Abstract

Orbitofrontal reality filtering (ORFi) is a memory mechanism that allows to sense if a memory pertains to reality or not. A deficit in ORFi is typically associated to amnesia, disorientation and behaviourally spontaneous confabulations. Electrophysiologically, it is characterized by a frontal positivity at 200-400ms. Anatomically, it is associated to an activation of the orbitofrontal cortex, the regions usually damaged in behavioural spontaneous confabulating patients. This thesis used electrophysiological and behavioural data to disentangle some aspects about ORFi that were still unclear. Taken together, our results showed that: 1) ORFi behaviourally and electrophysiologically dissociates from another memory mechanism, that is the ability to recall the temporal context of a memory; 2) ORFi electrophysiologically dissociates from memory's emotionality effects, and it is not modulated by the emotional value of a memory; 3) Children of 7 years old are already capable to perform the task, indicating that at this age ORFi is already functional.

Reference


URN : urn:nbn:ch:unige-844430
DOI : 10.13097/archive-ouverte/unige:84443
TITRE DE LA THÈSE

ORBITOFRONTAL REALITY FILTERING SPECIFIED: RELATION TO TEMPORAL ORDER MEMORY, ROLE OF EMOTION, AND DEVELOPMENT

THÈSE
Présentée à la
Faculté de Psychologie et des Sciences de l’Education
de l’Université de Genève
pour obtenir le grade de
Docteure en Neurosciences
par
Maria ChiaraLiverani

de Faenza, Italie

Thèse N° 170
Genève

Editeur ou imprimeur: Université de Genève
2016
DOCTORAT EN NEUROSCIENCES
des Universités de Genève et de Lausanne

Thèse de Maria Chiara LIVERANI

Intitulée : ORBITOFRONTAL REALITY FILTERING SPECIFIED: RELATION TO TEMPORAL ORDER MEMORY, ROLE OF EMOTION, AND DEVELOPMENT

*La Faculté de psychologie et des sciences de l'éducation, sur préavis du jury de thèse formé par Professeur Armin Schnider, directeur de thèse, Faculté de médecine, Université de Genève ; Docteur Radek Ptak, FPSE, Université de Genève ; Professeur David Sander, FPSE, Université de Genève ; Professeur Jean-Marie Annoni, Faculté de médecine, Université de Fribourg

autorise l'impression de la présente thèse, sans prétendre par là émettre d'opinion sur les propositions qui y sont énoncées.

Genève, le 22 avril 2016

Le doyen

Pascal Zesiger

Thèse No 170
Numéro d'immatriculation : 09.326.216

N.B. La thèse imprimée doit porter la déclaration précédente * et remplir les conditions énumérées dans les « Informations aux étudiants relatives aux thèses de doctorat à l'Université de Genève ». 
REMERCIEMENTS

Je tiens d’abord à remercier le Professeur Armin Schnider, pour son soutien, sa patience, son enthousiasme pour la recherche et pour m’avoir dévoilé tous les secrets et les mystères de l’ORFi. Merci d’avoir toujours pu trouver le temps de discuter de nos projets et pour tous vos dimanches passés à corriger mes articles et cette thèse.

Je remercie ensuite les Professeurs David Sander, Jean-Marie Annoni et le Docteur Radek Ptak pour avoir si gentiment accepté de faire partie de mon jury et de prendre en considération mon travail.

Merci à tous les participants et les patients qui on accepté de passer des heures dans une cage Faraday avec des électrodes sur la tête.

Merci à Mélanie Genetti pour sa disponibilité et de m’avoir appris les bases de l’EEG.

Un énorme merci à tous mes collègues de l’Hôpital de Beau-Séjour. Merci à Adrian Guggisberg pour sa disponibilité, ses conseils constructifs et ses consultations neurologiques gratuites. Merci à Radek Ptak (encore !!!) pour son soutien, ses mots de réconfort quand j’en avais besoin et ses conseils de personal trainer à la salle de sport. Un énorme merci à mes Aurélies: pour les fous rires pendant ces années, pour la patience avec laquelle vous avez répondu à mes mille questions sur l’EEG, et pour avoir toujours su me conseiller la bonne couleur de vernis à ongles. En particulier merci à Aurélie Bouzerda-Whalen qui m’a convaincue qu’il n’y a jamais de bons moments ni de mauvais moments, mais que des moments, et à Aurélie Manuel Stocker pour toutes les « frogs » qu’elle m’a fait manger. Vous êtes les meilleures collègues et copines que j’aurais pu avoir. Un grand merci à Elena Pedrazzini pour toutes ses attentions et ses commentaires « frontaux » qui nous ont fait beaucoup rire. Merci à Raphaël Thézé et à ses 1600 mots par minute qui m’ont permis d’apprendre le français beaucoup plus rapidement, et mes félicitations pour avoir survécu à nos pauses de midi « girly ». Merci à Julia Fellrath pour avoir partagé avec moi ce chemin pendant le doctorat, et pour m’avoir rendue un peu plus suisse. Merci à Tatiana Aboulafia Braka pour son énorme soutien et pour notre to-do list qui n’a jamais marché. Merci à Louis (Gigi) Nahum pour son aide et sa disponibilité, pour Sapristi, les farfadets, son imitation de Gollum et ses « Bonjour ! ». Merci à Anaïs Mottaz pour ses bons conseils de maman expérimentée. Merci à Pierre Nicolo pour être si drôle. Merci à Madame Fuchs pour sa patience et pour avoir toujours su régler nos bêtises administratives avec un grand sourire. Merci à Sabrina Sergi pour nos pauses de midi bio-ecolo-kilometre-zéro et pour être la seule personne à Beau-Séjour qui mange plus lentement que moi. Merci à Cristina Silva pour toutes nos rigolades et pour m’avoir toujours rappelé que sa fille dormait 10 heures par nuit alors que mon fils dormait 10 heures par semaine.

C’était un honneur de travailler avec vous et je ne vous oublierai jamais.
Un grand merci à Carlo Tomasetto, pour son amitié, ses mails de soutien et pour m’avoir donné le meilleur conseil de ma vie.

Merci à mes magnifiques copines : Floriana Di Stefano et son énergie positive, Martina Franchini et sa confiance en moi, Nele Dael et sa compréhension, Lara Lordier pour être ma choubidoux préférée, Benedetta Maimone pour être toujours présente et Ilaria Vitali pour être la meilleure « motivational coach » que je connaisse.

Un énorme merci à Mario et Carol Casini, pour nous avoir accueilli dans votre famille, pour aimer David comme un des vos petits enfants et pour tous les délicieux plats que vous m’avez préparé quand j’étais trop fatiguée pour cuisiner. Vous êtes les meilleurs grands-parents adoptifs du monde entier.

Merci à mes parents d’avoir toujours cru en moi et de me soutenir même à 700 kilomètres de distance. Merci pour toutes les heures passées en voiture pour me sauver la vie quand David était malade (quasi tous les jours) et de n’avoir jamais oublié de remplir votre voiture de gourmandises italiennes pour me réconforter. Sans vous, j’aurais pas pu arriver jusqu’ici.

