A second issue addressed in that study was to determine whether there was an overlap between the network involved in the basic task and the one involved in the corresponding metacognitive task. We previously (cf. *Part II*) observed that the presence of masked primes did not affect the performance of the subjects in the task-selection paradigm, but they influenced the awareness of the subjects' performance. On the basis of that pattern of results, we made the hypothesis that masked primes could nevertheless influence the action selection processes and then could either give rise to less accurate responses, or leave a trace that biased the immediate retrospective self-evaluation of the subjects.

Thus, basically, we tried to visualize the network involved in the metacognitive task, by removing the sensory, motor, and memory components of the task (cf below, paradigm) and independently of the factors manipulated for the basic task, that is to say independently of the prime visibility, compatibility and of the cognitive load.

We expected to find regions involved in both response and task selection, including anterior cingulate, medial premotor cortex and lateral prefrontal cortex.

Note that, on the basis of the outcome of the previous behavioral experiment, whereby we were unable to compute a meta-d' for each subject and each condition because of the too few errors, we expected the same situation (too few errors). However we do consider that analyses with meta-d' as regressor would have been relevant in the present study.

3.2 Hypothesis and Expectations



[**Figure 3.1**: trial Procedure]

3.2.1 Behavior

Because of practical and technical constraints due to neuroimaging protocols (the participants cannot stay for more than one hour within the scanner; necessity to introduce jittered long time intervals between the trials). The study involved fewer trials compared with our previous behavioral studies, namely half as many. Moreover, for practical reasons it was possible to recruit 20 subjects only, instead of the 24 that we aimed to test. Finally, it seems that the effects we were tracking are small (Dehaene, 2008; Van Gaal et al, 2008; Sackur and Dehaene, 2009), though significant.

In spite of these constraints, we expected a pattern similar to the one obtained in our previous behavioral study (cf. *Part II*). Regarding accuracy, we expected an effect of prime compatibility, but in the long SOA condition only, and possible effects of congruence regardless of SOA. Regarding the reaction times, we expected to obtain a main effect of the congruence, since incongruent trials involve a necessary additional serial mental operation.

3.2.2 Neuroimaging

Ideally we were interested in analyzing the signal that was as "response -locked" as possible, limiting as far as possible any prefrontal activations that could be due to the executive components of the previous perceptual or attentional selection stages (Ghuman et al., 2008). We thus chose to convolve the HRF function with a (boxcar) function of which onset was defined as the *target onset* and of which duration as the response latency.

We were expecting at least an effect of congruency, at both SOA, associated with a dorsolateral prefrontal (BA46/9) activation (Koechlin et al, 2003). We also were expecting a compatibility effect in the long SOA trials associated with medial activations involved in response selection, namely anterior cingulate cortex (BA 24), possibly medial premotor cortex (BA6).

Note that we did not exclude the possibility of compatibility effects in short SOA trials even ones not associated with accuracy or reaction times effects, but with meta-accuracy or meta reaction times. In our previous behavioral experiment (cf. part II, section 2.5), we had observed that invisible primes did not influence the performance *per se* but the awareness of the performance (giving rise to significant effects of compatibility in short SOA trials only, for both hits and false alarms in an error detection task). We thus made the hypothesis that some trace could or should have remained in the regions involved in monitoring response selection/response conflict.

3.3 Materials and Methods

3.3.1 Participants.

20 right-handed healthy volunteers (9 females; 24,8 ± 3.6 years) participated in the study. All participants had no existing neurological or psychiatric illness, and gave written informed consent. The study was approved by the independent Ethics Committee of the "Santa Maria della Misericordia" Udine Hospital.

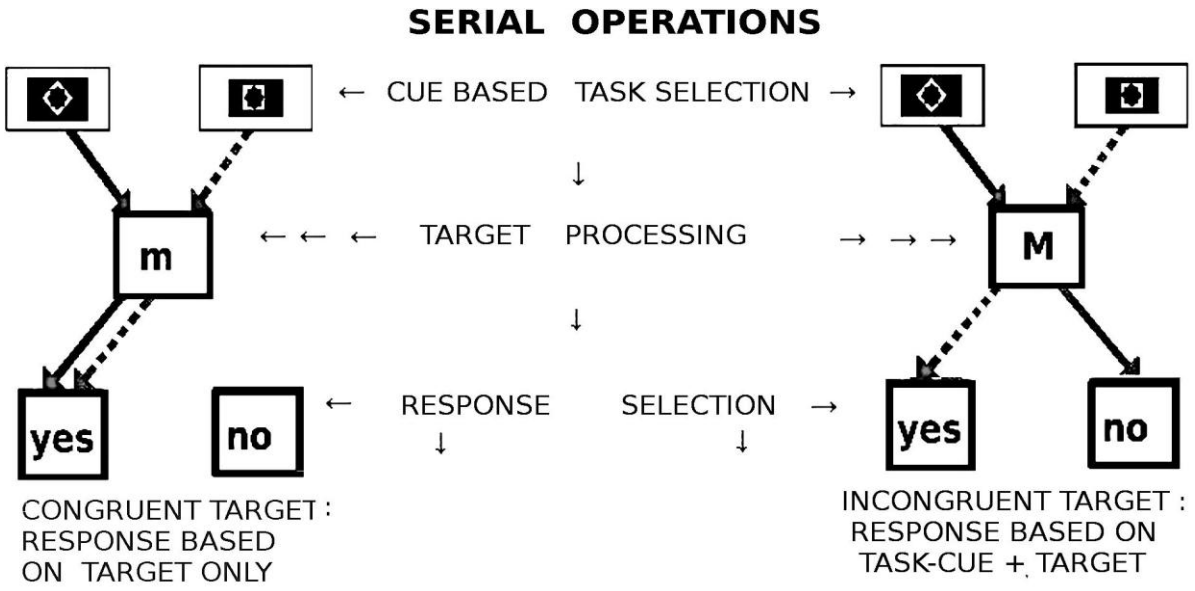
3.3.2 Stimuli and design.

Basic Task

On every trial, (Figure 3.1) the participants were displayed a sequence of three stimuli after a fixation cross followed by a fixed time interval of 1000 msec. The first stimulus could be a little square or diamond, situated at 2 degrees above or below the fixation cross. The prime and the mask were displayed at the periphery in order to increase the masking effect (Vorberg, 2003), and that could be clearly visible or invisible/nearly invisible. The subjects were instructed to ignore this first figure when they managed to see it. The second stimulus was another geometrical shape, that could also be a diamond or a square. That shape had two roles (Mattler, 2003 ; Lau & Passingham, 2007) : it masked the previous prime when the SOA was short (16ms) and second, it indicated which task participants had to perform regarding the upcoming letter –namely a consonant judgment (*"era una consonante?"* meaning *"was it a consonant?"*) or a lowercase judgment (*"era minuscola ?"* meaning *"was it in lowercase?"*).
After a blank interval of 300 ms, a target letter was displayed for 400ms. According to the task cue and the letter, the subject had to press a right or left key to say if YES or NO, the current letter was a consonant or not, or a lowercase or not.

The *cue/task* mapping and *key/response* mapping were both counterbalanced across subjects.
A set of 16 letters was used and perfectly balanced in terms of proportions of consonant/vowels, uppercase/lowercase, congruent/incongruent. In each block, each letter appeared 4 times in a random

order, so that one block comprised 64 trials. Each block was balanced in terms of compatible/incompatible primes, short/long SOA, task-cues, congruent/incongruent trials.



[**Figure 3.2**: distinction between congruent and incongruent trials]

The cognitive control load associated with response selection was controlled by manipulating the congruence of the targets. As showed in figure 3-2, some target letters (on the right), namely *consonants in uppercase* and *vowels in lowercase*, always elicit a different response according to the task cue, so that two signals are necessary for the response selection. On the contrary, other targets (left), namely *consonants in lowercase* and *vowels in uppercase*, always elicit the same response independently of the task-cue, so that the response selection is conditioned by only one signal (the target) and the operation is formally equivalent to a simple stimulus response association. Incongruent trials thus a priori entailed a higher cognitive load, and the congruent trials a lower one.

Meta-task

After having responded and following a short jittered blank interval (1000-2500ms), the participants had to answer a second question that could consist of either remembering a property of the target (two-thirds of the time, a *memory task*) or self-evaluating (one third of the time, metacognitive or *meta-task*).

Thus three possible questions could be displayed on the screen for 1000 msec: *"era una consonante?" ("was the letter a consonant?")* , *"era minuscola?" ("was the letter in lowercase?")* (corresponding to the memory tasks) or *"ha risposto giusto?" ("did you answer correctly?")* (meta-task). As for the memory tasks, the nature and the proportion of the task-cues in the same trial were controlled and balanced, in such a way that half of the time they were asked a question referring to a property of the letter that was *relevant* for the task, or *irrelevant* for the task. The relevance of the memory task was thus controlled and entered as a factor afterwards.

The participants could respond *"yes"*, *"no"* or *"I don't know"* by pressing a third key with the mid finger, and were encouraged to respond *"I don't know"* if they had any doubt. That allowed us to minimize the number of guessed correct responses and to take into account only confident responses.

After a second longer jittered blank interval (3500-6000ms), the next trial was displayed.

3.3.3 Procedure: training phase and experimental phase

In order to minimize the prefrontal activations because of lack of practice of the basic task, the participants had to carry out an intense training whereby the performance was controlled in terms of accuracy and response speed. The training consisted in two steps.

During the first step, after being explained the rules by the experimenter and having a demonstration displayed, each participant practiced the basic task only. The training program comprised 90 trials with a feedback given immediately after each response. The participant stet informed that to pass to the second step, they had to reach a quite high performance level, corresponding to an accuracy greater than 90 percent of correct responses and a mean response less than 1000 msec. They were also informed that, although speed was important, accuracy was the priority, for both the training and the experimental phase as well.

Early in the first training phase, after the 30th trial, the program stopped and gave the participant his/her mean accuracy and response speed computed upon the first 30 responses, then went on for 60 additional trials. Before ending, the program gave a final estimate of the performance and decided whether the participant had to carry on the practice or not. Participants generally repeated that phase twice or three times (hence they reached the requested performance level after 180-270 trials of practice).

The second step of the training consisted in a block of 64 trials separated from each other by a jittered time interval, *without feedback* and including either meta-task or memory questions after each response, so that the subject was already familiar with the test environment before entering the scanner.

Between each trial, there were the same jittered blank intervals as those of the upcoming experiment. If the participant successfully passed that second step of the training (more than 90% of correct responses on the basic task), he/she could enter into the scanner for the experimental phase. If not, he had to repeat it until reaching the requested performance for one block. All the participants successfully passed that second step, after one or two blocks.

The experimental procedure in the scanner involved 256 trials in total, divided into four fMRI runs of 64 trials each. Each run began and finished with a fixation period of 15 seconds. Each trial comprised two responses: namely the response to the basic task (that one will name *Accuracy* and *Reaction Times*) then the response to the meta-task/memory questions (that one will named *Meta-Accuracy* and *Meta- Reaction Times* for no confusion).

3.3.4 fMRI methods: acquisition and analysis.

Images were acquired using a 3-T MRI scanner (Achieva 3.0 T Philips Medical Systems, Netherlands) equipped with a standard quadrature head coil and for echo-planar imaging (EPI). Head movement was minimized by mild restraint and cushioning. Thirty-four slices of functional MR images were acquired using blood oxygenation level-dependent imaging (3.59 mm × 3.59 mm, 4 mm thick, repetition time = 2 s, time echo = 35 ms; flip angle: 90; field of view, FOV: 23 cm, acquisition matrix: 64 × 64; SENSE factors: 2 in anterior–posterior direction), covering the entire cortex. At the beginning of the scanning session, anatomical scans were also acquired for each participant, using a T1-weighted MP-RAGE (magnetization-prepared, rapid acquisition gradient echo).

The experimental stimuli were controlled using the Presentation software (Neurobehavioral Systems,

Inc.) and delivered within the scanner by means of MR-compatible goggles mounted on the coil.

SPM8 (Wellcome Department of Cognitive Neurology; www.fil.ion.ucl.ac.uk/spm/software/spm8/) was used for both data preprocessing and statistical analyses. About 1600 volumes were acquired on average for each participant (400 volumes on average for each fMRI-run).

The last volume was discarded for each run. All images were corrected for head movement; slice-acquisition delays were corrected using the middle slice as reference. All images were then normalized to the standard SPM EPI template and spatially smoothed using an 8 mm FWHM Gaussian filter set to the cut-off value of 128 s.

All subsequent analysis of the functional images were performed using the general linear model implemented in SPM8.

For the correct trials of the basic task, blocks were epochs, one epoch for each of the 8 conditions (i.e. SOA (2) x Prime Compatibility (2) x Congruency (2)).
A box-car function was defined for each trial of the basic task to convolve the hemodynamic response function (HRF). Since we were interested in the possible effect of the cognitive control load associated with the response selection, the onset of each epoch was determined by the onset of the *target*. The duration of each boxcar function corresponded to the latency of the first button press generated by the participant in response to the target letter.

For the correct trials of the Meta-task, we determined the onset of each epoch by the onset of the question (*"was the letter a consonant?"* or *"was the letter a lowercase?"* or *"did you answer correctly?"*) display and convolved with the hemodynamic response function (HRF) as well. Each of the 5 conditions (namely relevant semantic memory, irrelevant semantic memory, relevant case memory, irrelevant case memory, meta-task) was modeled using a boxcar function. As previously, the duration of each boxcar function corresponded to the latency of the first button pressed. Thus, for the Basic task and Meta-task, the resulting beta values represented an estimate of the neural response per unit time spent in selecting the response.

The first-level analysis also included the parameters of the realignment (motion correction), errors in

basic task, and errors in each meta-task, as covariates of no interest.

Finally, the Basic task analysis and Meta-task analysis were performed independently. As for the Basic Task, the 8 different conditions were entered as regressors in a second-level ANOVA full factorial *SOA(2)xCompatibility(2)xCongruency(2)* model, while for the Meta-Task, 5 different conditions were entered as regressors in a second level linear model *((memory task(2)*Relevance(2) + Meta-task(1))).*

A standard procedure was used, so Statistical threshold was set to p-corr. < 0.05 corrected at the cluster level using FWE (cluster size estimated using a p-uncorr < 0.001). In addition, we followed the same procedure as we did in the chapter two, by splitting the data by SOA and used a Bonferroni correction.

3.4 Results

3.4.1 Basic Task

3.4.1.1 Behavioral data

The data were cleaned (the reaction times greater than 3000ms or associated with an incorrect response were eliminated), then submitted to a Shapiro-Wilk normality test. All the measures, including log transformed reaction times and arcsine transformed accuracy failed the test. Hence a pairwise Wilcoxon test was systematically used for the comparisons. All the behavioral data were analyzed with R software.

Accuracy

The analysis (pairwise Wilcoxon test) performed on the whole dataset revealed a main significant effect of compatibility ($p<.031$). The congruency factor was not significant ($p>.87$), neither the SOA factor, which failed to reach significance ($p>.08$).

Short SOA:

A pairwise Wilcoxon (Bonferroni correction, p set at 0.025) test performed on short SOA trials revealed a significant effect of Compatibility ($p< 0.01$) -- below. The congruency factor was not significant ($p>.62$).

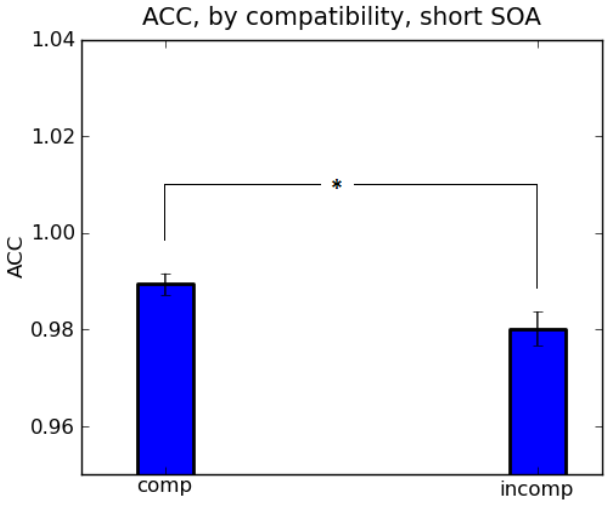


Figure 3-3: accuracy, compatibility effect in SHORT SOA trials, bars represent standard Errors

Long SOA:

A pairwise Wilcoxon test performed on long SOA trials did not reveal any significant effect of Compatibility (p> 0.20), nor of Congruence (p > 0.59).

Reaction times

The analysis (pairwise Wilcoxon test) performed on the whole dataset revealed a main significant effect of congruency (p<.002). The compatibility factor was not significant (p>.20), neither the SOA factor (p>.70).

Short SOA:

A pairwise Wilcoxon test performed on short SOA trials revealed no effect of compatibility (p>.43), but a significant effect of congruency (p<.002) – figure 3-4 below.

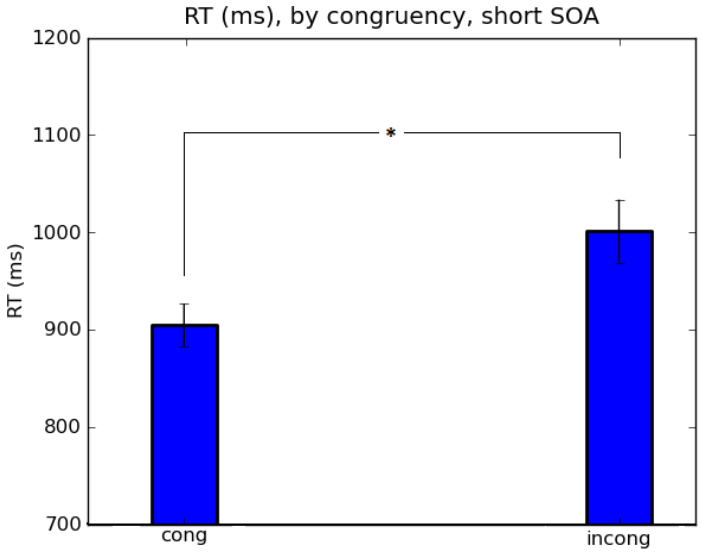


Figure 3-4: RT, congruency effect in SHORT SOA trials, bars represent standard Errors

Long SOA:

A pairwise Wilcoxon test performed on long SOA trials revealed a quite similar pattern, namely a non significant effect of Compatibility ($p > 0.10$), and a significant effect of Congruence ($p < 0.002$).

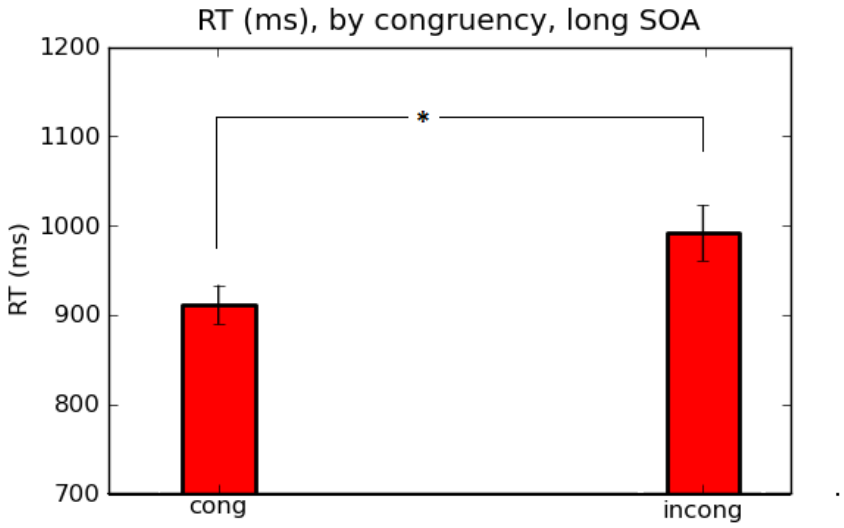


Figure 3-5: RT, congruency effect in LONG SOA trials, bars represent standard Errors

SOA	CONGRUENCE	COMPATIB	RT (ms) Mean	RT (ms) SD	ACC Mean	ACC SD
Short SOA	congruent	Compatible	914.83	±240.51	0.99	±0.02
		Incompatible	941.07	±206.07	0.98	±0.04
	congruent Total		927.95	±221.46	0.99	±0.03
	incongruent	Compatible	1008.74	±280.35	0.99	±0.02
		Incompatible	997.45	±263.67	0.98	±0.03
	incongruent Total		1003.10	±268.69	0.99	±0.02
Short SOA Total			**965.52**	**±247.55**	**0.99**	**±0.03**
Long SOA	congruent	Compatible	918.59	±170.73	0.99	±0.01
		Incompatible	948.09	±244.31	0.98	±0.03
	congruent Total		933.34	±208.57	0.99	±0.03
	incongruent	Compatible	1000.07	±262.57	0.99	±0.02
		Incompatible	1014.65	±249.25	0.99	±0.02
	incongruent Total		1007.36	±252.80	0.99	±0.02
Long SOA Total			**970.35**	**±233.26**	**0.99**	**±0.02**
Total			**967.94**	**±239.77**	**0.99**	**±0.02**

Table 3-1: Reaction Times and Accuracy by condition

3.4.1.2 Neuroimaging data

3.4.1.2.1 Three factor model: SOA(2), compatibility(2) and congruency(2)

We first tested for the main effects, 2-way and 3-way interactions of the 3-way ANOVA, with *SOA(2), compatibility(2)* and *congruency(2*) as factors.

(1) Main effects:

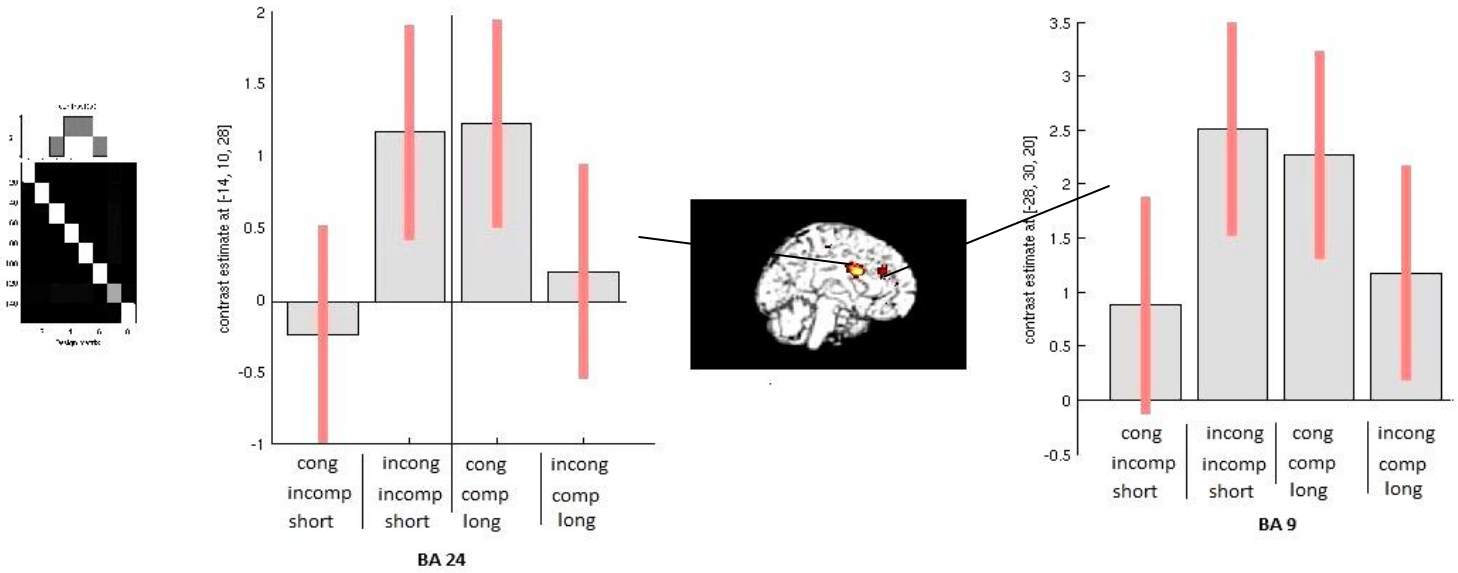
The 3way ANOVA revealed a nearly significant main effect of congruency, associated with an activation within the (left) ventrolateral prefrontal cortex [BA 11, (-40; 46;-10), 218 voxels, Z= 4.00, p-FWE-corr =0.067, plots are visible at the end of the current chapter, supplemental data].

(2) 2-way interaction(s):

The 3way ANOVA revealed a nearly significant congruency*compatibility interaction, associated with an activation of the Left ventral Anterior cingulate [BA 24, (-8; 2;32), 212 voxels, Z= 4.73, p-FWE-corr =0.07, plots are visible at the end of the current chapter, supplemental data].

(3) 3-way interactions:

The 3way ANOVA revealed a significant congruency*compatibility*SOA interaction (figure 3-6, table 3-2) associated with the activation of the left ventral anterior cingulate (BA 24) and left medial frontal gyrus (BA9).



[**Figure 3-6** : prime compatibility *congruence *soa Interaction ,
Left ventral Anterior cingulate, BA 24(-8;2;32), Left Medial Frontal Gyrus, BA 9 (-16, 34, 28),
see **table 3-2** for the corresponding labels, coordinates, cluster size, z-score, p-values]

*

Since (i) we observed a three-way interaction, (ii) we considered as relevant the comparison of the different effects in short SOA and in long SOA conditions, on the basis of our previous behavioral study, (iii) we had hypothesized that one would observe brain activity associated with an effect of prime compatibility on the accuracy or on meta-accuracy, in the short SOA condition (cf. hypothesis 2, introduction of the current chapter), we intended to explore the Congruency and Compatibility effects, including two-factor interactions, in low (short SOA) versus high (long SOA) visibility conditions.

We therefore carried out two 2-way ANOVA with *compatibility(2)* and *congruency(2)* as factors, as we did previously (cf. Part II).

3.4.1.2.2 Two factor models (by SOA) : compatibility(2) and congruency(2)

We were expecting different effects o according to the visibility (cf. PART II), and split the dataset into short SOA / long SOA, to see whether we could observe different effects in each condition, as we did in our previous experiment.

Short SOA :

The 2-way ANOVA carried out in short SOA trials revealed a clear *compatibility*congruence* interaction associated with a medial prefrontal network including left ventral anterior cingulate (BA 24) and frontopolar (BA 10) cortices, and marginally, the dorsal anterior cingulate cortex (BA 32) (Figure 3-7a).

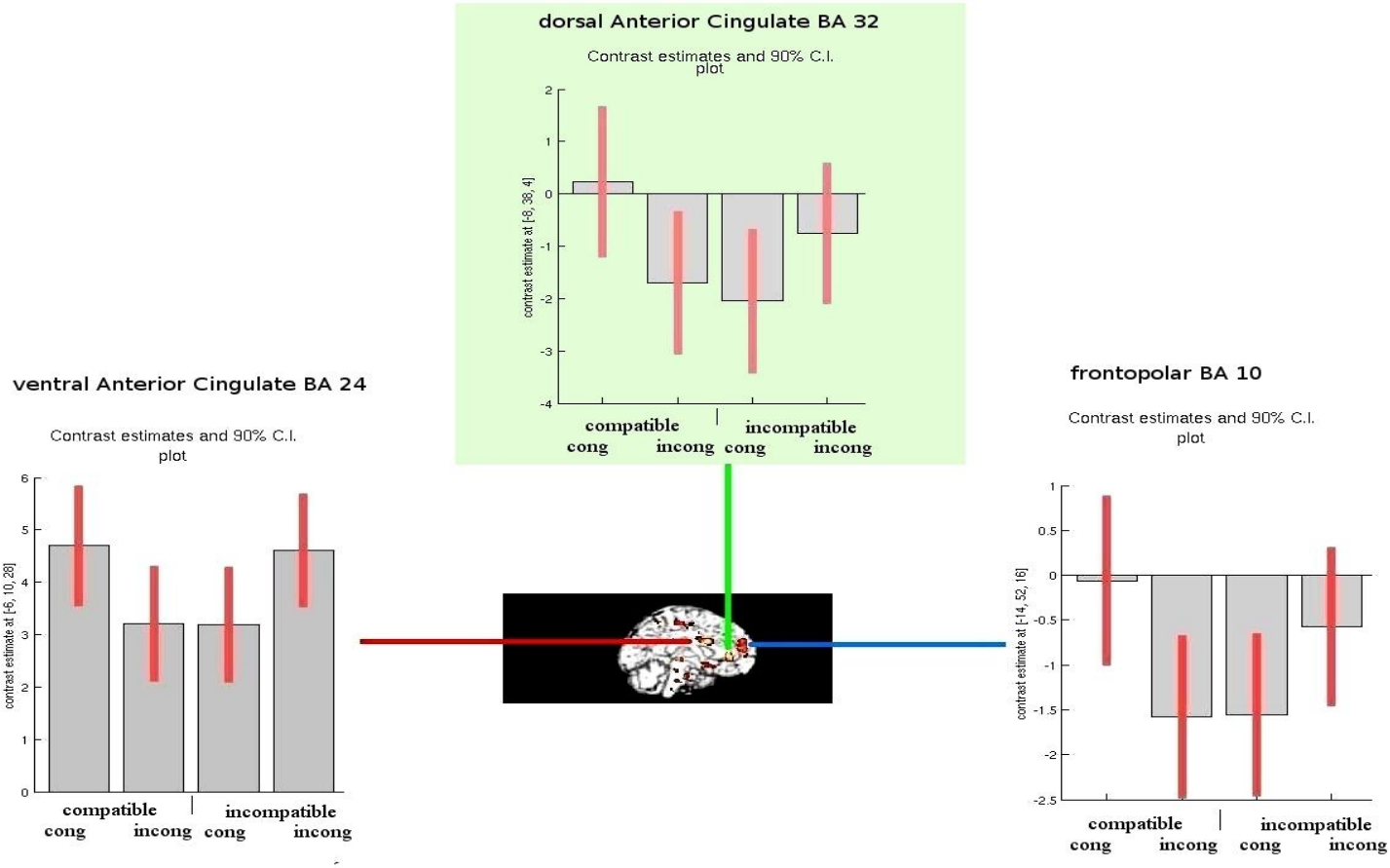


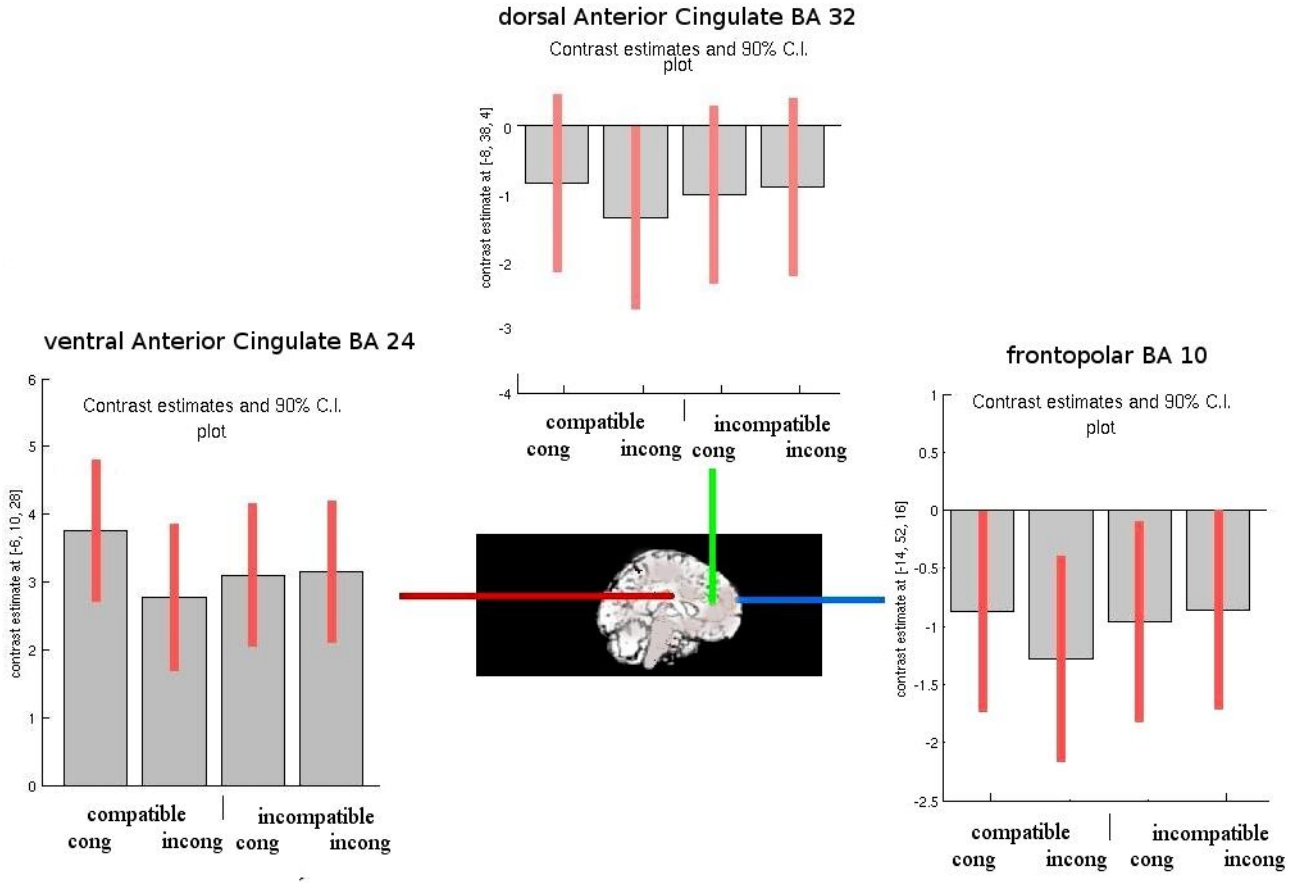
Figure 3-7a: prime compatibility *congruence Interaction,
SHORT SOA condition,
ventral ACC (-6, 10, 28) ; medial frontopolar (-8, 38, 4);
NEARLY SIGNIFICANT dorsal ACC (-14, 52, 16)]

Label	x,y,z	cluster size	Z-score	P-value (FWE-corr)
COMPATIB*SOA*CONGRUENCY :				
Left ventral Anterior cingulate, BA 24	(-8;2;32)	462	5.55	0.003
Left Medial Frontal Gyrus, BA 9	(-16, 34, 28)	241	4.02	0.049
COMPATIB*CONGRUENCY , short SOA:				
Left ventral Anterior cingulate, BA 24	(-6, 10, 28)	245	4.55	0.04
Left Medial Frontal Gyrus, BA 10	(-14, 52, 16)	283	4.70	0.02
Left dorsal Anterior Cingulate, BA 32	(-8, 38, 4)	229	4.04	0.056

[table 3-2 : labels, coordinates cluster size, z-score, p-values (FWE-corr) in contrasts of the basic task, GREEN indicates nearly significance**]**

Long SOA :

We did not observe any significant effect, nor interaction, in the long SOA (high visibility) condition– see figure 3-7b, not significant, just for the comparison with the short SOA condition.



[**Figure 3-7b** : **NOT SIGNIFICANT** prime compatibility *congruence Interaction**,***
LONG SOA condition, ventral ACC (-6, 10, 28); medial frontopolar (-8, 38, 4); dorsal ACC (-14, 52, 16)]

Summary of the behavioral and neuroimaging outcomes for the basic task:

On the basis of the whole dataset, we observed a main effect of compatibility for accuracy, and of congruency for the reaction times (pairwise Wilcoxon).

By splitting the dataset according to SOA, pairwise Wilcoxon tests allowed us to observe:

- a significant *compatibility* effect *restricted in short SOA trials* for the accuracy,

- a significant *congruency effect at both SOA* for the Reaction Times.

In addition to this behavioral pattern of results, we observed the following neuroimaging outcomes:

- A nearly significant main effect of *congruency* and associated with a prefrontal ventrolateral (BA 11) activation[5] --This region tended to be less active in incongruent trials.

-A nearly significant *compatibility*congruency 2-way interaction*, associated with a left ventral anterior cingulate (BA 24) activation

- A *compatibility*congruency*soa 3-way interaction,* associated with activations in left ventral anterior cingulate (BA 24) and left medial frontal gyrus (BA9).

By splitting the dataset according to SOA:

- a *significant interaction between congruency and compatibility in the short SOA condition only*, associated with a *medial prefrontal* network, including differential activations in BA24, and differential deactivations BA10 and marginally the dorsal Anterior cingulate cortex (BA32).

[5] In view of the widespread errors of the fMRI reporting of statistics (Vul and Paschler, 2009), we do not comment on nearly significant results unless other analysis motivate it.

3.4.2 Meta-Tasks:

3.4.2.1 Behavioral data

Note that two orthogonal aspects of the metacognitive performance were of interest.

(i) We were essentially interested in the eventual effects of the factors manipulated for the basic task (namely *compatibility* and *congruency,* in each SOA condition) on the metacognitive performance in the error detection task.

As a secondary interest, we looked for the differences between the meta-task and the memory tasks, in order to dissociate them (and justify the contrast carried out for the neuroimaging analysis).

For the memory tasks, note we also considered the possible effects of the *relevance* factors – that refers to whether the information the subjects had to recollect was relevant or not to select the right response. The other factors were also considered.

The data were cleaned (reaction times superior to 2000ms or associated with an incorrect response or unconfident response were eliminated). Both raw data and transformed data (including log transformed meta reaction times, arcsine transform meta accuracy) were beforehand submitted to a Shapiro-Wilk normality test. According to the outcome they could be entered either into a pairwise Wilcoxon test, or into an ANOVA. All the behavioral data were analyzed with R software. One subject was excluded of the sample for the following analysis because he turned out to be an outlier.

Meta-accuracy

For the meta-accuracy, a response was considered as correct when it was confident and correct. Unconfident responses (corresponding to *"I don't know"* responses) were not removed, but considered as incorrect ones.

Shapiro normality tests indicated that meta-accuracy, even arcsine transformed, was not normal ($p<.001$). We thus carried out pairwise Wilcoxon tests.

(i) As we did for the previous behavioral experiments, we split the dataset according to the SOA:

Short SOA:

A pairwise Wilcoxon test performed on short SOA trials revealed no significant effect of Compatibility ($p>.22$). The congruency factor was not significant either ($p> .35$).

Long SOA:

A pairwise Wilcoxon test performed on long SOA trials revealed no significant effect of compatibility ($p>.21$), nor of congruency ($p>.85$).

(ii) We then compared the meta accuracy in the so called meta-task with the memory tasks (Wilcoxon pairwise tests).

The performance in the meta-task was not significantly different from the performance in the memory tasks ($p>.60$).

Considering the accuracy in the memory tasks only, it was not influenced by the SOA ($p>.58$), congruency ($p>.61$), compatibility ($p>.15$) nor relevance ($p>.17$).

Meta Reaction times

The Shapiro normality test revealed that the distribution of log transformed meta reaction times likely were normal (p>.43). Thus they were entered in an ANOVA.

The within subject 3-way ANOVA (SOA*congruency*compatibility) did not reveal anything significant.

(i) By splitting the dataset according to the SOA we found:

Short SOA:

In short SOA trials, the within subject 2-way ANOVA with *compatibility (2*) and *congruency (2)* as factors revealed no significant effect of compatibility (F=0.03; p>.85), nor of congruency (F=0.80; p>.37).

Long SOA:

In long SOA trials, the within subject 2-way ANOVA with *compatibility(2*) and *congruency(2)* as factors revealed that neither the compatibility (F=0.87 ; p>.37), nor the congruency effects (F=0, p>.99) were significant.

(ii) We then compared the reaction times in the meta-task with those of the memory tasks. A pairwise two-tailed t-test indicated that the reaction times in meta-task significantly differed from the reactions times in the memory task (t=11.48; p<0.001) -- the reaction times of the memory tasks being greater.

The reaction times in the two memory tasks (*'post-lowercase'* versus *'post-consonant'*) did not differ one from each other(t=1.23 ; p>.22), and were not significantly influenced by the relevance factor (t=0.17 ; p>.86).

SOA	CONGRUENCY	COMPATIB	MRT(ms) Mean	SD	Meta ACC Mean	SD
Short SOA	congruent	compatible	834.97	±167.22	0.94	±0.20
		incompatible	835.82	±184.46	0.93	±0.21
	congruent Total		835.40	±173.78	0.93	±0.20
	incongruent	compatible	794.09	±132.26	0.92	±0.17
		incompatible	806.68	±139.54	0.92	±0.18
	incongruent Total		800.22	±134.21	0.92	±0.17
Short SOA Total			818.03	±155.53	0.93	±0.19
Long SOA	congruent	compatible	842.92	±147.76	0.93	±0.23
		incompatible	811.92	±136.45	0.92	±0.17
	congruent Total		827.42	±141.26	0.93	±0.20
	incongruent	compatible	835.98	±106.23	0.93	±0.22
		incompatible	813.56	±142.08	0.92	±0.23
	incongruent Total		824.77	±124.26	0.92	±0.22
Long SOA Total			826.13	±132.39	0.93	±0.21
Total			822.05	±144.09	0.93	±0.20

table 3-3 : meta-ACC and meta-RT by condition, metacognitive task only

META TASK	CONGRUENCY	Mean_MRT	SD_MRT(ms)	Mean M_ACC	SD_M_ACC
meta_task	**Congruent**	826.70	±120.62	0.98	±0.05
	Incongruent	813.71	±118.03	0.97	±0.05
meta_task Total		820.21	±118.71	0.97	±0.05
post_consonant	**Congruent**	1040.68	±238.07	0.98	±0.05
	Incongruent	1139.32	±256.21	0.98	±0.05
post_consonant Total		1090.00	±250.61	0.98	±0.05
post_lowercase	**Congruent**	1021.86	±223.42	0.95	±0.07
	Incongruent	1097.18	±189.22	0.97	±0.05
post_lowercase Total		1059.52	±209.11	0.96	±0.06
Total		989.91	233.40	0.97	±0.05

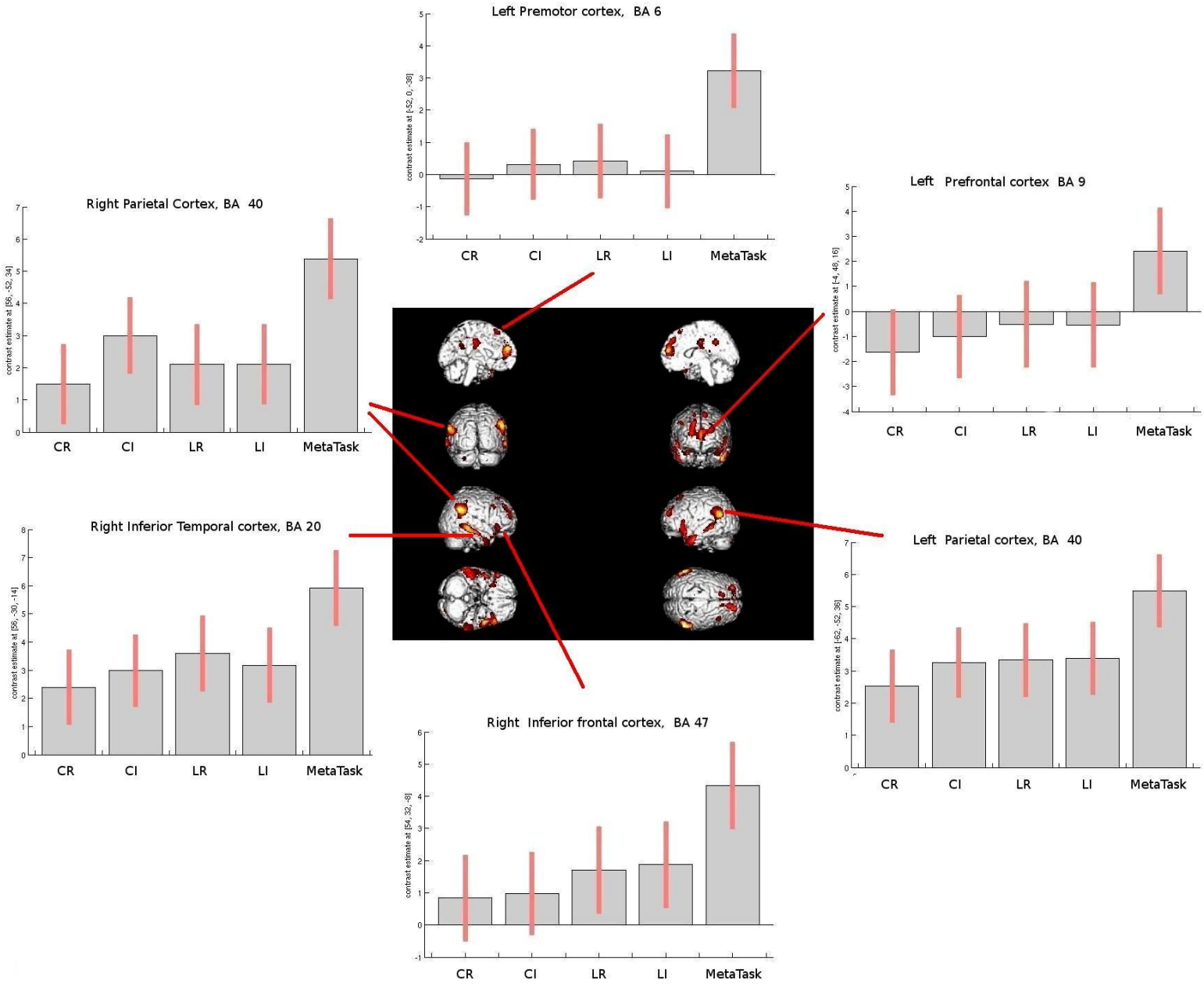
table 3-4 : meta-ACC and meta-RT by metacognitive and memory tasks, and congruency

3.4.2.2 Neuroimaging data

We were mainly interested in identifying a network recruited during metacognition (retrospective self-evaluation). We contrasted the meta-task with the memory tasks (metacognitive task minus memory task). That contrast allowed us to remove the perceptual and motor components of the activations, but also, and importantly, of some diverse memory components.