Et pour finir, le plus grand merci va aux deux hommes de ma vie: Massi, pour être ma force, ma certitude et le meilleur des maris et des pères et David, pour tes yeux qui savent remplir notre vie de lumière.
ABSTRACT (ENGLISH)

Orbitofrontal Reality Filtering (ORFi) is a thought control mechanism that allows distinguishing between memories and thoughts that are related to the ongoing reality and memories that are not. The existence of this mechanism has been derived from the observation of patients affected by reality confusion, as evidenced by disorientation and confabulations that patients act upon. These patients usually have a lesion of the posterior medial orbitofrontal cortex or of a region directly connected with it. The memory disorder emanating from deficient ORFi has been experimentally verified using repeated runs of a continuous recognition task that reliably separated reality confusing patients from non-confabulating amnesic patients and controls. Electrophysiologically, ORFi is characterized by a frontal positivity between about 200 and 300 ms, indicating that it is a preconscious process. In healthy subjects, it is associated with activation of the posterior medial orbitofrontal cortex, the brain region that is usually damaged in reality confusing patients.

ORFi dissociates from other memory monitoring mechanisms like content monitoring and source monitoring from a behavioral, electrophysiological and clinical point of view. Conversely, its relation to temporal source monitoring, the ability to recall the correct temporal context in which a memory was acquired (temporal order judgment, TOJ), is still unclear.

In the first study of this thesis (Liverani, Manuel, Bouzerda-Wahlen, et al., 2015), we investigated commonalities and differences between ORFi and TOJ. We tested both mechanisms within a combined experimental paradigm. Healthy subjects performed the task while high-density evoked potentials were recorded. We found a behavioral dissociation, with lower accuracy and slower reaction time in the TOJ condition. Electrophysiologically, ORFi and TOJ share commonalities at an early stage of information processing, between 240 and 280 ms. They rapidly dissociate between 310 - 360 ms and again at 530 - 560 ms, and they induce activation of different brain areas. Thus, we concluded that ORFi and TOJ are two distinct mechanisms, an observation in line with the different clinical manifestations of the dysfunction of these two processes.

In the second study (Liverani, Manuel, Guggisberg, Nahum, & Schnider, 2016), we investigated whether ORFi is modulated by the positive emotionality of memory, based on studies claiming that confabulations have a positive emotional bias. We performed an electrophysiological study with healthy subjects using two runs of a continuous recognition
task with positive and neutral images. The ERP effects associated with items’ emotionality appeared in the same time window as ORFi, that is, between 260-350 ms. However, ORFi and emotionality had dissociated expressions in term of amplitudes and topography, confirming the dissociation between ORFi and emotionality.

In the last study (Liverani, Manuel, Nahum, et al., 2015), we explored the development of ORFi in school age children, to understand at what age this memory mechanism is functioning. An adapted version of the continuous recognition task was administered to 7, 9 and 11 years old children. The main finding was that ORFi develops in parallel with recognition capacity. Younger children were already capable to perform the task above chance, indicating that at this age ORFi is already functioning at a level corresponding to the memory capacity at this age. Until the age of 9 recognition memory and ORFi correlate. At 11 year of age on they do not correlate anymore, similar to adults.

These studies allowed to better understand basic aspects of ORFi that were still unclear. They show that it is independent from other memory monitoring mechanisms like source monitoring, that it is not influenced by the positive valence of memories, and that it develops early in childhood, allowing children to base their actions on thoughts that relate to ongoing reality.
ABSTRACT (FRENCH)

Le filtre de la réalité (Orbitofrontal reality filtering, ORFi) est un mécanisme de monitoring de la mémoire permettant de distinguer les souvenirs et pensées se référant, ou non, à la réalité. Ce mécanisme de mémoire est proposé sur la base de l’observation de patients souffrant de confusion de la réalité. Il s’agit d’un trouble caractérisé par la désorientation et l’amnésie ainsi qu’une tendance aux confabulations et à agir sur la base de celles-ci. Ces patients présentent souvent des lésions au cortex orbitofrontal ou à des régions qui y sont directement connectées. L’ORFi a été testé expérimentalement avec une tâche de reconnaissance continue permettant d’isoler avec fiabilité les patients souffrant d’une confusion de la réalité des autres patients amnésiques ou des sujets contrôles. D’un point de vue électrophysiologique, l’ORFi est caractérisé par une positivité frontale entre 200 et 300 ms, indiquant qu’il s’agit d’un processus précoce. Anatomiquement, il est associé avec l’activation du cortex orbitofrontal médial postérieur, la région du cerveau notamment lésée chez les patients souffrant de confusion de la réalité.

L’ORFi, d’un point de vue comportemental, électrophysiologique et clinique, se dissocie des autres mécanismes de monitoring de la mémoire tels que le monitoring de contenu ou de source. Par contre, son rapport avec le monitoring de la source temporelle d’un souvenir – c'est-à-dire la capacité de se rappeler correctement son contexte temporel (Temporal Order Judgment, TOJ) – n’est pas encore clair.

Dans la première étude de cette thèse (Liverani, Manuel, Bouzerda-Wahlen, et al., 2015) nous avons investigué les similarités et différences entre l’ORFi et le TOJ. Nous avons testé les deux mécanismes dans un même paradigme expérimental au cours duquel nous avons demandé à des sujets sains d’effectuer cette tâche tout en enregistrant des potentiels évoqués à haute densité. Nous avons d’abord trouvé une dissociation comportementale, avec une performance moindre et des temps de réaction plus lents dans la condition TOJ. Électrophysiologiquement, l’ORFi et le TOJ partagent des similarités pendant l’étape précoce du traitement de l’information, entre 240 et 280 ms. Les deux mécanismes se dissocient rapidement entre 310 et 360 ms puis encore entre 530 et 560 ms. De plus ils se démarquent par l’activation de différentes régions du cerveau. En conséquence, nous avons conclu que l’ORFi et le TOJ sont deux mécanismes différents. Cette proposition est en accord avec les différentes observations cliniques associées au dysfonctionnement de deux processus de mémoire.
Dans la deuxième étude (Liverani et al., 2016) nous avons investigué si l’ORFi est modulé par l’émotionalité positive des souvenirs. Nous avons effectué une étude électrophysiologique avec des sujets sains qui ont complété une tâche de reconnaissance continue avec des images positives et neutres. Les effets ERP associés à l’émotionalité des images étaient visibles dans la même fenêtre temporelle que celle où l’ORFi s’exprime habituellement, c'est-à-dire entre 260 et 350 ms. Cependant, les deux conditions se dissociaient en termes d’amplitudes et de topographie, confirmant ainsi la dissociation entre l’ORFi et l’émotionalité des images.

Dans la dernière étude (Liverani, Manuel, Nahum, et al., 2015) nous avons exploré le développement de l’ORFi chez les enfants en âge scolaire, pour comprendre à quel âge ce mécanisme de mémoire se développe. Nous avons administré une version adaptée de la tâche de reconnaissance continue à des enfants de 7, 9 et 11 ans. Nous avons trouvé que les plus jeunes enfants étaient déjà capables de compléter la tâche avec un taux de réussite au-dessus du hasard, indiquant qu’à cet âge l’ORFi s’était déjà développé. Une analyse de corrélation a montré que l’ORFi se développe en parallèle avec la capacité de reconnaissance jusqu’à l’âge de 9 ans, et qu’à partir de 11 ans il devient indépendant.

Ces études ont permis de mieux comprendre certains aspects de l’ORFi qui n’étaient alors pas encore clarifiés, démontrant ainsi son indépendance des autres mécanismes de monitoring de la mémoire, tel que le monitoring de source. De plus il a été montré que l’ORFi n’est pas influencé par la valence positive des souvenirs et qu’il se développe tôt au cours de l’enfance, permettant aux enfants de baser leurs actions sur des pensées pertinentes avec la réalité.
# TABLE OF CONTENTS

ABSTRACT (ENGLISH) .................................................................................................................. 5
ABSTRACT (FRENCH) ................................................................................................................... 7

1. INTRODUCTION...................................................................................................................... 11
   1.1 Memory systems .................................................................................................................. 11
   1.2 Orbitofrontal reality filtering and reality confusion ............................................................ 13
      1.2.1 Task .............................................................................................................................. 14
      1.2.2 Electrophysiological correlates ..................................................................................... 15
      1.2.3 Anatomical correlates from neuroimaging ..................................................................... 17
      1.2.4 Model ........................................................................................................................... 18
      1.2.5 ORFi and extinction ...................................................................................................... 19
   1.3 ORFi and other memory monitoring mechanisms ............................................................... 20
      1.3.1 Orbitofrontal reality filtering vs. Working with memory ............................................... 21
      1.3.2 Orbitofrontal reality filtering vs. Source monitoring ..................................................... 24
         1.3.2.1 Electrophysiological correlates of source monitoring and its relation to ORFi ........ 26
   1.4 Questions explored in this thesis: Background ................................................................. 27
      1.4.1 Orbitofrontal reality filtering and temporal order judgment ........................................ 28
      1.4.2 ORFi and positive emotions ........................................................................................ 29
      1.4.3 The early development of ORFi .................................................................................. 31
   1.5 Evoked potential methods .................................................................................................. 32

2. RESULTS.................................................................................................................................. 35

3. DISCUSSION............................................................................................................................. 38
   3.1 ORFi and Temporal order judgment .................................................................................... 38
   3.2 ORFi and positive emotions .............................................................................................. 40
   3.3 The early development of ORFi ......................................................................................... 44
   3.4 Is ORFi important for psychosis and schizophrenia? ......................................................... 47

4. CONCLUSION............................................................................................................................ 49
   4.1 Limitations and future perspectives .................................................................................... 49

5. REFERENCES............................................................................................................................ 51

6. ARTICLES................................................................................................................................. 61
LIST OF ABBREVIATIONS

ORFi = orbitofrontal reality filtering
ms = millisecond
OFC = orbitofrontal cortex
PFC = prefrontal cortex
DRM = Deese-Roediger-McDermott paradigm
DLPFC = dorso lateral prefrontal cortex
EEG = electroencephalography
ERP = event-related potential
FOR = feeling of rightness
MTL = medial temporal lobe
PET = H2O positron emission tomography
WWM = working with memory
VLPFC = ventro lateral prefrontal cortex
VMPFC = ventro medial prefrontal cortex
SMF = source monitoring framework
fMRI = functional magnetic resonance
TOJ = temporal order judgment
1. INTRODUCTION

Think of your last summer holidays. You were on a wonderful beach, wearing your swimsuit, sunbathing on a beach chair and drinking a fresh juice. The smell of the sea, the heat of the sun on your skin, the sound of the waves against the cliffs are all very vivid memories to you...almost real. But thanks to your brain, and more precisely thanks to your orbitofrontal cortex, you are completely aware that these memories belong to the past; this is not now! Thus, you are not going to wear your swimsuit and thongs to go to the beach. You will continue to read this thesis, seated on your chair, in cloudy and windy Geneva, in the month of February 2016. But how does our brain prevent memories about the past from wrongly guiding our behaviour? This is the topic of this thesis. But let’s go step by step. First of all: what is memory?

1.1 Memory systems

Memory entails the abilities to encode, store, and subsequently recall events and experiences. It allows us to remember past experiences, but it is also the essence of our thoughts that guide current behaviour and influence our future projects.

Memory is composed of different systems mediated by distinct neural structures that cooperate together in the process of memorization (Tulving, 1985). The first distinction consists in the different duration of information retention: from 10-15 seconds up to some minutes for short-term memory, or until the whole lifetime for long-term memory. Another differentiation is between implicit and explicit memory (Schacter & Graf, 1986). Implicit memory allows the unconscious acquisition of skills through practice and experience. Explicit memory refers to memories of facts and events that can be consciously acquired and recalled. It is further divided in two subtypes: episodic memory, which represents our experiences and
specific events in time, and *semantic memory*, which is a record of concepts, facts and knowledge about the external world independent of our personal experience or of the context in which they have been acquired (Schnider, 2008b; Squire & Zola, 1998; Tulving & Craik, 2000; Tulving & Schacter, 1990).

In this thesis, I will focus on episodic memory and its implications for our behaviour.

Many theories have tried to explain how our brain encodes new information and how information is stored in memory, consolidated and then, eventually, retrieved. Evocation of a memory is not a simple reproduction of the past, but instead a constructive process in which various information from different sources are bound together. This reconstructive capacity also permits individuals to simulate and anticipate future experiences, events and scenarios (Schacter & Addis, 2007).

A fundamental but rarely studied aspect of memory is its role in the interpretation of the present, the "now". How can the brain distinguish between memories (thoughts) that relate to the ongoing reality and thoughts that do not? How does it shape behaviour? It appears reasonable to assume that, without this ability, memories that are inappropriate guides of present actions could interfere and erroneously guide our behaviour or shape our future plans. Patient with this failure exist: they are amnesic, confabulate, act according to these confabulations, and are disoriented in time, space and regarding their current obligation. The disorder has been called "behaviorally spontaneous confabulation" (Schnider, 2008a, 2013).

The *orbitofrontal reality filtering theory* (ORFi) has been derived from the observation of such patients and concerns specifically this aspect of memory: the reality check of upcoming memories (thoughts), at the basis of the adaptation of behaviour to ongoing reality (Schnider, 2008a).
The aim of this thesis is to disentangle aspects of ORFi that are still unclear. In the following paragraphs, I will explain the basis of the theory, juxtapose it with competing theories, and lay out the aspects of ORFi explored in this thesis work.

1.2 Orbitofrontal reality filtering and reality confusion

The memory mechanism of orbitofrontal reality filtering is derived from the observation of patients affected by reality confusion. This disorder is characterized by disorientation regarding time, space and current situation, amnesia and confabulations that patients act upon. The false actions typically correspond to previous habits (Schnider, 2008a, 2013). For example, a dentist hospitalized for a rupture of an aneurysm of the anterior communicating artery repeatedly attempted to leave the hospital to go to his clinic in the certainty that patients were waiting for him (Ptak & Schnider, 1998). A hospitalized 58-year-old woman tried to leave the hospital to go to feed her baby, who at the time was 35 years old (Schnider, von Daniken, & Gutbrod, 1996b). Experimental exploration indicated that these patients fail to verify the relation of a memory or thought with the “now”.