Such a contrast (figure 3-8, table 3-6) clearly revealed a fronto-parieto-temporal network, including an important prefrontal cluster (including BA 9 and BA 6) and another one including parietal cortices bilaterally (BA 40).

An appealing aspect of this pattern of activation, when looking at the plots of each cluster, is *that BA 9 and BA 6 were recruited during the meta-task only* (note even a decrease of activation is observed for BA 9 during memory post questions), while the parietal and temporal regions are also activated, though less importantly, by the other (memory) conditions.



[**Figure 3-8**: META-TASK minus all others p-unc 0,001, , see **table 6 below** for the corresponding labels, coordinates cluster size, z-score, p-values]

Label	x,y,z	cluster size	Z-score	P-value (FWE-corr) on the cluster size
Left medial prefrontal cortex (BA 9)	(-4, 48, 16)	1790 voxels	4.86	0.000
Left Supramarginal Gyrus, BA 40	(-62, -52, 36)	1302 voxels	5.15	0.000
Right Supramarginal Gyrus, BA 40	(56, -52, 34)	1200 voxels	6.18	0.000
Right Inferior Temporal Gyrus, BA 20,	(56, -30, -14)	1194 voxels	5.30	0.000
Left Inferior Frontal Gyrus, BA 47,	(-48, 22, -14)	578 voxels	4.86	0.001
Left Premotor cortex, SMA, BA 6,	(-52, 0, 38)	427 voxels	6.34	0.006
Right Inferior Frontal Gyrus, BA 47,	(54, 32, -8)	304 voxels	5.46	0.02

[table 3-6 : metaTask-minus memory tasks
labels, coordinates, cluster size, z-score, p-values for the metatask**]**

Summary of the behavioral and neuroimaging outcomes for the metacognitive task:

(i)*No effect* of compatibility or congruency was observed on meta-accuracy, neither on the corresponding reaction times. We could not look for effects at the neuroimaging level because of the very few trials by condition (8 conditions, about 80 trials by subjects – a minimum of 24 trials by subject and condition would have been necessary).

(ii) No significant difference between the accuracy obtained in the memory tasks, versus meta-task was observed. However, the reaction times showed a significant difference (slower reaction times in memory tasks).

That difference in reaction times was associated with more activation in a *fronto-parieto-temporal network during metacognitive task*. We also observed that the *premotor (BA6)* and *medial prefrontal (BA9)* cortices were recruited only during the metacognitive task.

3.5 Discussion

3.5.1 Basic task: compatibility effects in the short SOA condition, medial prefrontal activation

A central issue of the study regarding the basic task was thus the question of whether one could observe some effects of compatibility on brain activity in the short SOA condition –let it be associated with a compatibility effect on accuracy or on meta-accuracy. In our previous behavioral experiment, we had observed no effect of invisible primes (no compatibility effect in the short SOA condition), but only an effect of visible primes (compatibility in the long SOA condition) on the accuracy of the subjects. However, we had observed that invisible primes considerably influenced the subjective rating of their own accuracy (*i.e* meta-accuracy).

We had therefore hypothesized, that (i) at least in these conditions of very high learning level, invisible primes do not influence overt behavior, (ii) invisible primes can nevertheless influence action selection processes (therefore brain activity) and leave a trace that could exert a bias on the retrospective self-evaluation of the subjects.

We thus were expecting the same behavioral pattern as the previous one (a compatibility effect in the LONG SOA condition only), and finally observed a compatibility effect on accuracy in the short SOA condition only. It is noteworthy that, compatibility of the primes did not influence the reaction times, and consequently did not trigger any additional serial mental operations –contrary to the congruency factor that gives rise to greater reaction times in incongruent trials. Therefore, invisible primes presumably influenced bottom up parallel processes.

The behavioral effects of compatibility were associated with a trend of congruency-dependent activation in anterior cingulate (BA24), and also a significant three-way interaction (congruency*compatibility*soa) associated with differential activation of the ventral Anterior cingulate BA 24 and medial prefrontal cortex BA 9.

These latter points, together with the prediction of our behavioral study, lead us to consider the pattern of activation in each SOA condition separately. The simple trend toward a *congruency*compatibility* interaction turned out to be significant when one considered the short SOA condition only, and as expected, involved a *medial* prefrontal network, including anterior cingulate

cortex (BA24 and marginally BA32) and medial frontopolar cortex (BA10).

Finally and importantly, as far as the three-way interaction *(congruency*compatibility*soa,* associated with ventral Anterior cingulate BA 24 and medial prefrontal BA 9) is concerned, one would need a more refined and quantitative model to account for the shape of this interaction. However, the fact that, some properties of an external signal (prime visibility and compatibility) do interact with an intrinsic property of the executive control system (higher or lower levels of cognitive control that are recruited for the response selection) is sufficient to claim that generally non visible primes do influence task and response selection.

The activation of a cluster centered on the anterior cingulate cortex BA 24 was expected and suggests the existence of a response conflict, modulated by the cognitive load but also by the SOA. Interestingly, and importantly, the activity of medial prefrontal BA9 is consistent with the notion that the tonic activity of prefrontal neurons that lasts until response selection.

Activation of a medial prefrontal network was expected, especially of the ventral anterior cingulate, since it has been typically involved in situations of (unpredicted) response conflict, risk of error during response selection. Importantly, considering the fact that we had defined the onset of the boxcar function as the target onset, it is most likely that these activations are attributable to mechanisms related to response selection – even if not exclusively.

Furthermore, the activation of that medial prefrontal network is by the way compatible with the most recent theories of hierarchical cognitive and motivational control with the prefrontal cortex (Kouheiner et al, 2009; Taren et al, 2011) that stipulate a parallel between medial and lateral regions, the medial ones energizing the lateral ones.

Consistent with the cascade model proposed by Koechlin in 2003, it has recently been proposed (Kounehier, Charron, Koechlin, 2009) that different medial networks can be individualized on the basis of the hierarchical level whereby motivational signals are integrated to drive the behavior : *"In the posterior sector, medial regions (pre-SMA) evaluate immediate contextual incentives for action and energize (or inhibit) lateral prefrontal resources that guide action selection according to immediate*

contextual signals. In the more anterior sector, medial regions (dACC) retain incentive values of past events and energize/inhibit lateral prefrontal resources that guide action selection according to past events". According to these authors, the pre-SMA would be involved in contextual motivational control and would energize/inhibit BA 44/45, while the dorsal ACC (BA32) would be involved in the so-called episodic motivational control, at higher temporal scale, and would energize BA 46/9.

Within this framework, the activation of the dorsal ACC makes sense, considering the nature of our paradigm, whereby subjects are cued to select a task-set. However, an open question is why the ventral ACC (BA 24) and the medial frontopolar (BA10) cortices are involved : the model does not relate to these regions, nor to the network involved in the "sensory" control, what would correspond to the level inferior to task setting.

To return to our results, it is likely that the activation of ventral and medial orbitofrontal cortices reflects a conflict of response –as we were expecting if the conflict could propagate from task- to response selection. Moreover, the fact that the compatibility factor (motivational factor) interacts with the congruence factor (cognitive factor) while they are a priori independent, tends to confirm that masked or invisible stimuli can influence the cognitive control mechanisms at a task setting stage, and propagate downstream, onto response selection.

Some question marks...

A non trivial question is why we did not obtain the same behavioral pattern as before (in our previous study). A possible reason why we obtained effects of invisible primes on accuracy is the fact that the necessary introduction of jitters increases the difficulty of the basic task (task-cueing paradigm) considerably. The temporal irregularity of the trials has effects on the preparatory states of the subjects, and reduces the 'routinization'. In such conditions, of relative uncertainty and impossibility of setting up a regular routine, one can conceive that the executive system is more sensitive to noise or to external perturbation.

However, the corresponding medial prefrontal activations do not exclude a link with an impaired meta-accuracy (instead of an exclusive link with accuracy in the basic task). In effect, since we have fewer

trials in the current neuroimaging experiment (half as many subjects and half as many trials by subject) compared with the previous behavioral study, we also have fewer data points for the meta-accuracy.
In reality, *the whole dataset of meta-accuracy in the meta-task* (if one combines across SOA) *showed a trend of compatibility effects (pairwise Wilcoxon, p<.07).* So it is difficult to affirm that the medial prefrontal activations in the short SOA condition would not or could not have been associated with a possible impaired meta-accuracy if we had more data.

3.5.2 Metacognition: which networks are involved

The second issue of the current study was to identify the network implicated in the retrospective self-evaluation processes (metacognition versus simple memory). On the basis of overt performance, we have been able to dissociate the metacognitive task from the simple memory tasks: the reaction times in metacognitive task were faster than those in memory tasks. By subtracting the (collapsed) memory tasks from the metacognitive task, we observed a large fronto-temporo-parietal network. Within this network, only two clusters were recruited by the metacognitive task only, namely the premotor cortex BA 6 and the prefrontal cortex BA9.

Within Koechlin's model, the premotor cortex BA 6 (including the pre-SMA),is indeed supposed to be recruited during the performance itself, that is to say during target based action selection, and BA 9 (dorsomedial part of prefrontal cortex), typically associated with rule-based action selection and task setting was activated independently of the motor, perceptual and memory of the target components. Note that medial BA9 was also present in the 3-way interaction (cf figure 3-6, table 3-2).

A relevant question that follows is whether and to which extent the metacognition-related activation depended on the previous task (that is to say on the mechanisms involved in the task set in order to produce a correct response) or whether they are instantiating a more general purpose self-evaluation network.

Previous studies (Fleming et al., 2010 ; Fleming and Dolan, 2012 for a review) reported slightly more anterior (BA 10) activations associated with *retrospective* metacognitive performance (confidence). But their report referred to metacognitive performance in a *perceptual task*, and do not report any

premotor/SMA activation.

Independently of these prefrontal networks of interest, we also observed different regions that were recruited by the other memory tasks (e.g parietal cortex, BA40 bilaterally, left ventrolaeral prefrontal BA 47/11).

I will not consider the question of the functional significance of that overlap for two reasons. First it is outside the scope of the hypothesis we had set beforehand, and consequently out of the scope of the present study. Secondly, I would be only able to speculate without any solid empirical nor theoretical basis.

*

To conclude about the outcome of the present neuroimaging study, despite some differences between our expectations and actual outcome, as far as we know, this is *the first demonstration of medial prefrontal activation* :

(v) during response conflict created by invisible primes

(vi) in such a task-cueing paradigm whereby the conflict is propagated in a top down way from an upper level.

Our demonstration rests on the fact that two independent factors supposed to affect response selection interacted at the level of brain activity: namely the nature of the prime and the cognitive control load. Moreover they modulated the brain activity in cortices typically involved in situations of response conflict.

The interpretation of the present results is nevertheless contingent on the fact that, in the scanner, whereby subjects were displayed the stimuli with goggles, the short and long SOAs effectively gave rise to low and high visibility condition, respectively. Although we considered as more likely that we obtained a type-A masking (short SOA associated with a low visibility, long SOA with a high visibility), we acknowledge that further visibility tests inside the scanner should be done to remove the doubt regarding our interpretation.

3.6 Summary of overall theoretical conclusions and further research

In our previous behavioral study (cf. *Part II : replicating and exploring*), we had observed that, in our task-cueing paradigm, unseen primes did not influence the performance of the subjects, but affected the *awareness* of their own performance, that we indirectly measured through their metacognitive ability to evaluate their trial-by-trial performance. We had thus hypothesized that unseen primes could nevertheless have influenced response selection mechanisms, and left a trace that would have been used afterwards, during metacognitive judgment.

In our subsequent neuroimaging study (cf. *Part III*), we tested this hypothesis, by tracking brain activity during response selection, and during metacognitive judgment. The outcome of this neuroimaging study seemed consistent with our initial hypothesis: we observed a prefrontal region active during both response selection and metacognitive task, namely medial prefrontal cortex BA9 (associated with a three way *SOA*compatibility*Congruency* interaction during response selection). We also observed that a medial prefrontal network, including the anterior cingulate cortex (BA24), showed a pattern of activation dependent of both cognitive control load and prime compatibility, only when the primes were generally not visible (*compatibility*Congruency* interaction in short SOA trials).

Both experiments are consistent with the current literature about metacognition, which suggests a critical role of prefrontal cortex in metacognition and access to consciousness (Fleming et Dolan, 2012). They are also consistent with the two-stage model proposed by Pleskac & Busemeyer (2010), in which the same network is involved in both first-order and second-order decisions : after the first-order decision (response selection) is made, the accumulation of evidence continues, then the network is re-accessed for second-order decision (confidence judgment). On this model (see figure 3-9 below), confidence (or second-order judgment) *does not* depend only on parameters of response selection (or first-order decision). However, it does not assume anything regarding the mechanisms by which the network is re-accessed, it only assumes the existence of a 'judge'.

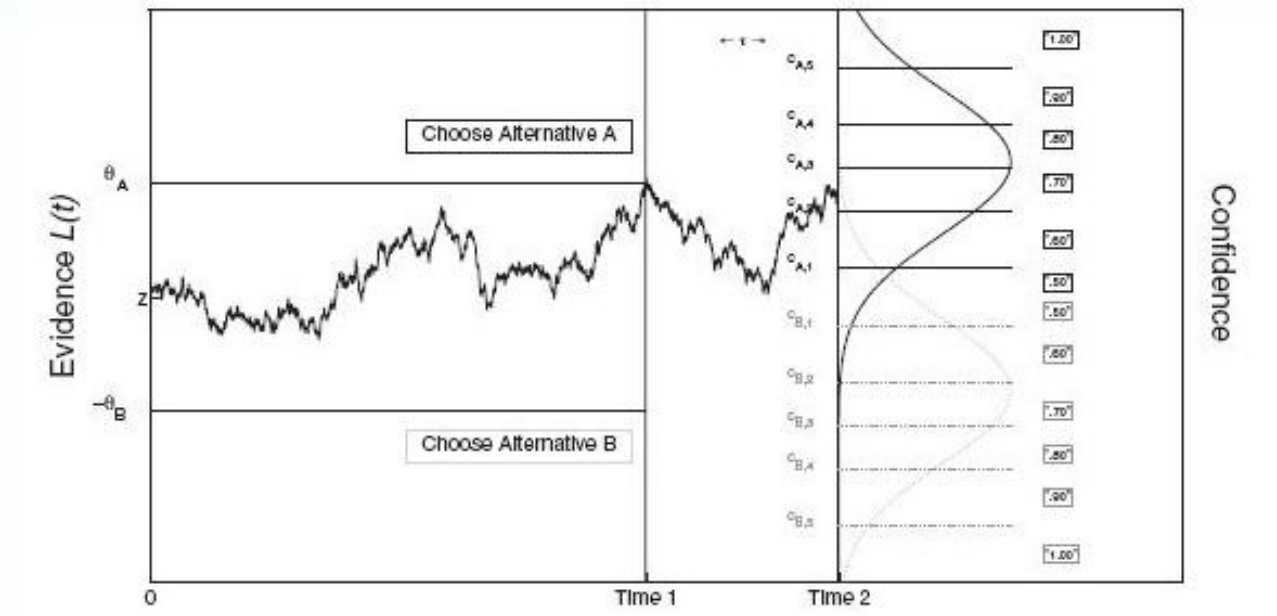


Figure 1. A realization of evidence accumulation in the two-stage dynamic signal detection (2DSD) interrogation model of confidence. The black jagged line depicts the accumulation process when a judge correctly predicts Response Alternative A. To produce a confidence estimate the model assumes after a fixed time interval passes (or interjudgment time τ) more evidence is collected and an estimate (e.g., .50, .60, . . . , 1.00) is chosen based on the location of the evidence in the state space. The solid black normal curve on the right-hand side of the figure is the distribution of evidence at the time of confidence t_C when a judge correctly chooses Alternative A. The dashed normal curve is what the distribution would be if a judge would have incorrectly chosen Alternative B. θ_A and $-\theta_B$ = choice thresholds for Alternatives A and B, respectively; t_D = predicted decision time; z = starting point; $c_{choice\,k}$ = confidence criteria.

Figure 3-9: from Pleskac and Busemeyer, 2010

Yet, hardwiring this model (in terms of neural mechanisms) may shed some light on the links existing between access to consciousness, cognitive control and metacognition. A possibility would be that an internal loop, which enables recurrent connections and allows the network to take its own output state as input –in that case the overlap between the 'judge' and the 'actor' would be total. Another possibility would be that the first-order networks are accessed by an upstream network, situated at a higher level of processing. In that second scenario, judge and actor would be implemented in different networks, and the overlap would be *partial*.

We did observe an activation of the regions implementing the task set (namely medial prefrontal BA9 and premotor BA6 cortices) during metacognitive judgment, whereas the activity in these regions remained constant or even decreased during memory judgment. This point is thus consistent with the assumption of Pleskac & Busemeyer (2010). It is also consistent with previous studies of metacognition which suggest a critical role for prefrontal cortex (Fleming et al, 2012, for a review), although different

prefrontal networks have been reported, in particular dorsolateral prefrontal cortex BA46/9 (e;g Rounis et al, 2010) and orbitofrontal cortex BA10(Del Cul et al, 2010; Fleming et al, 2010; Yokohama et al, 2010). Nevertheless the specific contribution of these prefrontal regions to metacognition is unknown, as are the brain mechanisms underpinning confidence judgment. In our study, we did not observe any activation of frontopolar BA 10 during metacognitive judgment, despite it has often been reported. It is unclear whether this anterior prefrontal region plays a general purpose functional role in metacognition, independently of the first order task.

At that point, an hypothesis can be drawn, according to which metacognitive processes would actually be *relative* to first-order processes. In that scenario, the networks involved in the first-order decision would be managed and re-accessed by a second-order network, *situated at a superior level within the cognitive control hierarchy*. This would be consistent with different empirical and theoretical reports, especially (1) that cognition and metacognition seem to be domain-specific, (2) that cognition and metacognition can be selectively impaired, (3) that in the mathematical psychology literature about decision, the parameters of the first-order decision (starting point of accumulated evidence, drift rate, choice threshold) are necessary but not sufficient to account for second-order decisions.

This hypothesis could also explain why these two prefrontal networks, namely BA46/9 and BA10, have been often reported by studies of metacognition. Their situation, respectively in the midst or at the top of the cognitive control hierarchy, is such that they are involved in most of experimental tasks.

Our previous neuroimaging results could also be consistent with this hypothesis. The contrast we used to remove the sensory, motor and memory components of the metacognitive judgment may explain why BA10 did not appear, since it has also been reported during memory tasks (Rugg et al, 1996, Pochon et al, 2002).

The next chapter reports a study which tries to carry out an additional step to exploring cognitive control and metacognitive functions, but in schizophrenia, a psychiatric condition known to display robust abnormalities within the anterior cingulate (BA24) and prefrontal (BA9) cortices –see below section 4.1.

Two aspects are relevant in that study. The first one is related to the issue of the brain networks (and mechanisms) involved in metacognition. It considers the issue of whether the 'metacognitive judge' could be implemented in a network situated at a higher level of the module(s) involved in the current response selection. The other is specific, related to schizophrenia by itself, and concerns the issue of whether there exists a basic metacognitive deficit, which could account for the specificity of some psychotic symptoms in that pathology. That second aspect can be considered independently of the previous one, and presents an interest by itself. Nevertheless, considering the hypothesis of *relative* metacognitive processes leads us to formulate more specific hypotheses regarding the nature of second-order decisions (metacognitive judgments) that population should produce.

Some common empirical facts that models of decision/confidence have to account for are the following. First, faster reaction times generally correlate with higher confidence level. Second, confidence is generally higher for correct than for incorrect first-order responses. Pleskac & Busemeyer account for both points by pointing the fact that they are natural consequences of the inner mechanism by which the second-order decision is made, and which is function of the quality and the quantity of accumulated evidence during first-order decision (for more details see Pleskac & Busemeyer, 2010).

Thus, patients with schizophrenia could display two different kinds of impaired metacognitive profiles.

Either (1) : if BA9 (that we know is involved in first-order decision in our paradigm) but not any other upstream prefrontal region is involved in metacognitive judgment, then schizophrenic patients should produce more errors in first-order decisions, and also more errors in second order decisions. But it is not sufficient because the nature of the second-order errors matters. They should not be significantly less confident than control or other type of patients. They should detect less often their errors (while being confident), and produce equally frequent false alarms. Moreover, they should not display the common negative correlation between level of confidence and reaction time.

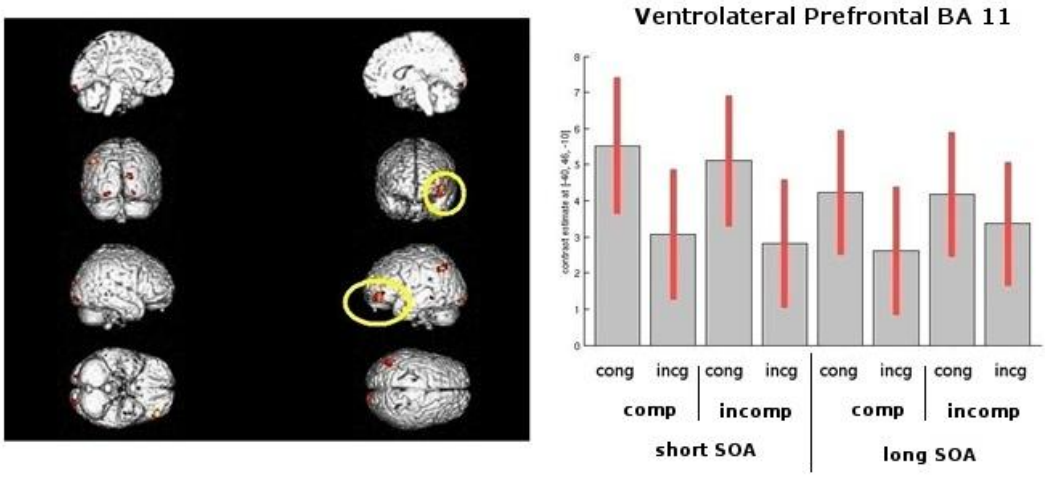
Or (2) : if BA9 is only *partially* involved in the metacognitive judgment, assuming that an upstream network has an actual access to it, then despite more errors in first-order decisions, one should observe only less confident second-order responses, not necessarily confident incorrect second-order decisions. In that perspective, the general level of confidence should be lowered. And thus, it should be possible to

observe equally frequent error detections (incorrect first-order decisions being anyway less confident even for control subjects) but more false alarms (since the confidence would be lower even when the first-order decision is correct), since it would give rise only to a decreased confidence level. Importantly, the fact of not observing that pattern will not exclude the corresponding hypothesis, but observing it will exclude the first alternative (of BA9 playing both the role of both actor and judge).

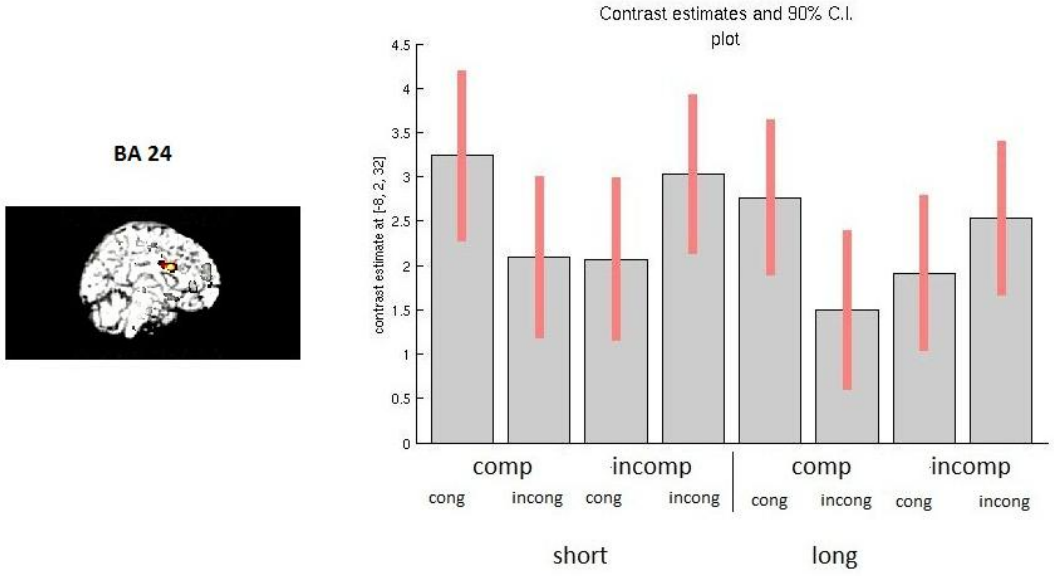
Moreover, schizophrenia patients should show the same negative correlation between reaction times and confidence, as do healthy subjects.

Appendix :

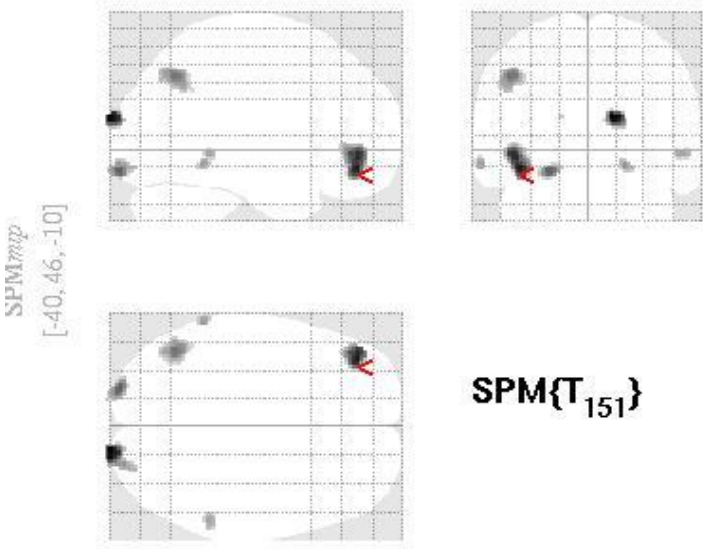
1) Plot for the nearly significant main effect of congruency, associated with an activation within the (left) ventrolateral prefrontal cortex [BA 11, (-40;46;-10), 218 voxels, Z= 4.00, p-FWE-corr =0.067].



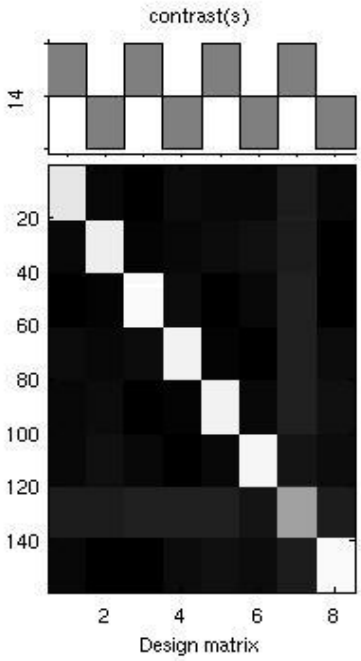
2) Plot for the nearly significant congruency*compatibility interaction, associated with an activation of the Left ventral Anterior cingulate [BA 24, (-8;2;32), 212 voxels, Z= 4.73, p-FWE-corr =0.07].



cong-incong



SPM{T_{151}}



SPMresults: /third analysis 01 05 2012
Height threshold T = 3.145081 (p<0.001 (unc.))
Extent threshold k = 0 voxels

Statistics: *p-values adjusted for search volume*

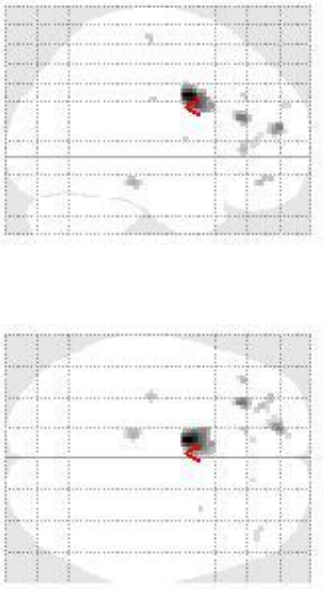
set-level		cluster-level				peak-level					mm mm mm		
p	c	$p_{FWE-corr}$	$q_{FDR-corr}$	k_E	p_{uncorr}	$p_{FWE-corr}$	$q_{FDR-corr}$	T	$(Z_{=})$	p_{uncorr}			
0.541	**8**	**0.469**	**0.212**	**85**	**0.079**	**0.278**	**0.316**	**4.27**	**4.14**	**0.000**	**14**	**-100**	**20**
		0.067	**0.070**	**218**	**0.009**	**0.418**	**0.316**	**4.11**	**4.00**	**0.000**	**-40**	**46**	**-10**
						0.698	0.414	3.86	3.76	0.000	-44	42	-2
		0.698	**0.301**	**55**	**0.150**	**0.792**	**0.434**	**3.77**	**3.68**	**0.000**	**-24**	**-94**	**-12**
		0.157	**0.086**	**159**	**0.021**	**0.877**	**0.483**	**3.66**	**3.58**	**0.000**	**-46**	**-62**	**44**
		0.844	**0.374**	**37**	**0.233**	**0.962**	**0.542**	**3.50**	**3.43**	**0.000**	**22**	**-92**	**-10**
		0.952	**0.465**	**20**	**0.381**	**0.973**	**0.542**	**3.46**	**3.39**	**0.000**	**-62**	**-46**	**-8**
		0.961	**0.465**	**18**	**0.407**	**0.980**	**0.542**	**3.43**	**3.36**	**0.000**	**56**	**-42**	**-2**
		0.999	**0.877**	**1**	**0.877**	**0.997**	**0.737**	**3.27**	**3.21**	**0.001**	**-16**	**-102**	**20**

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 3.15, p = 0.001 (1.000)
Extent threshold: k = 0 voxels, p = 1.000 (1.000)
Expected voxels per cluster, <k> = 28.033
Expected number of clusters, <c> = 7.96
FWEp: 4.792, FDRp: Inf, FWEc: Inf, FDRc: Inf

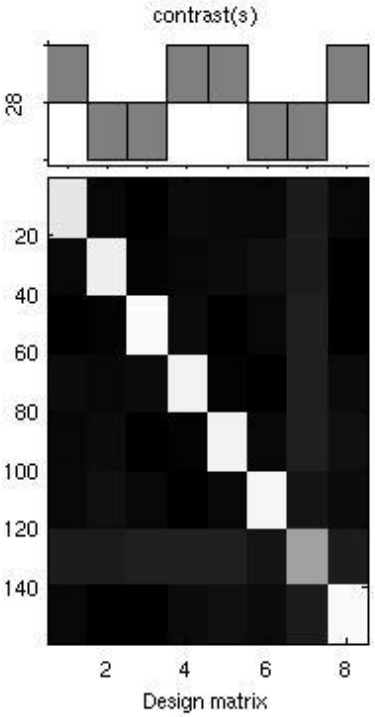
Degrees of freedom = [1.0, 151.0]
FWHM = 13.0 13.0 11.8 mm mm mm; 6.5 6.5 5.9 (voxels)
Volume: 1509344 = 188668 voxels = 704.1 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 248.14 voxels)

comp*cong



SPM*mip*
[-8, 2, 32]

SPM{T_{151}}



SPMresults: /third analysis 01 05 2012
Height threshold T = 3.145081 (p<0.001 (unc.))
Extent threshold k = 0 voxels

Statistics: *p-values adjusted for search volume*

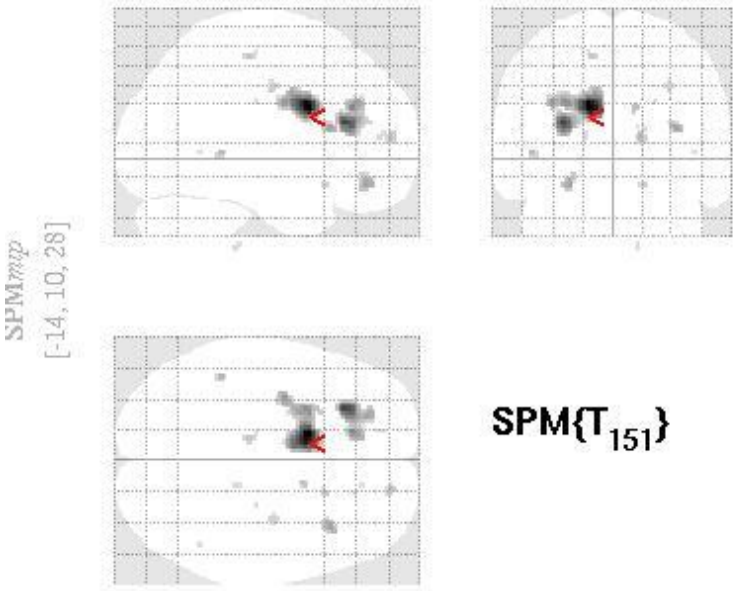
set-level		cluster-level				peak-level					mm mm mm		
p	c	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	k_E	p_{uncorr}	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	T	(Z_E)	p_{uncorr}			
0.004	**17**	**0.073**	**0.161**	**212**	**0.009**	**0.075**	**0.216**	**4.68**	**4.51**	**0.000**	**-8**	**2**	**32**
						0.375	0.650	4.16	4.04	0.000	-4	6	26
						0.912	0.942	3.61	3.53	0.000	-6	14	26
		0.859	**0.877**	**35**	**0.246**	**0.724**	**0.942**	**3.83**	**3.74**	**0.000**	**-28**	**32**	**20**
		0.888	**0.877**	**31**	**0.274**	**0.805**	**0.942**	**3.75**	**3.66**	**0.000**	**-14**	**52**	**14**
		0.961	**0.877**	**18**	**0.407**	**0.976**	**0.942**	**3.45**	**3.38**	**0.000**	**-10**	**-30**	**-16**
		0.987	**0.877**	**10**	**0.544**	**0.990**	**0.942**	**3.37**	**3.30**	**0.000**	**-40**	**32**	**2**
		0.989	**0.877**	**9**	**0.567**	**0.991**	**0.942**	**3.36**	**3.30**	**0.000**	**-26**	**42**	**-16**
		0.997	**0.877**	**4**	**0.719**	**0.995**	**0.942**	**3.31**	**3.25**	**0.001**	**-32**	**-20**	**62**
		0.999	**0.877**	**1**	**0.877**	**0.996**	**0.942**	**3.30**	**3.24**	**0.001**	**-12**	**34**	**30**
		0.998	**0.877**	**2**	**0.812**	**0.998**	**0.942**	**3.26**	**3.20**	**0.001**	**-32**	**-18**	**30**
		0.999	**0.877**	**1**	**0.877**	**0.998**	**0.942**	**3.26**	**3.20**	**0.001**	**-14**	**32**	**32**
		0.999	**0.877**	**1**	**0.877**	**0.998**	**0.942**	**3.24**	**3.18**	**0.001**	**30**	**8**	**34**
		0.998	**0.877**	**2**	**0.812**	**0.999**	**0.942**	**3.21**	**3.15**	**0.001**	**-12**	**56**	**26**
		0.999	**0.877**	**1**	**0.877**	**0.999**	**0.942**	**3.20**	**3.14**	**0.001**	**-14**	**0**	**8**
		0.998	**0.877**	**3**	**0.761**	**0.999**	**0.942**	**3.19**	**3.14**	**0.001**	**42**	**40**	**8**
		0.999	**0.877**	**1**	**0.877**	**0.999**	**0.942**	**3.19**	**3.13**	**0.001**	**-8**	**38**	**6**
		0.999	**0.877**	**1**	**0.877**	**0.999**	**0.942**	**3.19**	**3.13**	**0.001**	**40**	**42**	**10**
		0.999	**0.877**	**1**	**0.877**	**0.999**	**0.942**	**3.18**	**3.12**	**0.001**	**14**	**38**	**4**

table shows 3 local maxima more than 8.0mm apart

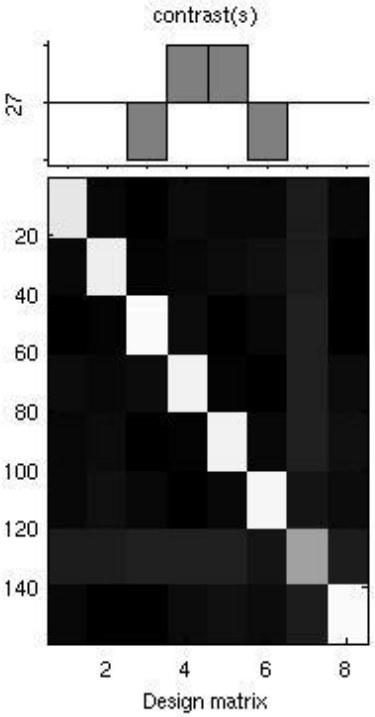
Height threshold: T = 3.15, p = 0.001 (1.000)
Extent threshold: k = 0 voxels, p = 1.000 (1.000)
Expected voxels per cluster, <k> = 28.033
Expected number of clusters, <c> = 7.96
FWEp: 4.792, FDRp: Inf, FWEc: Inf, FDRc: Inf

Degrees of freedom = [1.0, 151.0]
FWHM = 13.0 13.0 11.8 mm mm mm; 6.5 6.5 5.9 (voxels)
Volume: 1509344 = 188668 voxels = 704.1 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 248.14 voxels)

cong*comp*soa



SPM{T_{151}}



SPMresults: /third analysis 01 05 2012
Height threshold T = 3.145081 (p<0.001 (unc.))
Extent threshold k = 0 voxels

Statistics: *p-values adjusted for search volume*

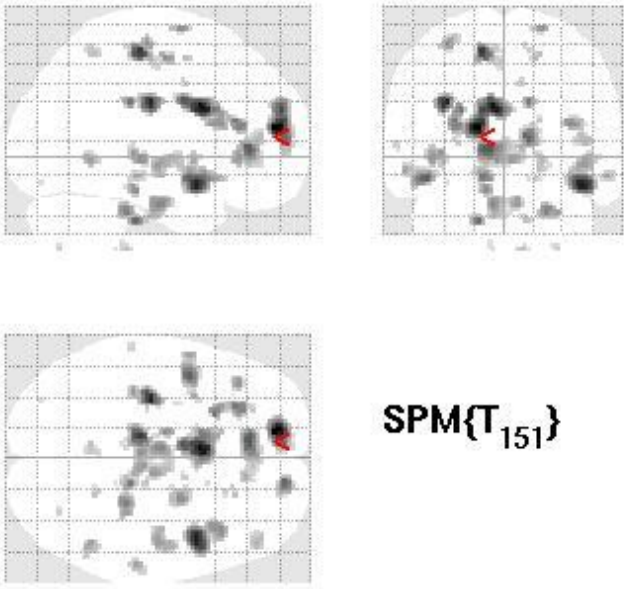
set-level		cluster-level				peak-level					mm mm mm		
p	c	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	k_E	p_{uncorr}	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	T	(Z_E)	p_{uncorr}			
0.002	**18**	**0.003**	**0.007**	**462**	**0.000**	**0.001**	**0.002**	**5.85**	**5.55**	**0.000**	**-14**	**10**	**28**
						0.468	0.396	4.07	3.95	0.000	-28	8	32
						0.581	0.422	3.96	3.86	0.000	-34	-4	32
		0.049	**0.056**	**241**	**0.006**	**0.031**	**0.047**	**4.92**	**4.73**	**0.000**	**-28**	**30**	**20**
						0.390	0.396	4.14	4.02	0.000	-16	34	28
		0.925	**0.877**	**25**	**0.326**	**0.713**	**0.422**	**3.84**	**3.75**	**0.000**	**36**	**22**	**18**
		0.937	**0.877**	**23**	**0.347**	**0.716**	**0.422**	**3.84**	**3.75**	**0.000**	**-24**	**42**	**-14**
		0.993	**0.877**	**7**	**0.619**	**0.896**	**0.620**	**3.64**	**3.56**	**0.000**	**-46**	**-40**	**4**
		0.984	**0.877**	**11**	**0.523**	**0.928**	**0.663**	**3.58**	**3.50**	**0.000**	**16**	**56**	**12**
		0.996	**0.877**	**5**	**0.682**	**0.958**	**0.735**	**3.51**	**3.44**	**0.000**	**12**	**6**	**28**
		0.997	**0.877**	**4**	**0.719**	**0.972**	**0.768**	**3.47**	**3.40**	**0.000**	**-12**	**-24**	**56**
		0.999	**0.877**	**1**	**0.877**	**0.991**	**0.937**	**3.37**	**3.30**	**0.000**	**-26**	**16**	**28**
		0.998	**0.877**	**3**	**0.761**	**0.993**	**0.946**	**3.34**	**3.27**	**0.001**	**18**	**20**	**-14**
		0.993	**0.877**	**7**	**0.619**	**0.996**	**0.981**	**3.30**	**3.24**	**0.001**	**32**	**-12**	**38**
		0.998	**0.877**	**2**	**0.812**	**0.999**	**0.998**	**3.23**	**3.17**	**0.001**	**14**	**-32**	**-48**
		0.998	**0.877**	**2**	**0.812**	**0.999**	**0.998**	**3.22**	**3.16**	**0.001**	**-16**	**50**	**12**
		0.999	**0.877**	**1**	**0.877**	**0.999**	**0.998**	**3.20**	**3.14**	**0.001**	**24**	**-20**	**32**
		0.999	**0.877**	**1**	**0.877**	**0.999**	**0.998**	**3.18**	**3.12**	**0.001**	**18**	**36**	**16**
		0.999	**0.877**	**1**	**0.877**	**1.000**	**0.998**	**3.16**	**3.10**	**0.001**	**46**	**-52**	**4**
		0.999	**0.877**	**1**	**0.877**	**1.000**	**0.998**	**3.15**	**3.09**	**0.001**	**14**	**40**	**18**
		0.999	**0.877**	**1**	**0.877**	**1.000**	**0.998**	**3.15**	**3.09**	**0.001**	**-14**	**-20**	**58**

table shows 3 local maxima more than 8.0mm apart

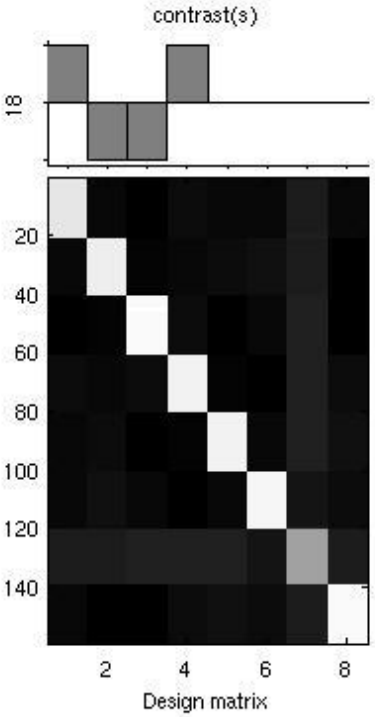
Height threshold: T = 3.15, p = 0.001 (1.000)
Extent threshold: k = 0 voxels, p = 1.000 (1.000)
Expected voxels per cluster, <k> = 28.033
Expected number of clusters, <c> = 7.96
FWEp: 4.792, FDRp: 4.925, FWEc: 241, FDRc: 462