Figure 1. Brain regions that are usually damaged in reality confusing patients. Lesions are usually in the posterior medial orbitofrontal cortex and basal forebrain (darker region), but can also involve hippocampus, amygdala and perirhinal cortex (gray regions). Adapted from Schnider (2003).
The disorder is typically caused by traumatic brain injury, encephalitis, rupture of an aneurysm of the anterior communicating artery or due to a deficiency in vitamine B1 (thiamine) inducing the Wernicke-Korsakoff Syndrome (Schnider, 2008a).

The critical variable is the site of brain damage (Figure 1): reality confusing patients, who are out of a confusional state, most commonly have a lesion of the posterior medial orbitofrontal cortex or of a brain region that is directly connected with it (Schnider, 2008a; Schnider & Ptak, 1999; Schnider, von Daniken, & Gutbrod, 1996a).

1.2.1 Task

![Figure 2. Continuous recognition task used to test ORFi.](image)

Figure 2. Continuous recognition task used to test ORFi. The same set of images is presented in the two passages (first and second run). In both runs, participants are asked to indicate picture recurrences only within the ongoing run, answering to the question “Have you already seen this image in this run?”. Distracters 1 (d1) and distracters 2 (d2) are images seen for the first time in the first and in the second run, respectively. Targets 1 (T1) and targets 2 (T2) are images already seen within the first and second run, respectively. Adapted from Schneider (2008).

Figure 2 shows the experimental paradigm that reliably separated amnesic patients who confused reality and other amnesics. It consists of two runs of a continuous recognition task.
in which the same line drawings are repeated (Schnider, von Daniken, et al., 1996b). Subjects are asked to indicate picture recurrences only within the ongoing run. In the first passage, which measures recognition capacity, familiarity alone is sufficient to correctly perform the task. In the second run, when all the pictures are known from the previous run, familiarity is not enough to discriminate the first presentation of an image. A mechanism that permits to “filter” the upcoming memories of the previous occurrence of an item, and to sense whether familiarity stems from the ongoing run (the "ongoing reality") or not is needed to correctly accomplish the task (Schnider, 2008a).

In comparison to non-confabulating amnesic patients, reality confusing patients typically fail in this task: they perform the first run as well - or badly - as other amnesics, but they have a performance drop in the second one, which is specifically due to a failure to reject images seen for the first time within the second run. This increase in false positives strongly correlates with the level of disorientation (Nahum, Bouzerda-Wahlen, Guggisberg, Ptak, & Schnider, 2012a; Schnider, von Daniken, et al., 1996b). Recovery from reality confusion with cessation of confabulations is accompanied by the normalization of false positives in the second run of the task (Schnider, Ptak, von Daniken, & Remonda, 2000).

1.2.2 Electrophysiological correlates

In 2002, Schnider and colleagues investigated the electrophysiological correlates of the orbitofrontal reality filtering in healthy subjects (Schnider, Valenza, Morand, & Michel, 2002). They wanted to know when precisely ORFi occurred, in relation to the recognition of learned material. To do this, they used high-resolution event-related-potentials (ERPs), a method that, thanks to its excellent temporal resolution, can record electrical activity of neuronal cells on a millisecond time scale (Michel et al., 2004; Slotnick, 2010).
The task used was a variation of the task in which reality confusing patients had failed. Specifically, participants performed two runs of the continuous recognition task, but with fewer repetitions, shorter interval between the runs, and separated in two blocks (to avoid effects of fatigue). As explained above, the first run of the task was thought to test learning and recognition, while in the second run, the images presented for the first time within the run (called *distracters 2*) was used to test ORFi. Simple waveform analysis (a type of analysis of the electrophysiological signal that will be detailed below) revealed that learning and recognition were characterized by a posterior positivity peaking at about 400-500 ms after stimulus onset. In the second run, the electrophysiological response associated to distracters 2 - the stimuli on which confabulating patients had failed - revealed the absence of a negative deflection, common to all other stimuli, in the frontal region at about 200 - 300 ms (Fig.3). This finding was interpreted as the electrophysiological signature of ORFi (Schnider et al., 2002) and has been replicated in several other studies (Bouzerda-Wahlen, Nahum, Liverani, Guggisberg, & Schnider, 2014; Liverani, Manuel, Bouzerda-Wahlen, et al., 2015; Liverani et al., 2016; Wahlen, Nahum, Gabriel, & Schnider, 2011).

![Figure 3. Evoked potential responses in response to the four conditions of the task in frontal (Fz), central (Cz) and posterior (Pz) electrodes.](image-url)

The red arrow indicates the electrophysiological response to distracters 2, the stimuli critical for ORFi. Adapted from Schnider et al. (2002).
Furthermore, segmentation analysis (a technique that allows to determine periods of stable configuration of electrocortical activity characterizing a defined cognitive process (Michel et al., 2004; Murray, Brunet, & Michel, 2008)) revealed that at 200 - 300 ms, distracters 2 skipped an electrocortical configuration map that was present for all the other types of stimuli. Inverse solution analysis, which estimates the brain sources of the electrophysiological signal (Michel et al., 2004) showed that this stage was associated with an extended activation of neocortical association areas, as it occurred in response to all stimuli except those requiring ORFi, namely, the distracters of run 2 (Schnider, 2008a). The results suggested that reality filtering works by momentary inhibition of extended neocortical synchronization at around 200 - 300 ms after evocation of a thought that does not relate to the "now" (Schnider, 2008a; Schnider et al., 2002).

1.2.3 Anatomical correlates from neuroimaging

A neuroimaging study on healthy subjects using an adapted version of the same task further investigated the neural correlates of ORFi in healthy subjects. It used H_2O positron emission tomography (PET) (Schnider, Treyer, & Buck, 2000). The first run of the task was associated with a medio-temporal (parahippocampal) activation that reflected learning and recognition. In the subsequent third run, which requires ORFi, there was activation of the posterior medial orbitofrontal cortex, Brodmann area 13. This finding remained consistent both using pictures (Schnider, Treyer, et al., 2000; Treyer, Buck, & Schnider, 2003) and meaningful sounds (Treyer, Buck, & Schnider, 2006). This area corresponds to the region typically damaged in patients suffering from reality confusion (Schnider & Ptak, 1999; Schnider, Ptak, et al., 2000; Schnider, von Daniken, et al., 1996a).
1.2.4 Model

These electrophysiological and neuroimaging findings allowed delineating a functional model to describe the ORFi (Schnider, 2008a) (Fig. 4). Orbitofrontal reality filtering is an early process, which starts even before the content of an upcoming memory is consciously recognized. The model posits that 200 - 300 ms after its evocation, the upcoming memory (thought) passes through the orbitofrontal reality filter, which determines whether the evoked memory has a correlate in (refers to) ongoing reality or not, in the latter case representing a fantasy. According to the electrophysiological data, a memory that does not relate to reality skips the processing stage associated with extended neocortical synchronization at 200 - 300 ms. The memory thus has the format of a thought relating to reality or a fantasy. At 400 - 600 ms the memory is recognized and again encoded.

In reality confusing patients ORFi is dysfunctioning. After the activation by a cue, memories do not pass the filter at 200 - 300 ms, and thus are all labelled as “reality”, even if they do not pertain to the present situation. Patients thus confuse reality and act according to this false reality. Their verbal statements about their perceived reality are evident as confabulations.

Figure 4. Functional model of the orbitofrontal reality filtering. Adapted from Schnider (2008).
1.2.5 ORFᵢ and extinction

The behaviour of reality-confusing patients suggests a failure of extinction capacity, the ability to learn that a previously valid outcome is no longer valid. This hypothesis was put to the test in diverse studies.

Schnider et al. (2007) and Nahum et al. (2011) performed EEG studies in which healthy participants performed a reversal-learning task, a task that includes learning of an association and extinction. The ERP correlate of this process was a frontal positivity at 200 - 300 ms, similar to the period in which ORFᵢ in the continuous recognition task is expressed (Nahum, Gabriel, & Schnider, 2011; Schnider, Mohr, Morand, & Michel, 2007). An fMRI study showed that extinction capacity in this task indeed activated the orbitofrontal cortex (Nahum, Simon, Sander, Lazeyras, & Schnider, 2011).

A simplified version of the same reversal-learning task was used with disoriented, confabulating patients and with non-confabulating amnesic patients. In contrast to the non-confabulating amnesics, reality-confusing patients specifically failed in extinction capacity (Nahum, Ptak, Leemann, & Schnider, 2009). The finding strongly supported the claim that ORFᵢ depends on extinction capacity.

Furthermore, ORFᵢ appears to depend on structures of the reward system. In the PET study cited above (Treyer et al., 2003) authors demonstrated that the performance of the continuous recognition task was associated with the activation of a subcortical loop connecting the OFC with the striatum, substantia nigra and medial thalamus, suggesting the involvement of the dopaminergic system in the reality filtering mechanism. These structures are the anatomical core of the reward system. Primate experiments demonstrated the existence of orbitofrontal neurons that specifically increase their firing when an anticipated reward does not occur; they process negative prediction errors (Rosenkilde, Bauer, & Fuster, 1981). Dopaminergic neurons in the substantia nigra and striatum decrease firing in response to such events.
Thus, Schnider and colleagues hypothesized that a condition of augmented dopamine would impair ORFi. This appears indeed to be the case: when healthy subjects under the effect of L-DOPA, a dopamine precursor and a crucial transmitter in the reward system, performed a very challenging version of the continuous recognition task described above, they had a significantly higher rate of false positives compared to when they received risperidone, a dopamine antagonist (Schnider, Guggisberg, Nahum, Gabriel, & Morand, 2010). In addition, a confabulating patient significantly decreased his false memories and confabulations after a treatment with risperidone (Pihan, Gutbrod, Baas, & Schnider, 2003). These observations suggest that ORFi is a capacity conveyed by the reward system and is under dopaminergic control (Schnider, 2008a, 2013).

1.3 ORFi and other memory monitoring mechanisms

The ORFi theory is in competition with different memory monitoring hypotheses, which tried to explain how the brain verifies the accuracy of memories about the past. Several models have been conceived to describe the processes involved in the retrieval of the content of memories or the context under which they were acquired. As memory is not a simple replay of the past but a constructive process, these theories have also tried to investigate to which extent memory retrieval is prone to errors, leading to memory distortions, false memories or confabulations.

The specificity of ORFi in the explanation of reality confusion and its independence from other memory monitoring models have been questioned (Gilboa et al., 2006), motivating our research group to perform a series of studies - detailed below - aimed to test the specificity of ORFi as a memory mechanism. In the following, I will introduce the two main models about the monitoring of the content and the context of a memory. These studies showed that ORFi is indeed independent of these mechanisms.
1.3.1 Orbitofrontal reality filtering vs. Working with memory

![Diagram of WWM model]

**Figure 5. The WWM model.** The red arrow indicates the direct route of retrieval, the blue arrow indicates the indirect route. VLPFC = Ventrolateral Prefrontal Cortex, VMPFC = Ventromedial Prefrontal Cortex, DLPFC = Dorsolateral Prefrontal Cortex. Adapted from Gilboa (2009).

A model about the retrieval of the content of a memory is the *Working with memory model* (WWM, Fig. 5) (Gilboa et al., 2006; Moscovitch & Melo, 1997; Moscovitch & Winocur, 2002). Moscovitch and Wincour conceived this theoretical framework to describe the role of prefrontal cortex in memory retrieval.

According to this model, recuperation of memory traces can pass through two different routes. The first one, the *associative cue-dependent retrieval*, is a direct route in which a cue directly activates the memory representation (*ecphory*) associated to the hippocampal-
neocortical ensemble (Figure 5, red arrow). The second route (Figure 5, blue arrow) is indirect and involves strategic processes mediated by structures of the prefrontal cortex. If the retrieval cue is inadequate to reactivate the memory trace, a search strategy is established through the dorsolateral PFC, according to the goals of the task (strategy formulation). Then the ventrolateral PFC would specify the congruous retrieval cues necessary to the specific memory task (cue specification) (Gilboa, 2009). The search process to generate retrieval cues is a reiterative process that continues until the correct cue is found. These steps constitute what are called the “constitutional mechanisms” of the theory. Once the memory trace is activated, a verification process signals the acceptance or rejection of the retrieved memory on the basis of an intuitive “feeling of rightness” (FOR). This feeling is mediated by the ventromedial and orbitofrontal PFC. The next step consists of a “problem-solving” process controlled by the dorsolateral PFC and neocortical regions. Here, the compatibility of the retrieved memory with other memories and knowledge such as the retrieval context, semantic knowledge, and so on is checked (monitoring) (Gilboa, 2009). Finally, the dorsolateral PFC would select the most adapted and pertinent response (response decision). These three steps represent the “core mechanisms”. According to this model, a deficit in the core mechanisms could be responsible for the occurrence of confabulations.

In 2006 Gilboa and colleagues argued that the ORFi mechanism was not sufficient to explain the origin of confabulations, and hypothesized that it could be a sub-component of the WWM framework (Gilboa et al., 2006). They based their argument on the observation that confabulating patients not only failed in the task developed by Schnider et al. (1996b), but also in an additional continuous recognition task in which the repeated images were either true repetitions or pictures only resembling the previously presented images. According to the authors, ORFi could correspond to the preconscious FOR process (Feeling of Rightness).
The main limitation of this study was that patients were not matched regarding the degree of amnesia or any other cognitive measure. In fact, the confabulating patients also performed worse on other cognitive tasks of executive control and memory, which casts doubt on the specificity of the authors' claims.

In an additional study, Gilboa and colleagues (2009) investigated the electrophysiological correlates of the WWM model. Using a recognition task with familiar, famous and non-famous faces, they suggested that the FOR process was associated with an early positive modulation in the frontal region of the scalp at about 230 - 260 ms. On the basis of the similarity between the potential specific to the FOR and the one specific to the ORFi, the authors suggested that the mechanism that allows to filter pertinent memories could simply be a sub-component of the more general monitoring of the content of a memory (Gilboa et al., 2009). However, this study did not include active confabulators.