Degrees of freedom = [1.0, 151.0]
FWHM = 13.0 13.0 11.8 mm mm mm; 6.5 6.5 5.9 (voxels)
Volume: 1509344 = 188668 voxels = 704.1 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 248.14 voxels)

comp*cong SHORT



SPM{T_{151}}



SPMresults: /third analysis 01 05 2012
Height threshold T = 3.145081 (p<0.001 (unc.))
Extent threshold k = 0 voxels

Statistics: *p-values adjusted for search volume*

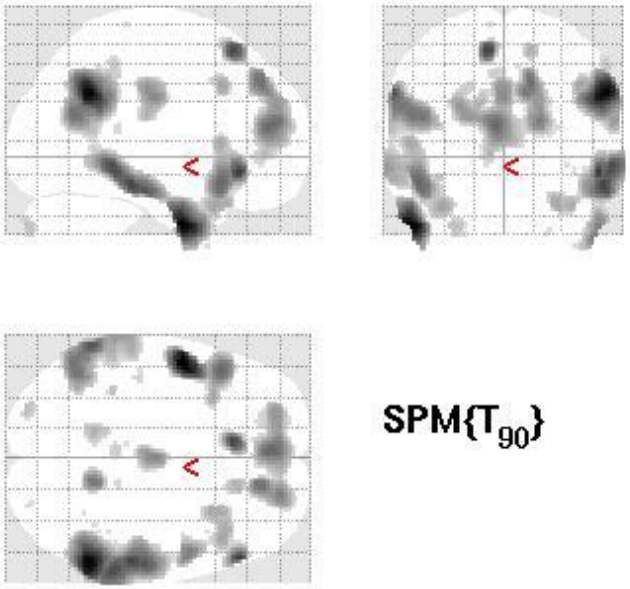
set-level		cluster-level				peak-level					mm mm mm		
p	c	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	k_E	p_{uncorr}	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	T	(Z_E)	p_{uncorr}			
0.000	**51**	**0.028**	**0.126**	**283**	**0.004**	**0.035**	**0.264**	**4.89**	**4.70**	**0.000**	**-14**	**52**	**16**
						0.274	0.484	4.28	4.15	0.000	-12	56	26
		0.046	**0.126**	**245**	**0.006**	**0.065**	**0.264**	**4.72**	**4.55**	**0.000**	**-6**	**10**	**28**
						0.351	0.484	4.18	4.06	0.000	-8	0	32
		0.087	**0.147**	**199**	**0.011**	**0.101**	**0.264**	**4.59**	**4.44**	**0.000**	**44**	**8**	**-16**
		0.406	**0.400**	**95**	**0.065**	**0.115**	**0.264**	**4.55**	**4.40**	**0.000**	**-32**	**-18**	**30**
		0.412	**0.400**	**94**	**0.067**	**0.200**	**0.402**	**4.38**	**4.24**	**0.000**	**-12**	**-24**	**58**
						0.959	0.752	3.51	3.44	0.000	-6	-18	54
		0.057	**0.126**	**229**	**0.007**	**0.370**	**0.484**	**4.16**	**4.04**	**0.000**	**-8**	**38**	**4**
						0.957	0.752	3.52	3.44	0.000	10	36	6
						0.998	0.909	3.25	3.19	0.001	2	44	-2
		0.211	**0.253**	**139**	**0.030**	**0.416**	**0.484**	**4.12**	**4.00**	**0.000**	**-44**	**4**	**-12**
						0.926	0.703	3.59	3.51	0.000	-56	4	-8
		0.554	**0.415**	**73**	**0.101**	**0.485**	**0.543**	**4.05**	**3.94**	**0.000**	**12**	**58**	**12**
		0.881	**0.729**	**32**	**0.267**	**0.505**	**0.543**	**4.03**	**3.92**	**0.000**	**-14**	**-26**	**-34**
		0.430	**0.400**	**91**	**0.071**	**0.607**	**0.613**	**3.94**	**3.84**	**0.000**	**-24**	**18**	**26**
						0.623	0.613	3.93	3.82	0.000	-28	30	20
		0.925	**0.729**	**25**	**0.326**	**0.627**	**0.613**	**3.92**	**3.82**	**0.000**	**14**	**-30**	**30**
		0.510	**0.415**	**79**	**0.090**	**0.640**	**0.613**	**3.91**	**3.81**	**0.000**	**-12**	**12**	**-10**
						0.997	0.909	3.27	3.21	0.001	-12	4	-6
		0.698	**0.511**	**55**	**0.150**	**0.699**	**0.661**	**3.86**	**3.76**	**0.000**	**24**	**-32**	**-30**
		0.166	**0.233**	**155**	**0.023**	**0.755**	**0.665**	**3.80**	**3.71**	**0.000**	**-6**	**-14**	**-26**
						0.890	0.665	3.65	3.56	0.000	8	-16	-26
		0.829	**0.707**	**39**	**0.222**	**0.799**	**0.665**	**3.76**	**3.67**	**0.000**	**-32**	**-20**	**62**

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 3.15, p = 0.001 (1.000)
Extent threshold: k = 0 voxels, p = 1.000 (1.000)
Expected voxels per cluster, <k> = 28.033
Expected number of clusters, <c> = 7.96
FWEp: 4.792, FDRp: Inf, FWEc: 245, FDRc: Inf

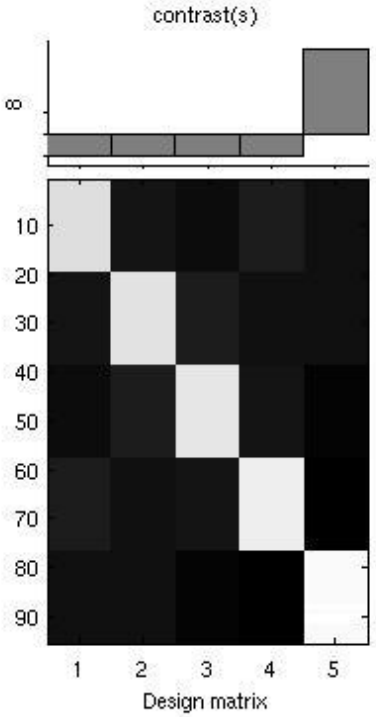
Degrees of freedom = [1.0, 151.0]
FWHM = 13.0 13.0 11.8 mm mm mm; 6.5 6.5 5.9 (voxels)
Volume: 1509344 = 188668 voxels = 704.1 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 248.14 voxels)
Page 1

meta-all



SPM{mip} [0, 0, 0]

SPM{T_{90}}



SPMresults: /block_related/all blocks
Height threshold T = 3.183271 (p<0.001 (unc.))
Extent threshold k = 0 voxels

Statistics: *p-values adjusted for search volume*

set-level		cluster-level				peak-level					mm mm mm		
p	c	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	k_E	p_{uncorr}	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	T	(Z_E)	p_{uncorr}			
0.000	**27**	**0.006**	**0.003**	**427**	**0.001**	**0.000**	**0.000**	**7.15**	**6.34**	**0.000**	**-52**	**0**	**-38**
		0.000	**0.000**	**1200**	**0.000**	**0.000**	**0.000**	**6.92**	**6.18**	**0.000**	**56**	**-52**	**34**
						0.275	0.093	4.36	4.14	0.000	64	-54	20
						0.955	0.544	3.56	3.44	0.000	60	-38	50
		0.024	**0.011**	**304**	**0.003**	**0.001**	**0.003**	**5.96**	**5.46**	**0.000**	**54**	**32**	**-8**
						0.520	0.165	4.09	3.90	0.000	30	20	-26
						0.529	0.165	4.08	3.90	0.000	38	24	-26
		0.319	**0.136**	**114**	**0.050**	**0.002**	**0.004**	**5.82**	**5.35**	**0.000**	**-10**	**28**	**58**
		0.000	**0.000**	**1194**	**0.000**	**0.002**	**0.004**	**5.76**	**5.30**	**0.000**	**56**	**-30**	**-14**
						0.003	0.005	5.68	5.23	0.000	60	-38	-8
						0.005	0.005	5.58	5.16	0.000	58	-22	-18
		0.000	**0.000**	**1302**	**0.000**	**0.005**	**0.005**	**5.57**	**5.15**	**0.000**	**-62**	**-52**	**36**
						0.039	0.021	4.99	4.68	0.000	-46	-60	22
						0.040	0.021	4.98	4.67	0.000	-58	-56	28
		0.001	**0.001**	**578**	**0.000**	**0.018**	**0.014**	**5.22**	**4.86**	**0.000**	**-48**	**22**	**-14**
						0.417	0.135	4.19	3.99	0.000	-34	18	-26
						0.489	0.163	4.12	3.93	0.000	-54	22	8
		0.000	**0.000**	**1790**	**0.000**	**0.018**	**0.014**	**5.21**	**4.86**	**0.000**	**-4**	**48**	**16**
						0.030	0.021	5.07	4.74	0.000	16	46	40
						0.059	0.027	4.87	4.57	0.000	20	56	20
		0.153	**0.065**	**165**	**0.022**	**0.042**	**0.021**	**4.97**	**4.66**	**0.000**	**14**	**-50**	**34**
		0.014	**0.007**	**346**	**0.002**	**0.122**	**0.050**	**4.64**	**4.38**	**0.000**	**0**	**-16**	**32**
		0.578	**0.254**	**71**	**0.113**	**0.404**	**0.135**	**4.20**	**4.01**	**0.000**	**44**	**22**	**40**
		0.711	**0.337**	**54**	**0.162**	**0.648**	**0.213**	**3.96**	**3.80**	**0.000**	**14**	**30**	**58**

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 3.18, p = 0.001 (1.000)
Extent threshold: k = 0 voxels, p = 1.000 (1.000)
Expected voxels per cluster, <k> = 29.292
Expected number of clusters, <c> = 7.65
FWEp: 4.918, FDRp: 4.641, FWEc: 304, FDRc: 304

Degrees of freedom = [1.0, 90.0]
FWHM = 13.3 13.2 12.2 mm mm mm; 6.7 6.6 6.1 (voxels)
Volume: 1510856 = 188857 voxels = 652.8 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 267.93 voxels)
Page 1

PART IV :

Cognitive Control, Access to Consciousness and Metacognition in Psychosis, an Observational Study

4.1 Introduction: schizophrenia from a cognitive neurosciences point of view

4.1.1 Symptomatology of schizophrenia

Schizophrenia is a mental disorder that can be considered as one of the most crippling psychiatric pathologies, for both the patients and their relatives. Its global lifetime prevalence is of about 0.3–0.7% (Os and Kapur, 2009). It is commonly characterized by two kinds of symptoms, namely negative (referring to a loss or an impairment compared with healthy subjects) and positive (referring to behaviors absent in healthy subjects). Negative symptoms mainly include a loss or a decrease of cognitive and motivational skills. Regarding the cognitive aspect, one observes an impaired ability to make decision, impaired concentration, impaired ability to produce coherent and organized speech and behavior, impaired logical thinking (associative thought, "jumping to conclusion"). On the motivational side, one observes an emotional flattening, the loss or decrease of expression of emotions, a poor spontaneous motor activity (catatonia) and action initiation, a loss of feeling of pleasure (giving rise to a loss of reward seeking behaviors).

Positive symptoms include auditory hallucinations (hearing voices), delusional beliefs or systems of beliefs with a paranoid or persecutory dimension. It is noteworthy that these hallucinations and delusions do not remain abstractions that the patients report to the psychiatrist. They drive and

influence the actual behavior of the patients, this is why positive symptoms are particularly burdensome.

Diagnosis is made on the basis of overt behavior, including the patient's reported experiences. The key symptoms must be confirmed and any other differential diagnosis eliminated (drug or alcohol abuse, metabolic illness, neurological condition).

Explaining and finding a treatment for such an a priori heterogeneous set of symptoms is very challenging for Neurosciences and Psychiatry. There generally exists no term-to-term correspondence between behaviors (by extension symptoms) and anatomical brain systems, so that it is difficult to define an efficient target structure. One presumably lacks an intermediary level of cognitive explanation that would account for how these overt symptoms are generated, namely a level of explanation able to decipher the "distribution of neuronal work" and to account for both types of symptoms.

4.1.2 Accounting for positive symptoms: dopamine dysregulation or specific cognitive disorders?

Dopamine

Schizophrenia is also characterized by an abnormal dopamine activity, this is even such a core characteristic of the pathology at brain level, that it has been hypothesized as the cause of the illness. This hypothesis has evolved since its first version, taking into account brain region and receptor specificity (Howles and Kapur, 2009). Dysregulation of dopaminergic activity in schizophrenia has thus been characterized by an increased baseline dopamine activity conjoint with prevalent D2-receptor binding in basal ganglia (AbiDargham et al., 2000), decreased D2-receptor binding in anterior cingulate (Suhara et al., 2002), decreased dopaminergic activity with prevalent D1-receptor binding in dorsolateral prefrontal cortex, correlating with impaired working memory performance (AbiDargham et al., 2002, 2003).

Dopamine dysregulation might be more involved in psychosis than in schizophrenia per se (Howles and Kapur, 2009). Nevertheless, the therapy (for schizophrenia) mainly consists in containing positive (i.e psychotic) symptoms pharmacologically, with drugs that block dopamine D2 receptors (haloperidol is a typical first generation antipsychotic) but cause collateral motor symptoms and are not effective for the

negative symptoms. New generations of antipsychotic drugs (e.g clozapine, risperidone) that tap also onto serotonin receptors are effective to restrain positive symptoms, less often cause motor side effects, but their efficacy for the negative symptoms has not been borne out.

As a matter of fact, if dopamine dysregulation can trigger psychotic episode, as it can also be observed during manic phases in bipolar patents (Berk et al, 2007), it does not provide by itself a sufficient account for some psychotic (positive) symptoms observed in schizophrenia. In effect, patients with bipolar disorders, when they present psychotic symptoms, display a psychotic profile that differs in schizophrenia (Dunayevich and Keck, 2000). Despite common characteristics, such as hallucinations and more frequently delusions, no self-agency disturbances, nor intrusive thoughts are observed. The content, but also the complexity and the temporal pattern of occurrence of delusions differ one from each other. Bipolar patients also do not present cognitive impairments such as those observed in schizophrenia –at least in the classical neuropsychological tests. The variability seems to be more important, and cognitive impairments may be absent in bipolar patients with psychotic history.

Thus, dopamine dysregulation is not sufficient to account for some specific positive symptoms in schizophrenia. Could it be a cause or an effect? Might altered dopamine receptor densities within striatum and prefrontal cortex also be an effect of another morphological cortical alteration? This could give rise to cognitive impairments, which then would bias reward based learning --known to involve dopaminergic pathways. This is an hypothesis to explore.

Sensorimotor efference copy mechanism

Some of the positive symptoms (loss of self-agency, voices) have been explained in terms of abnormal awareness of action. Feinberg (1978) seems to be the first to have hypothesized a mechanism of efference copy to explain the disruption of self-agency in schizophrenia, although in rather purely speculative way. This idea was empirically developed by Blakemore, Wolpert and in Frith, (2002), in a framework comprising a 'forward model' (prediction of sensory outcome according to a motor command) and an 'inverse model' (prediction of the motor command according to a desired outcome).

In that framework, when a motor command is activated, a signal -- an *efference copy*, is sent to the

corresponding sensory areas. This signal modulates the excitability of the sensory networks, so that the (predicted) effects of the movement are inhibited. According to this model, this neural attenuation following the motor command contributes to, and is even the signature of, the perception of the movement as self-caused. Then, in a second stage, actual sensory effects of the movement are sent to a comparator, where an error signal is calculated. This error signal is then exploited in order to correct the movement.

One must keep in mind that this framework rests on the *predictive coding* theory (see Friston, 2008 or Friston et al, 2011), according to which neural networks, based on previous sensory evidence, keep updating an estimation of the pending input. Actual local brain activity would thus correspond to error signals, as far as the actual neural activity would reflect the difference between estimated and actual input. Thus, if no efference copy is combined with sensory input during a self-caused movement, then one should observe a greater error signal in sensory networks. In other words, during sensorimotor control paradigm, one should observe stronger (parietal) activations in schizophrenic patients compared with control subjects.

Empirical evidence indeed is consistent with the existence of a forward modelling mechanism involving a motor parietal network (Desmurget et al., 2009). Neuroimaging studies (Frith et al., 2000; Fourneret et al., 2001, 2002 ; Blakemore et al., 2002; Frith, 2005) report that, compared with healthy volunteers, schizophrenic patients show stronger parietal activations during processing the sensory effects of self-caused events. They also typically fail to maintain the level of performance by sensorimotor adjustment. Importantly, patients seem to present such impairments only when they have to consciously attend to their actions (Knoblich et al., 2004; see Frith, 2005 for a review). Within this framework, delusions of control, including auditory hallucinations, are held to stem from an unawareness of the initiation of action and a disruption of the forward modelling system, because subjects perceive their own movements/inner speech as if it was externally caused.

Versus 'Central' cognitive impairments

The disruption of an efference copy mechanism, which would impair the inhibition of sensory effects, seems plausible in schizophrenia, but might stem from processes situated at a higher level than sensorimotor level (involving interactions between motor and parietal cortices). Instead, it might

originate from an executive control level and could consist in : (i) impaired integration of diverse error signals to monitor the behavior and adjusting cognitive control resources, (ii) impaired top-down modulation by lateral prefrontal networks, giving rise to cognitive deficits associated with this network, namely impaired working memory, impaired attentional selection, switched up threshold of access to consciousness, impaired cognitive control of action selection including inhibition of sensory effects of self-caused events. Empirical evidence supports these points.

(i) In effect, a first argument for a central origin of the disorder relies on evidence that schizophrenic patients present impaired performance in cognitive control paradigms, including error based performance monitoring. It has been observed, in healthy participants, that both commission and prediction error signals are exploited and combined in order to recruit more cognitive control of the ongoing movement/action/task (Holroyd and Coles, 2002; Krigolson and Holroyd, 2007; Baker and Holroyd, 2011). Thus, one typically observes response-locked and reward locked activity in anterior cingulate cortex, which is typically followed by a adjustment of the cognitive control resources. This adjustment can be observed through overt behavior: according to the paradigm one can observe an error or just a slower reaction time in the current trial, a slower reaction time in the following trial, improvement of the performance or change of strategy. It manifests itself also through brain activity. In anterior cingulate cortex one can observe response-locked error related negativity peaking at about 70-100 ms after response onset, or feedback-locked error related negativity peaking at about 300ms after feedback onset. In some cases, especially when subjects are aware of having made an error, one can also observe stronger activation in lateral prefrontal cortex.

This might be a critical point, since schizophrenic patients have been demonstrated to present impaired performance conjoint with abnormal error-locked brain activity in both anterior cingulate and dorsolateral prefrontal cortices (Mathalon et al., 2002 ; Dehaene et al., 2003; Morris et al., 2008; Polli et al., 2008; Laurens et al., 2003 ; see Adams & David, 2007 for a review). Few studies report an absence of post-error slowing and behavioral adjustment (Alain et al., 2002), whereas others report a post-error slowing without any significant difference between patients and control (Polli et al., 2008; Laurens et al., 2003). However, despite some divergences regarding behavior, abnormalities in anterior cingulate activity during error/conflict processing are consistent across studies, and compatible with impaired recruitment of dorsolateral prefrontal cortex during contextual executive control (Barch et al., 2001;

Perlstein et al, 2003; Chambon et al., 2008 ; Barbalat et al., 2009).

Thus, patients fail to monitor and adjust their own behavior (or their anterior cingulate/lateral prefrontal cortices fail to respond) not only upon sensory prediction error signals, but upon *all* kinds of signal, including motor selection error, and reward error signals (Heerey et al., 2008). For more details about Anterior cingulated functions*, see Part I, section 3.4.1 medial prefrontal cortex : Motivational Control, the Anterior cingualte Cortex as a Cornerstone*.

(ii) Furthermore, the hypothesis of deficits situated at a more central level can account in the same vein for various disorders, including those of awareness of action that are not considered by the model proposed by Frith, Blakemore and Wolpert. It does not explain, for example, why patients are impaired in adjusting their movements only when they have to *consciously* attend to them. It is unclear whether it is an issue of awareness of action initiation, or of cognitive control of action.

Impaired performance on neuropsychological assessments of executive functions is well documented in schizophrenia (Neil and Rossell, 2013). Studies based on most recent models in cognitive neuroscience have reported more fine-grained evidence about very specific impairments held to pertain to superior cognitive functions. Thus, a higher threshold of access to consciousness has been reported in patients with schizophrenia– as far as visual information is concerned at least (Del Cul et al., 2006). Impaired *contextual,* but not *episodic* (cf Part I for the definition of these notions) cognitive control of action has been observed as well (Barbalat et al., 2008). A common denominator of these deficits is that they both involve lateral prefrontal cortex.

Abnormalities of dorsolateral prefrontal have been well documented in schizophrenia. This prefrontal network has received a strong focus of attention because of structural and morphological alterations (Lewis et al, 2012), in particular a decreased density of a subset of GABA interneurons. A particularly important feature of these interneurons rests on their connectivity with pyramidal neurons and their role in the generation of theta and gamma oscillations, and phase coherence (Cardin et al., 2009; Benchenane et al., 2011). Gamma oscillations, with or without long range phase coherence, have been shown to underlie several cognitive processes: attentional selection (Lakatos et al., 2008), working memory (Gonzalez-Burgos et al., 2010), preparatory cognitive control (Cho et al., 2006; Minzenberg et

al., 2010), access to consciousness (Sergent et al, 2003; Luo et al, 2009). Kihara et al (2012) recently reported reduced gamma intertrial phase coherence, increased theta amplitude, despite intact cross-frequency coupling in schizophrenia patients, relative to healthy control subjects.

The decreased density of these GABA neurons may explain why antipsychotics attenuate positive symptoms but remain inefficient for the negative ones. Moreover, GABA-interneurons alterations may conceivably lead to altered reinforcement learning within prefrontal cortex, in which dopamine is critically involved (Schultz, and Dayan, MJ Franck), and act as another factor contributing to dopamine dysregulation in the long term.

4.2 Hypothesis and scope of the study:

The present study was carried out in two populations of patients with history of psychotic disorders: (i) a group including patients with a diagnosis of schizophrenia, (ii) a second group including patients with a diagnosis of bipolar disorders (see appendix 2 for DSM-4 criteria). It aimed to investigate some cognitive functions that have typically been associated with the anterior cingulate cortex (BA24) and prefrontal cortex, including medial and lateral BA9. These cognitive functions include cognitive control of behavior (consisting in selecting a motor response on the basis of a cascade of signals, cf. Koechlin et al., 2003; Badre, 2007; Kouneiher et al., 2009), conflict monitoring, access to consciousness (input or output information, cf. Dehaene and Changeux, 2011) and metacognition (previous chapters).

Although this study was exploratory, it was motivated by the "central" account of positive symptoms in schizophrenia. More specifically, it was driven by two objectives (i) demonstrating the existence of both cognitive and metacognitive impairments in schizophrenia patients, and (ii) putting in evidence that the cognitive and metacognitive profiles observed in schizophrenia qualitatively and quantitatively differ from the profile observed in bipolar disorders (with history of psychosis).

Considering this set of hypotheses, then, more specifically :

(1) The schizophrenia group should show a globally impaired performance in the basic task compared

to bipolar and control groups (Slower reaction times than bipolar).

(2) The schizophrenia group should be particularly influenced by the cognitive control load, that we manipulated with the congruency factor, or more than the other groups. This would manifests itself through cognitive performance (being less accurate in incongruent trials) and metacognitive confidence (more false alarms).

(3) In the basic task, the schizophrenia group might either (i) not show any priming effects (*id est* no compatibility effect), because of the functional abnormalities of response conflict related activity in Anterior cingulate cortex or the hypoactivation within prefrontal regions; or (ii) show priming effects in SHORT SOA trials only, as it has already been observed, though in a different paradigm (Dehaene et al., 2003).

4.2.1 Paradigm

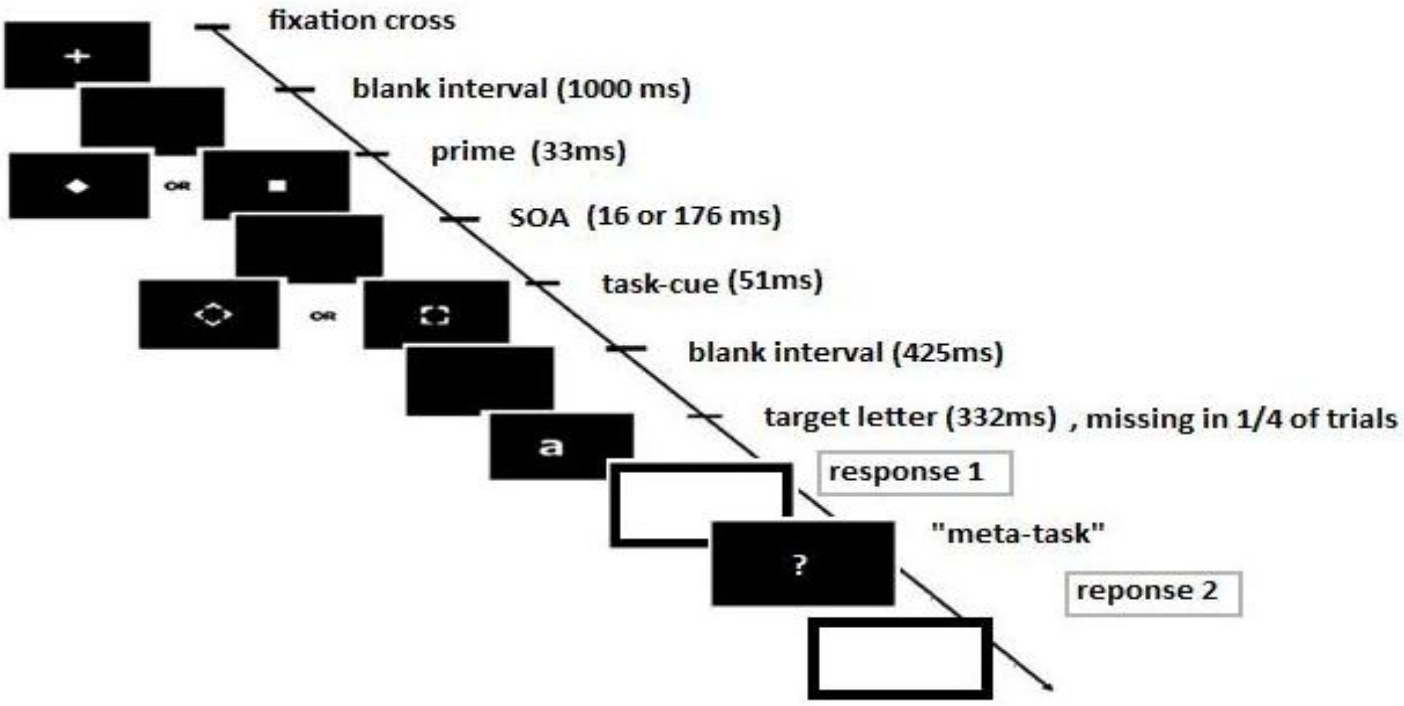
We used the same basic procedure as in our previous studies, but we adapted it for patients.

4.2.1.1 Double Staircase Algorithm

At the beginning of the study, since a more elevated threshold of visual awareness had been reported in schizophrenia patients (Del Cul et al, 2006), we intended to define the SOA values using a staircase algorithm (see appendix 1). For practical (time available for the patient) and clinical reasons (the experiments could not be so long because of a difficulty of concentration or motivation of the patients), we carried out a preliminary study with two patients with schizophrenia, for whom the staircase converged on 16 ms and 176ms, and 16ms and 160ms. We chose the SOAs values with the bigger interval, namely 16 ms and 176 ms for the other patients and control subjects.

4.2.1.2 Task-cueing paradigm

We used a paradigm similar to the one of our previous (unreported) EEG study (see figure 4-B), adapted for patients (different SOA values, time interval between task-cue and target, less constraining training). The factor manipulated remained the same: SOA (2), prime compatibility (2) and congruency (2). The paradigm is presented and explained below (figure 4-B), and the critical notion of congruency is explained immediately after, see below (figure 4-C).



[**Figure 4-B** Paradigm Trial :
task cueing : At each trial, subjects are displayed a prime (little square or little diamond) – then a task cue (big square or big diamond) which also plays the role of mask, and finally a target letter.
The priming can be *compatible* (prime and cue identical) or *incompatible* (prime and cue different), *invisible* (SOA = 16. ms) or *visible* (SOA = 176ms). The task cue indicates which of the two tasks to perform regarding the upcoming letter. The subject must answer *yes* or *no* as fast as possible, by pressing a left or right key. **Metacognitive task:**, After having answered, the subjects were asked about the correctness of their response. They could answer *'yes', 'no', 'don't know/not sure'*.
'I do not know/I am not sure' responses were considered as incorrect]

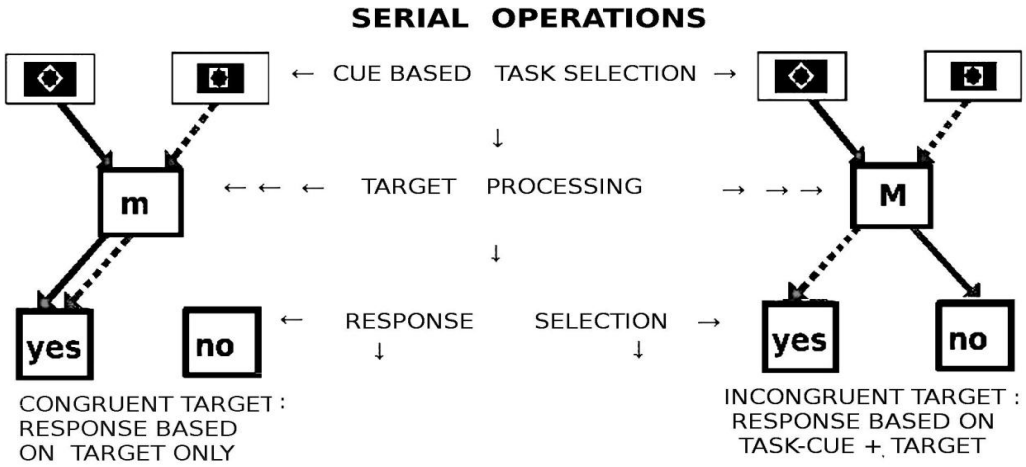


Figure 4-C : manipulating cognitive control load with target congruence
according to the task to perform, some targets give rise to a single (left, congruent) or two (right, incongruent) possible responses, so that in the first case, only 1 bit information is necessary to select the response, and in the other case, 2 bits information are needed.
In a theoretical point of view, a critical difference between these trials consists in that response selection is conditioned by task cue *and* the target only in incongruent trials, whereas congruent trials are equivalent to a simple target-based response selection. Lateral prefrontal activity is a priori more important in incongruent trials, because the cognitive control load (information necessary to select the correct response) is more important.

4.2.2 Subjects :

Patients :

Patients recruited in the CSM (Centro di Salute) of Udine Nord, a center for mental health situated in the main city of the Friuli region, diagnosed using DSM-IV criteria (see Appendix 6)

Bipolar (N=6; 3 males ; mean age 47,3 ± 12,86)

They were taking (one of) the following drugs : Acid Valproic,+Paroxetine, Carbolitium, Lamotrigin.

- Schizophrenia (N=9; 7 males; mean age 38, 7 ± 6,26)

They were taking (one of) the following drugs: Haloperidol, Risperidone, Paliperidone

Control subjects;

Subjects (N=7, 3 males ; age 33 ± 6,95) were recruited in Trieste, being mainly selected to match the patients, stet on the age (25-63 years old), gender, handedness. Exclusion criteria were: scholarship length (more than 3 years at the University), no medical treatment, no past or present history of psychiatric disorders for themselves or the members of their family. All gave informed consent, and were informed that they could abandon the experiment at any time, without giving any justification.

4.2.3 Procedure:

Day 1 : The first day consisted in a training and the experiment proper.

The training ended when **mean accuracy was superior to 90%,** and when **mean Response Times were inferior to 2000 ms.** The schizophrenia group reached that level of performance within about 3/4 training sessions (about 300 trials). The bipolar group reached that performance within 2 sessions (about 200 trials). The control group reached that performance within about one training session (about 100 trials). See below for between group statistics about the training length.

Day 2 : A clinical assessment and neuropsychological battery tests were carried out (Iowa Gambling Task, Wisconsin Sorting Cards Test, Raven Matrices, digit short term memory span). This aspect is not reported because the study is incomplete in this respect .

The whole experiment standardly comprised 7 blocks each comprising 64 trials (+16 trials without target), each trials being separated by an interval of 1 second. The key/response mapping was counterbalanced across subjects, thus splitting them into 2 groups.

Subjects were instructed to ignore the prime whenever they saw it, and to pay attention to the mask/cue and to the letter. They had to answer a yes/no question about the letter according to the shape of the cue. The *square* always indicated the question '*is it a consonant?*' whereas the diamond always indicated *'is it in lowercase?*'. The subjects answered **yes** or **no** by pressing one of two keys as fast as possible (the mapping key/answer was constant during all the experiment, but differed across groups).

4.3. Results:

Since the study is incomplete, only the outcome of the paradigm are considered at the moment (clinical and demographic variables will not be considered). The data were cleaned (reaction times superior to 6000 ms for patients, 4000ms for controls and inferior to 300ms). Both raw and log transformed reaction times (including meta reaction times), and raw and arcsine transformed accuracy (including meta accuracy) were submitted to a Shapiro Wilk normality test. Since all these measures failed to pass the normality test, non parametric tests were systematically used.

As a general procedure a Kruskal-Wallis test was used to compare the three groups. If there was a significant effect, each group was compared with the others to attempt to pin down the source of the effect. If no significant effect was found in the initial Kruskal-Wallis test then the schizophrenic group was compared with each of the other two groups using a Bonferroni-based criterion of $p<0.025$ for significance. For the within group analysis, pairwise Wilcoxon tests were used. For multiple comparisons in each SOA condition, we set up a Bonferroni-based criterion of $p<0.025$ for significance.

Importantly, we used the same criteria as in our previous neuroimaging study to determine correct metacognitive judgment. Therefore, **unconfident** ('I am not sure/do not know') **metacognitive responses were considered** as **incorrect.**

4.3.1 Basic Task : task_cueing combined with masked priming

SOA	CONG	COMP	Mean ACC	SD	Mean RT	SD
Long	**cong**	**comp**	0,95	0,07	1475,06	665,85
		incomp	0,96	0,07	1524,64	691,68
	Total cong		0,95	0,07	1499,85	671,42
	incong	**comp**	0,94	0,10	1468,21	625,45
		incomp	0,94	0,10	1540,96	749,35
	Total incong		0,94	0,10	1504,59	683,11
Total long			0,95	0,08	1502,22	673,39
Short	**cong**	**comp**	0,96	0,07	1540,52	701,14
		incomp	0,95	0,07	1479,57	712,26
	Total cong		0,96	0,07	1510,04	699,14
	incong	**comp**	0,93	0,09	1546,66	783,45
		incomp	0,93	0,10	1562,46	812,41
	Total incong		0,93	0,09	1554,56	788,77
Total short			0,94	0,08	1532,30	741,35
Total			0,94	0,08	1517,26	706,32

[**TABLE 1 : GLOBAL PERFORMANCE** IN THE BASIC TASK,
see at the end of the basic task results for performance in each group**]**

The training length : Kruskal-Wallis rank sum test was performed on mean training length (expressed in number of trials) with clinical group as factor. It revealed a significant effect of **clinical group (p<0.006).** Binary comparisons (p-value set at 0.025) revealed no significant difference between the schizophrenia and bipolar groups (p<0.07). The **schizophrenia group, however, significantly differed from the control group (p< 0.004)**. The bipolar group also differed from the control group **(p<0.05), marginally though.**

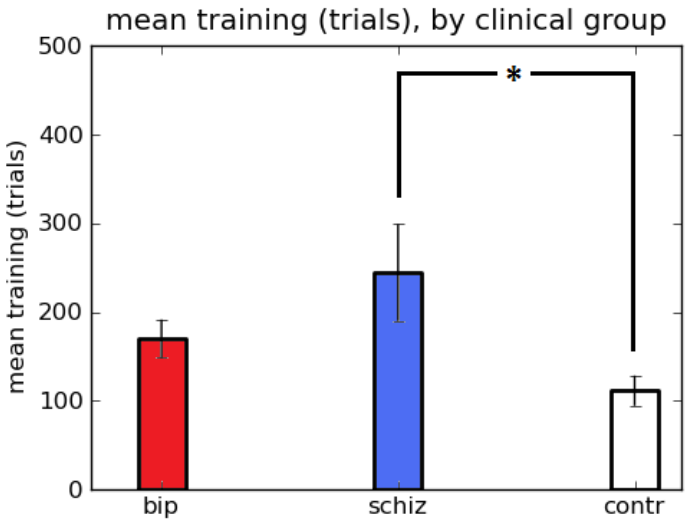
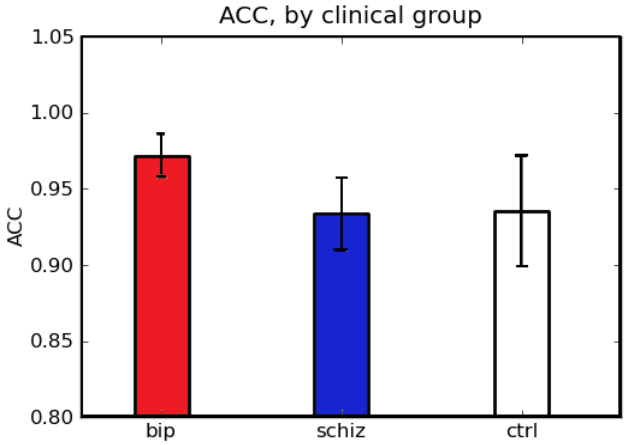


figure 4.0 : mean training length (trials) by clinical group,
(error bars represent standard errors)]

4.3.1.1 Accuracy

4.3.1.1. a Between group

Kruskal-Wallis rank sum test performed on global accuracy with clinical group as a factor did not reveal any significant difference between the three groups (p> 0.23). Binary comparisons (p-value set at 0.025) revealed no significant difference between schizophrenia and bipolar groups (p>0.07), between bipolar and control groups (p>0.88), between schizophrenia and control groups (p>0.26).

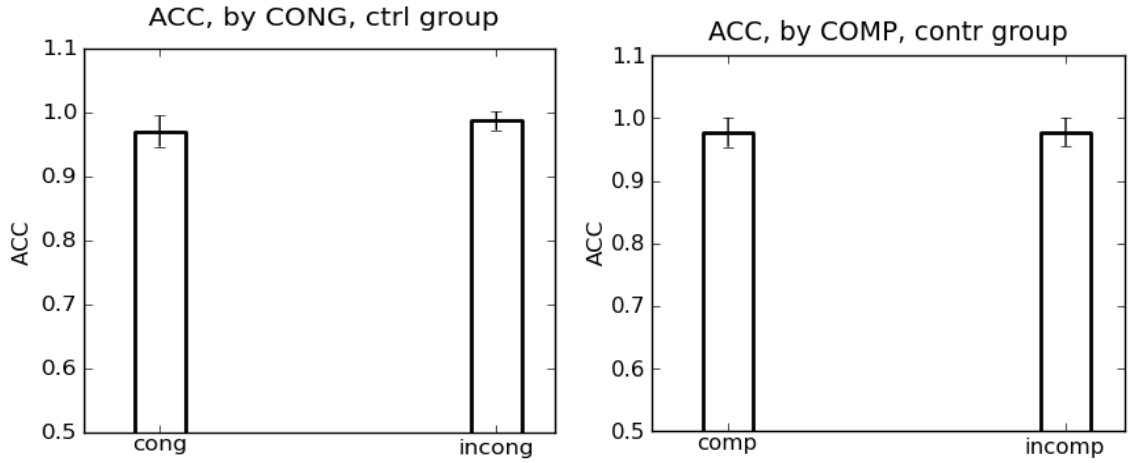


[**figure 4.1 :** accuracy by clinical group, no significant difference (error bars represent standard errors)]

4.3.1.1. b Within group

- Control group

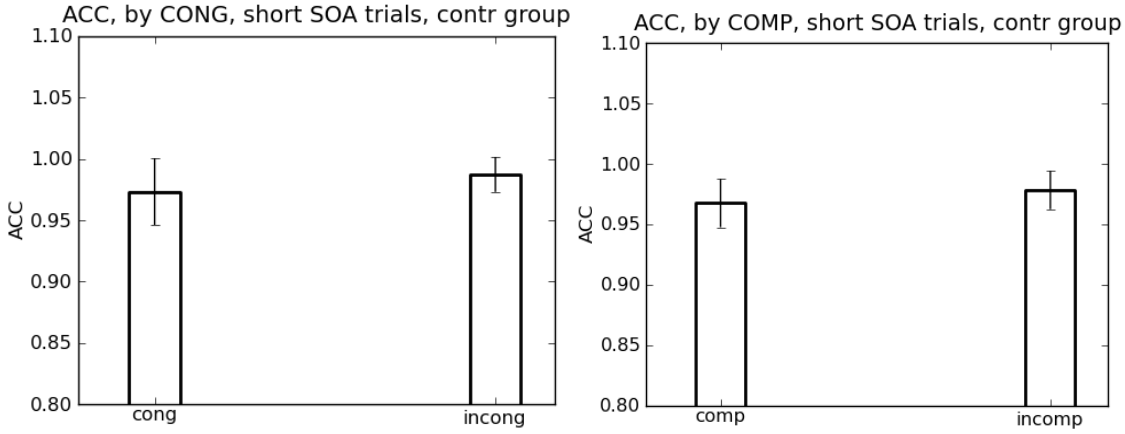
No congruency effect (p>0.82), no effect of compatibility (p> 0.59) were observed in the control group.



[**figure 4.2 :** accuracy in the control group, no significant effect of congruency (left) nor compatibility (right), (error bars represent standard errors)]

Short SOA:

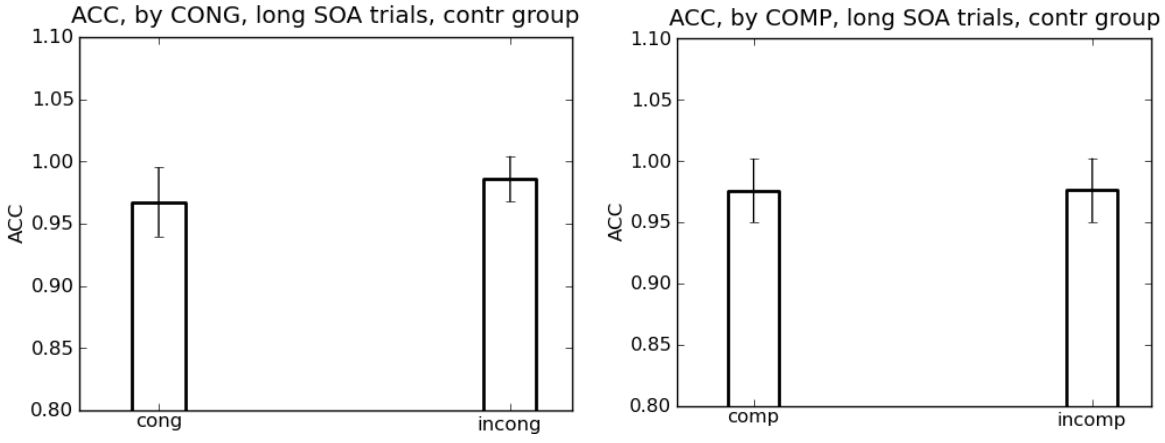
We observed no effect of congruency (pairwise Wilcoxon, p>0.88) in the short SOA trials neither of compatibility (pairwise Wilcoxon, p>0.9).



[**figure 4.3 :** accuracy in the control group, SHORT SOA trials, no significant effect of congruency (left) nor compatibility (right), (error bars represent standard errors)]

Long SOA:

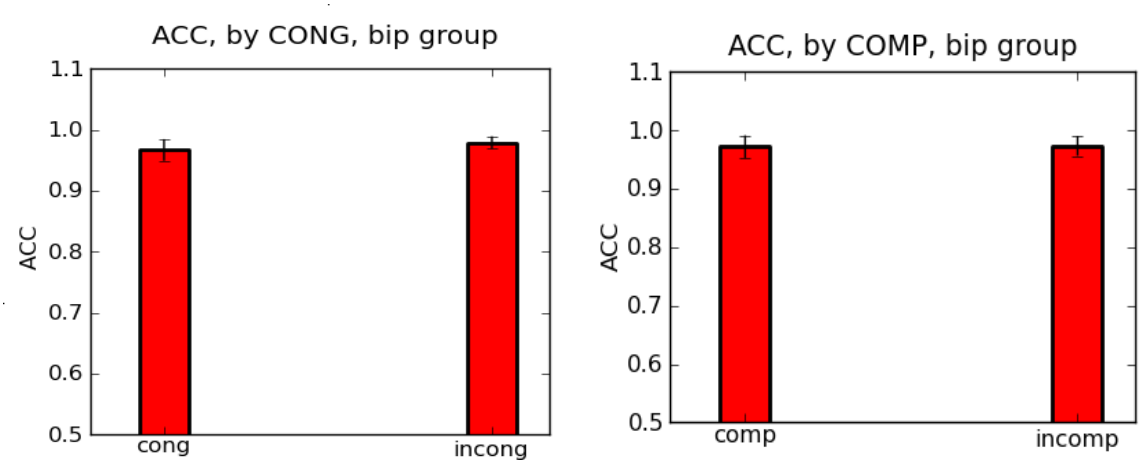
We observed no effect of congruency (p>0.86), and no effect of compatibility (p>0.86) in long SOA trials.



[**figure 4.4 :** accuracy in the control group, LONG SOA trials, no significant effect of congruency (left) nor compatibility (right), (error bars represent standard errors)]

- Bipolar group :

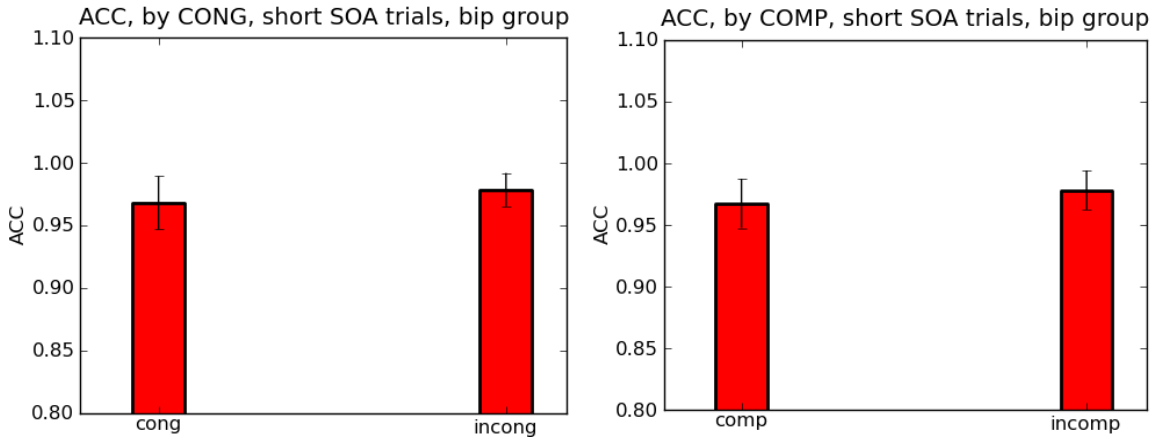
Bipolar group showed no significant effect of compatibility (p>0.42), and no effect of congruency (p>0.89).



[**figure 4.5 :** accuracy in the bipolar group, no significant effect of congruency (left) nor compatibility (right), (error bars represent standard errors)]

Short SOA:

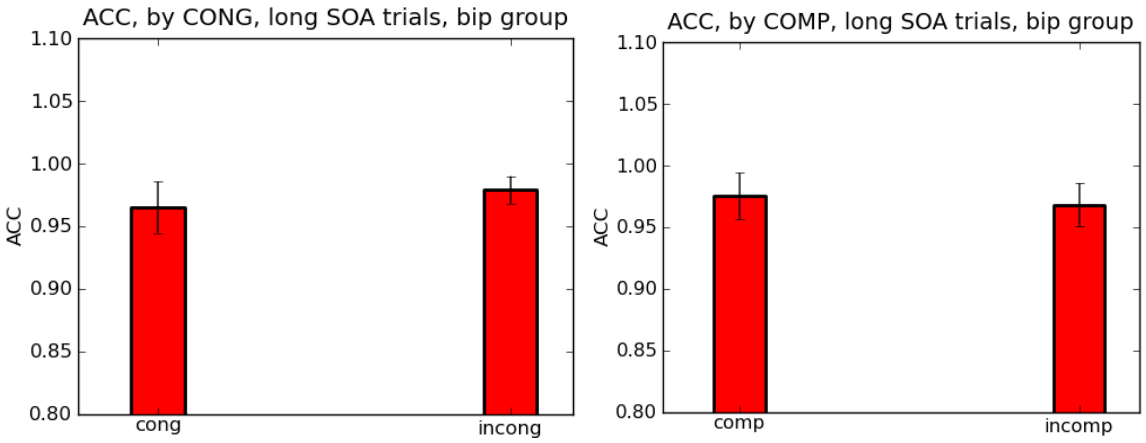
No significant effect of congruency was observed (p> 0.89) in short SOA trials; No significant effect of compatibility (p> 0.13)



[**figure 4.6 :** accuracy in the bipolar group, SHORT SOA trials, no significant effect of congruency (left) nor compatibility (right), (error bars represent standard errors)]

Long SOA:

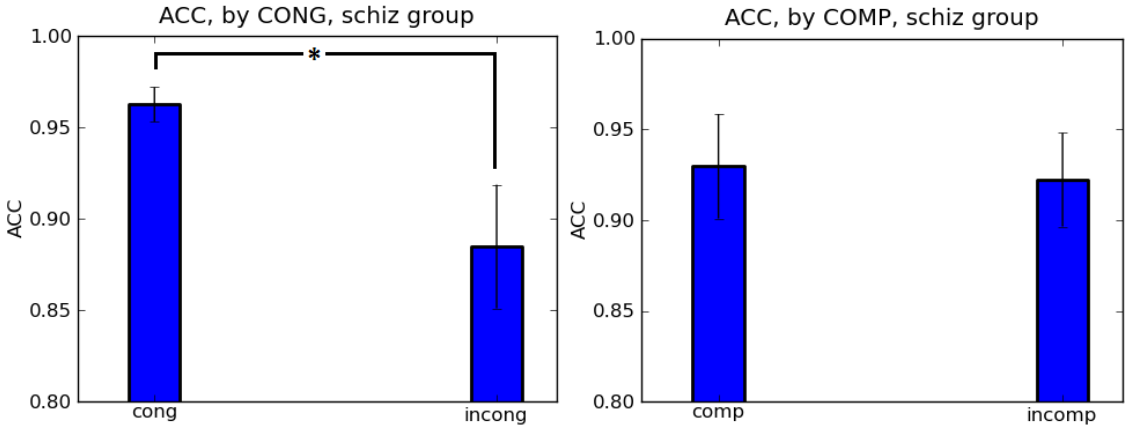
No congruency effect was observed (p> 0.78) and no compatibility effect (p> 0.28) were observed in long SOA trials.



[**figure 4.7 :** accuracy in the bipolar group, LONG SOA trials, no significant effect of congruency (left) nor compatibility (right), (error bars represent standard errors)]

- Schizophrenia group

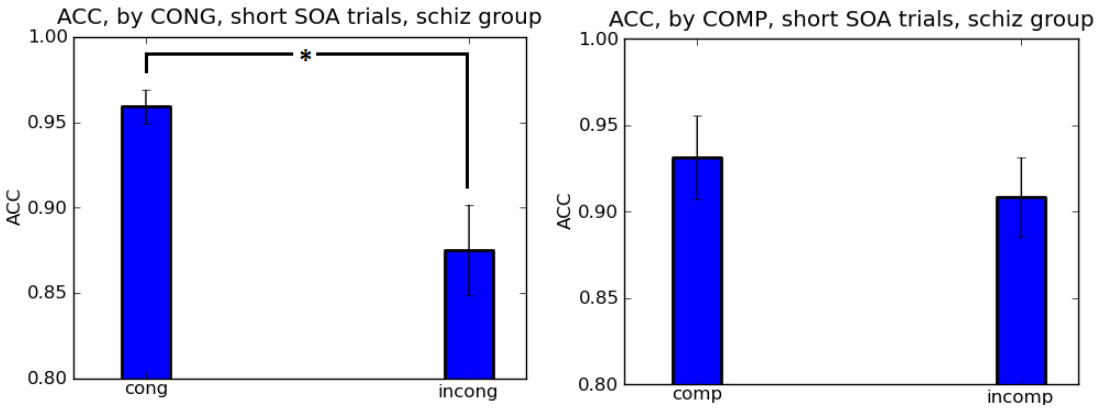
The schizophrenia group showed significant effect of **congruency (p<0.04),** but no significant effect of compatibility (p>0.65).



[**figure 4.8 :** Accuracy in the schizophrenia group by congruency (left), by compatibility (right),]

Short SOA:

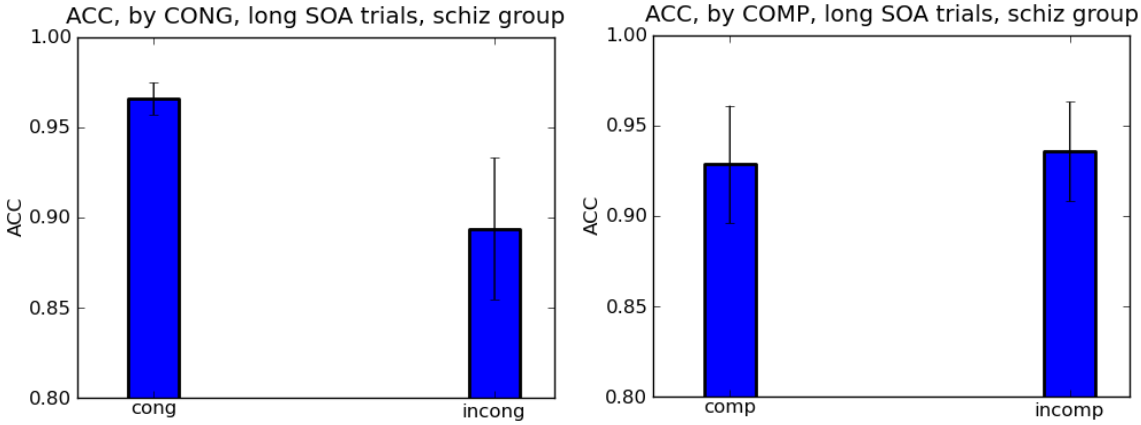
In short SOA trials, we observed a significant effect of **congruency** (**p <0.008**.) but no effect of compatibility (p>0.16).



[**figure 4. 9:** Accuracy in the schizophrenia group
by congruency (left), by compatibility (right), in SHORT SOA trials]

Long SOA:

In long SOA trials, we observed no significant effect of compatibility (p>0.9), neither of congruency (p>0.49).



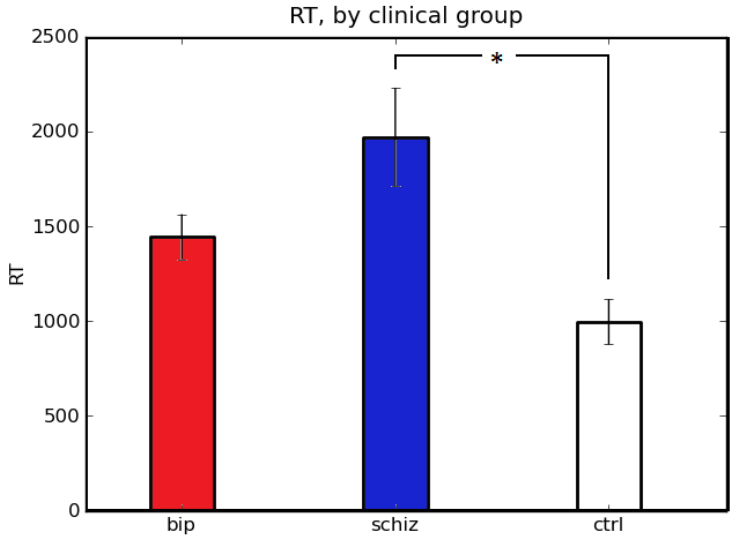
[**figure 4.10 :** Accuracy in the schizophrenia group
by congruency (left), by compatibility (right), in LONG SOA trials]

4.3.1.2 Reaction Times :

4.3.1.2.a Between group

Kruskal-Wallis rank sum test revealed a significant effect of clinical group (p<0.029). Two-by-two

comparisons between the groups showed that the schizophrenia group did not differ from bipolar patients (p>0.27), that the **schizophrenia group was significantly slower than the control group (p< 0.02)**. The bipolar group was not significantly slower than the control group **(p<0.07).**

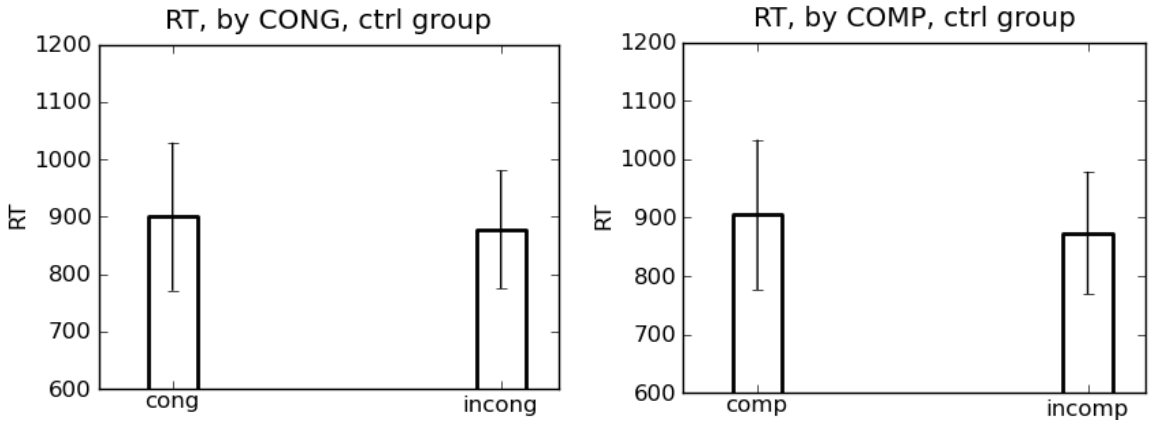


[**figure 4-11** : Reaction times by clinical group, bars represent standard errors]

4.3.1.2.b Within group

- Control group

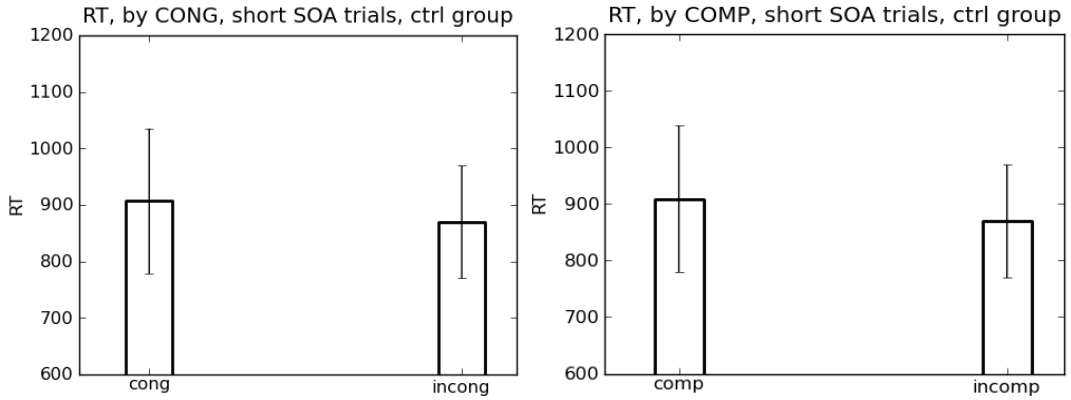
No significant effect of congruency (p>0.9), and no significant effect of compatibility (Wilcoxon, p>0.29) were observed in the control group.



[**figure 4.12** : RT, control group, by congruency (left) and compatibility(right)]

Short SOA:

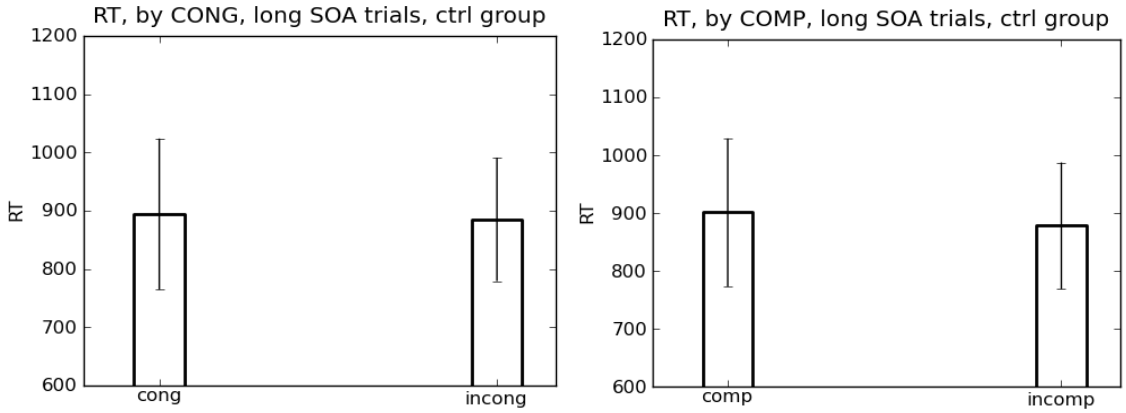
In short SOA trials, we observed no effect of congruency (pairwise Wilcoxon, p>0.32), nor of compatibility (p>0.21).



[**figure 4.13** : RT, control group, by congruency (left) an compatibility(right), SHORT SOA trials]

Long SOA:

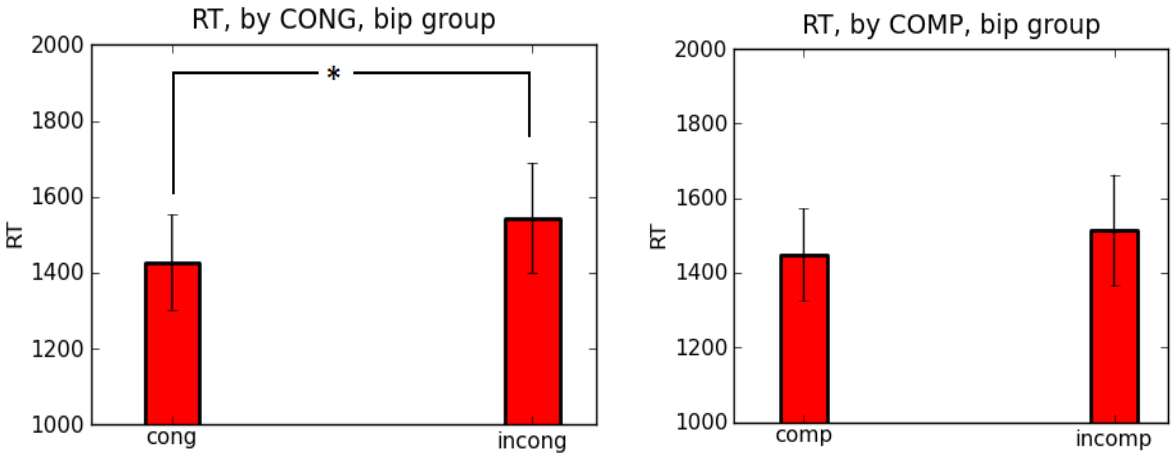
In long SOA trials, we observed no significant effect of congruency (p>0.57), no significant effect of compatibility (p>0.46),



[**figure 4. 14** : RT, control group, by congruency (left) an compatibility(right) in LONG SOA trials]

- Bipolar group :

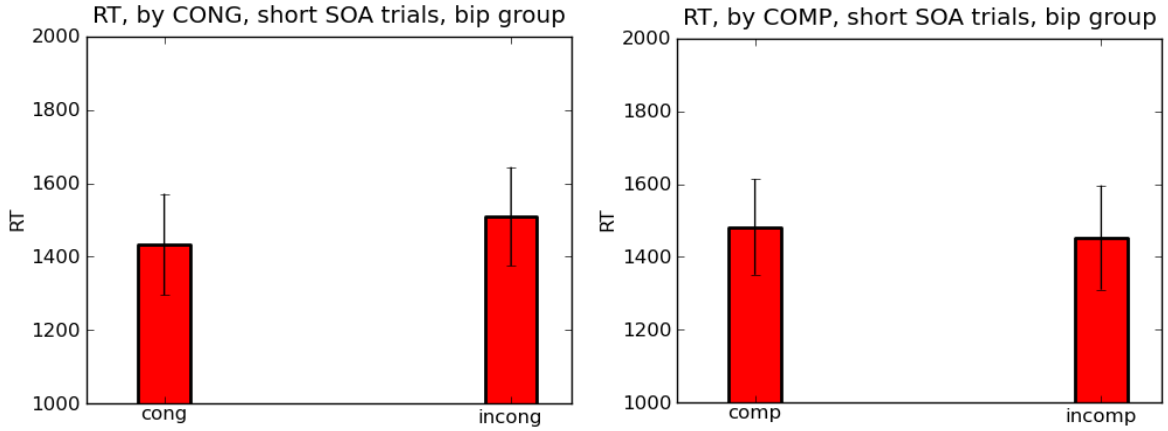
Reaction times of the bipolar group showed a significant effect of **congruency (pairwise Wilcoxon test, p<0.008),** but no global significant effect of compatibility (p>0.16).



[**figure 4-14**: Reaction times, bipolar group,
congruency effect (left), no compatibility effect (right), bars represent standard errors]

Short SOA:

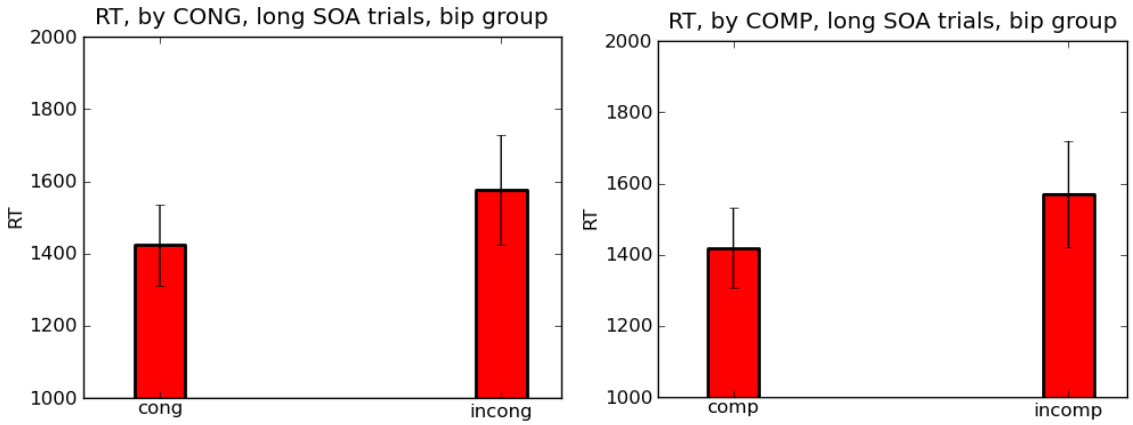
At short SOA, no effect of congruency (p>0.15) and no compatibility was observed (p>0.43),



[**figure 4-15**: Reaction times, bipolar group,
congruency effect (left), no compatibility effect (right), in SHORT SOA trials,
bars represent standard errors]

Long SOA:

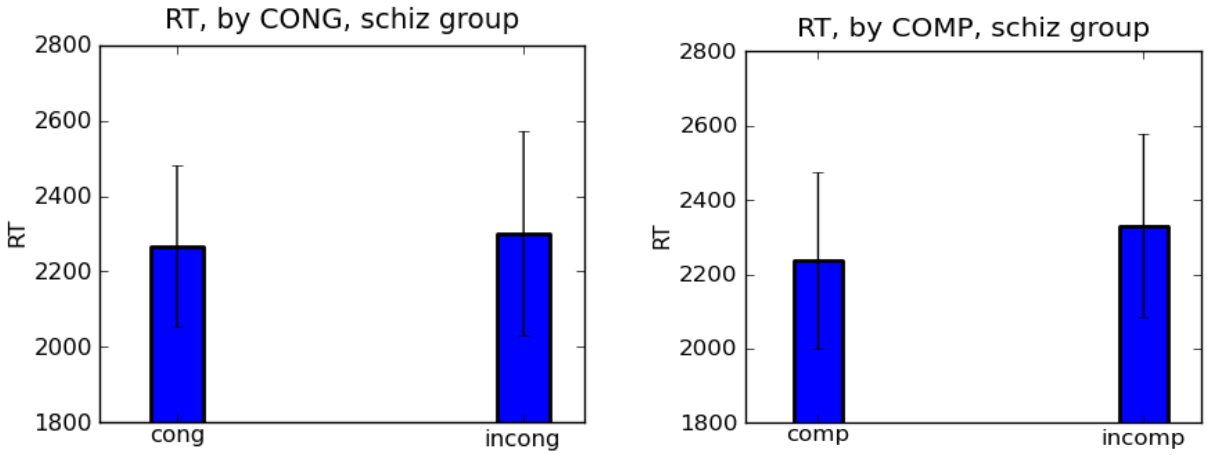
At long SOA, reaction times did not show any effect of congruency (p>0.15), nor of compatibility (p>0.15).



[**figure 4-16**: Reaction times, bipolar group, congruency effect (left), no compatibility effect (right), in LONG SOA trials, bars represent standard errors]

- Schizophrenia group

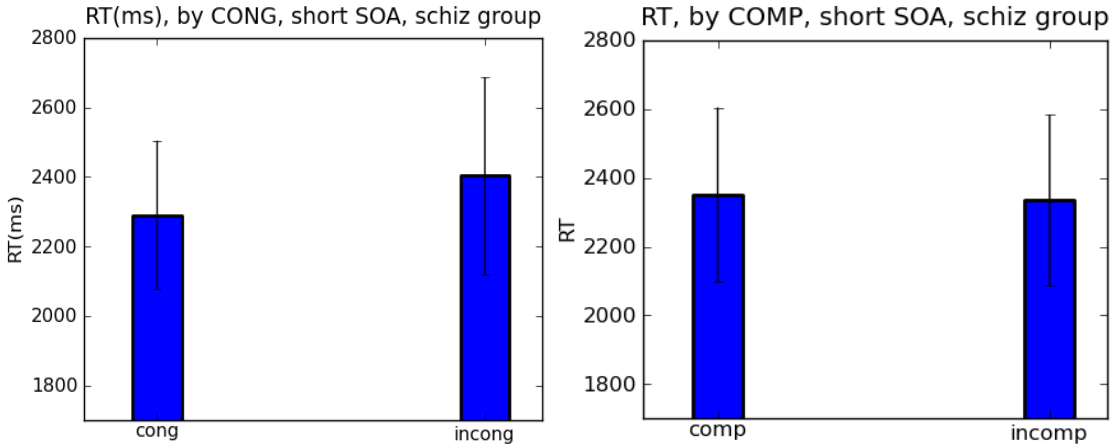
Globally, reaction times of the schizophrenia group were not significantly influenced by congruency (p>0.48) nor by compatibility (p>0.35).



[**figure 4-17**: Reaction times, schizophrenia group, No congruency effect (left), no compatibility effect (right), bars represent standard errors]

Short SOA:

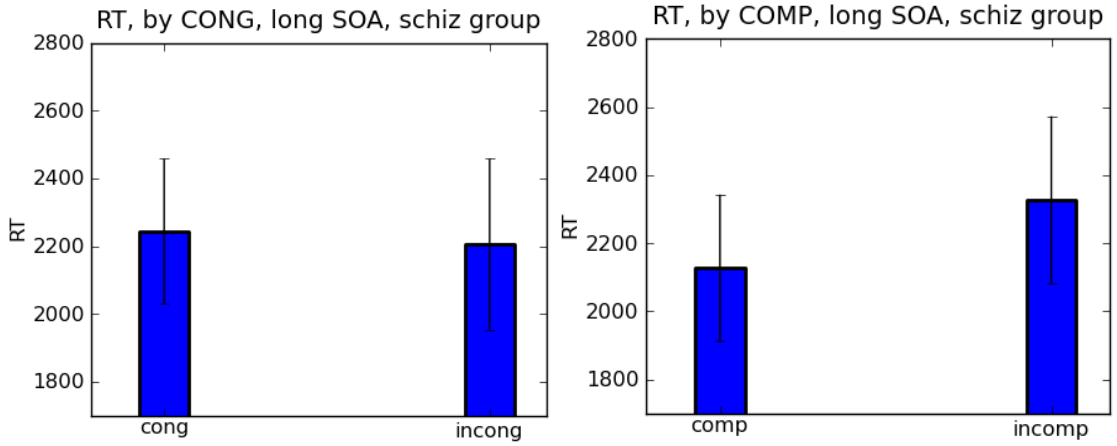
At short SOA, reaction times of the schizophrenia group did not show significant effects of compatibility (p>0.73), nor of congruency (p>0.57)



[**figure 4-18**: Reaction times, schizophrenia group, No congruency effect (left), no compatibility effect (right), SHORT SOA trials, bars represent standard errors]

Long SOA:

At long SOA, reaction times did not show significant effects of congruency (p>0.35), nor of compatibility (p>0.15).



[**figure 4-19**: Reaction times, schizophrenia group, No congruency effect (left), no compatibility effect (right), LONG SOA trials, bars represent standard errors]

Summary of the results in the basic task (task-cueing + masked priming)

Between group : Accuracy did not differ significantly between the three clinical groups. However, despite this equal performance level among the clinical groups, the schizophrenia group was significantly slower than control group and reached the requested performance with a significantly longer training than the control group. The training of bipolar group tended to be longer than the training of the control group.

Within group : Congruency significantly influenced the performance of the schizophrenia group, with more errors in incongruent trials. Further analyses showed that this effect was significant in short SOA trials, but not in long SOA trials. Congruency also influenced the reaction times of the bipolar group only, which was slower in incongruent than in congruent trials.

We did not observe any significant effect of compatibility.

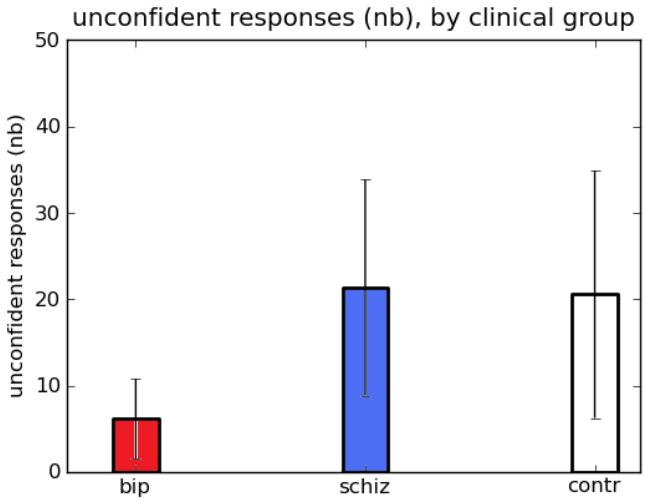
Neither congruency, nor compatibility significantly influenced the performance of the control group.

4.3.3 Metacognitive task

SOA	CONG	COMP	Mean meta-ACC	SD	Mean meta-RT	SD
Long	**Cong**	**comp**	0,92	0,09	551,41	158,40
		incomp	0,93	0,08	544,54	144,86
	Total cong		0,92	0,09	547,97	149,96
	incong	**comp**	0,91	0,11	532,95	134,74
		incomp	0,91	0,12	561,48	149,50
	Total incong		0,91	0,11	547,21	141,31
Total long			0,92	0,10	547,59	144,82
Short	**cong**	**comp**	0,93	0,09	576,88	154,02
		incomp	0,93	0,10	545,91	126,02
	Total cong		0,93	0,09	561,40	139,88
	incong	**comp**	0,89	0,13	535,41	146,53
		incomp	0,90	0,12	560,86	138,39
	Total incong		0,89	0,12	548,14	141,36
Total short			0,91	0,11	554,77	139,93
Total			0,91	0,11	551,18	142,01

[TABLE 5: **GLOBAL** PERFORMANCE IN THE METACOGNITIVE TASK, see at the end of the of the metacognitive task results the performance in each clinical group]

With respect to the number of unconfident responses ("I do not know") : A Kruskal-Wallis rank sum test performed on unconfident responses (expressed in absolute number of such response) with clinical group as factor revealed no significant effect of clinical group ($p > 0.43$). Binary comparisons (Bonferroni corrected p-value set at 0.025) revealed no difference between the schizophrenia and bipolar groups ($p>0.19$), between the schizophrenia and control groups ($p>0.83$), between the bipolar and control group ($p>0.38$)**.** Splitting these reports by correct versus incorrect first-order responses revealed no significant difference, in any group (pairwise Wilcoxon, all p-values >0.17). See Appendix 5 for more detailed data (by group, subject and correct/incorrect trial).



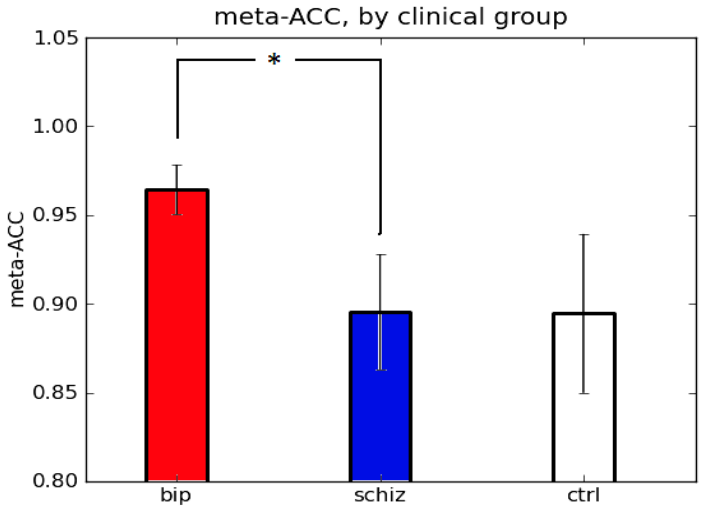
[figure 4-20 : unconfident responses by clinical group, bars represent standard errors]

4.3.3.1 Meta-accuracy

Note that "I do not know" responses given by the participants were not removed from the dataset, but were considered as **incorrect metacognitive responses**.

4.3.3.1.a Between group

Kruskal-Wallis rank sum test revealed no significant difference between the three groups ($p>0.23$). Two-by-two comparisons between groups (Mann-Whitney) revealed that both bipolar and schizophrenia groups did not differ from the control group ($p>0.51$). However we observed a significant difference between the s**chizophrenia and bipolar groups. ($p<0.025$).**

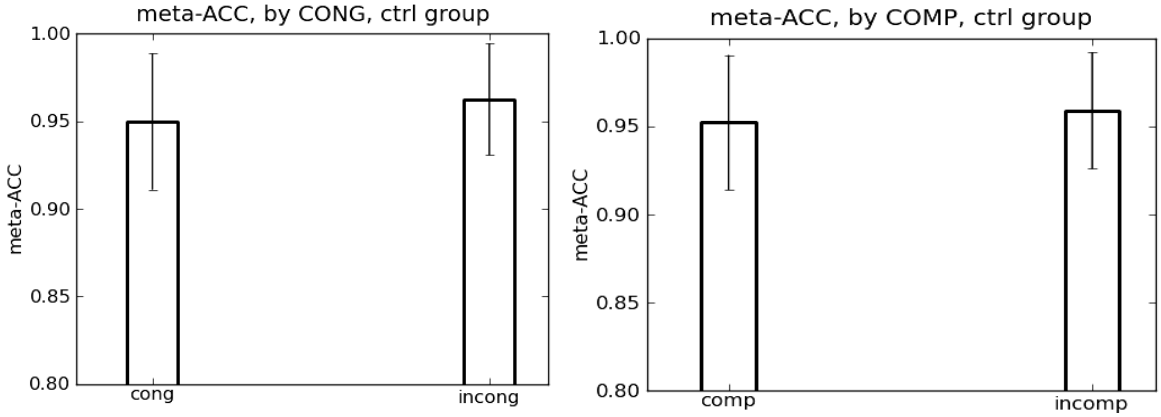


[figure 4-21 : Meta Accuracy by group, bars represent standard errors]

4.3.3.1.b within group

- Control group

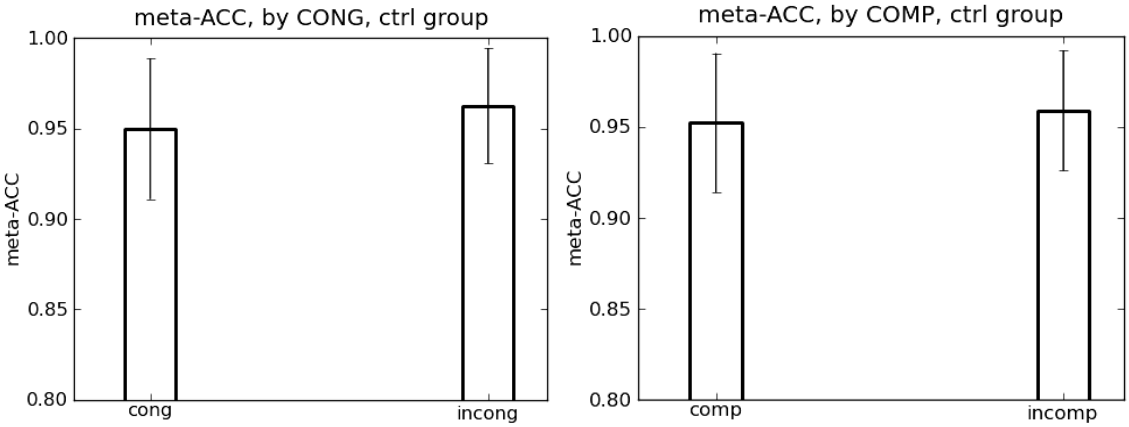
The meta accuracy of the control group did not show any significant effect of congruency (p>0.69) nor of compatibility (p>0.78).



[figure 4-22 : Meta Accuracy in the control group, by congruency (left) and compatibility right); bars represent standard errors]

<u>Short SOA:</u>

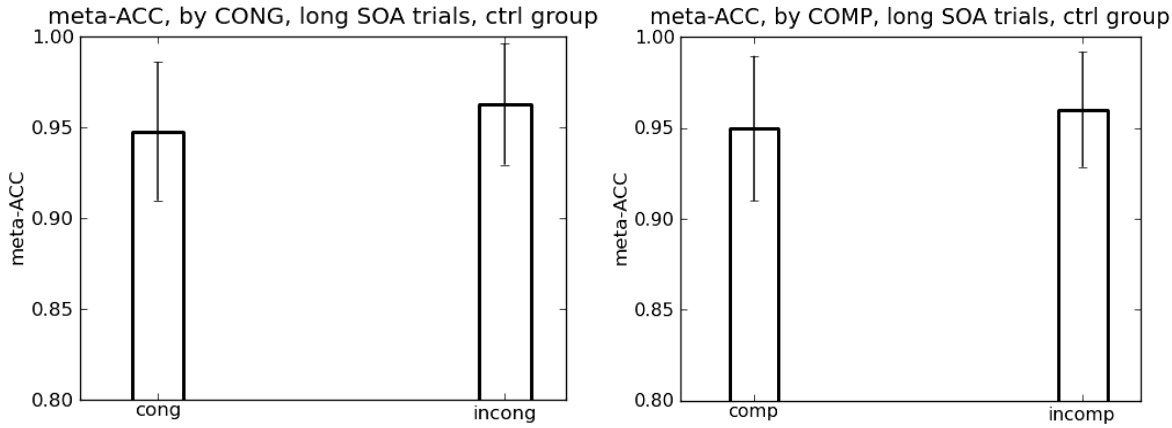
We observed no significant effect of congruency (p>0.83), nor of compatibility (p>0.91).