Bouzerda-Wahlen and colleagues (Bouzerda-Wahlen, Nahum, Ptak, & Schnider, 2013; Wahlen et al., 2011) directly compared strategic content monitoring and ORFi. They conceived an electrophysiological and a clinical study in which the two mechanisms were juxtaposed in the same continuous recognition task. Half of the repeated images were exactly the same already presented (true repetitions, critical to test ORFi), while the other half only resembled previous ones (similar repetition, critical to test content monitoring).

In the first study (Wahlen et al., 2011) healthy subjects performed the task while high-resolution evoked potentials were recorded. The authors found a behavioural and electrophysiological dissociation. Behaviourally, subjects were slower and less accurate in response to stimuli demanding content monitoring than to stimuli testing ORFi. Electrophysiologically, the two processes dissociated at 200 - 300 ms: while content monitoring was associated with a strong negative frontal potential, ORFi was characterized by
a particularly weak negative frontal potential (the most positive one of all stimuli), similar to the previous study (Schnider et al., 2002).

In addition, true repetitions were associated with a stronger positive frontal potential compared to all the other stimuli and with activation of the ventro-medial prefrontal cortex (Wahlen et al., 2011). The authors interpreted this result as reflecting the FOR process, supposed to be responsible for signalling whether a retrieved memory is adequate to the task’s goal or not (Gilboa et al., 2006; Moscovitch & Melo, 1997).

In a clinical study (Bouzerda-Wahlen et al., 2013), non-confabulating amnesics and reality confusing patients performed the same task as above. The degree of amnesia and executive failure was similar in both groups. The study confirmed the dissociation between the two mechanisms: reality confusion, as measured by the degree of disorientation, significantly correlated with performance on stimuli measuring ORFi but not with the processing of stimuli measuring strategic content monitoring.

A further piece of evidence comes from a single case observation. Nahum et al. (2010) described a woman suffering from severe reality confusion due to encephalitis. While her performance was impaired in the continuous recognition task testing ORFi, she did not produce any confabulation in tasks supposed to test strategic content monitoring.

These finding underscore the idea that ORFi is a distinct memory control mechanism that behaviourally, electrophysiologically and clinically dissociates from strategic content monitoring.

1.3.2 Orbitofrontal reality filtering vs. Source monitoring

In 1981, Johnson and Raye proposed a model for the capacity to attribute a memory or a thought to an external or an internal source, the reality monitoring framework (Johnson & Raye, 1981). The concept of reality monitoring later became a sub-set of the more extended
source monitoring framework (SMF) (Johnson, 1997; Johnson, Hashtroudi, & Lindsay, 1993; Mitchell & Johnson, 2009), which encompasses processes required in the identification of the source of the information.

Memories have various characteristics that are the results of different cognitive processes during the encoding of the event and when the memory is reactivated.

The term source refers to all the characteristics that define the modalities and conditions under which a memory was acquired, including the temporal, spatial or social context of an event, the modalities under which it was perceived, the perceptual, semantic, emotional information, and the cognitive processes engaged during memory formation (Johnson et al., 1993; Mitchell & Johnson, 2009). Thus, source monitoring concerns all processes needed to make attributions about the origin of thoughts, memories or knowledge. When all these specific characteristics are bound together, an event can be differentiated from another, and differences between these attributes are used to identify the origin of different memories (Johnson, 1997).

The SMF holds that recollection of a memory depends on the amount and the distinctness of remembered details specific to this memory. Therefore, on the basis of the number of details, the subjective experience may range from a relatively undifferentiated feeling of familiarity to a precise sense of recollection (Leynes & Phillips, 2008). The more a memory characteristic is specific to an event, the better the memory will be retrieved. Conversely, a high degree of similarity between two sources may lead to source errors. A deficit in source monitoring, similar to orbitofrontal reality filtering, has been proposed as a cause of confabulations (Johnson et al., 1993; Mitchell & Johnson, 2009), suggesting possible commonalities between the two mechanisms.

In the following, I will first describe studies that attempted to characterize the electrophysiological signature of source monitoring, and secondly, the studies directly
comparing a particular type of source monitoring - context monitoring - with orbitofrontal reality filtering.

1.3.2.1 Electrophysiological correlates of source monitoring and its relation to ORFi

Johnson et al. (1997) presented a series of words and pictures to healthy subjects. During the test phase, they asked them either to discriminate old stimuli from new ones, or to indicate the source of the presented items (whether the stimulus had been presented as a word or a picture). Electrophysiological recording during the task revealed a dissociation between the old/new discrimination task and the source discrimination task at about 1400 ms after stimulus presentation in frontal regions.

Other studies tried to disentangle the ERP components associated with source monitoring. The parietal old-new effect, which occurs at about 600 - 800 ms and is largest over left parietal regions, is thought to reflect recollection because it has higher amplitude for recollected than non-recollected items (Leynes & Phillips, 2008; Tsivilis, Otten, & Rugg, 2001). Another characteristic ERP component is a frontal effect starting at 800 ms and reflecting post-retrieval decision processes described in the source monitoring framework (Leynes & Phillips, 2008; Tsivilis et al., 2001). Thus, processes of source monitoring appear to be expressed at the earliest after 600 ms.

Impaired source monitoring has been suggested as an explanation of confabulations (Johnson et al., 1993; Mitchell & Johnson, 2009). The only experimental study performed by Johnson et al. (1997) did, however, not support the claim: the confabulating patient did not differ from non-confabulating amnesics on tasks of source monitoring.

The claim that source monitoring deficits may explain confabulations has persisted. For this reason, Bouzerda-Wahlen et al. (2014) directly compared a facet of source monitoring - the monitoring of context - with ORFi using EEG. In their study, the continuous recognition task
previously used to test ORFi was adapted to test also context source monitoring, which is the verification of the context under which a memory was acquired. Images were presented with one of two different visual contexts, namely, a frame consisting either of red and blue circles, or of green and yellow squares. During each run, healthy participants were asked to indicate if they had already seen the picture within the same context and within the ongoing run. While in the first run images were always presented with the same context (to ensure encoding), they could have the same or the alternate context during the second presentation in the second run. Analysis indicated a dissociation between the two mechanisms. Behaviourally, the retrieval of the context associated with an item was effortful, as indicated by slower reaction times and higher error rates compared to reality filtering. Electrophysiologically, ORFi was characterized by the typical absence of a negative potential in the frontal region, reflecting posterior orbitofrontal activation, at 200 - 300 ms, similar to previous studies (Schnider et al., 2002; Wahlen et al., 2011). By contrast, source monitoring was expressed by a prolonged frontal positivity from 400 ms on. The authors hypothesized that this late potential reflected elaborate and effortful processing during the retrieval of the memory’s source. They concluded that, like the comparison between ORFi and strategic content monitoring, the ability to filter memories that are not pertinent to reality is independent of the ability to recollect the source of a memory (Bouzerda-Wahlen et al., 2014).