[figure 4-23 : Meta Accuracy in the control group, SHORT SOA trials, by congruency (left) and compatibility right); bars represent standard errors]

Long SOA:

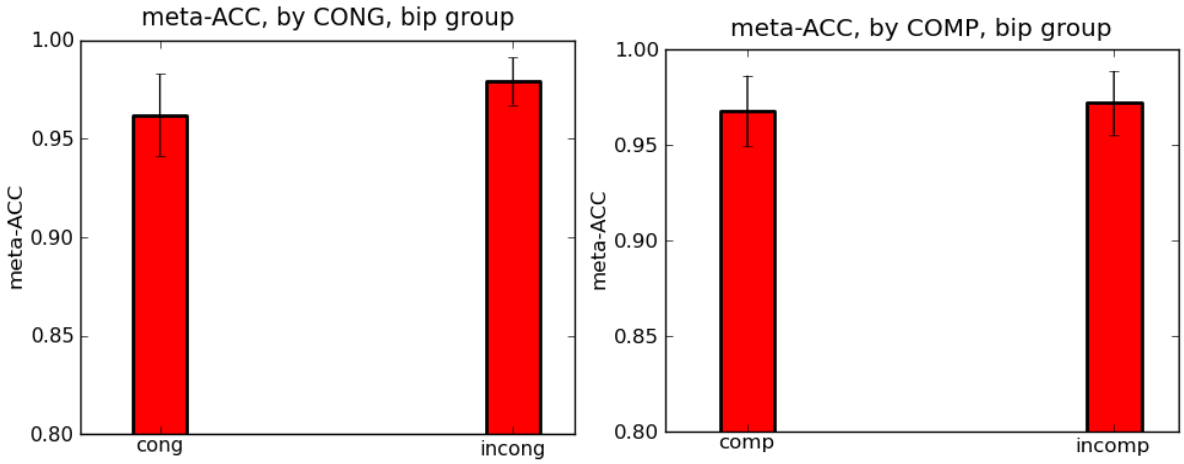
We observed no significant effect of congruency (p>0.85), nor of compatibility (p>0.58).



[figure 4-24 : Meta Accuracy in the control group, LONG SOA trials, by congruency (left) and compatibility right); bars represent standard errors]

- Bipolar group :

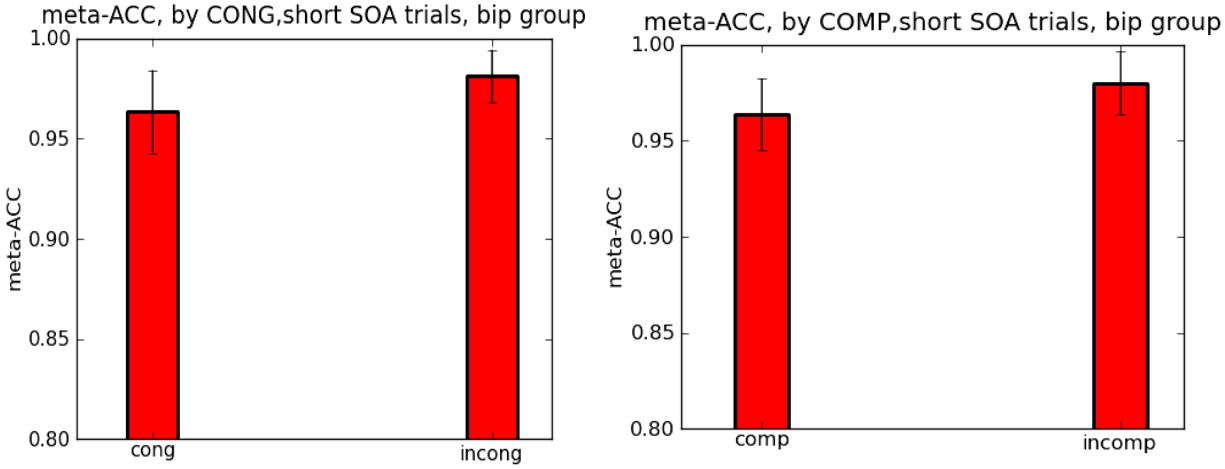
We observed no effect of congruency (p>0.33), nor of compatibility(p>0.28)



[figure 4-25 : Meta Accuracy in the bipolar group, by congruency (left) and compatibility (right); bars represent standard errors]

Short SOA :

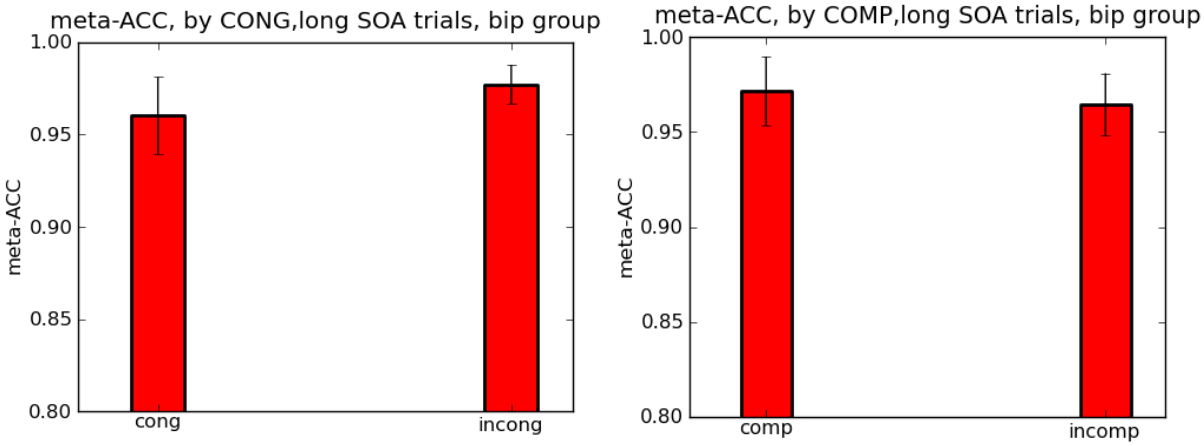
We observed no significant effect of congruency (p>0.8), no of compatibility (p>0.10)



[figure 4-26 : Meta Accuracy in the bipolar group, SHORT SOA trials, by congruency (left) and compatibility (right); bars represent standard errors]

Long SOA :

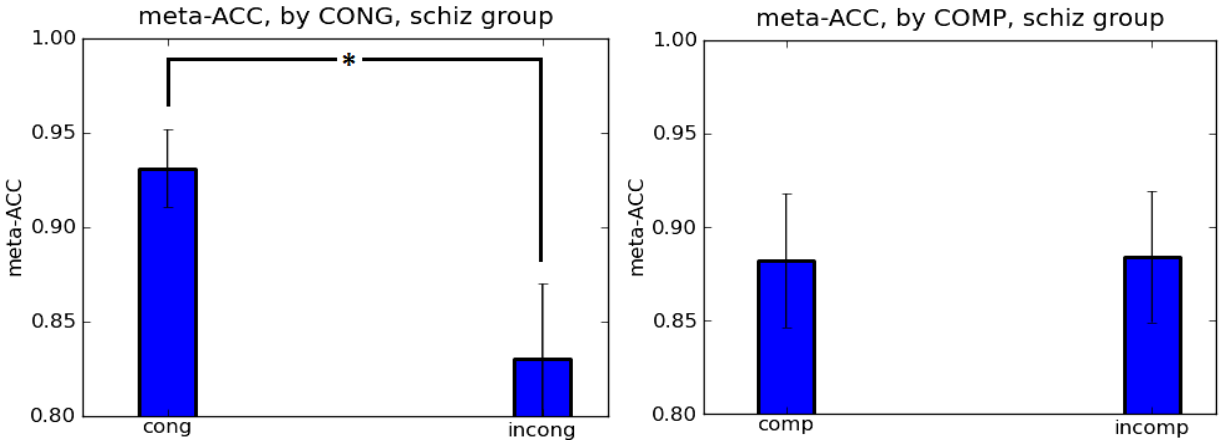
We observed no significant effect of congruency (p>0.24), nor of compatibility (p>0.81).



[figure 4-27 : Meta Accuracy in the bipolar group, LONG SOA trials, by congruency (left) and compatibility (right); bars represent standard errors]

- Schizophrenia group

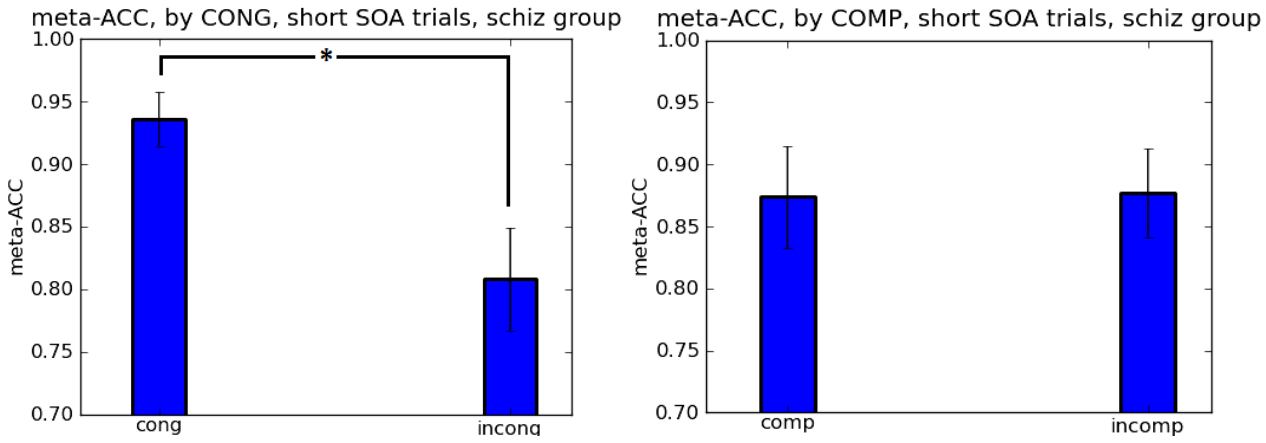
We observed a significant effect of congruency **(p<0.006),** but not of compatibility (p> 0.65)



[figure 4-28 : Meta Accuracy in the schizophrenia group, by congruency (left) and compatibility (right); bars represent standard errors]

Short SOA :

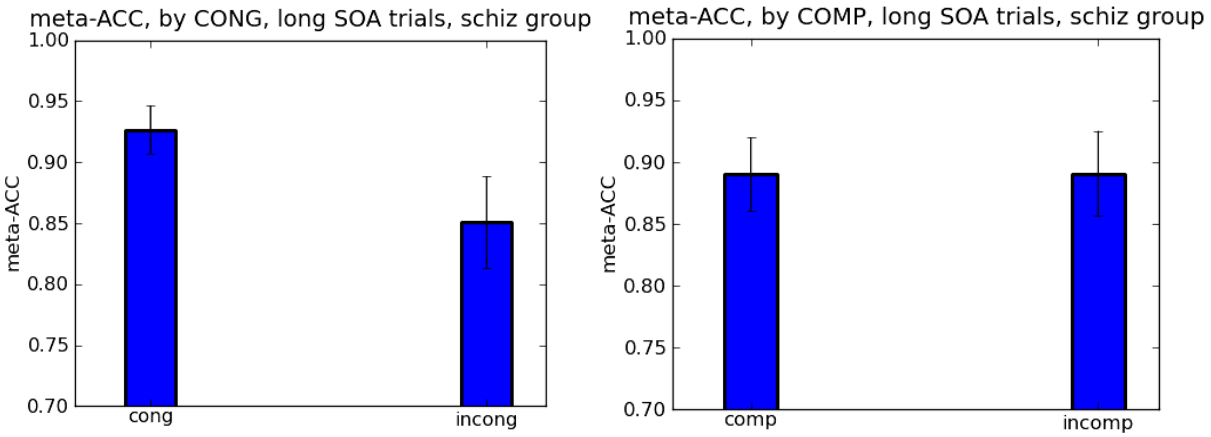
We observed a significant effect of **congruency (p< 0.023),** but not of compatibility (p> 0.72),



[figure 4-29 : Meta Accuracy in the schizophrenia group, SHORT SOA trials, by congruency (left) and compatibility (right); bars represent standard errors]

Long SOA :

In long SOA trials, the congruency effects no longer was significant($p>0.57$). We did not observe any effect of compatibility ($p>0.62$).



[figure 4-30 : Meta Accuracy in the schizophrenia group, LONG SOA trials, by congruency (left) and compatibility (right); bars represent standard errors]

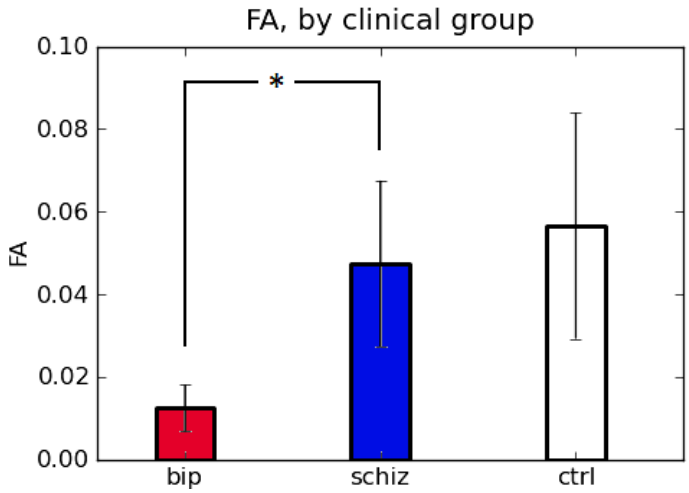
4.3.3.2 False Alarms (incorrect or unconfident second-order response after a correct first-order response):

SOA	CONG	COMP	mean FA	SD
Long	**cong**	**comp**	0,04	0,05
		incomp	0,04	0,05
	Total cong		0,04	0,05
	incong	**comp**	0,04	0,08
		incomp	0,03	0,06
	Total incong		0,04	0,07
Total long			0,04	0,06
Short	**cong**	**comp**	0,04	0,06
		incomp	0,04	0,06
	Total cong		0,04	0,06
	incong	**comp**	0,04	0,08
		incomp	0,04	0,08
	Total incong		0,05	0,08
Total short			0,04	0,07
Total			0,04	0,06

[TABLE 6: **GLOBAL** FALSE ALARMS IN THE METACOGNITIVE TASK, see the results by group at the end of the metacognitive task results]

4.3.3.2.a Between group

Kruskal-Wallis rank sum test revealed no significant difference between the three groups (p>0.19). Two-by-two comparisons between groups (Bonferroni corrected p-value set at 0.025) showed that both schizophrenia and bipolar groups did not differ from the control group (p> 0.42). However we observed significant differences **between the schizophrenia and bipolar groups (p< 0.025).**

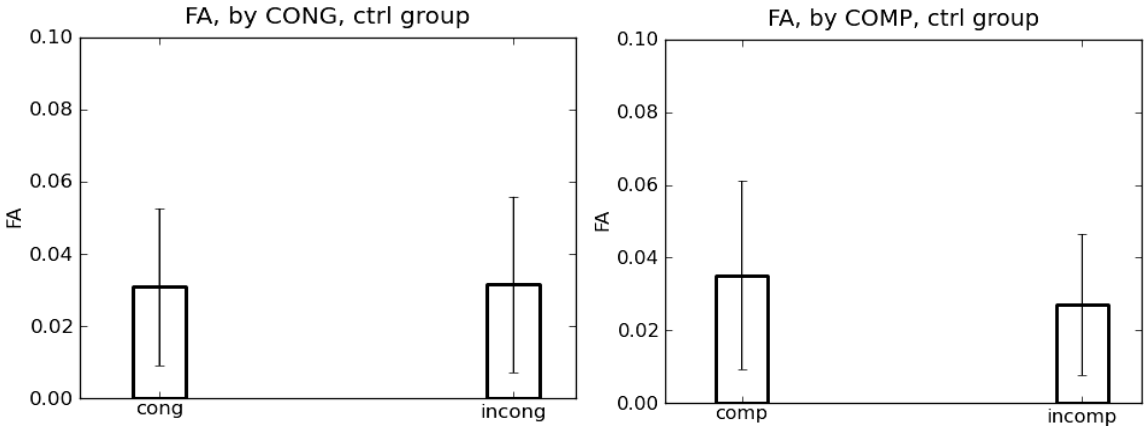


[figure 4-31 : False Alarms by clinical group, bars represent standard errors]

4.3.3.2.b within group

- Control group

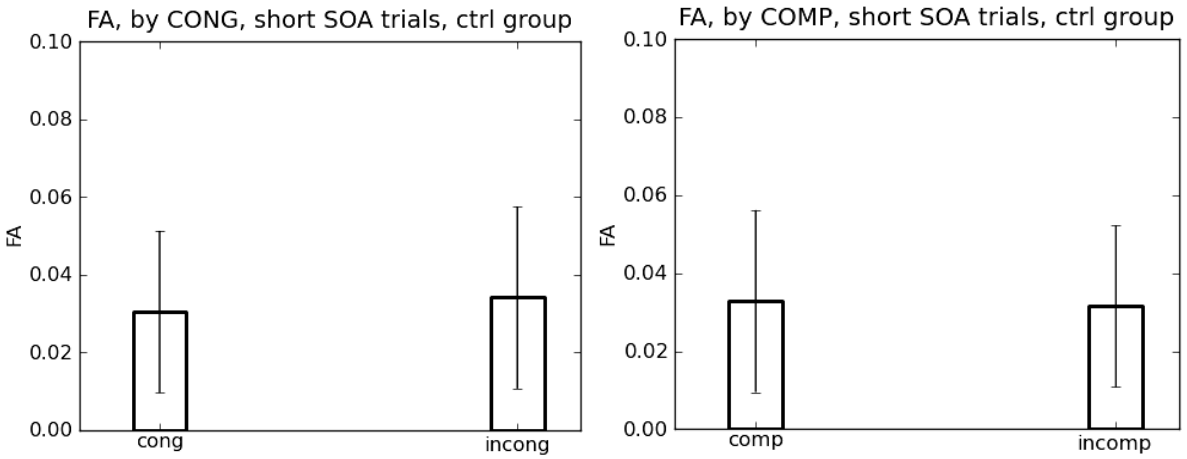
We did not observe any effect of congruency, (p>0.37) nor of compatibility, (p>0.67)



[figure 4-31 : False Alarms in the control group, by congruency (left) and compatibility (right); bars represent standard errors]

Short SOA :

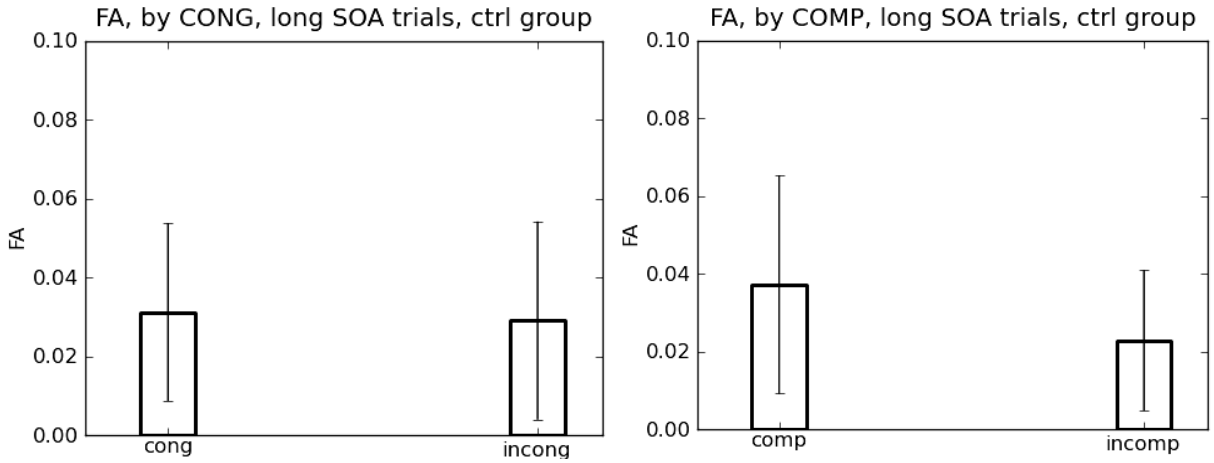
Congruency (Wilcoxon pairwise test, p>0.38); Compatibility (p>0.52)



[figure 4-32 : False Alarms in control group, SHORT SOA trials, by congruency (left) and compatibility (right); bars represent standard errors]

Long SOA :

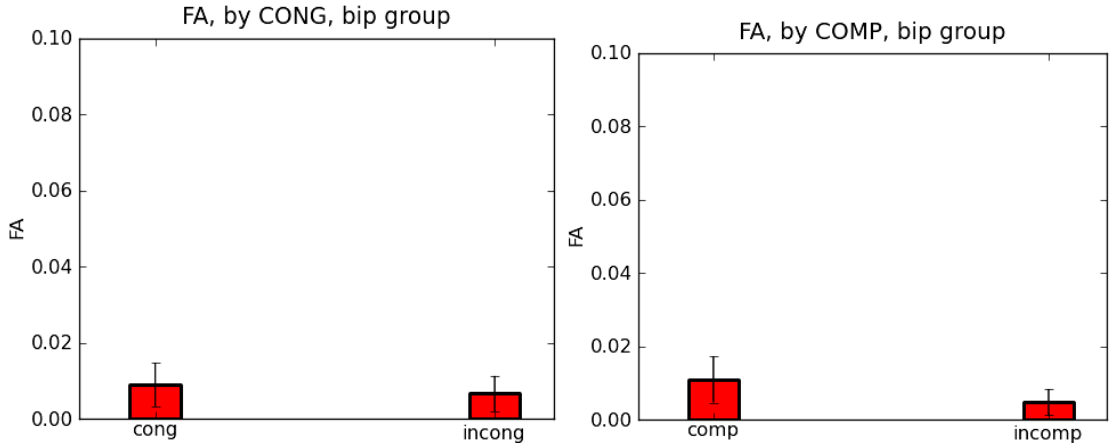
We did not observe any effect of congruency (p> 0.43), nor of compatibility (p>0.78)



[figure 4-33 : False Alarms in the control group, LONG SOA trials, by congruency (left) and compatibility (right); bars represent standard errors]

- Bipolar group :

We observed no significant effect of congruency (p>0.92), nor of compatibility (p<0.5)

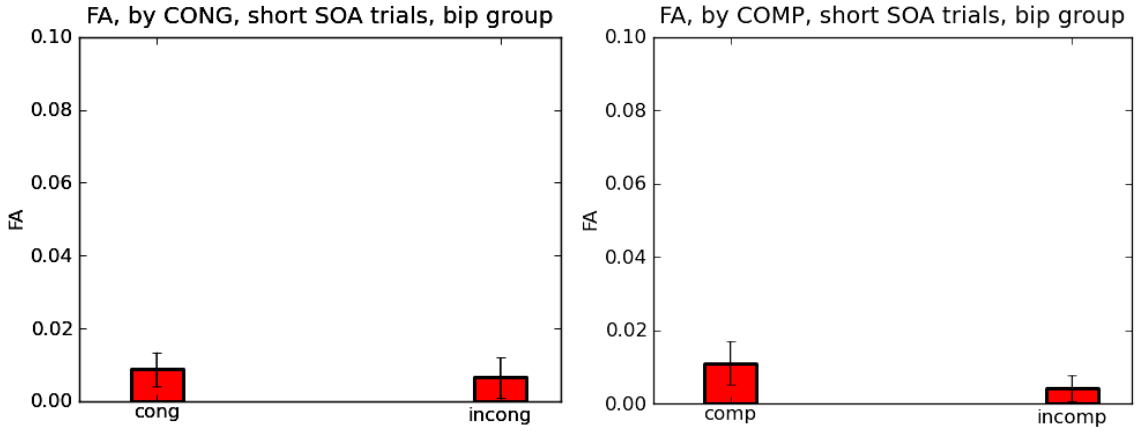


[figure 4-34 : False Alarms in the Bipolar group, by congruency (left) and compatibility (right); bars represent standard errors]

Short SOA :

In short SOA trials, we observed no significant effect of congruency (p>0.58), nor of ompatibility

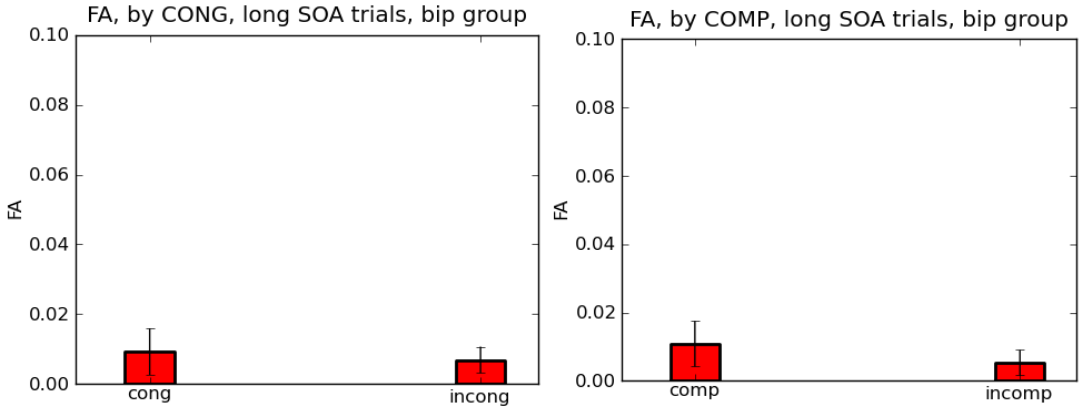
(p>0.11)



[figure 4-35 : False Alarms in the Bipolar group, SHORT SOA trials, by congruency (left) and compatibility (right); bars represent standard errors]

Long SOA :

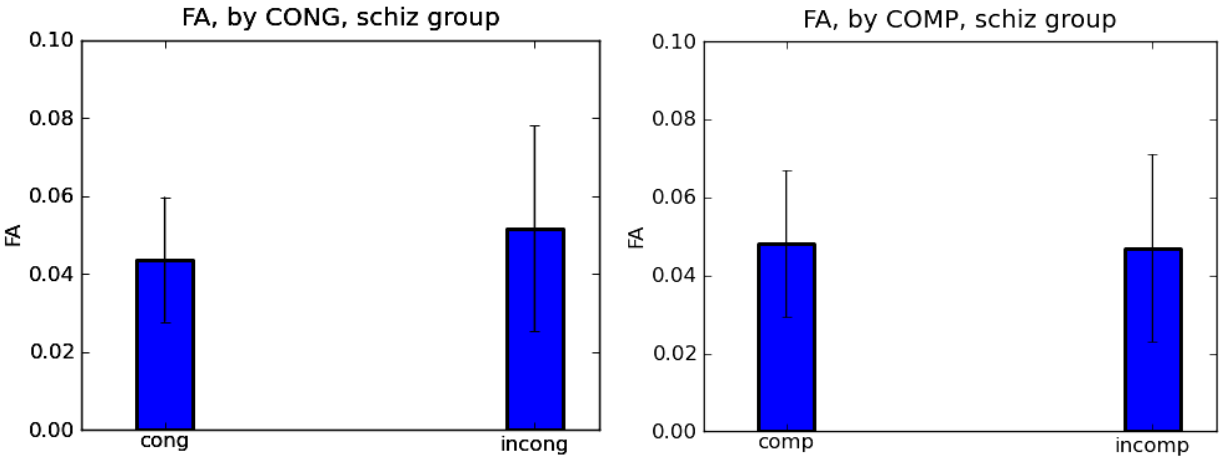
In long SOA trials, we observed no significant effect of congruency (p>0.43), nor of compatibility (p>0.21)



[figure 4-36 : False Alarms in the Bipolar group, LONG SOA trials, by congruency (left) and compatibility (right); bars represent standard errors]

- Schizophrenia group

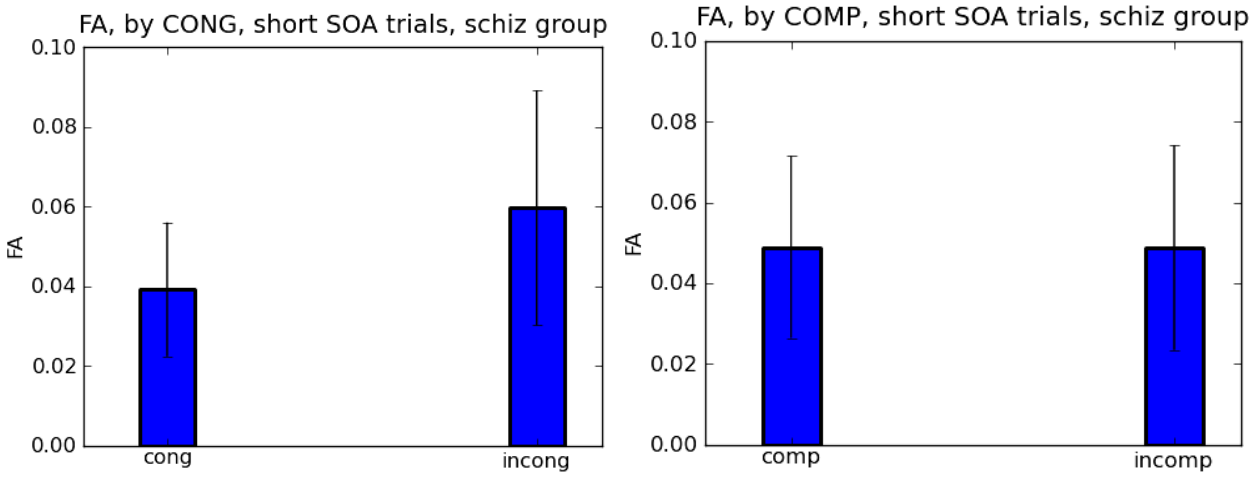
We observed no significant effect of congruency (p>0.90), nor of compatibility (p>0.57)



[figure 4-37 : False Alarms in the schizophrenia group, by congruency (left) and compatibility (right); bars represent standard errors]

Short SOA :

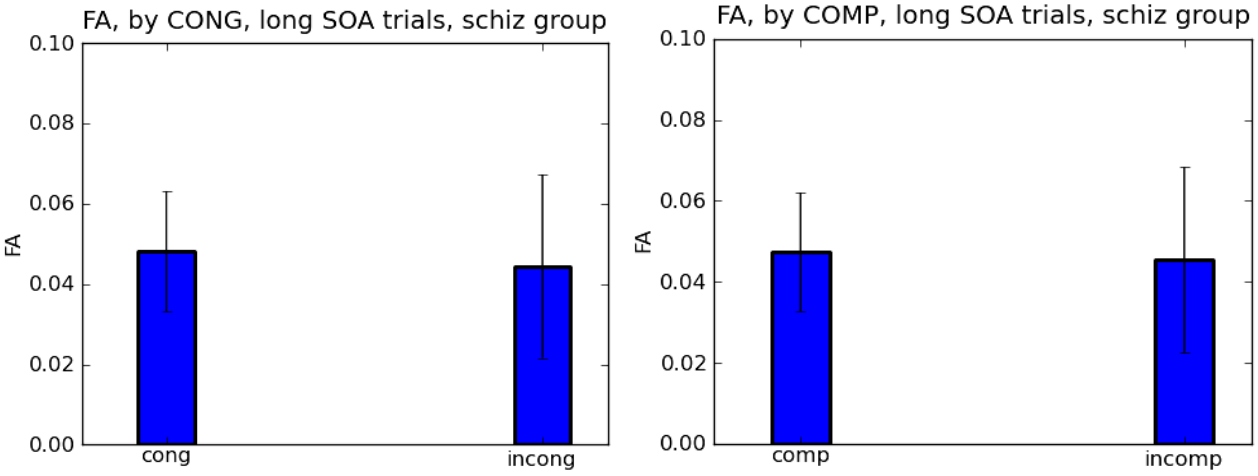
In short SOA trials, we observed no significant effect of congruency (p>0.38), nor of compatibility (p>0.74)



[figure 4-38 : False Alarms in the schizophrenia group, SHORT SOA trials, by congruency (left) and compatibility (right); bars represent standard errors]

Long SOA :

In long SOA trials, we observed no significant effect of congruency (p>0.38), nor of compatibility (p>0.38)



[figure 4-39 : False Alarms in the schizophrenia group, LONG SOA trials, by congruency (left) and compatibility (right); bars represent standard errors]

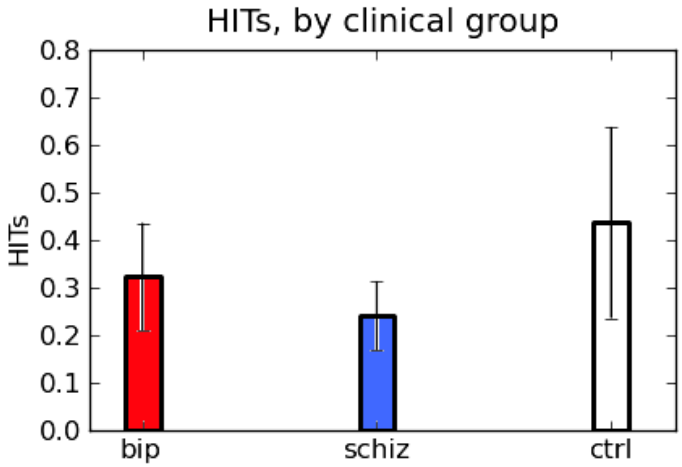
4.3.3.3 Hits (correct second order response after an incorrect first-order response)

SOA	CONG	COMP	mean HIT	SD
long	**cong**	**comp**	0,20	0,37
		incomp	0,29	0,37
	Total cong		0,24	0,37
	incong	**comp**	0,43	0,44
		incomp	0,19	0,29
	Total incong		0,30	0,39
Total long			0,27	0,37
short	**cong**	**comp**	0,25	0,34
		incomp	0,47	0,47
	Total cong		0,37	0,42
	incong	**comp**	0,47	0,42
		incomp	0,40	0,33
	Total incong		0,43	0,38
Total short			0,40	0,40
Total			0,33	0,39

TABLE 6 : HIT (errors successfully detected), GLOBAL PERFORMANCE

4.3.3.3.a Between group

A Kruskal-Wallis rank sum test performed on hits revealed no effect of group ($p<0.87$). Two-by-two comparisons (Bonferroni corrected p-value set at 0.025) revealed no difference between the control and schizophrenia groups ($p>0.87$), no difference between the control and bipolar groups ($p>0.46$) and no difference between the bipolar and schizophrenia groups ($p>0.68$).

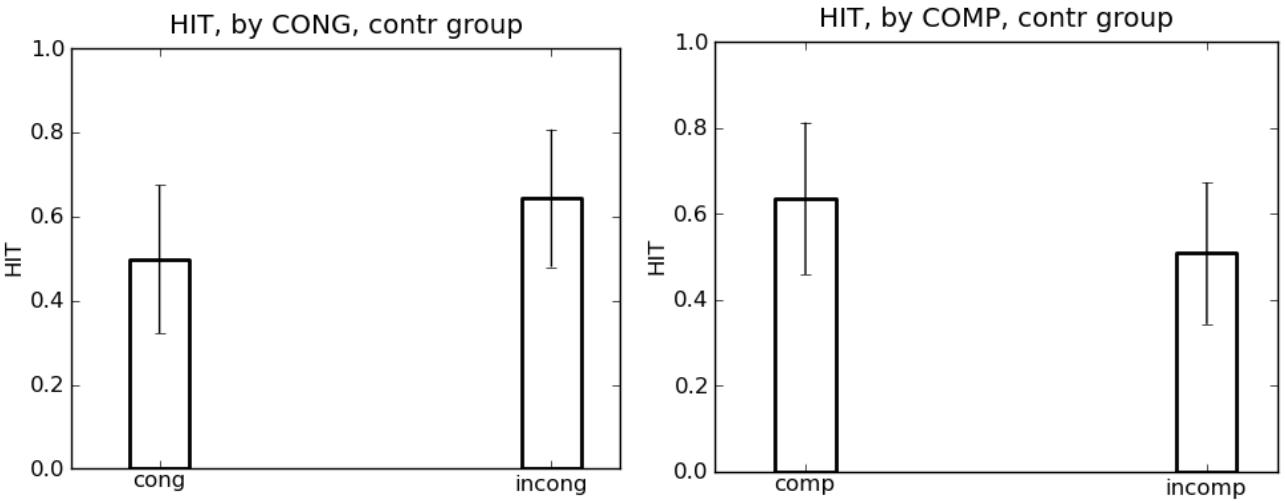


[**figure 4-40**: HITs by clinical group, bars represent standard errors]

4.3.3.3.b within group

- Control group

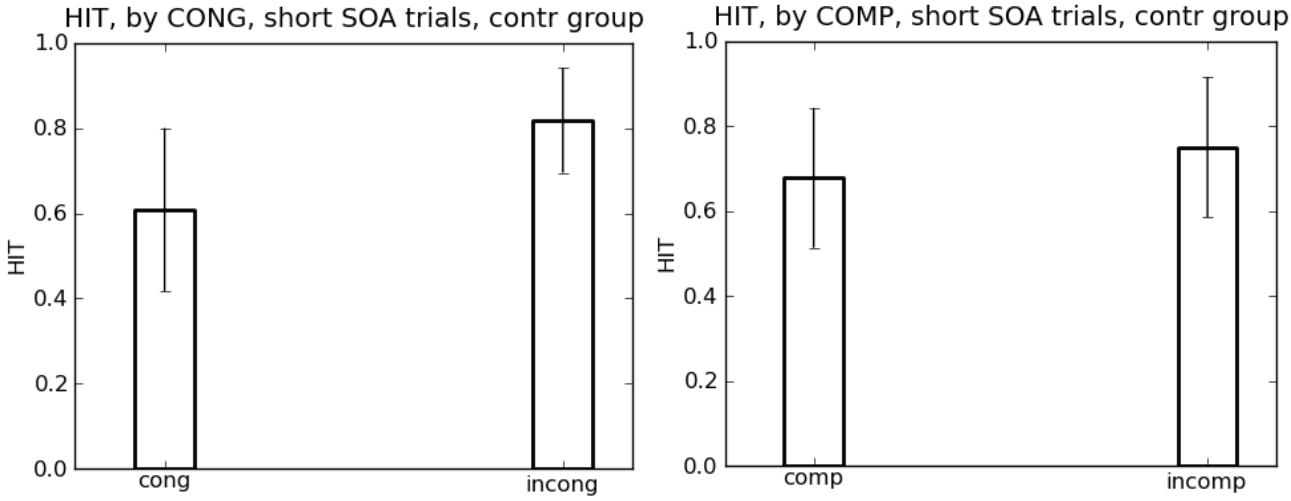
We observed no effect of congruency (p>0.71), nor of compatibility (p>0.12)



[**figure 4-41**: HITs in the control group, by congruency (left) and compatibility (right), bars represent standard errors]

Short SOA :

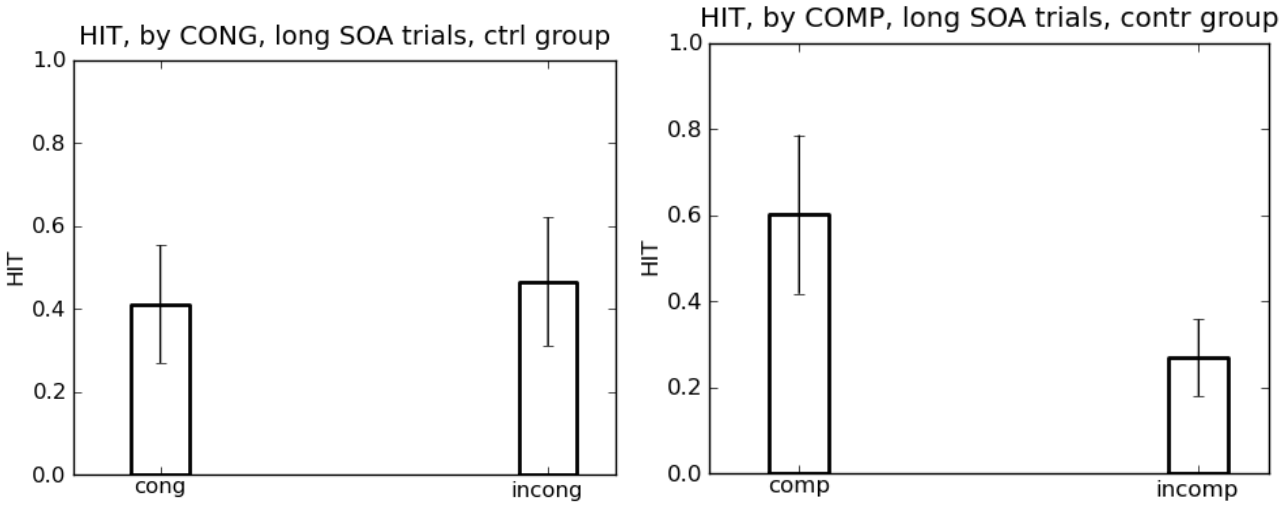
In short SOA trials, we did not observe any significant effect of congruency (p>0.9), nor of compatibility (p>0.37)



[**figure 4-42**: HITs in the control group, SHORT SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

Long SOA :

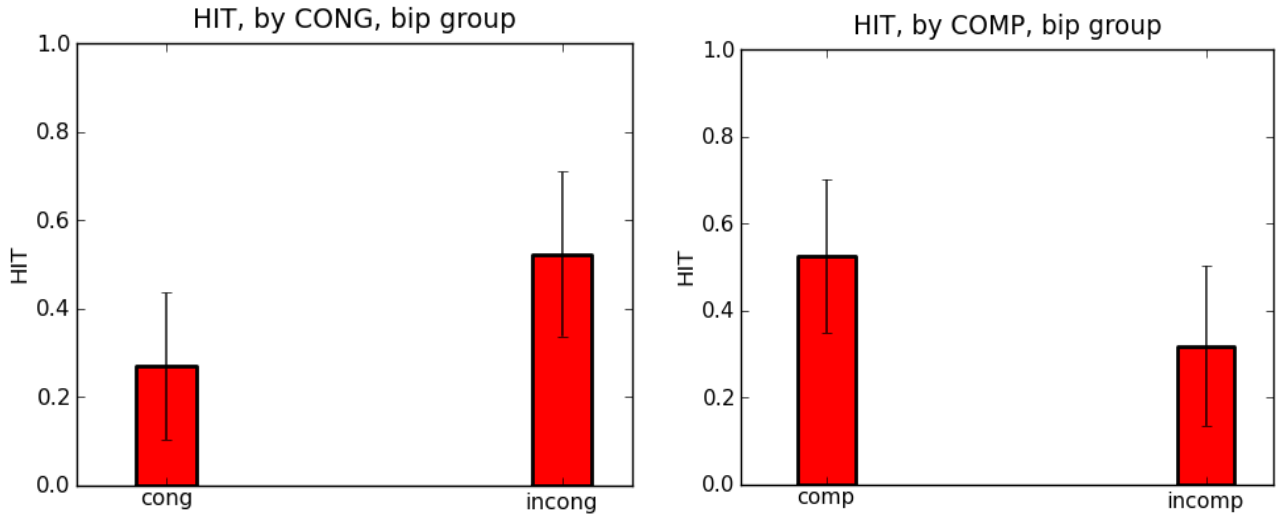
In long SOA trials, we observe no effect of congruency (p>0.9), nor of compatibility (p>0.79);



[**figure 4-43**: HITs in control group, LONG SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

- Bipolar group :

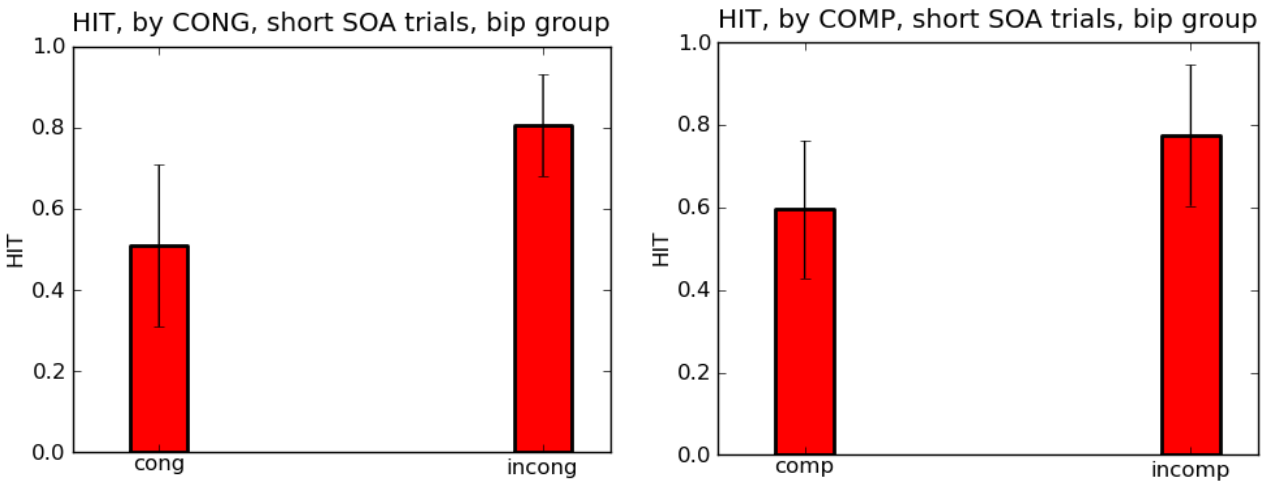
We observed no significant effect of congruency (p>0.87), nor of compatibility (p>0.36);



[**figure 4-44**: HITs in the bipolar group, by congruency (left) and compatibility (right), bars represent standard errors]

Short SOA :

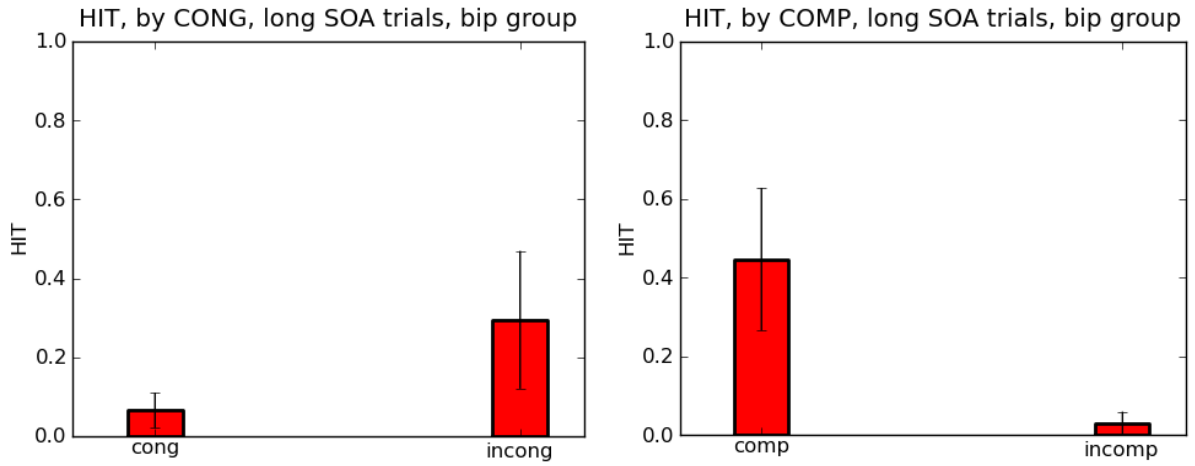
We observed no significant effect of congruency (p>0.9), nor of compatibility (p>0.9)



[**figure 4-45**: HITs in the bipolar group, SHORT SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

Long SOA :

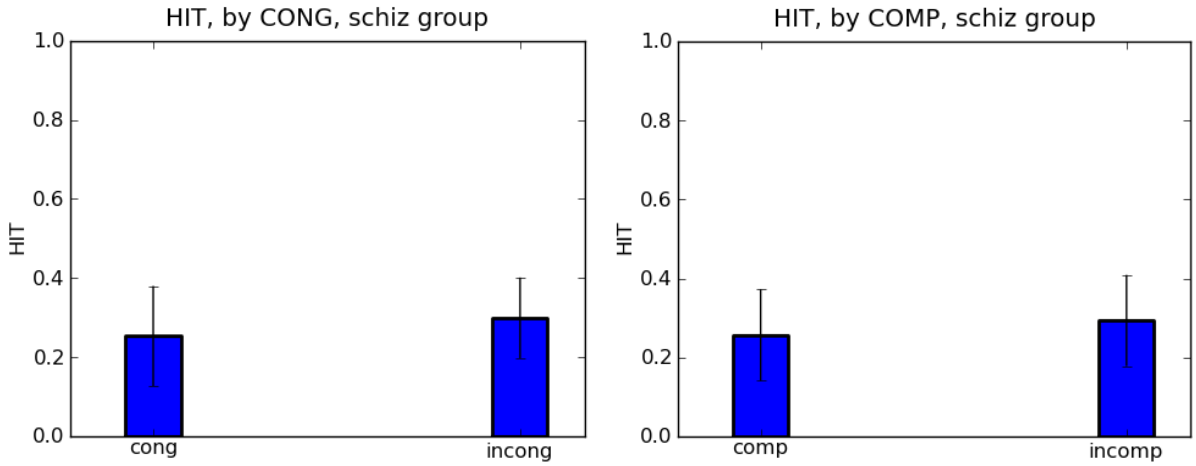
We observed no significant effect of congruency (p>0.37), nor of compatibility (p>0.37)



[**figure 4-46**: HITs in the bipolar group, LONG SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

- Schizophrenia group

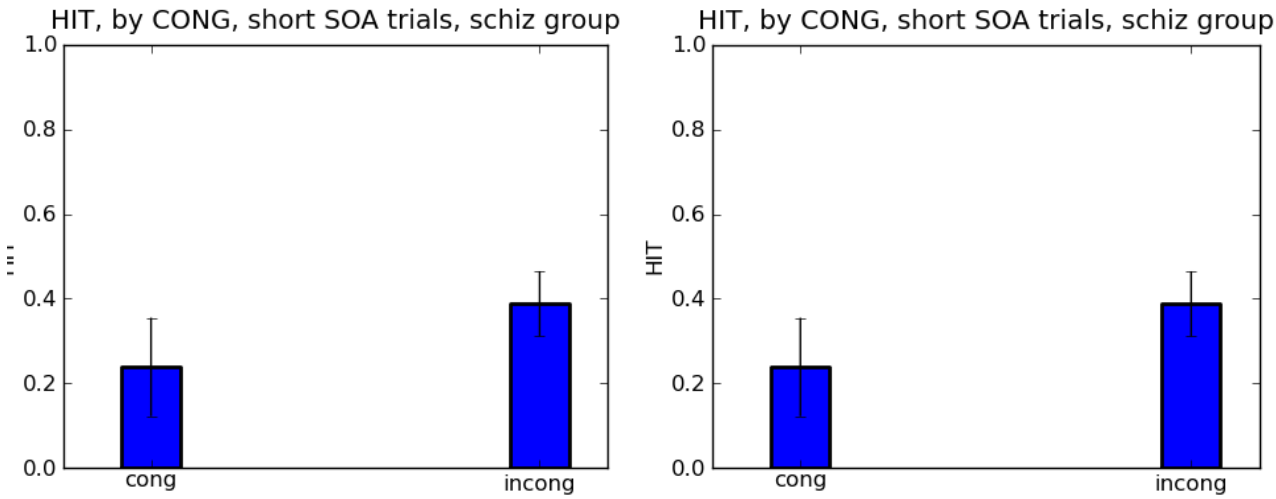
We observed no significant effect of congruency (p>0.84), nor of compatibility (p>0.40)



[**figure 4-47**: HITs in the schizophrenia group, by congruency (left) and compatibility (right), bars represent standard errors]

Short SOA :

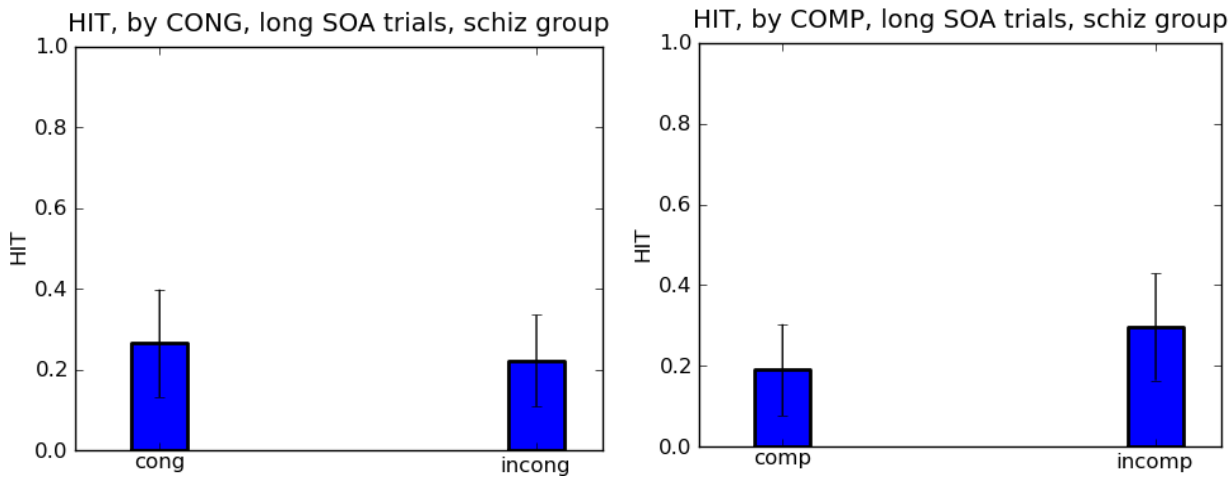
We observed no significant effect of congruency (p>0.9), nor of compatibility (p>0.58)



[**figure 4-48**: HITs in the schizophrenia group, SHORT SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

Long SOA :

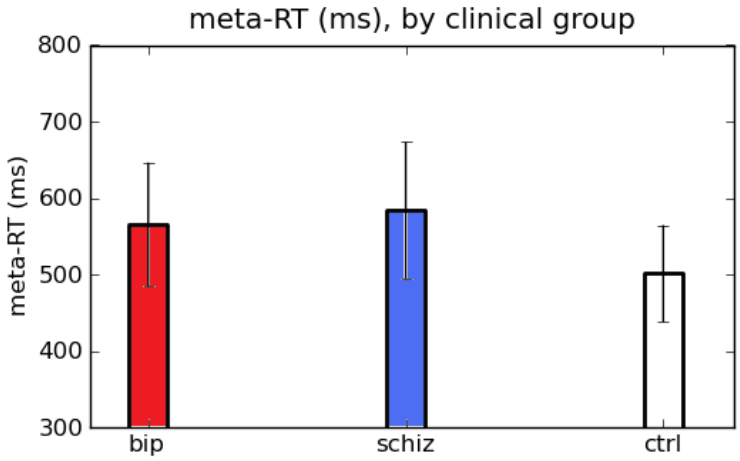
We observed no significant effect of congruency (p>0.36), nor of compatibility (p>0.58)



[**figure 4-49**: HITs in the schizophrenia group, LONG SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

4.3.3.4 Meta Reaction Times :

4.3.3.4.a Between group



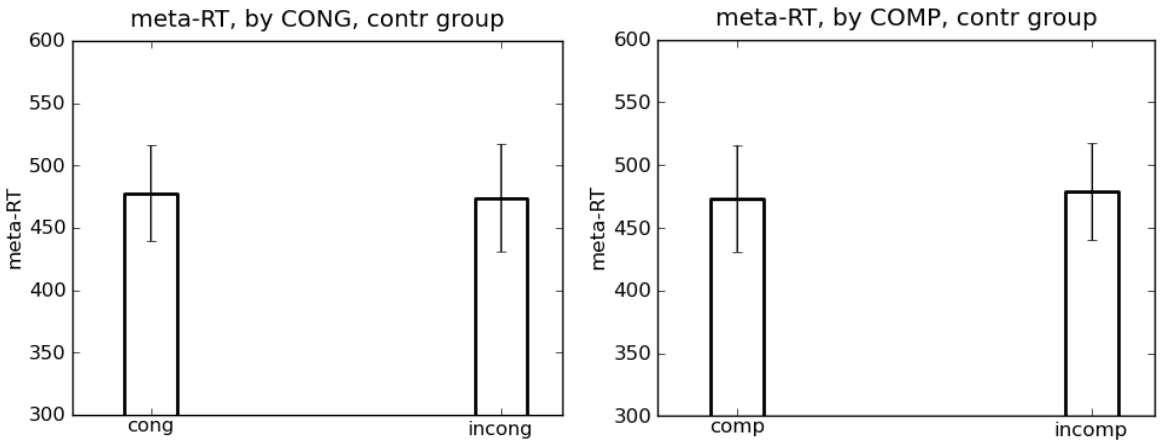
[**figure 4-50**: Meta-Reaction times (time to correctly self-evaluate one's performance) by clinical group, bars represent standard errors]

Kruskal-Wallis rank sum test revealed no significant effect of clinical group (p>0.49). Two-by-two comparisons (Bonferroni corrected p-value set at 0.025) revealed no significant difference between the schizophrenia and bipolar groups (p>0.69), between the schizophrenia and control groups (p>0.24), nor between the bipolar and control groups (p>0.57).

4.3.3.4.b within group

- Control group

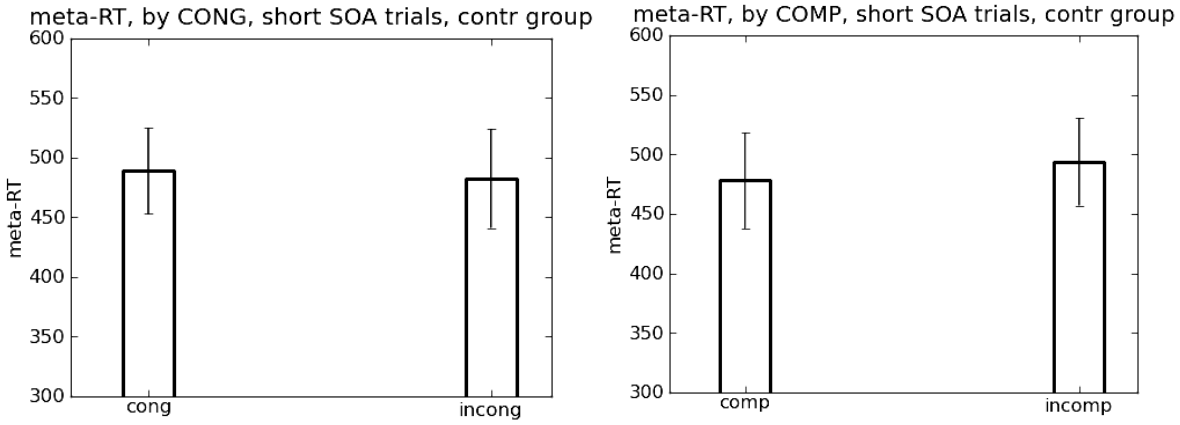
We observed no significant effect of congruency (p>0.93), and no effect of compatibility (p>0.57)



[**figure 4-51**: meta RT in the control group, by congruency (left) and compatibility (right), bars represent standard errors]

<u>Short SOA:</u>

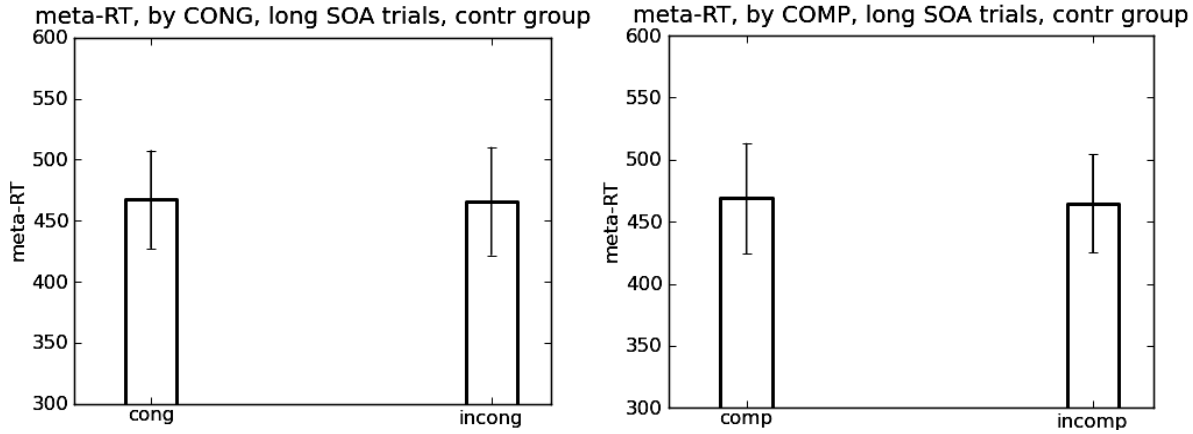
We observed no significant effect of congruency (p>0.81), neither of ompatibility (p>0.57)



[**figure 4-52**: meta RT in the control group, SHORT SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

Long SOA :

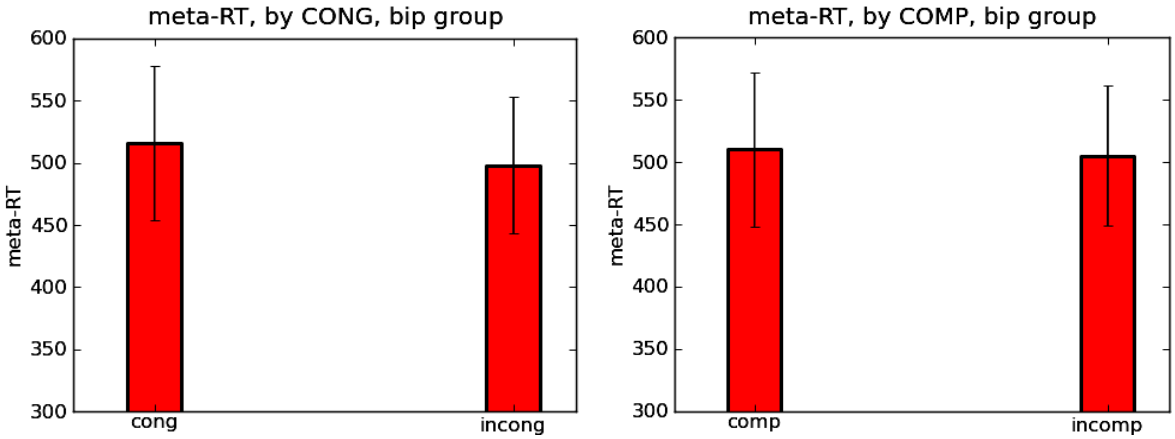
We observed no significant effect of congruency (p>0.57), neither of compatibility (p>0.68)



[**figure 4-53**: meta RT in the control group, LONG SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

- Bipolar group :

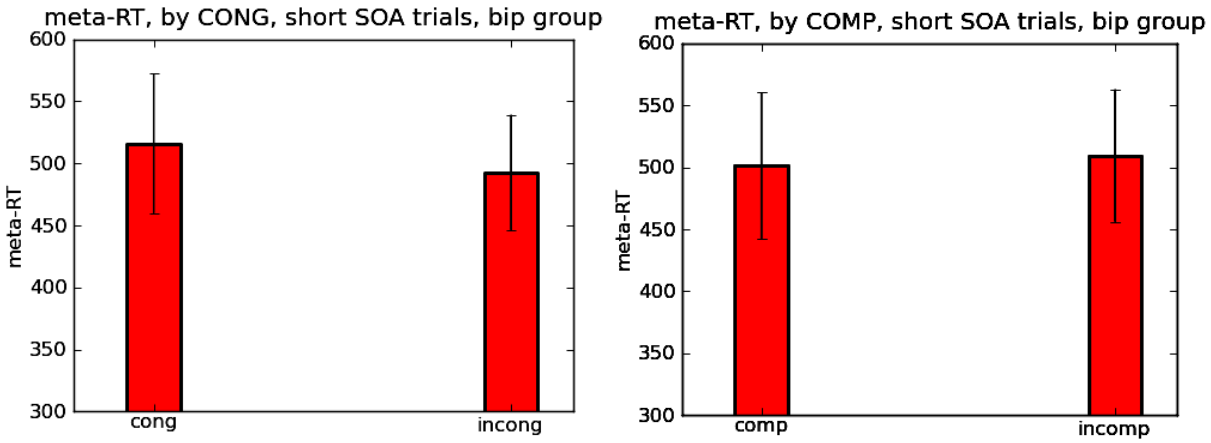
We observed no significant effect of congruency (p>0.21), nor of compatibility (p>0.95)



[**figure 4-54**: meta RT in the bipolar group, by congruency (left) and compatibility (right), bars represent standard errors]

Short SOA:

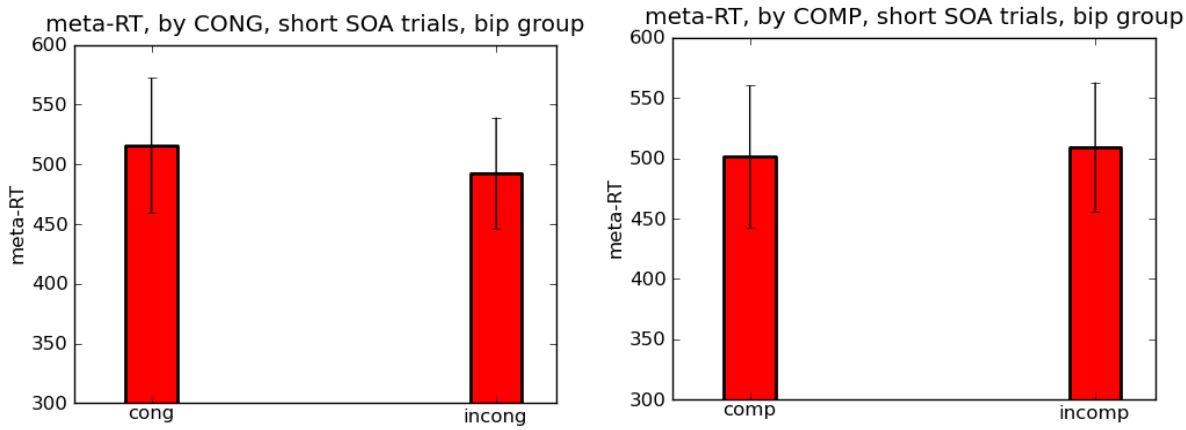
We observed no significant effect of congruency (p>0.21), nor of compatibility (p>0.84)



[**figure 4-55**: meta RT in the bipolar group, SHORT SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

Long SOA :

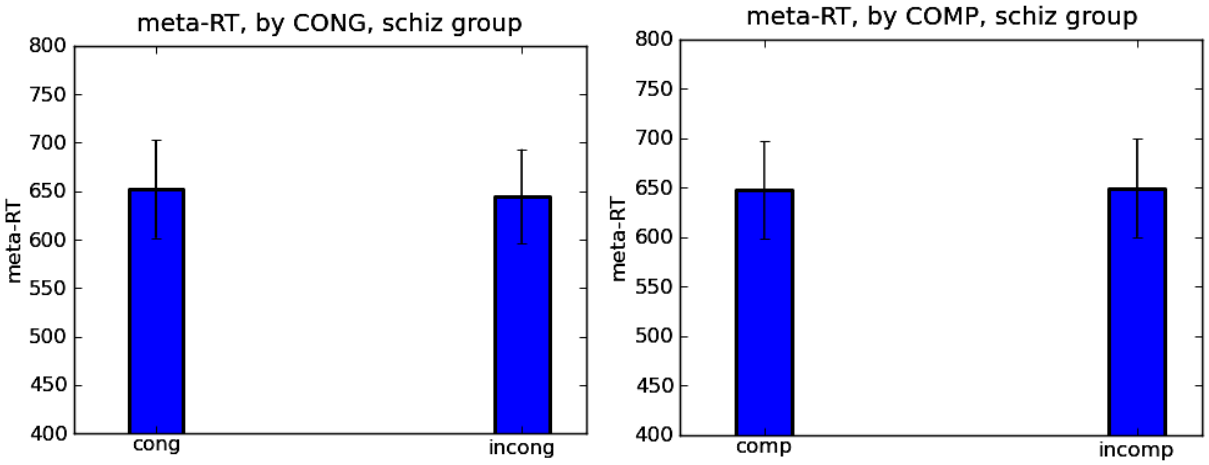
We observed no significant effect of congruency (p>0. 56), nor of ompatibility (p>0.56)



[**figure 4-56**: meta RT in the bipolar group, LONG SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

- Schizophrenia group

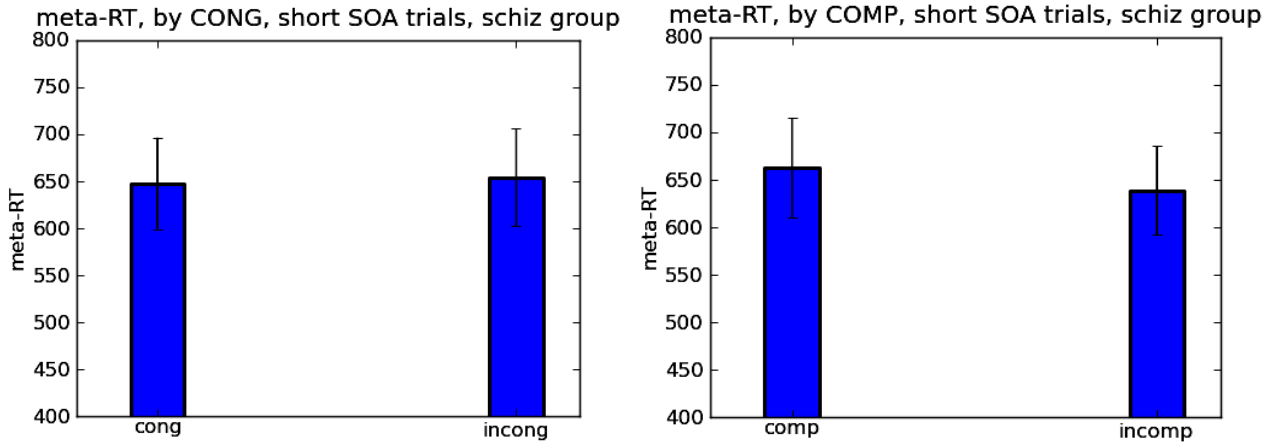
We observed no significant effect of congruency (p>0.46) nor of compatibility (p>0.74)



[**figure 4-57**: meta RT in the schizophrenia group, by congruency (left) and compatibility (right), bars represent standard errors]

Short SOA:

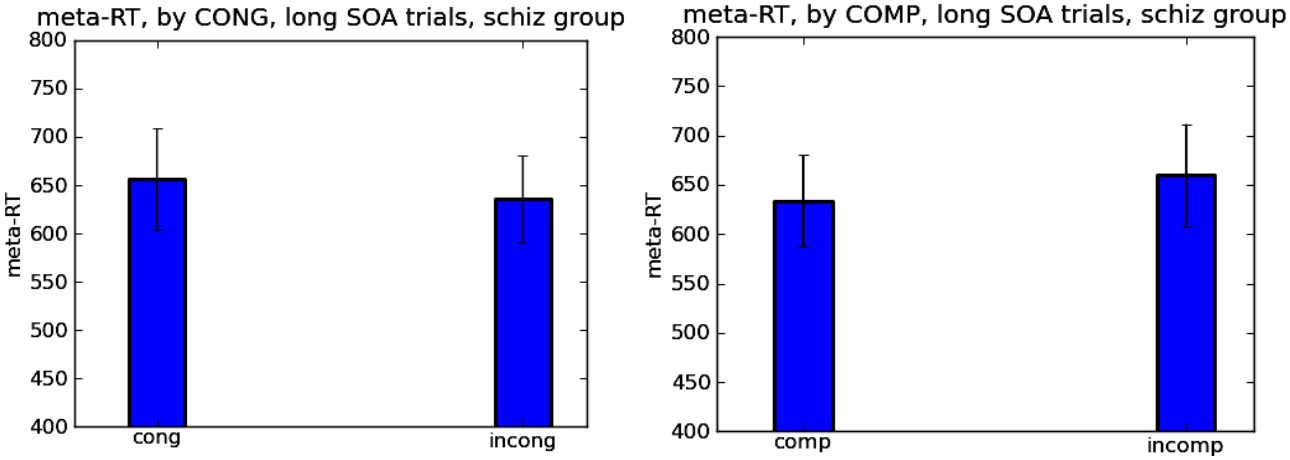
We observed no significant effect of congruency (p>0.84), nor of compatibility (p>0.31)



[**figure 4-58**: meta RT in the schizophrenia group, SHORT SOA trials, by congruency (left) and compatibility (right), bars represent standard errors]

Long SOA :

We observed no significant effect of congruency (p>0.84), nor of compatibility (p>0.25)



[**figure 4-59**: meta RT in the schizophrenia group, by congruency (left) and compatibility (right), bars represent standard errors]

Summary of the results of the metacognitive task :

Between group : We observed significant differences between bipolar and schizophrenia groups for global meta-accuracy and false alarms –the schizophrenia group were generally less accurate in self-evaluation and produced more false alarms than the bipolar group, but did not differ regarding the hits (error detection).

No difference was observed between the different groups regarding meta-reaction times and error detection (hits). The three groups did not differ significantly in the number of unconfident second-order responses reported.

Within group: Significant effects of factors were observed only within the schizophrenia group, namely a significant effect of congruency for meta-accuracy, which was significant on short SOA trials only. No significant effect of compatibility was observed.

Neither the bipolar nor th control group were significantly influenced by any factor in their metacognitive performance.

Splitting the unconfident second-order responses by correct versus incorrect first-order responses revealed no significant difference, in any group (pairwise Wilcoxon, all p-values >0.17), suggesting that all three group were a priori equally unconfident regardless of correct or incorrect responses.

4.3.4 Correlations between basic and metacognitive aspects of the tasks[6]

We assumed that the second–order decision (metacognitive judgment or confidence) is a function of parameters involved in the first-order decision (response selection). Consequently, one should observe some correlations between cognitive and metacognitive measures. Of importance for the present study was the negative correlation between reaction times and confidence (faster RT associated with higher confidence), although the relation between confidence and reaction times is not straightforward (see Pleskac & Busemeyer, 2010, for a review). We thus explored possible correlations between first-order decision measures (accuracy, reaction times) and second-order decision measures (meta-accuracy, FA, hits, meta reaction times, number of unconfident responses), in each clinical group. Finally, given the significant or nearly significant differences of training length despite equal accuracy level, we also tested for the correlations between cognitive and metacognitive measures with training length (measured in numbers of trials necessary to reach the requested performance). Given the few subjects per group, we also report the trends (star indicate the correlations which survive to a conservative Bonferroni threshold correction).

- **Control group :** we test for the following Pearson correlations :

- **Accuracy and meta-accuracy (p<0.007)**
- Accuracy and meta reaction times (p>0.98), see figure 4-60
- Reaction times and meta-accuracy (p>0.14)
- **Reaction times and Meta reaction times (p<0.026),** see figure 4-63
- **Reaction times and False alarms (p<0.033),** see figure 4-61
- Reaction times and Hits (p>0.12)
- Training and accuracy (p>0.92)
- Training and Reaction times (p>0.22), see figure 4-64
- Training and meta-accuracy (p>0.93)
- Training and False Alarms (p>0.95)

[6] From a conservative statistical perspective, it could well be argued that as 15 correlations are being carried out for each group the significance level used should take this into account with the .05 level being divided in Bonferroni fashion by 15. The 3 correlations that survive are indicated by an *. However, with 15 comparisons being made the mean number that would exceed the .05 level by chance is 0.75 and the Binomial probability of at least 10 being found by chance , as occurs in the schizophrenia group is exceedingly small. And even if the 3 that are accepted under Bonferroni are excluded then the mean probability for the number exceeding chance out of 12 is 0.6 and again the Binomial probability of an extra 7 occuring by chance is again exceedingly small. Thus application of the Bonferroni correction principle to this data set is probably too conservative. However we do consider the present results as trends to be confirmed by the inclusion of more patients.

- Training and Hits (p>0.48),), see figure 4-62
- Trainings and meta-Reaction times (p>0.38)
- **Unconfident responses and Reaction times (p<0.034)**
- Unconfident responses and hits (p>0.91)
- Unconfident responses and training length (p>0.95)

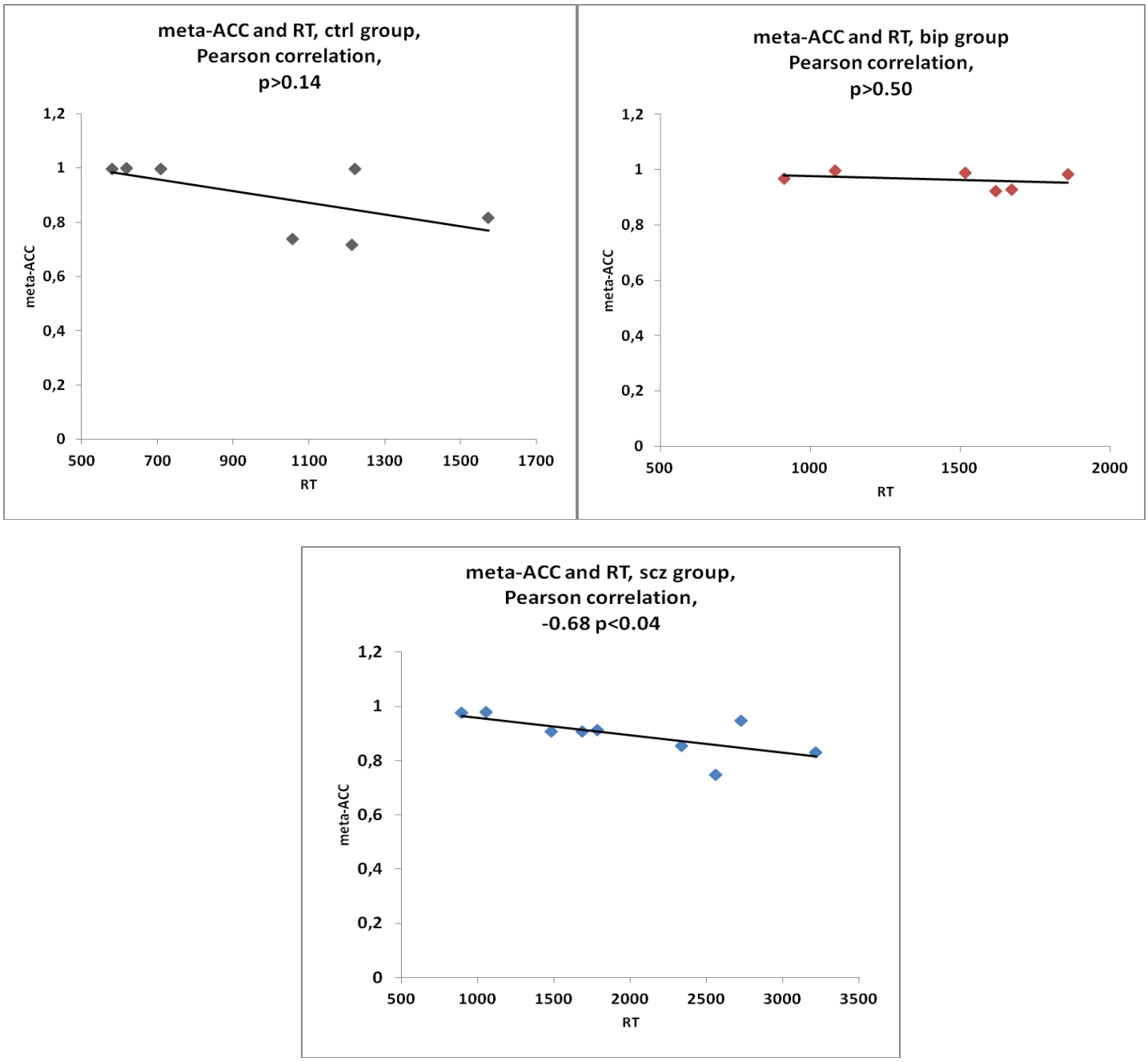
- Bipolar group: we test for the following Pearson correlations:

- **Accuracy and meta-accuracy (p<0.009)**
- Accuracy and meta reaction times (p>0.20), see figure 4-60
- Reaction times and meta-accuracy (p>0.50)
- Reaction times and Meta reaction times (p>0.96), see figure 4-63
- Reaction times and False alarms (p>0.40), see figure 4-61
- Reaction times and Hits (p>0.20),), see figure 4-62
- Training and accuracy (p>0.57)
- Training and Reaction times (p>0.35), see figure 4-64
- Training and meta-accuracy (p>0.53)
- Training and False Alarms (p>0.62)
- Training and Hits (p>0.48)
- Trainings and meta-Reaction times (p>0.20)
- Unconfident responses and Reaction times (p>0.79)
- Unconfident responses and hits(p>0.45)
- Unconfident responses and training length (p>0.74)

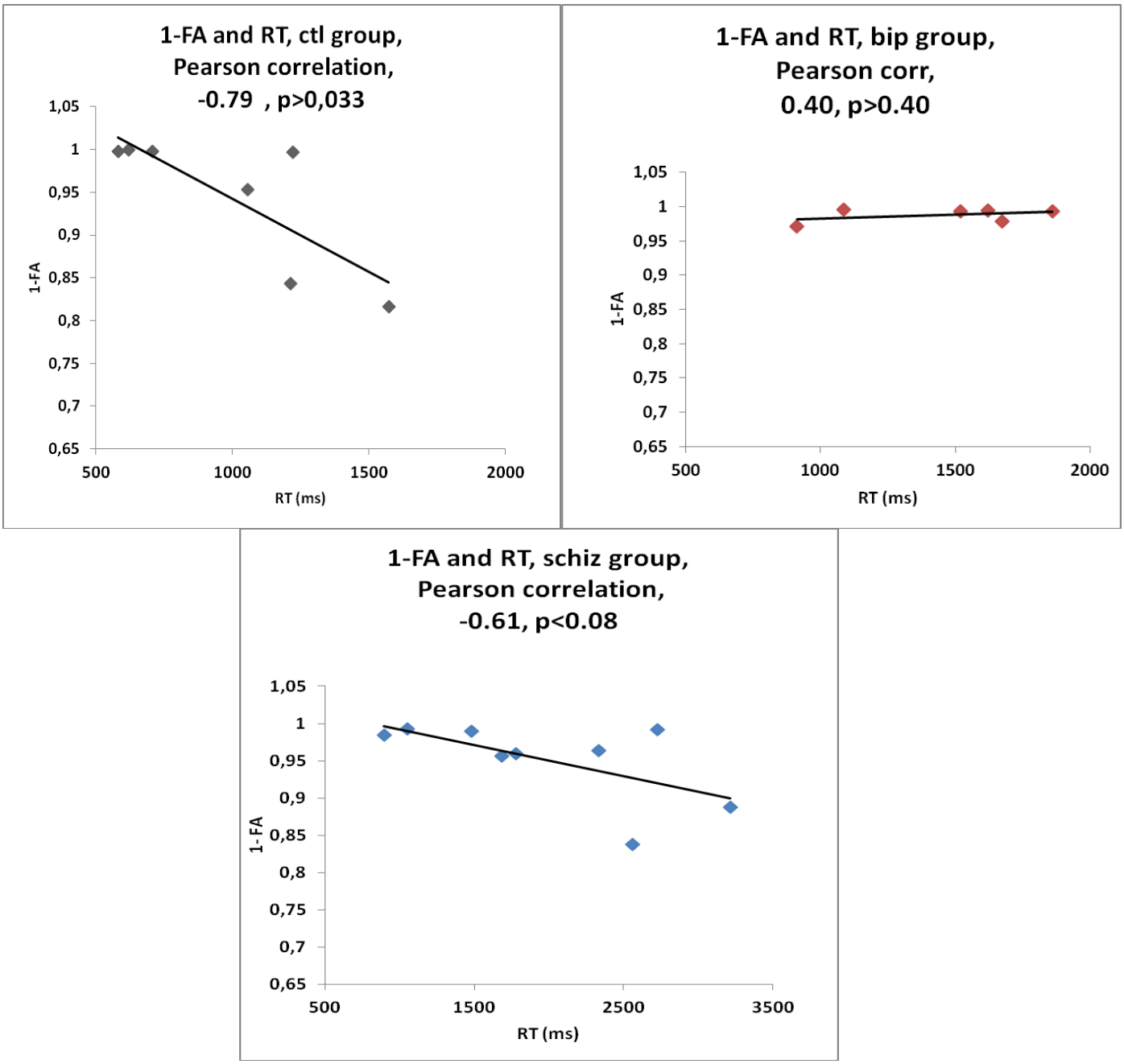
- Schizophrenia group: we test for the following Pearson correlations:

- **Accuracy and meta-accuracy (p<0.008)**
- Accuracy and meta reaction times (p>0.74), see figure 4-60
- **Reaction times and meta-accuracy (p<0.04),** see figure 4-63
- Reaction times and Meta reaction times (p>0.12)
- **Reaction times and False alarms (p<0.08),** see figure 4-61
- **Reaction times and Hits (p<0.036**), see figure 4-62
- Training and accuracy (p>0.33)
- **Training and Reaction times (p<0.008),** see figure 4-64
- **Training and meta-accuracy (p<0.002)*,** see figure 4-65

- **Training and False Alarms (p<0.003)*** see figure 4-65
- **Training and Hits (p<0.03),** see figure 4-65
- Training and meta-Reaction times (p>0.22),
- **Unconfident responses and Reaction times (p<0.05)**
- Unconfident responses and hits(p>0.33)
- **Unconfident responses and training length (p>0.001)*** see figure 4-65



[**figure 4-60**: Pearson correlation between meta-Accuracy and Reaction Times In control (top left), bipolar (top right) and schizophrenia (bottom) groups, **significant only in schizophrenia group**.]



[**figure 4-61**: Pearson correlation between 1-False Alarms and Reaction Times,
In control (top left), bipolar (top right) and schizophrenia (bottom) groups,
significant in control group, nearly significant in schizophrenia group, not significant in bipolar group.]