1.4 Questions explored in this thesis: Background

The studies described so far indicate that ORFi is distinct from other memory monitoring mechanisms, and is important in adults to maintain thinking and behaviour in phase with reality.
In this thesis, the following questions concerning ORFi were examined:

1. What is the relation between ORFi and the ability to consciously retrieve the temporal context of a memory?

2. Is there an influence of emotions on ORFi or is it independent from the emotionality of memories?

3. How does ORFi develop in children?

In the following, I will summarize current knowledge regarding these questions.

1.4.1 Orbitofrontal reality filtering and temporal order judgment

The ability to determine when in the past a memory was acquired - that is, knowledge about its temporal context - is considered a component of source monitoring. It is usually tested with recency or temporal order judgment tasks. Typically, two series of stimuli are presented to participants at different points in time. During the test phase they are asked to indicate which one of two items was encountered more recently (recency task), or they have to indicate whether an item belongs to the first or the second list (temporal order task) (e.g. (Duarte, Henson, Knight, Emery, & Graham, 2010; Tendolkar & Rugg, 1998)).

Temporal order deficits have been proposed as a possible explanation of confabulations (Dalla Barba, 1997; Talland, 1961; Van der Horst, 1932). Clinical evidence remains weak. In a single case study, Schnider et al (1996) found that a patient suffering from amnesia and confabulation indeed failed in a task were the retrieval of the temporal and spatial context of items was required. However, they did not test any amnesic control subjects. Johnson (1997) made the same observation in a confabulating patient. However, non-confabulating amnesics had the same difficulty, indicating that the deficit was not specific for confabulation. In fact, deficient temporal judgment can also be present in amnesic patients who do not confabulate.
(Hirst & Volpe, 1982; Kopelman, Stanhope, & Kingsley, 1997; Squire, Nadel, & Slater, 1981).

Many studies support the role of the prefrontal cortex in the retrieval of the temporal context: patients with frontal lesions may have impaired memory for recency and temporal order even if they have normal performance in item recognition (Milner, Corsi, & Leonard, 1991; Shimamura, Janowsky, & Squire, 1990).

These clinical results leave opened the question whether temporal order judgment shares commonalities with ORFi.

Previous studies have already investigated the independence of ORFi from strategic content monitoring (Bouzerda-Wahlen et al., 2013; Wahlen et al., 2011) and context monitoring (Bouzerda-Wahlen et al., 2014). Up to date, no study has investigated the relationship between temporal order judgment and ORFi within the same experimental paradigm and in the same temporal frame (around the present). Previous studies on temporal order memory presented items separated by minutes (Tendolkar et al., 2004); ORFi is tested in rapid succession in healthy subjects. In the first study of this thesis, we conceived a new paradigm in which both temporal order judgment and ORFi were tested within a similar time frame, using behavioural and electrophysiological analysis in healthy subjects. The hypothesis was that ORFi, an automatic and intuitive mechanism, would behaviourally and electrophysiologically dissociate from temporal order judgment, an effortful process.

1.4.2 ORFi and positive emotions

Another open question concerns the relationship between ORFi and emotions: is this mechanism influenced by memory emotionality?

The question has its roots both in clinical and anatomical observations.
Firstly, several studies claimed that the content of confabulations has a positive emotional bias (Fotopoulou, 2010; Fotopoulou, Conway, Griffiths, Birchall, & Tyrer, 2007; Fotopoulou et al., 2008). Fotopoulou et al. (2007) described a patient who confabulated after an aneurysm of the anterior communicating artery. When asked to provide both pleasant and negative self-defining autobiographical experiences, he produced confabulations characterized by a more positive valence compared to his true recollected memories and to memories produced by control subjects. Similar results have been found in a group of 10 confabulating patients when false memories were elicited: they had the tendency to distort past memories in a more self-enhancing way compared to controls (Fotopoulou et al., 2008). According to the authors, these positive contents in confabulations could be determined by emotional mechanisms that are under the control of the orbitofrontal cortex: impairment in this area is thought to lead to an exaggeration of self-serving memories (Conway & Tacchi, 1996).

These conclusions are, at least partially, in contrast with our clinical observation. Reality confusing patients mainly confabulate on non-emotional, common habits: for example, a lawyer was convinced to be expected at court and for weeks searched her files to prepare her advent at the tribunal (Nahum et al., 2010). It is thus possible that the particular type of confabulation that emanates from reality confusion is not characterized by this positive emotional bias as the other purely verbal confabulations. We, therefore, hypothesized that the mechanism of orbitofrontal reality filtering is not affected by the emotionality of memories.

The other observation comes from an anatomical point of view: the orbitofrontal cortex is part of the anterior limbic system, which is the most important brain region involved in emotional processes. Emotions influence the three components of episodic memory: encoding, consolidation and retrieval, especially throughout the modulation of attention and perception by the amygdala (Phelps, 1996).
Based on these observations, we wanted to investigate the relationship between memories’ emotionality and the limbic capacity to sense if an activated memory refers to the present reality or not. To do that, we adapted the continuous recognition task already used to test ORFi, but composed of pictures with neutral or positive valence. Healthy subjects performed the task while high-resolution evoked potentials were recorded, allowing investigating the behavioural and electrophysiological influence of emotions on recognition capacity (tested in the first run of the task) and ORFi.

1.4.3 The early development of ORFi

All the studies on ORFi cited above were done with adult subjects. Nothing is known about the development of ORFi in relation with other memory mechanisms during childhood. When are children able to handle memories that interfere with ongoing tasks or duties? Does the filter depend on recognition capacity or not? When does the concept of orientation to time, space, and current situation develop? Study number 3 of this thesis aimed to answer to these questions.

Several studies have demonstrated that distinct memory mechanisms develop at different stages during childhood. Item memory is already functional at the age of 4 years, as Marshall et al (2009) showed using an old/new task. Cycowicz and colleagues (2001) compared item and source memory in 7 - 8 years old children and young adults. During the study phase participants viewed pictures drawn in red or in green. During the test phase subjects had to indicate whether the presented image was new or had already been seen (item recognition). To test source memory, subjects also had to retrieve the colour under which the image had been presented. Not surprisingly, adults performed better than children. More interestingly, the greater difference in performance between the two groups was in the source memory task, while for the item recognition task
the difference was slight. Thus, 6-7 year old children have more difficulty in the recall of the source of a memory in comparison to the simple retrieval of information.

Another study investigating reality monitoring - a component of source monitoring - evidenced that at 6 year of age this memory process is still immature, leading children to make errors when asked to retrieve whether they have just imagined or rather said aloud a word (Foley, Johnson, & Raye, 1983).

Given the devastating consequences of deficient ORFi in adults (Nahum et al., 2012a; Schnider, 2008a; Schnider & Ptak, 1999; Schnider, von Daniken, et al., 1996b), we speculated that ORFi developed early in life, in parallel with explicit memory storage capacity. Another point supporting this hypothesis is the fact that extinction capacity, the presumed mechanism underlying ORFi (Nahum, Ptak, et al., 2009), is an archaic ability already present in primitive creatures like aplysia (Hawkins, Clark, & Kandel, 2006). This leads us to hypothesize that young children could already be capable to filter and reject memories not pertinent for ongoing reality.