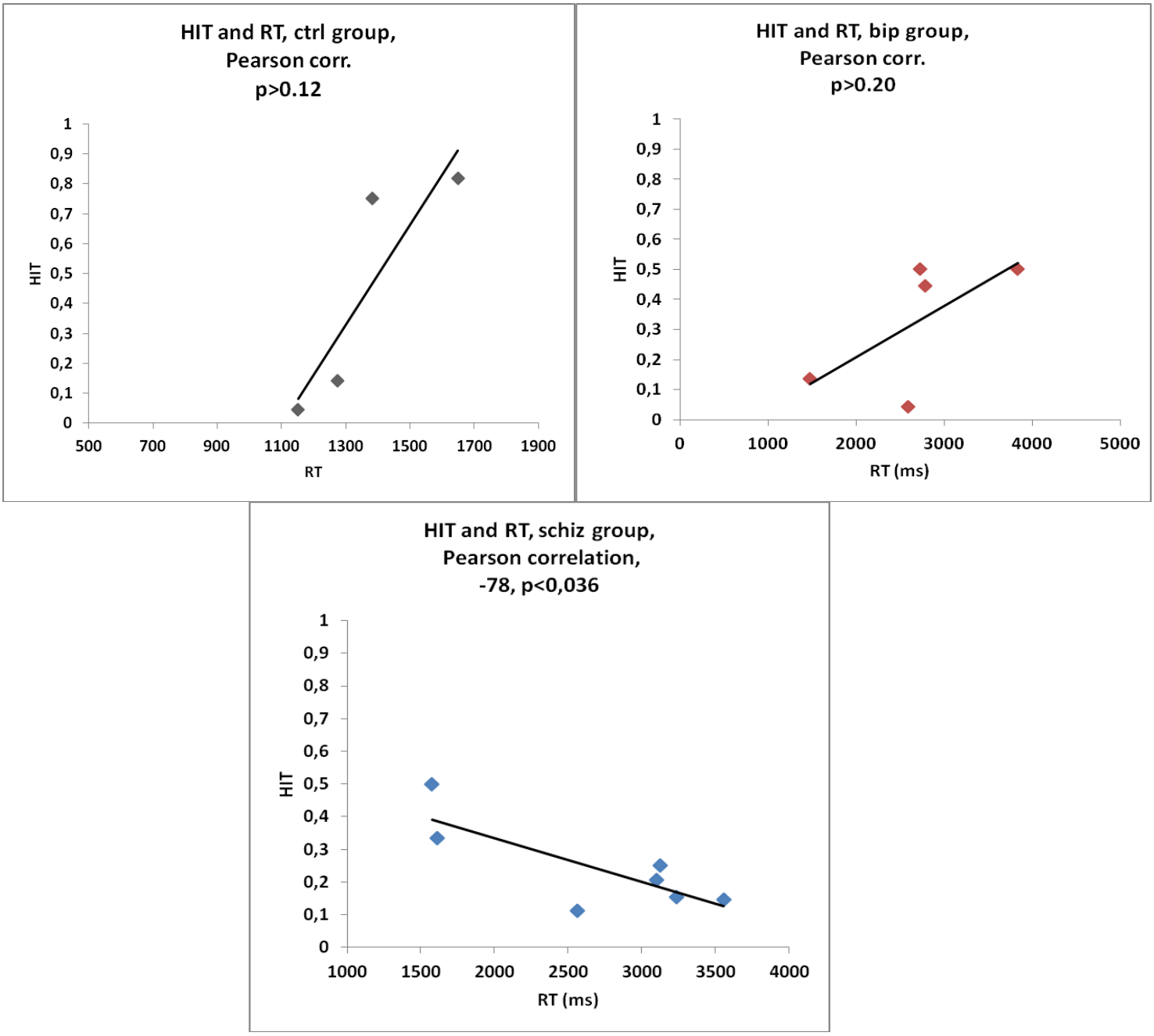


figure 4-62: Pearson correlation between HITs and Reaction Times, in each clinical group,
In control (top left), bipolar (top right) and schizophrenia (bottom) groups,
not significant in control and bipolar groups, significant and **negative in schizophrenia group**]

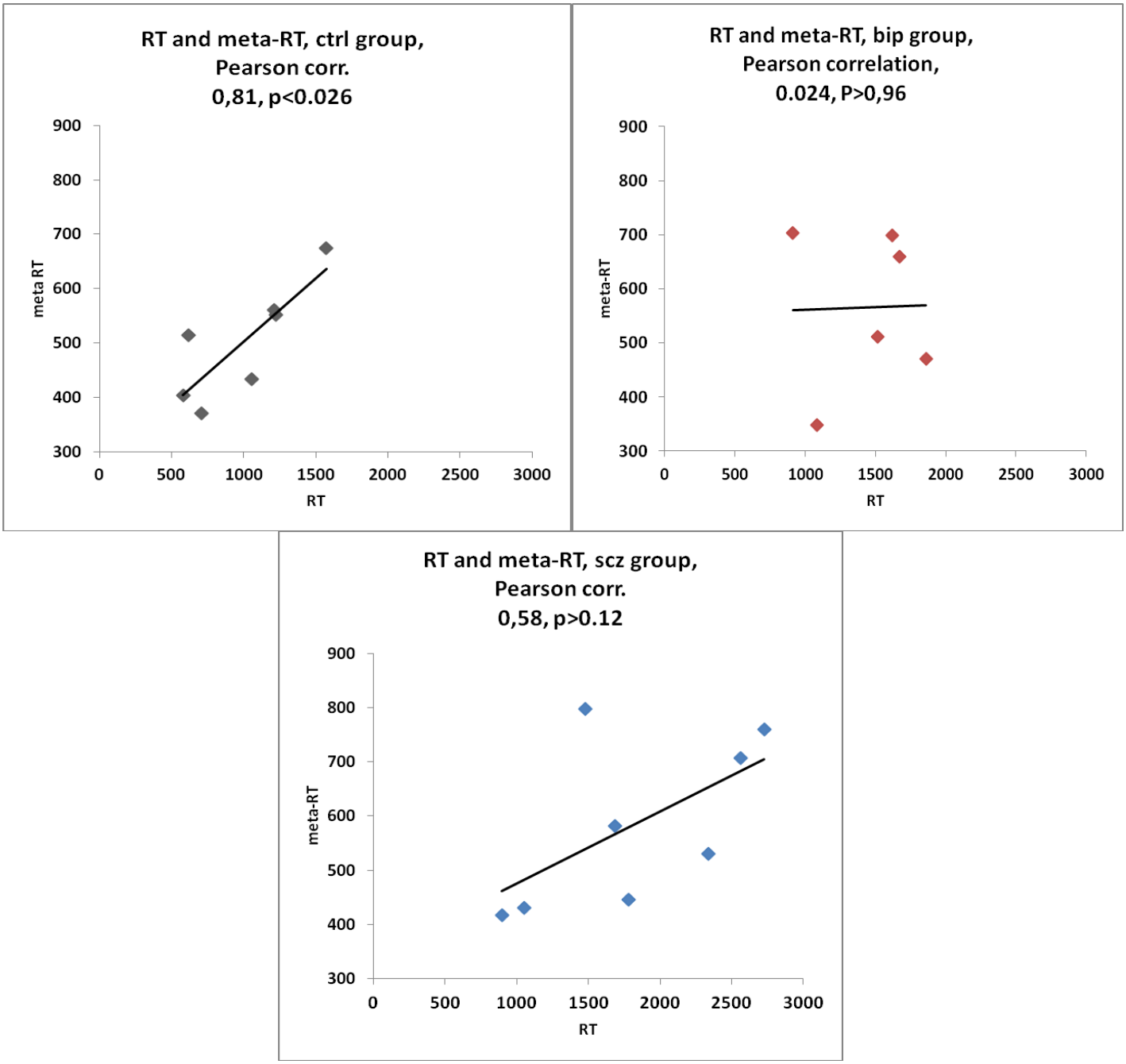


figure 4-63: Pearson correlation between **meta-Reaction Times** and **Reaction Times**
by clinical group,
In control (top left), bipolar (top right) and schizophrenia (bottom) groups,
positive and **significant in control group only**]

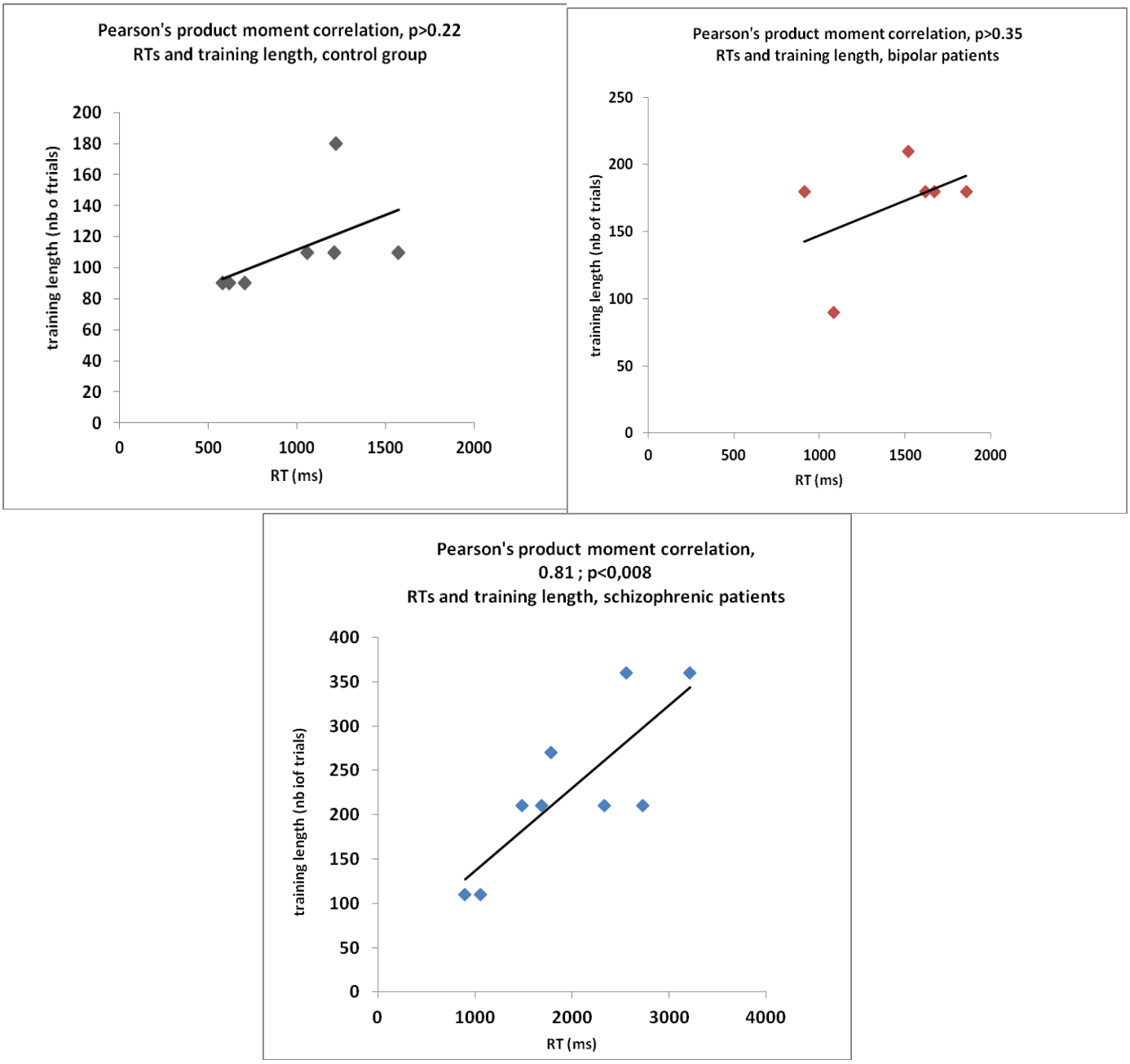


figure 4-64: Pearson correlation between **training length** and **Reaction Times** by clinical group, In control (top left), bipolar (top right) and schizophrenia (bottom) groups, **Significant in the schizophrenia group only]**

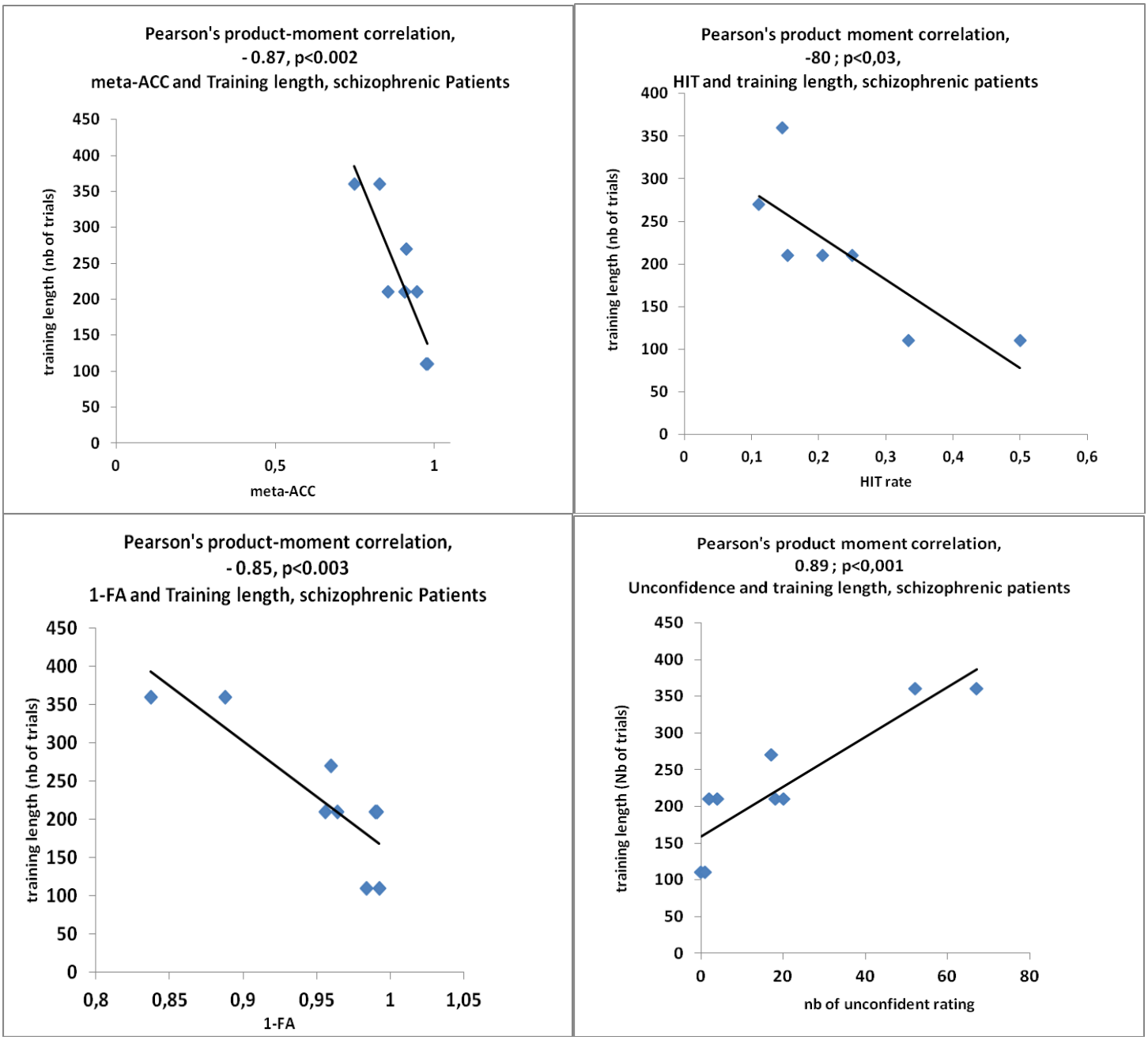
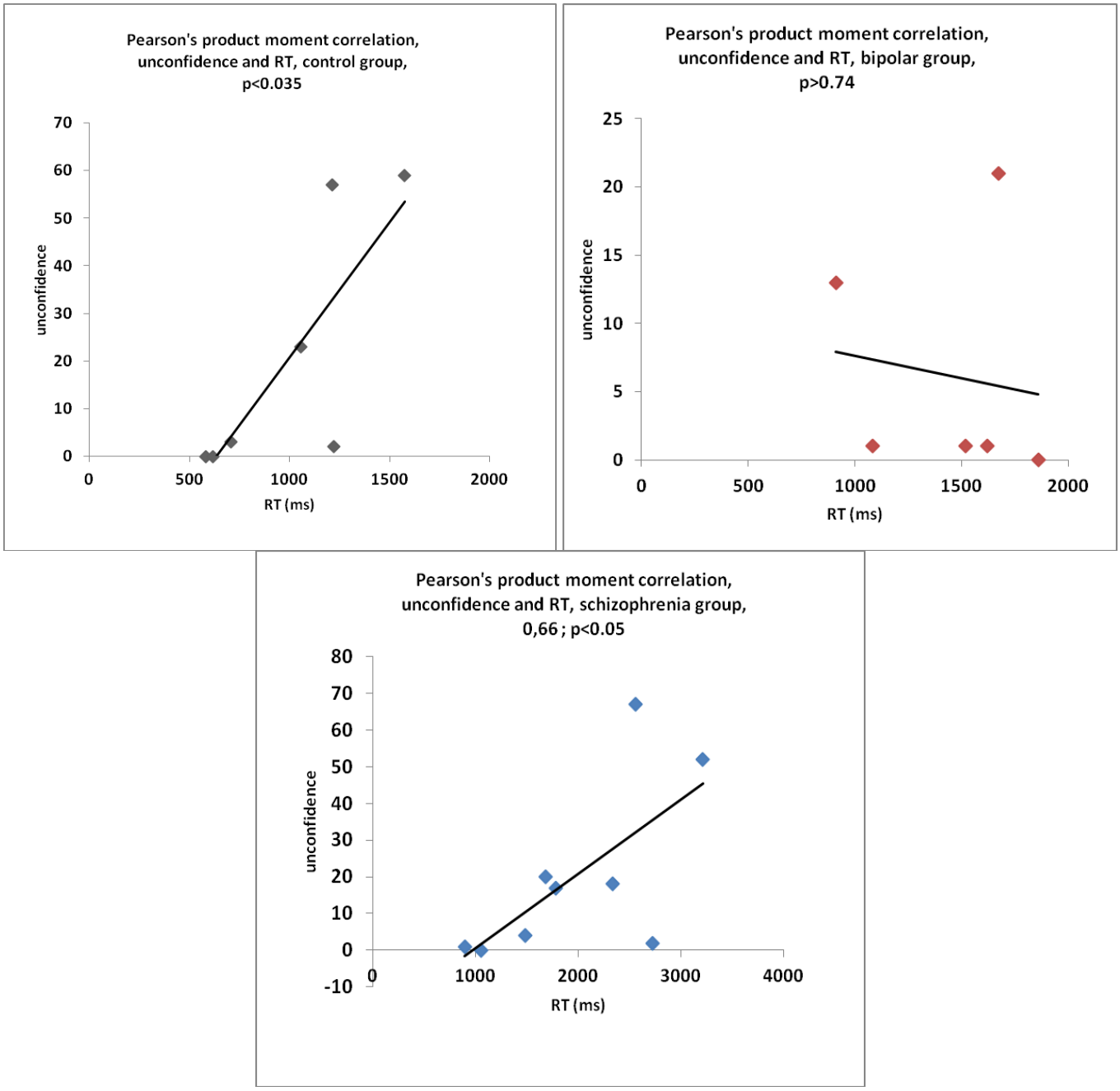


figure 4-65: Pearson correlation between Training length and **metacognitive measures (meta-ACC: top left ; HIT: top right ; unconfident responses: bottom left ; FA : bottom right)** In the schizophrenia group]



[figure 4-66: Pearson correlation between number of Unconfident responses and RT, by clinical group **(control group: top left ; bipolar : top right ; schizophrenia : bottom)**]

Summary of the correlations between cognitive and metacognitive performance:

Control group : The control group showed significant correlations, in particular the expected negative correlation between reaction times and confidence related measures, namely false alarms and unconfidence. One observed also a positive correlation between reaction times and meta reaction times. No significant correlations at all was observed between the cognitive or metacognitive performances and their respective training length.

Bipolar group : First-order and second-order decisions were generally not correlated. The exception was the correlation between accuracy and meta-accuracy, which is highly correlated in all three groups. We observed no significant correlation at all between the cognitive and metacognitive performance in the bipolar group. In particular, we did not observe any correlation between reaction times and confident-related measures.

Bipolar patients showed no significant correlation at all between the cognitive or metacognitive performances and their own training length.

Schizophrenia group: The schizophrenia group showed the most numerous and significant correlations between its cognitive and metacognitive performances. In particular, it showed significant correlations between reaction times and metacognitive and confidence-related measures (False alarms, hits, meta-accuracy, unconfidence). It also displayed significant correlations between training length and metacognitive and confidence-related measures. These correlations include a lower global meta-accuracy, errors less frequently detected , more frequent false alarms/unconfident responses associated with longer training.

They showed a significant correlation between number of unconfident second-order judgments and reaction times.

4.4 Discussion:

An appealing aspect of the results is that the control group was not influenced by any factor of interest at all. In our previous studies, the subjects were submitted to more stringent constraints regarding the required performance (90% correct, less than 1000 ms). In the present study, the control group was submitted to the same requisites as the patient groups regarding their performance. Although this has allowed us to correlate the training length with the metacognitive performance, in the same vein it has prevented the control group from reaching the fastest reaction times possible.

4.4.1 Cognitive aspects

A first aspect of that study was the demonstration of central impairment of schizophrenia group compared with the bipolar and control group. As expected, the schizophrenia group displayed a globally impaired performance in the first-order cognitive control task, compared to the control group. That manifested itself through the fact that, despite a homogenous accuracy level among groups, the schizophrenia group turned out to be significantly slower than the control group, and especially needed significantly more training to reach the requested accuracy level. There were trends in the same direction for the comparison between the schizophrenia and bipolar groups.

As expected a well, the schizophrenia group was significantly influenced by the cognitive control load. The patients with schizophrenia produced more errors in incongruent trials. An open question is the issue of why we observed a significant congruency effect in short SOA trials, and not long SOA trials. This suggests an interaction between access to consciousness and cognitive control load, but deciphering the nature of this interaction would require a quantitative model – as previously discussed in our neuroimaging study.

We were uncertain about the possible priming effects in the schizophrenia group. We actually observed no priming effect at all (*id est* no compatibility effect) in any group. However these effects are small and we had very few subjects compared with our previous experiments, which involved at least 20 subjects (who belonged to an a priori homogenous population). Previous studies of non conscious priming in schizophrenia are few, but priming effects in schizophrenia have been reported in SHORT SOA trials only, although with a different paradigm (Dehaene et al., 2003).

4.4.2 Metacognitive aspects

Regarding metacognition, we stet demonstrated that the schizophrenia group was impaired compared to other groups (cf. hypothesis 2). In that respect, it may be surprising that schizophrenia group did not differ significantly from the control group regarding meta-accuracy. However that may be due to the variability within both the control and schizophrenia group (cf. Appendices 4 and 5).

Our data are nevertheless consistent with the hypothesis of a metacognitive impairment in schizophrenia patients compared with bipolar patients, since analyses of both meta-accuracy and false alarms (Kruskal-Wallis rank sum test) revealed significant differences between the bipolar and schizophrenia groups : the bipolar group performed generally better than schizophrenia group, and made less false alarms. Importantly, it seems that schizophrenia group did not significantly differ from control nor bipolar groups regarding error detection (hits) – although that might also be due to the variability or the small size of our samples,

In any case, despite the lack of significant difference in unconfident responding per se between the bipolar and schizophrenia groups, schizophrenia patients differed significantly from bipolar patients on false alarms. Given the criteria we used for false alarms, it turns out that schizophrenia patients are generally less confident or accurate than bipolar. The bipolar group, in contrast to the schizophrenia group, was on the contrary very confident and accurate in self-evaluating their own performance. This constitutes a major difference between the bipolar and schizophrenia groups.

A second hypothesis regarding metacognition was the influence of cognitive control load (that is to say congruency factor) on metacognitive performance of schizophrenia patients. We indeed observed a significant effect of congruency in schizophrenia group only, with an impaired meta-accuracy in incongruent trials. Interestingly, this effects perfectly mirrors what we observed for accuracy : it turned out significant in short but not long SOA trials. However, we did not observe a lower confidence in incongruent trials. Neither the false alarms, nor the unconfidence responses was influenced by the congruency factor.

Finally, we did not observe that schizophrenia patients were more or less confident after correct than after incorrect first-order responses. But there also was a relatively high variability among patients

with respect to this issue.

4.4.3 Correlations between cognitive and metacognitive decisions

Although these correlations must be considered with caution and remain to be confirmed, setting up such correlations between cognitive and metacognitive measures was useful and relevant for different purposes.

First, regarding the common correlation between reaction times and confidence, the control group showed significant correlations between unconfident responses and reaction times (figure 4-66) and between false alarms (including both incorrect and unconfident second-order decision about a correct first-order decision). It consequently behaved as expected.

Second, the bipolar group showed no correlation at all between its mean reaction times and any confidence-related measure. Our results do not demonstrate the existence of cognitive control impairments in bipolar patients, and might need more subjects and patients to do so. Assuming that bipolar patients present cognitive and metacognitive abnormalities, they would very likely differ from those observed in schizophrenia.

Third, contrary to the bipolar group, the schizophrenia group showed significant or nearly significant correlations between reaction times and confidence related measures, including unconfident responding and false alarms. Interestingly, the schizophrenia group also showed significant correlations between training length and most of metacognitive or confidence-related measures. It is noteworthy that reaction times and training length were both significantly different from the control or bipolar groups, and allowed us to put in evidence a basic deficit regarding cognitive control functioning. The fact that these variables significantly correlate with all the metacognitive measures is in itself very relevant for our purposes, because it demonstrates a tight link between cognitive control impairments and metacognition in schizophrenia.

We were not in position to make any correlation with clinical variable, including the type of pharmacological treatment, but this surely opens new direction of research.

4-5 Conclusions:

In conclusion of the last chapter (Part 3), we formulated different set of hypotheses, because we

hoped that investigating metacognition and cognitive control skills in schizophrenia patients would shed some light on the metacognitive mechanisms in the brain and in particular on the question of the total or partial overlap between first-order and second-order decision processes. We therefore formulated different hypotheses regarding the metacognitive profile one might obtain in schizophrenia. On another independent issue, we wanted to investigate the issue of whether schizophrenia patients presented impairments in metacognition.

We first draw conclusions regarding the second aspect. Considered together, these results confirm the impairments of cognitive control functioning in schizophrenia, and demonstrate a link between the cognitive control impairments and impaired metacognition in schizophrenia. Moreover, they are consistent with the thesis of qualitatively and quantitatively different metacognitive profiles in schizophrenia and bipolar disorders. Schizophrenia patients are impaired at the metacognitive level compared with bipolar patients, even when these patients have had past psychotic episodes.
The last step for us will consist in testing the correlation between cognitive control performance (including training length), metacognitive performance (including unconfident reports) with the magnitude of positive symptoms and other relevant clinical variables.

4. 6 Overall Conclusions

Regarding the general question of the metacognitive mechanisms, we had hypothesized that metacognition is actually a relative process, whereby a first-order decision network is managed and accessed by a second-order one, situated at a level superior within the hierarchy of cognitive control.
Several points are compatible with that account (cf. introduction). The results we obtained in schizophrenia patients are perfectly consistent with that account as well.
The interest of such a population of patients stands in the fact they are known to display strong and robust abnormalities in a prefrontal network including anterior cingulate (BA24) and dorsolateral prefrontal (BA9) cortices. These cortices, in particular BA9, are involved in the cognitive control of behavior, and have been associated with metacognition or access to consciousness (Rounis et al, 2010). In a situation whereby schizophrenic patients have to perform a cognitive control task self-evaluating as the same time, one could expect a different metacognitive pattern if BA9 is critical for metacognitive

judgment.

The results are compatible with the idea that, in our paradigm, BA9 indeed is critical, but not for metacognitive judgment or confidence. It does not play the role of a judge, but seems more critical for the accessing or maintenance of the relevant information (in our case task and response selections). Patients were impaired in evaluating their performance, but their frequent reports of being unconfident suggests that they were aware of lacking awareness of what they actually did. They did not produce less hits than other groups. They only reported that they did not know whether they performed correctly or not.

This is consistent with a there existing distribution of metacognitive processes within the cognitive control hierarchy, depending on which type of information being manipulated (external or internal), and which role this information played in the driving or management of behavior (was it a signal to trigger a decision process or a reward ? for example).

Moreover, if such an account is appropriate, when a lesion is present within the cognitive control hierarchy, one should observe a more impaired metacognitive performance when the lesion is situated at more anterior site. In addition, one should observe greater neural activity in the regions situated upstream of the site of the lesion, in order to compensate. In the case of schizophrenia patients, in a situation of metacognitive judgment of their own performance, they should show decreased activation in dorsolateral and medial prefrontal cortex BA9, but also increased activation in more anterior prefrontal regions (BA10) for example, compared with control or bipolar group. A future neuroimaging (MRI or PET) study should be able to shed some light on these issues.

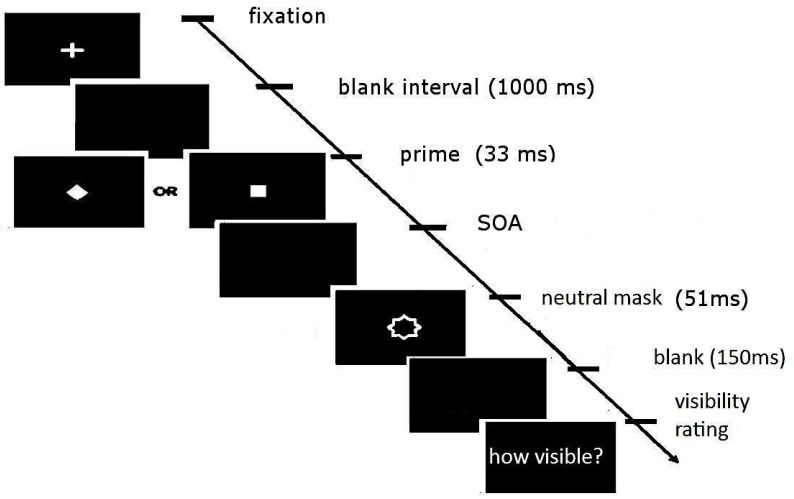
Appendix 1 : Double Staircase Algorithm.

Basically this preliminary test consisted for the subject in evaluating the visibility of a target followed by a mask. The SOAs were both initialized at 16ms . After 40 trials, the program began to increment (+16ms) or decrement (-16ms) the SOA values according to the mean visibility given by the subject. The program ended up when SOA values reached a mean visibility inferior to 0.20 for the low visibility condition, and a mean visibility superior to 0.80 for the high visibility condition.

Patients had to discriminate the prime (diamond versus square) by pressing a left or a right key. When they did not manage to see them, they had to randomly press one of the keys. After each response, they had to rate the visibility of the prime using the following scale.

Note that they were informed that sometimes there was no prime at all:

- 0 = niente (nothing)
- 0.2 to 0.4 = visto solo un flash (I saw just a flicker)
- 0.6 = visto, ma sono insicuro (I saw a shape, but I am very uncertain)
- 0.8-1 = visto abbastanza o molto bene (quite well or very well seen)



[**Figure 4-A:** Trial of visibility test, double staircase algorithm]

Appendix 2 : BEHAVIORAL PERFORMANCE IN THE FIRST ORDER COGNTIVE TASK :

SOA	CONG	COMP	Mean ACC	SD	Mean RT	SD
long	**cong**	**comp**	0,93	0,10	999,76	458,95
		incomp	0,93	0,11	986,49	394,82
	Total cong		**0,93**	**0,10**	**993,12**	**411,35**
	incong	**comp**	0,94	0,12	1026,79	387,54
		incomp	0,94	0,13	946,83	300,83
	Total incong		**0,94**	**0,12**	**986,81**	**335,87**
Total long			**0,93**	**0,11**	**989,97**	**368,51**
short	**cong**	**comp**	0,94	0,11	1070,52	484,21
		incomp	0,93	0,10	950,99	331,20
	Total cong		**0,94**	**0,10**	**1010,75**	**403,35**
	incong	**comp**	0,93	0,12	989,90	353,35
		incomp	0,94	0,14	995,01	320,25
	Total incong		**0,93**	**0,12**	**992,45**	**323,99**
Total short			**0,94**	**0,11**	**1001,60**	**359,11**
Total			**0,94**	**0,11**	**995,78**	**360,56**

[**TABLE 2** : PERFORMANCE OF **CONTROL GROUP** IN THE BASIC TASK]

SOA	CONG	COMP	Mean ACC	SD	Mean RT	SD
long	**cong**	**comp**	0,96	0,06	1349,27	265,72
		incomp	0,97	0,05	1426,89	312,59
	Total cong		**0,96**	**0,05**	**1388,08**	**279,55**
	incong	**comp**	0,98	0,02	1454,13	386,03
		incomp	0,97	0,04	1609,53	489,09
	Total incong		**0,98**	**0,03**	**1531,83**	**427,85**
Total long			**0,97**	**0,04**	**1459,95**	**360,99**
short	**cong**	**comp**	0,97	0,06	1388,89	350,69
		incomp	0,97	0,05	1381,61	384,74
	Total cong		**0,97**	**0,05**	**1385,25**	**351,00**
	incong	**comp**	0,97	0,04	1491,29	410,97
		incomp	0,98	0,03	1454,10	425,09
	Total incong		**0,98**	**0,03**	**1472,69**	**399,11**
Total short			**0,97**	**0,04**	**1428,97**	**370,27**
Total			**0,97**	**0,04**	**1444,46**	**362,09**

[TABLE 3 : PERFORMANCE **BIPOLAR GROUP** IN THE BASIC TASK]

SOA	CONG	COMP	Mean ACC	SD	Mean RT	SD
long	**cong**	**comp**	0,96	0,04	1928,59	721,64
		incomp	0,97	0,03	2008,37	744,36
	Total cong		**0,96**	**0,03**	**1968,48**	**712,38**
	incong	**comp**	0,92	0,11	1820,93	714,49
		incomp	0,92	0,10	1957,34	865,55
	Total incong		**0,92**	**0,10**	**1889,14**	**773,12**
Total long			**0,94**	**0,08**	**1928,81**	**733,78**
short	**cong**	**comp**	0,97	0,02	2007,17	763,03
		incomp	0,96	0,04	1955,99	806,17
	Total cong		**0,96**	**0,03**	**1981,58**	**761,92**
	incong	**comp**	0,90	0,08	2016,61	943,41
		incomp	0,89	0,08	2076,05	977,44
	Total incong		**0,89**	**0,08**	**2046,33**	**932,39**
Total short			**0,93**	**0,07**	**2013,95**	**839,82**
Total			**0,93**	**0,07**	**1971,38**	**784,19**

[TABLE 4: PERFORMANCE **SCHIZOPHRENIA GROUP** IN THE BASIC TASK]

Appendix 3 : BEHAVIORAL PERFORMANCE IN THE METACOGNTIVE TASK :

SOA	CONG	COMP	Mean meta-ACC	SD	Mean meta-RT	SD
Long	**cong**	**comp**	0,90	0,14	497,50	135,83
		incomp	0,90	0,12	478,51	81,19
	Total cong		**0,90**	**0,13**	**488,00**	**107,96**
	incong	**comp**	0,88	0,16	486,89	116,87
		incomp	0,90	0,15	521,60	135,09
	Total incong		**0,89**	**0,15**	**504,25**	**122,68**
Total long			**0,90**	**0,14**	**496,12**	**113,70**
Short	**cong**	**comp**	0,90	0,14	508,55	87,59
		incomp	0,90	0,15	509,24	122,02
	Total cong		**0,90**	**0,14**	**508,89**	**102,04**
	incong	**comp**	0,89	0,15	501,68	154,40
		incomp	0,89	0,15	507,17	86,15
	Total incong		**0,89**	**0,14**	**504,43**	**120,16**
Total_short			**0,89**	**0,14**	**506,66**	**109,41**
Total			**0,89**	**0,14**	**501,39**	**110,68**

[TABLE 7-1:PERFORMANCE **CONTROL GROUP** METACOGNITIVE TASK**]**

SOA	CONG	COMP	Mean FA	SD
long	**cong**	**comp**	0,05	0,09
		incomp	0,05	0,08
	Total cong		**0.05**	**0,08**
	incong	**comp**	0.07	0,12
		incomp	0.03	0,06
	Total incong		**0.06**	**0,09**
Total long			**0.05**	**0,08**
short	**cong**	**comp**	0,05	0,09
		incomp	0,05	0,06
	Total cong		**0,05**	**0,07**
	incong	**comp**	0,07	0,09
		incomp	0,06	0,10
	Total incong		**0,07**	**0,09**
Total short			**0,06**	**0,08**
Total l			**0,06**	**0,08**

[TABLE 7-2:**FA** in **CONTROL GROUP,** METACOGNITIVE TASK**]**

SOA	CONG	COMP	Mean HIT	SD
long	**cong**	**comp**	0,33	0,58
		incomp	0,28	0,19
	Total cong		**0,31**	**0,39**
	incong	**comp**	0,56	0,51
		incomp	0,17	0,29
	Total incong		**0,36**	**0,43**
Total long			**0,33**	**0,39**
short	**cong**	**comp**	0,11	0,07
		incomp	0,50	0,58
	Total cong		**0,37**	**0,49**
	incong	**comp**	0,62	0,46
		incomp	0,46	0,49
	Total incong		**0,55**	**0,44**
Total short			**0,47**	**0,45**
Total			**0,40**	**0,42**

[**TABLE 7-3**:HITS **CONTROL GROUP** METACOGNITIVE TASK**]**

SOA	CONG	COMP	Mean meta-ACC	SD	Mean meta-RT	SD
Long	**cong**	**comp**	0,95	0,06	595,47	180,50
		incomp	0,96	0,05	546,83	145,55
	Total cong		**0,95**	**0,05**	**571,15**	**158,38**
	incong	**comp**	0,98	0,02	546,18	128,19
		incomp	0,97	0,03	577,36	173,00
	Total incong		**0,97**	**0,03**	**561,77**	**146,08**
Total long			**0,96**	**0,04**	**566,46**	**149,08**
Short	**cong**	**comp**	0,96	0,05	578,95	175,22
		incomp	0,97	0,05	576,49	144,82
	Total cong		**0,96**	**0,05**	**577,72**	**153,27**
	incong	**comp**	0,96	0,05	541,29	129,95
		incomp	0,98	0,04	561,90	129,23
	Total incong		**0,97**	**0,04**	**551,59**	**124,03**
Total_short			**0,97**	**0,05**	**564,66**	**137,00**
Total			**0,96**	**0,04**	**565,56**	**141,64**

[TABLE 8-1 :PERFORMANCE **BIPOLAR GROUP** METACOGNITIVE TASK**]**

SOA	CONG	COMP	Mean FA	SD
long	**cong**	**comp**	0,02	0,03
		incomp	0,01	0,01
	Total cong		**0,02**	**0,02**
	incong	**comp**	0,01	0,01
		incomp	0.00	0,01
	Total incong		**0,01**	**0,01**
Total long			**0,01**	**0,02**
short	**cong**	**comp**	0,02	0,01
		incomp	0,02	0,01
	Total cong		**0,01**	**0,01**
	incong	**comp**	0,02	0,02
		incomp	0,02	0,01
	Total incong		**0,01**	**0,02**
Total short			**0,01**	**0,02**
Total			**0,01**	**0,02**

TABE 8-2 :FA in **BIPOLAR GROUP** METACOGNITIVE TASK**]**

SOA	CONG	COMP	Mean HIT	SD
long	**cong**	**comp**	0,08	0,14
		incomp	0,00	0,00
	Total cong		**0,04**	**0,10**
	incong	**comp**	0,67	0,58
		incomp	0,05	0,10
	Total incong		**0,31**	**0,47**
Total long			**0,19**	**0,37**
short	**cong**	**comp**	0,17	0,24
		incomp	0,67	0,58
	Total cong		**0,47**	**0,51**
	incong	**comp**	0,60	0,49
		incomp	0,50	0,71
	Total incong		**0,57**	**0,50**
Total short			**0,52**	**0,48**
Total			**0,34**	**0,45**

[TABLE 8-3 :HIT in **BIPOLAR GROUP** METACOGNITIVE TASK**]**

SOA	CONG	COMP	Mean meta-ACC	SD	Mean meta-RT	SD
Long	**cong**	**comp**	0,91	0,05	565,52	166,74
		incomp	0,93	0,06	600,60	176,69
	Total cong		0,92	0,06	583,06	166,95
	incong	**comp**	0,89	0,11	563,34	158,42
		incomp	0,88	0,12	584,45	156,31
	Total incong		0,88	0,11	573,89	152,42
Total long			0,90	0,09	578,48	157,32
Short	**cong**	**comp**	0,94	0,05	635,12	175,29
		incomp	0,92	0,08	555,05	124,07
	Total cong		0,93	0,06	595,09	152,42
	incong	**comp**	0,84	0,15	560,51	164,19
		incomp	0,86	0,11	607,07	176,37
	Total incong		**0,85**	**0,13**	**583,79**	**166,36**
Total_short			**0,89**	**0,11**	**589,44**	**157,05**
Total			**0,90**	**0,10**	**583,96**	**156,03**

[**TABLE 9-1:**PERFORMANCE **SCHIZOPHRENIA GROUP** METACOGNITIVE TASK**]**

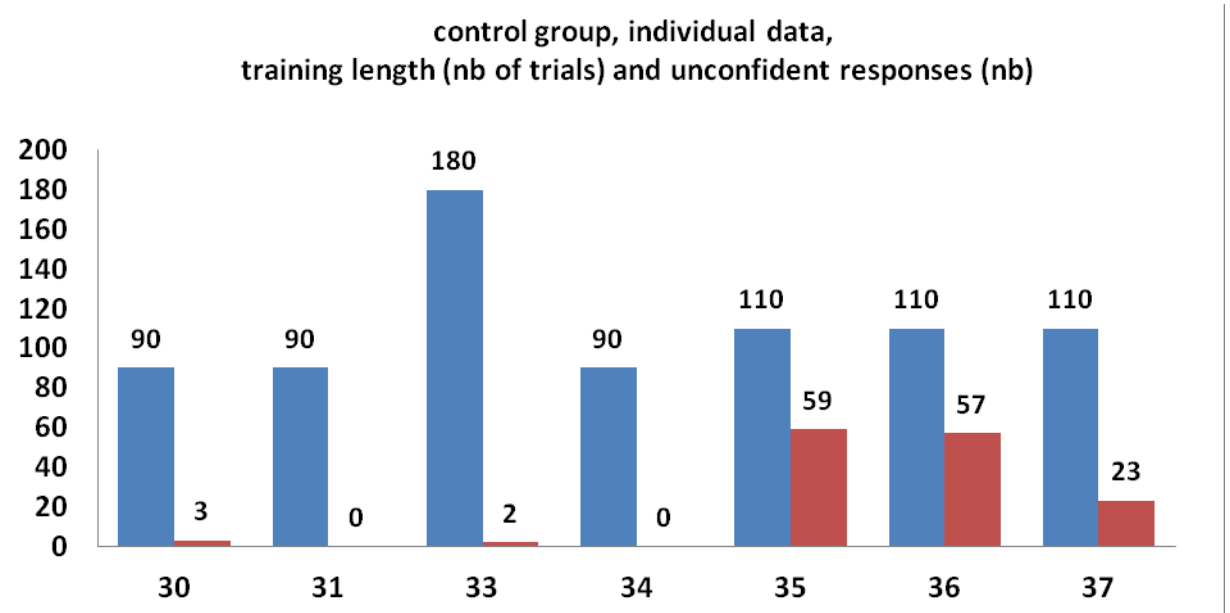
SOA	CONG	COMP	Mean FA	SD
long	**cong**	**comp**	0,05	0,04
		incomp	0,04	0,05
	Total cong		**0,05**	**0,04**
	incong	**comp**	0,04	0,05
		incomp	0,04	0,07
	Total incong		**0,04**	**0,06**
Total long			**0,04**	**0,05**
short	**cong**	**comp**	0,04	0,04
		incomp	0,05	0,07
	Total cong		**0,04**	**0,06**
	incong	**comp**	0,06	0,08
		incomp	0,05	0,08
	Total incong		**0,06**	**0,08**
Total short			**0,05**	**0,07**
Total			**0,05**	**0,06**

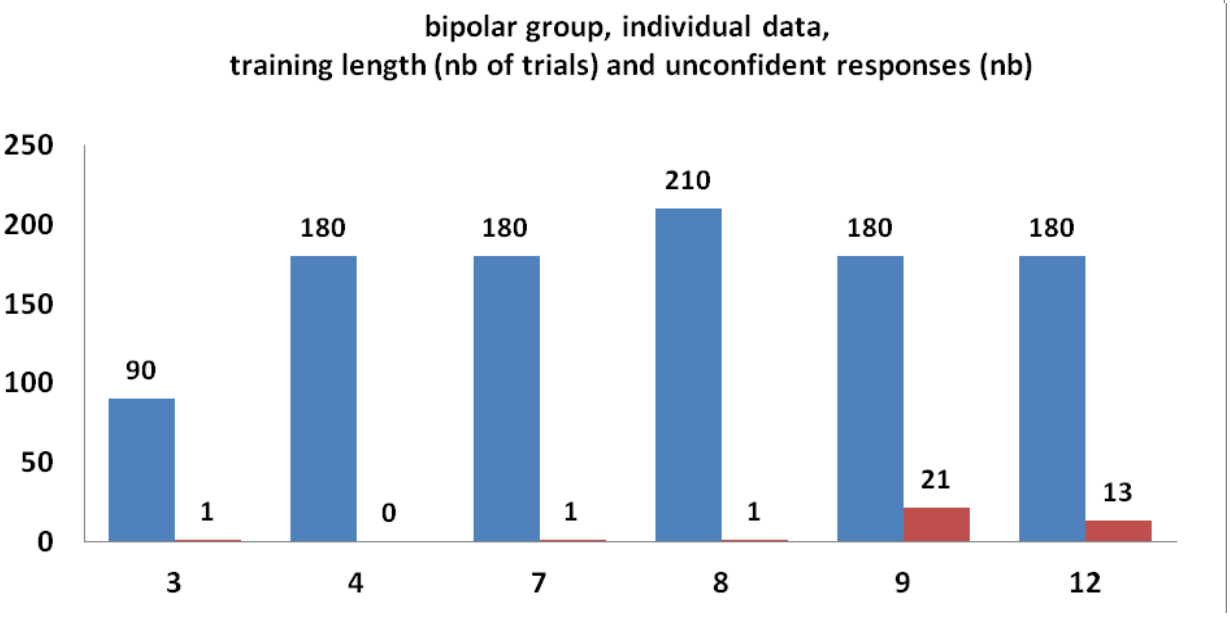
[TABLE 9-2 :FA in **SCHIZOPHRENIA GROUP** METACOGNITIVE TASK**]**

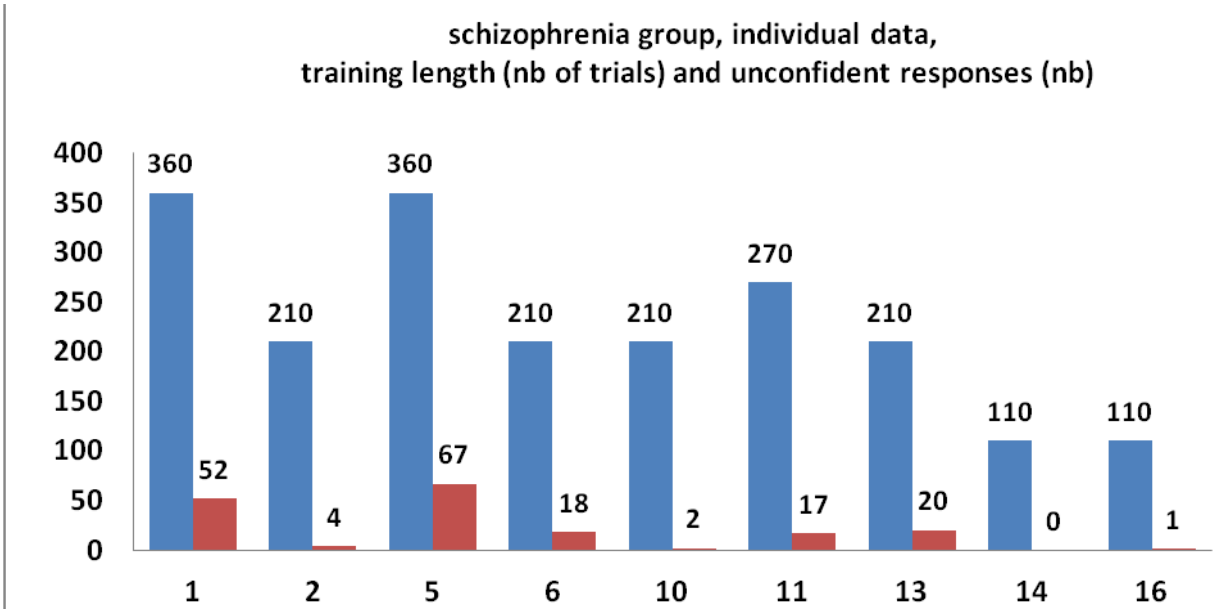
SOA	CONG	COMP	Mean FA	SD
long	**cong**	**comp**	0,19	0,37
		incomp	0,43	0,45
	Total cong		**0,30**	**0,41**
	incong	**comp**	0,28	0,37
		incomp	0,26	0,36
	Total incong		**0,27**	**0,35**
Total long			**0,28**	**0,38**
short	**cong**	**comp**	0,31	0,41
		incomp	0,34	0,40
	Total cong		**0,33**	**0,39**
	incong	**comp**	0,30	0,35
		incomp	0,34	0,17
	Total incong		**0,32**	**0,27**
Total short			**0,32**	**0,33**
Total			**0,30**	**0,35**

[TABLE 9-3 :HIT in **SCHIZOPHRENIA GROUP** METACOGNITIVE TASK**]**

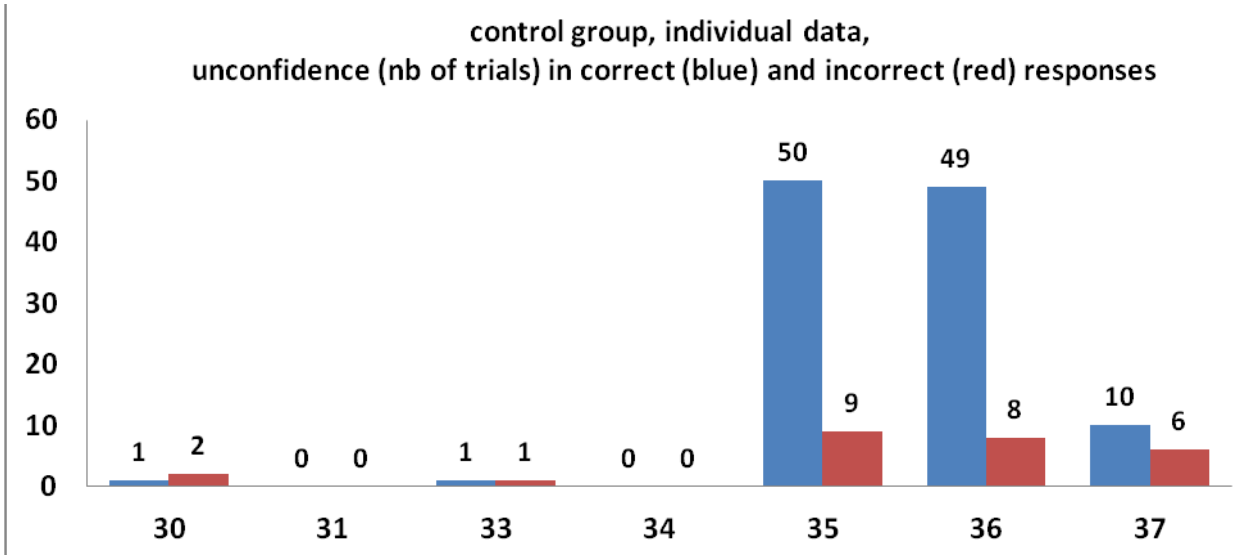
Appendix 4 : individual data, training length (blue) and unconfidence (red)

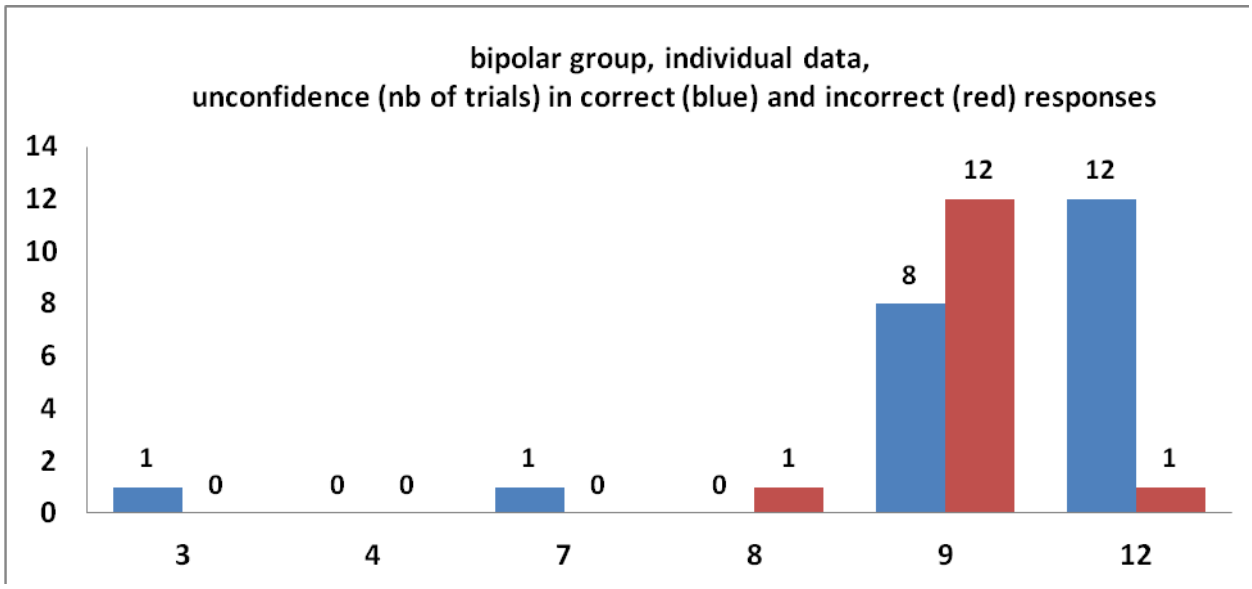


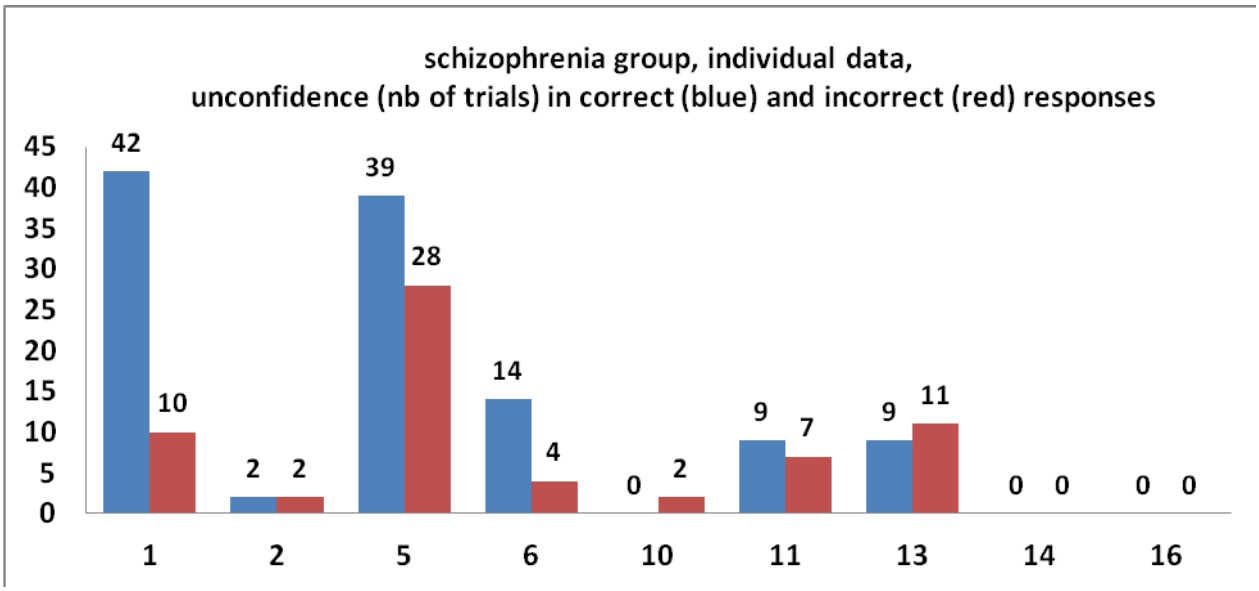




Appendix 5 : Unconfident responses in correct versus incorrect (first order) responses







Appendix 6 : Summary of DSM-IV criteria for Bipolar Disorders

Bipolar I Disorder: The essential feature of Bipolar I Disorder is a clinical course that is characterized by the occurrence of one or more Manic Episodes or Mixed Episodes. Often individuals have also had one or more Major Depressive Episodes. Episodes of Substance-Induced Mood Disorder (due to the direct effects of a medication, or other somatic treatments for depression, a drug of abuse, or toxin exposure) or of Mood Disorder Due to a General Medical Condition do not count toward a diagnosis of Bipolar I Disorder. In addition, the episodes are not better accounted for by Schizoaffective Disorder and are not superimposed on Schizophrenia, Schizophreniform Disorder, Delusional Disorder, or Psychotic Disorder Not Otherwise Specified.

Bipolar II Disorder: The essential feature of Bipolar II Disorder is a clinical course that is characterized by the occurrence of one or more Major Depressive Episodes accompanied by at least one Hypomanic Episode. Hypomanic Episodes should not be confused with the several days of euthymia that may follow remission of a Major Depressive Episode. Episodes of Substance- Induced Mood Disorder (due to the direct effects of a medication, or other somatic treatments for depression, a drug of abuse, or toxin exposure) or of Mood Disorder Due to a General Medical Condition do not count toward a diagnosis of Bipolar I Disorder. In addition, the episodes are not better accounted for by Schizoaffective Disorder and are not superimposed on Schizophrenia, Schizophreniform Disorder, Delusional Disorder, or Psychotic Disorder Not Otherwise specified.

Criteria for a Manic Episode

A. A distinct period of abnormally and persistently elevated, expansive, or irritable mood, lasting at least 1 week (or any duration if hospitalization is necessary):

B. During the period of mood disturbance, three (or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:

1. inflated self-esteem or grandiosity
2. decreased need for sleep (e.g., feels rested after only 3 hours of sleep)
3. more talkative than usual or pressure to keep talking
4. flight of ideas or subjective experience that thoughts are racing
5. distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)
6. increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation
7. excessive involvement in pleasurable activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)

C. The symptoms do not meet criteria for a Mixed Episode.

D. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others, or there are psychotic features.

E. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication, or other treatments) or a general medical condition (e.g., hyperthyroidism).

Note: Manic-like episodes that are clearly caused by somatic antidepressant treatment (e.g., medication, electroconvulsive therapy, light therapy) should not count toward a diagnosis of Bipolar I Disorder.

Criteria for a Mixed Episode

A. The criteria are met both for a Manic Episode and for a Major Depressive Episode (except for duration) nearly every day during at least a 1-week period:

B. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others, or there are psychotic features.

C. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g., hyperthyroidism).

Criteria for a Hypomanic Episode

A. A distinct period of persistently elevated, expansive, or irritable mood, lasting throughout at least 4 days, that is clearly different from the usual nondepressed mood:

B. During the period of mood disturbance, three (or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:

1. inflated self-esteem or grandiosity

2. decreased need for sleep (e.g., feels rested after only 3 hours of sleep)
3. more talkative than usual or pressure to keep talking
4. flight of ideas or subjective experience that thoughts are racing
5. distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)
6. increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation
7. excessive involvement in pleasurable activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)

C. The episode is associated with an unequivocal change in functioning that is uncharacteristic of the person when not symptomatic.
D. The disturbance in mood and the change in functioning are observable by others.
E. The episode is not severe enough to cause marked impairment in social or occupational functioning, or to necessitate hospitalization, and there are no psychotic features.
F. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g., hyperthyroidism).
Note: Hypomanic-like episodes that are clearly caused by somatic antidepressant treatment (e.g., medication, electroconvulsive therapy, light therapy) should not count toward a diagnosis of Bipolar II Disorder.

[From www.intermountainhealthcare.org]

PART VI:

Overall Conclusions, future research

The main scope of that work was *in fine* to sketch an account of metacognition. This objective entailed clarifying several interrelated issues, not only empirical, but also conceptual.

6.1 Conceptual issues:

Minor clarifications may be necessary regarding the notion of consciousness when one deals with metacognition as the same time: when for instance subjects are displayed some subliminal primes or distractors before responding, and are then asked to evaluate their performance. In such paradigms, it is obviously critical to distinguish *the awareness of the distractor and the awareness of one's performance*. Moreover, both types of awareness can be measured with metacognitive judgment, but in the first case, it is a matter of *perceptual (visual) metacognition*; in the second case it is a matter of *executive* or *response-related* metacognition. They are perfectly dissociable and can be selectively impaired. Reaching a consensus about a common terminology for each metacognitive domain may help avoiding confusions.

The main conceptual difficulty stood in defining metacognitive process, in such a way that it is possible to differentiate it from another type of process. That conceptual problem arose with the issue of error-related or conflict-related activity in Anterior Cingulate cortex. In the case of error, one can observe qualitative and quantitative differences according to whether the subject is aware of having made an

error or not. When the subject is aware of it, the amplitude of the rERN signal is greater, and importantly, it is followed by an increase of activity within the neighboring (upstream) network situated in the lateral prefrontal cortex, and overt behavioral changes. In brief, one can observe significant changes in neighboring brain activity, and overt behavior (Endgrass et al., 2007).

That phenomenon has given rise to the thesis of that metacognition can be deployed non-consciously (Charles et al., 2013). Yet accurate metacognitive judgment has been the only experimental method for a demonstration of conscious processing, since the beginning of scientific studies of Consciousness. It is thus simply incoherent and nonsensical to use (the accuracy level of) metacognitive judgments to demonstrate the existence of non conscious processing, and claiming at the same time that accurate metacognition is not a marker of consciousness.

I therefore estimated that a conceptual refinement is necessary and consider that:

- *Metacognition* is a cognitive control process whereby the output of a given network reaches consciousness, and is sent to another upstream network. It involves a global transfer of the information outside the network of origin which is observable through brain activity and through behavior if the paradigm allows it.
- It differs from *Metaprediction*, which is a process whereby a network predicts or learns to predict the errors of a downstream network. The information computed by the so-called metapredictor does not necessarily reach consciousness, and does not necessarily influence the activity of other networks, nor the behavior, in a significant way.

These definitions may be incorrect or vague, but at the moment they are consistent with the literature about consciousness and metacognition (as far as I know) and my data.

6.2 Empirical issues:

The first empirical issue concerned the effects of consciousness (of external information) and cognitive control on first-order decisions (accuracy, reaction times). In particular, whether we could observe significant effects of priming on first-order performance when the primes were generally not visible (that is to say, a compatibility effect in short SOA trials).

The response is not straightforward and certainly not simple, since we obtained inconsistent results across our different experiments. We did not obtain any effect of prime compatibility (in short SOA trials) in the first behavioral study (cf. *Part II, section 2.4.4*), we did in our neuroimaging study (cf. *Part III, section 3.4 and section 3.5*) and the control group of our observational study with patients did not show any such effects (cf. *Part IV, section 4.3*).

I suggest, since I did carry out all these experiments and numerous unreported pilots before, that the effects are an inverted U-shaped function of the training, or the learning level. The subjects of the first behavioral study were extremely trained, until they reached an almost perfect routinization of the paradigm. The situation was less demanding for the subjects of the neuroimaging study, for practical and statistical reason (the time was limited, and I knew that incorrect trials would be lost so I insisted more on the accuracy). Finally, the control group of the last study was clearly under trained – they had to meet the same training standards as schizophrenia patients, which were far below their capacity.

Moreover, these effects (of prime compatibility in short SOA trials) can interact with cognitive control load, at brain level. In effect, in our neuroimaging study, although we observed no interaction between prime compatibility and congruency factors at the behavioral level, we observed it at brain level. That difference may be explained by the fact that compatibility and congruency tap into different mechanisms, in our paradigm at least. Prime compatibility (especially in short SOA trials) influenced bottom-up parallel mechanisms of decision, whereas congruency clearly influenced top-down serial processes. It sounds intuitive that they interact at some point of convergence –that we identified as being the anterior cingulate cortex.

A second empirical issue concerned the possible effects that consciousness (of external information) and cognitive control load have on the awareness of one's first-order performance – in other words on metacognitive performance. Subjects had to produce second-order judgments, that to say metacognitive judgments, and had to say whether they answered correctly, whether they made a mistake, or whether they did not know (unconfident responding). In particular, we were interested in determining whether significant effects of compatibility could be observed on metacognitive accuracy when the primes were generally not visible (that is to say in short SOA trials), and when the cognitive control load was higher.

Again, the response is certainly not simple, since we did not observe the same patterns of results across our different studies. The effect of non-consciously perceived stimuli on metacognitive performance may also depend on the training length or learning level, in a non linear fashion. In any case, considering only confident responses, we have been able to observe that unseen primes could influence the metacognitive performance without influencing the cognitive performance itself. Interestingly, (cf. *Part II, section 2.4.4 and section 2.5*) hits and false alarms seemed to behave in a very similar way, and both showed a compatibility effect (more error reported in incompatible condition) in short SOA and incongruent trials, and not in long SOA nor congruent trials. This led us to suppose (i) that the metacognitive "judge" mechanism was sensitive to the level of noise/conflict/smoothness of the response selection – or, in other words, to the quality of evidence accumulated during first-order decision ; (ii) that it was also sensitive to the quantity of evidence *to be* accumulated during response selection – or, in other words, to the threshold of first-order decision (since in incongruent trials, 2 bits of information are necessary to select the response, while only 1 bit is necessary in congruent trials).

At that point we had dissociated the awareness of *external* information (visual, for instance) from the awareness of *internal* information (noise during response-selection and task-selection etc..). We also carried out an additional step toward the third empirical issue, namely the question of whether metacognitive processes also involve first-order processes.

The third issue thus concerned the question of whether the first-order decision networks are also involved in the second order decision. Our neuroimaging study showed that a prefrontal network (BA9) was involved in both first and second-order decisions (*Part III* for more details).
In the same vein we showed that response selection networks were indeed recruited in short SOA only, and showed an activity that reflected an interaction between compatibility and congruency – that is to say between the quality of evidence accumulated and the quantity of evidence to be accumulated).

Finally, a fourth issue was whether BA9 was critical for metacognition in itself or whether its involvement was mainly due to our paradigm. We tried to resolve this issue by studying patients with

schizophrenia, which are known to display robust abnormalities in this prefrontal network. We formulated some specific hypotheses regarding the cognitive and metacognitive performance of schizophrenic patients on the basis of what we already knew/considered as plausible the more specific question of the metacognitive mechanisms. We had hypothesized that metacognition is actually a *relative* process, whereby a first-order decision network is managed and accessed by a second-order one, situated at a level superior within the hierarchy of cognitive control. Several points were compatible with that account (cf. *Part I, section 5*) and the results we obtained in schizophrenia patients are perfectly consistent with that account as well.

First, the measures (reaction times and training length) which allowed us to confirm the existence of basic impairments regarding cognitive control functioning both significantly correlated with general meta-accuracy (which include only confident metacognitive responses) and the number of unconfident reports. The more impaired they were in the cognitive control task, the more impaired they were in the metacognitive task, and the more unconfident they were.

Secondly, as said earlier, schizophrenic patients were indeed impaired in evaluating their performance but the pattern of their metacognitive impairments suggests a bias toward a lowered confidence level, not a blindness to their impairments. They did not produce fewer hits than other groups, and the number of unconfident responses do not correlate with the hits rate. They either reported having made an error (while they actually produced a correct response), or reported that they did not know whether they performed correctly or not. Thus, their frequent reports of being unconfident or inaccurate suggest that they were conscious of their impairment in the cognitive control task, but also of their lack of awareness of what they actually did on a trial-by-trial basis.

Consequently, it would appear that BA9 might be not critical as a "metacognitive judge" *per se* and in a absolute way. This prefrontal network, situated upstream the premotor cortex and Anterior Cingulate cortex within the cognitive control hierarchy, might be critical only for the awareness of our (rule-based) action selection and response conflict.

Much more generally, considerations of this type might explain why psychotic episodes in schizophrenia involve a decrease or a loss of the sense of authorship, while those of bipolar patients for instance do not.

References

Alain C, McNeely H. E., He Y, B,. Christensen B.K., West R., Neurophysiological evidence of error-monitoring deficits in patients with schizophrenia., Cerebral Cortex, 2002, 12, 840-846.

Atkinson, R.C.; Shiffrin, R.M. (1968). "Chapter: Human memory: A proposed system and its control processes". In Spence, K.W.; Spence, J.T. The psychology of learning and motivation (Volume 2). New York: Academic Press. pp. 89–195.

Baars, B, A cognitive theory of consciousness, NY: Cambridge University Press 1989.

Badre David, Cognitive Control, Hierarchy, and the rostro-caudal organization of the frontal lobes, Trends in Cognitive Sciences, 2008, 12(5), 193-201.

Baker TE, Holroyd CB., Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200., Biol Psychol. 2011;87(1):25-34.

Barbalat G, Chambon V, Franck N, Koechlin E, Farrer C, Organization of Cognitive Control Within the Lateral Prefrontal Cortex in Schizophrenia , Arch Gen Psychiatry, 2009, 66 (4), 377-386.

Basar and Güntekin, A review of brain oscillations in cognitive disorders and the role of neurotransmitters, Brain Research, 2008 , 172-193.

Benes F.M., Neurobiological investigations of cingulate cortex in brain with schizophrenia, Schizophrenia Bulletin, 1993, 19 (3), 537-549.

Berk M, Dodd S, Kauer-Sant'anna M, Malhi GS, Bourin M, Kapczinski F, Norman T., Dopamine dysregulation syndrome: implications for a dopamine hypothesis of bipolar disorder., Acta Psychiatr Scand Suppl. 2007;(434):41-9.

Blakemore SJ, Frith CD, Wolpert DM. Spatio-temporal prediction modulates the perception of self-produced stimuli. J Cogn Neurosci. 1999 ;11(5):551-559.

Blakemore SJ, Wolpert DM, Frith CD. Abnormalities in the awareness of action. Trends Cogn Sci. 2002 ;6(6):237-242.

Brass M, Haggard P, To Do or Not to Do: The Neural Signature of Self-Control, The Journal of Neuroscience, 27(34):9141–9145.

Broadbent D., Perception and communication, Pergamon Press, 1958

Cardin JA, Carle´n M, Meletis K, Knoblich U, Zhang F, Deisseroth K, Tsai LH, Moore CI: Driving fast-spiking cells induces gamma rhythm and controls sensory responses. Nature 2009, 459:663-667.

Charles L, Van Opstal F, Marti S, Dehaene S., Distinct brain mechanisms for conscious versus subliminal error detection, Neuroimage. 2013;73:80-94.

Chambon V, Franck N, Koechlin E, Fakra E, Ciuperca G, Azorin J-M, Farrer C, The architecture of cognitive control in schizophrenia, Brain (2008), 131, 962-970.

Colebatch J. M., Bereitschaftspotential and Movement-Related Potentials: Origin, Significance, and Application in Disorders of Human Movement , Movement Disorders , 22 (5), 2007, 601– 610.

Costa E., Dong E., Grayson D.R., Ruzicka W.B., M.V. Simonini, Veldic M., Guidotti A., Epigenetic Targets in GABAergic Neurons to Treat Schizophrenia, Advances in Pharmacology, Volume 54, 2006, Pages 95-117

Cunnington R, Windischberger C, Deecke L and Moser E, The preparation and readiness for voluntary movement: a high-field event-related fMRI study of the Bereitschafts-BOLD response**,** NeuroImage, 20 (1), 2003, 404-412.

E. Daprati, N. Franck, N. Georgieff, J. Proust, E. Pacherie, J. Dalery, M. Jeannerod , Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients, Cognition 65 (1997) 71–86.

David, A. S., Bedford, N., Wiffen, B. & Gilleen, J. 2012 Failures of metacognition and lack of insight in neuropsychiatric disorders. Phil. Trans. R. Soc. B 367, 1379–1390.

Dehaene, S. et al. (2003) Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: the role of the anterior cingulate. Proc. Natl. Acad. Sci. U. S. A. 100, 13722 – 13727 .

Del Cul A, Dehaene S, Leboyer M, Preserved Subliminal Processing and Impaired Conscious Access in Schizophrenia , Arch Gen Psychiatry, 2006;63:1313-1323.

Eagleman DM, The Where and the When of Intention, Science 303 (2004), 1144-1146.

Exner C,Weniger G, Schmidt-Samoa C, Irle E, Reduced size of the pre-supplementary motor cortex and impaired motor sequence learning in first-episode schizophrenia . Schizophrenia Research 84 (2006) 386 – 396.

Fallona James H., Opolea Isaac O., Potkinc Steven G., The neuroanatomy of schizophrenia: circuitry and neurotransmitter systems, Clinical Neuroscience Research, Volume 3, Issues 1–2, 2003, 77–107.

Feinberg I., Efference copy and corollary discharge: implications for thinking and its disorders , Schizophrenia Bulletin, 1978;4(4):636-40.

Fleming, S.M., Dolan, R.J. & Frith, C.D. (2012) Metacognition: Computation, biology and function. Phil Trans R Soc B 367(1594): 1280-6.

Fleming, S.M. & Dolan, R.J. (2012) The neural basis of accurate metacognition. Phil Trans R Soc B 367(1594): 1338-49.

Fleming, S.M., Huijgen, J. & Dolan, R.J. (2012) Prefrontal contributions to metacognition in perceptual decision-making. Journal of Neuroscience, 32(18): 6117-25.

Ford, Judith M., Jorgensen, Kasper W., Roach, Brian J., Mathalon, Daniel H., Error detection failures in schizophrenia: ERPs and fMRI, International Journal of Psychophysiology (2009)

Fornito A, Yücel M,Dean B, Wood SJ, Pantelis C, Anatomical Abnormalities of the Anterior Cingulate Cortex in Schizophrenia: Bridging the Gap Between Neuroimaging and Neuropathology, Schizophr Bull, 2008

Fourneret, N. Franck, A. Slachevsky4 and M. Jeannerod, Self-monitoring in schizophrenia revisited , Neuroreport, 2001, 1203-1208.

Frith CD, The neural basis of hallucinations and delusions, C. R. Biologies 328 (2005) 169–175.

Frith CD, Blakemore SJ, Wolpert DM., Abnormalities in the awareness and control of action. Philos Tans R Soc Lond B Biol Sci. 2000 ;355(1404):1771-88.

Frith CD, Blakemore S, Wolpert DM. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. Brain Res Rev. ;31(2-3):357-63.

Gallagher S. Neurocognitive models of schizophrenia: a neurophenomenological critique. Psychopathology, 2004 ;37(1):8-19.

Gallinat J, Winterer G, Herrmann CS, Senkowski D.,Reduced oscillatory gamma-band responses in unmedicated schizophrenic patients indicate impaired frontal network processing., Clin Neurophysiol. 2004, 115(8):1863-74.

Galvin SJ, Podd JV, Drga V, Whitmore J., Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions., Psychon Bull Rev. 2003 ;10(4):843-76.

Haggard P., The Sources of Human Volition, Science, 2009, 324, 731-733.

Haggard P, Eimer M, On the relation between brain potentials and the awareness of voluntary movements , Exp Brain Res (1999) 126:128–133.

Hayden B. Y, Platt M. L., Cingulate Cortex, Encyclopaedia of Neuroscience, 2009, 887-892.

Erin A. Heerey, Kimberly R. Bell-Warren, and James M. Gold, Decision-Making Impairments in the Context of Intact Reward Sensitivity in Schizophrenia , Biological Psychiatry, 2008;64:62– 69.

Holender, D. & Duscherer, K. (2004). Unconscious Perception : The need for a paradigm shift. Perception and Psychophysics, 66, 872-881.

Holroyd, C. B., Nieuwenhuis, S., Mars, R., & Coles, M. G. H. (2004). Anterior cingulate cortex, selection for action, and error processing. In M. Posner (Ed.), Cognitive Neuroscience of Attention, (pp. 219-231). New York: Guilford Publishing, Inc.

Hong E. L, Ann Summerfelt, Robert McMahon, Helene Adami, Grace Francis, Amie Elliott, Robert W. Buchanan, Gunvant K. Thaker , Evoked gamma band synchronization and the liability for schizophrenia, Schizophrenia Research 70 (2004) 293 – 302.

Isomura Y, Ito Y, Akazawa T, Nambu A, and Takada M (2003) Neural coding of ''attention for action'' and ''response selection'' in primate anterior cingulate cortex. The Journal of Neuroscience 23: 8002–8012.

Jack, A.I., Shallice, T. Introspective physicalism as an approach to the science of consciousness. Cognition 2001 (79)1-2:161-196

Jeannerod M., The sense of agency and its disturbances in schizophrenia: a reappraisal , Exp Brain Res (2009) 192:527–532.

Kean C, Silencing the Self: Schizophrenia as a Self-disturbance, Schizophrenia Bulletin Advance Access published May 28, 2009.

Kerns, Keith H. Nuechterlein, Todd S. Braver, and Deanna M. Barch, Executive Functioning Component Mechanisms and Schizophrenia , Biological Psychiarty, 2008;64:26 –33.

Koch, Michele Ribolsi, Francesco Mori, Lucia Sacchetti, Claudia Codecà, Ivo Alex Rubino, Alberto Siracusano, Giorgio Bernardi, and Diego Centonze , Connectivity Between Posterior Parietal Cortex and

Ipsilateral Motor Cortex Is Altered in Schizophrenia ,

Krigolson, O. E., & Holroyd, C. B. (2007). Hierarchical error processing: Different errors, different systems. Brain Research, 1155, 70-80.

Kunde, W. (2003) Sequential modulations of stimulus-response correspondence effects depend on awareness of response conflict. Psychon. Bull. Rev. 10, 198 – 205

Lafargue G, Franck N, Sirigu A., Sense of motor effort in patients with schizophrenia, Cortex, 2006, 42(5), 711-719.

Hashimoto T, Volk DW, Eggan SM, Mirnics K, Pierri JN, Sun Z, Sampson AR, Lewis DA., Gene expression deficits in a subclass of GABA neurons in the prefrontal cortex of subjects with schizophrenia., J Neurosci. 2003 Jul 16;23(15):6315-26.

Heerey EA, Bell-Warren KR, Gold JM., Decision-making impairments in the context of intact reward sensitivity in schizophrenia., Biol Psychiatry. 2008 ;64(1):62-9.

Howes OD, Kapur S., The dopamine hypothesis of schizophrenia: version III--the final common pathway., Schizophr Bull. 2009 May;35(3):549-62.

Izaute M, Bacon E, Specific effects of an amnesic drug: effect of lorazepam on study time allocation and on judgment of learning., Neuropsychopharmacology. 2005; 30(1):196-204.

Kirihara K, Rissling AJ, Swerdlow NR, Braff DL, Light GA., Hierarchical organization of gamma and theta oscillatory dynamics in schizophrenia., Biol Psychiatry. 2012, 15;71(10):873-80.

Koriat A, Levy-Sadot R., Conscious and unconscious metacognition: A rejoinder, Conscious Cogn. 2000 ;9(2 Pt 1):193-202.

Kristin R. Laurens, Elton T. C. Ngan, Alan T. Bates, Kent A. Kiehl and Peter F. Liddle, Rostral anterior cingulate cortex dysfunction during error processing in schizophrenia , Brain (2003), 126, 610-622.

Light GA, Jung Lung Hsu, Ming H. Hsieh, Katrin Meyer-Gomes, Joyce Sprock, Neal R. Swerdlow, and David L. Braff , Gamma Band Oscillations Reveal Neural Network Cortical Coherence Dysfunction in Schizophrenia Patients, Biological Psychiatry, 2006;60:1231–1240.

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE., Entrainment of neuronal oscillations as a

mechanism of attentional selection., Science. 2008 Apr 4;320(5872):110-3

Lau CI, Wang HC, Hsu JL, Liu ME., Does the dopamine hypothesis explain schizophrenia? Rev Neurosci. 2013;24(4):389-400.

Lau Hakwan C., Rogers Robert D., Haggard, Patrick and. Passingham Richard E, Attention to Intention, Science 303 (2004), 1208-1210.

Lau HC, Passingham RE., Unconscious activation of the cognitive control system in the human prefrontal cortex., J Neurosci. 2007, 23;27(21):5805-11.

Lewis DA, Hashimoto T., Deciphering the disease process of schizophrenia: the contribution of cortical GABA neurons. Int Rev Neurobiol. 2007;78:109-31.

Lewis DA, Curley AA, Glausier JR, Volk DW., Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia., Trends Neurosci. 2012 ;35(1):57-67.

Luck SJ, Fuller RL, Braun EL, Robinson B, Summerfelt A, Gold JM.,The speed of visual attention in schizophrenia: electrophysiological and behavioural evidence., Schizophrenia Research, 2006;85(1-3):174-95.

Luck S., Impaired response selection in schizophrenia: Evidence from the P3 wave and the lateralized readiness potential , Psychophysiology, 2009, 46.

Luo Q, Mitchell D, Cheng X, Mondillo K, Mccaffrey D, Holroyd T, Carver F, Coppola R, Blair J., Visual awareness, emotion, and gamma band synchronization., Cereb Cortex. 2009 Aug;19(8):1896-904.

Luu P, Flaisch T, Tucker DM, Medial Frontal Cortex in Action Monitoring, The Journal of Neuroscience, January 1, 2000, 20(1):464–469.

MacDonald AW, Cohen JD, Stenger VA, Carter CS., Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control, Science (2000) 288, 1835-1841.

Marcel A. J., Conscious and unconscious perception: Experiments on visual masking and word recognition, Cognitive Psychology 15:197-237 (1983)

Maniscalco B, Lau H., A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings., Conscious Cogn. 2012;21(1):422-30.

Marois R., Ivanoff J., Capacity limits of information processing in the brain, Trends in Cognitive

Science, 2005, 9(6), 296-306.

Marquardt R., Levitt J. G., Blanton R. E., Caplan R., Asarnow R., Siddarth P., Fadale D., McCracken J.T., Toga A.W., Abnormal development of the anterior cingulate in childhood-onset schizophrenia: a preliminary quantitative MRI study, Psychiatry Research: Neuroimaging 138 (2005) 221 – 233.

McCoy AN and Platt ML, Risk-sensitive neurons in macaque posterior cingulate cortex. Nature Neuroscience (205) 8: 1220–1227.

Morris SE., Heerey EA, Gold JM., Holroyd CB., Learning-related changes in brain activity following errors and performance feedback in schizophrenia , Schizophrenia Research 99 (2008) 274 – 285.

Neill E, Rossell SL., Executive functioning in schizophrenia: The result of impairments in lower order cognitive skills?, Schizophr Res., 2013.

Norman, D. A. and Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In Davidson, R. J., Schwartz, G. E., and Shapiro, D., editors, Consciousness and Self-Regulation: Advances in Research and Theory. Plenum Press.

Quraishi S, Frangou S., Neuropsychology of bipolar disorder: a review., J Affect Disord. 2002;72(3):209-26.

Pashler, Harold (1994). "Dual-task interference in simple tasks: Data and theory.". Psychological Bulletin 116 (2): 220–244.

Pleskac TJ, Busemeyer JR., Two-stage dynamic signal detection: a theory of choice, decision time, and confidence., Psychol Rev. 2010 ;117(3):864-901

Polli FE., Barton Jason J. S., Thakkar Katharine N., Greve Douglas N., Goff Donald C.,. Rauch Scott L and Manoach Dara S., Reduced error-related activation in two anterior cingulate circuits is related to impaired performance in schizophrenia , Brain (2008), 131, 971-986.

Posner MI, Snyder CRR (1975). "Attention and cognitive control". In Solso RL. Information processing and cognition: the Loyola symposium. Hillsdale, N.J: L. Erlbaum Associates.

Rounis E., Maniscalco B., Rothwell J., Passingham R., Lau H. 2010. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. Cogn. Neurosci. 1, 165–175.

Rugg MD, Fletcher PC, Frith CD, Frackowiak RS, Dolan RJ., Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval., Brain. 1996 ;119 :2073-83.

Ruzicka WB, Zhubi A, Veldic M, Grayson DR, Costa E, Guidotti A., Selective epigenetic alteration of layer I GABAergic neurons isolated from prefrontal cortex of schizophrenia patients using laser-assisted microdissection., Mol Psychiatry. 2007 Apr;12(4):385-97.

Selva G, Salazar J, Balanzá-Martínez V, Martínez-Arán A, Rubio C, Daban C, Sánchez-Moreno J, Vieta E, Tabarés-Seisdedos R, Bipolar I patients with and without a history of psychotic symptoms: Do they differ in their cognitive functioning?, Journal of Psychiatric Research, Volume 41, 3–4, 2007, 265-272.

Sirigu A, Daprati E, Ciancia S, Giraux P, Nighoghossian N, Posada A, Haggard P, Altered awareness of voluntary action after damage to parietal cortex, Nature Neuroscience, 7 (1), 2004, 80-84.

Schneider, W. & R. M. Shiffrin. (1977). Controlled and automatic human information processing: Detection, search, and attention. Psychological Review, 84, pp1-66.

Shalgi S, Deouell LY., Is any awareness necessary for an Ne?, Front Hum Neurosci. 2012;6:124.

Shimamura A.P. Toward a cognitive neuroscience of metacognition. Conscious Cogn.(2000), **9**, 313–323.

Song C, Kanai R, Fleming SM, Weil RS, Schwarzkopf DS, Rees G, Relating inter-individual differences in metacognitive performance on different perceptual tasks., Conscious Cogn. 2011 Dec; 20(4):1787-92

S. A. Spence, D. J. Brooks, S. R. Hirsch,P. F. Liddle, J. Meehan4and P. M. Grasby, A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control) , Brain (1997), 120, 1997–2011.

Spencer KM, Nestor PG, Niznikiewicz MA, Salisbury DF, Shenton ME, McCarley RW, Abnormal neural synchrony in schizophrenia. Journal of Neurosci ence 23, (2003), 7407–7411.

Spencer KM, Nestor PG, Perlmutter R, Niznikiewicz MA, Klump MC, Frumin M, et al, Neural synchrony indexes disordered perception and cognition in schizophrenia. Proc Natl Acad Sci U S A 10, (2004), 17288 –17293.

Stephan KE, Friston KJ, and Frith CD, Dysconnection in Schizophrenia: From Abnormal Synaptic Plasticity to Failures of Self-monitoring , Schizophrenia Bulletin, 2009, 35 (3), 509–527.

Suhara T. et al, Decreased Dopamine D2 Receptor Binding in the Anterior Cingulate Cortex in Schizophrenia , Arch Gen Psychiatry. 2002;59:25-30.

Swick, Turken, Dissociation between conflict detection and error monitoring in the human anterior cingulate cortex , PNAS, 2002, 99 (25), 16354 –16359.

Tanji, Mushiake, Comparison of neuronal activity in the supplementary motor area and primary motor cortex, Cognitive Brain Research, 1996, 3 (2), 143-150.

Van Schouwemburg M., Aarts E., Cools R., Dopaminergic Modulation of Cognitive Control: Distinct Roles for the Prefrontal Cortex and Basal Ganglia, Current Pharmaceutical Design, 2010, 16, 2026-2032

Veldic M, Caruncho HJ, Liu WS, Davis J, Satta R, Grayson DR, Guidotti A, Costa E, DNA-methyltransferase 1 mRNA is selectively overexpressed in telencephalic GABAergic interneurons of schizophrenia brains., PNAS, 2004, 6;101(1):348-53.

Veldic M, Guidotti A, Maloku E, Davis JM, Costa E., In psychosis, cortical interneurons overexpress DNA-methyltransferase 1., PNAS 2005 8;102(6):2152-7.

Wokke ME, van Gaal S, Scholte HS, Ridderinkhof KR, Lamme VAF (2011) The Flexible Nature of Unconscious Cognition. PLoS ONE 6(9): e25729.