To investigate the development of ORFi we conceived a new version of the continuous recognition task described above, and asked 7, 9 and 11 years old children and a control group of adults to perform it. We adapted the difficulty of the task for children, and we used coloured cartoon images of animals to make the task more attractive and funny. The first run the task allowed us to test memory recognition ability, while the second run tested ORFi.

1.5 Evoked potential methods

To investigate the neural correlates underlying ORFi in studies 1 and 2, we recorded high-resolution evoked potentials extracted from continuous electroencephalography (EEG). This is a non-invasive method to record the electrical activity of the brain. More specifically, the sum of the post-synaptic potentials in the brain generates an electric field measured by the
electrodes on the scalp (Michel & Murray, 2012). It is a valid neuroimaging technique that provides a very good temporal resolution (sub-milliseconds) and allows determining how different experimental conditions differ in their temporal sequence and in terms of underlying neurophysiological mechanisms (Murray et al., 2008).

We focused our analyses on the event-related-potentials (ERPs), which are the measured electrocortical brain responses following a specific stimulus or event. ERPs represent a good method to study the psychophysiological correlates of mental processes and to compare different cognitive mechanisms over time.

A classical way to analyze ERPs is based on the morphology (latency and amplitude) of the associated waveforms over time and in correspondence of one or more electrodes. Waveforms of the individual subjects are averaged together and statistics are calculated on the differences between distinct conditions of the task. A disadvantage of the waveform analysis is that it is reference-dependent. The EEG measure is based on the difference of the scalp potentials between an electrode and a reference. If the reference is changed, the voltage difference - and thus the shape of the ERP waveform - changes as well. Thus, the choice of the reference is an important source of bias introduced by the experimenter, which has an impact on the results and on the data interpretations (Murray et al., 2008).

To overcome this “reference-problem”, we also performed topographic ERP analyses, which compare how the distribution of the magnetic field over the scalp differs between conditions. Contrary to the waveforms analysis, the configuration of the scalp topography is independent of the reference electrode (Michel et al., 2004; Murray et al., 2008). Furthermore, topographic analyses allow estimating the location of the brain generator at the origin of the different scalp topographies. On the basis of the assumption that different map topographies designate different source configurations in the brain (Michel & Murray, 2012), source localization
allows to directly indicate the origin of the neuronal activity in real time and in a non-invasive manner (Michel & Murray, 2012).

The most important advantage of EEG is the excellent temporal resolution. Since it directly measures neuronal activity, it has a sub-millisecond resolution (Michel et al., 2004). This allows investigating and comparing when different cognitive processes take place in the brain with a precision that imaging techniques like functional magnetic resonance (fMRI) or positron emission topography (PET) do not permit. However, the main disadvantage of this technique is that the signal measured on the scalp cannot unequivocally indicate the location of activated neurons (Michel et al., 2004). This is the so-called “inverse problem”: different sources can generate the same electrical fields in the scalp (Michel et al., 2004). Consequently, a maximal activity at a specific electrode does not mean that the generators are unequivocally located in the brain region underlying it. Conversely, different scalp topographies are necessarily generated by different sources. Thus, thanks to the use of specific algorithms, source localization allows underlying brain sources at the origin of a specific map configuration.
2. RESULTS

Study 1: Memory in time: electrophysiological comparison between reality filtering and temporal order judgment.


Contribution of M.C.L.: conceived and designed the experiment, acquired and analyzed data, wrote the paper.

Abstract:
Orbitofrontal reality filtering (ORF) denotes a little know but vital memory control mechanism, expressed at 200-300 ms after stimulus presentation, that allows one to sense whether evoked memories (thoughts) refer to present reality and can be acted upon, or not. Its failure induces reality confusion evident in confabulations that patients act upon and disorientation. In what way ORF differs from temporal order judgment (TOJ), that is, the conscious knowledge about when something happened in the past, has never been explored. Here we used evoked potential analysis to compare ORF and TOJ within a combined experimental task and within a comparable time frame, close to the experienced present. Seventeen healthy human subjects performed an experiment using continuous recognition tasks that combined the challenges of ORF and TOJ. We found that the two mechanisms dissociated behaviorally: subjects were markedly slower and less accurate in TOJ than ORF. Both mechanisms evoked similar potentials at 240-280 ms, when ORF normally occurs. However, they rapidly dissociated in terms of amplitude differences and electrical source from 310 ms to 360 ms and again from 530 to 560 ms. We conclude that the task of consciously ordering memories in the immediate past (TOJ) is effortful and slow in contrast to sensing memories' relation with the present (ORF). Both functions invoke similar early electrocortical processes which then rapidly dissociate and engage different brain areas. The results are consistent with the different consequences of the two mechanisms’ dysfunction: while failure of ORF has a known clinical manifestation (reality confusion as evident in confabulation and disorientation), the failure of TOJ, as tested here, has no such known clinical correlate.
Study 2: The influence of emotion on orbitofrontal reality filtering


Contribution of M.C.L.: conceived and designed the experiment, acquired and analyzed data, wrote the paper.

Abstract:
Orbitofrontal reality filtering is a mechanism that allows us to keep thought and behavior in phase with reality. Its failure induces reality confusion with confabulations and disorientation. Confabulations have been claimed to have a positive emotional bias, suggesting that they emanate from a tendency to embellish the situation of a handicap. Here we tested the influence of positive emotion on orbitofrontal reality filtering using a paradigm validated in reality confusing patients and with a known electrophysiological signature, a frontal positivity at 200-400 ms after memory evocation. High-density evoked potentials were recorded from nineteen healthy subjects as they performed two runs of a continuous recognition task, both composed of the same set of pictures, which were either neutral or positive. The first run measures learning and recognition. The second run, when all items are familiar, is a measure of reality filtering. Performance was more accurate and faster on neutral than positive pictures in both runs and all conditions. Evoked potential correlates of emotion and reality filtering occurred at 260-350 ms but dissociated in terms of amplitudes and topography. While positive stimuli evoked a more negative frontal potential than neutral ones, reality filtering induced a frontal positivity superimposed on the potential difference induced by stimuli’s emotional valence. We conclude that the electrophysiological expression of reality filtering is not influenced by emotional positivity of the processed material. The result is consistent with the observation that reality-confusing patients with defective reality filtering usually enact common habits rather than emotionally salient activities.
Study 3:  Children's sense of reality: the development of orbitofrontal reality filtering

M. C. Liverani, A. L. Manuel, L. Nahum, V. Guardabassi, C. Tomasetto, A. Schnider

Child Neuropsychology, 2015b, article ahead of print.

Contribution of M.C.L.: conceived and designed the experiment, analyzed data, wrote the paper.

Abstract:
Orbitofrontal reality filtering denotes a memory control mechanism necessary to keep thought and behavior in phase with reality. In adults, it is mediated by the orbitofrontal cortex and subcortical connections and its failure induces reality confusion, confabulations, and disorientation. Here we investigated for the first time the development of this mechanism in 83 children from ages 7 to 11 years and 20 adults. We used an adapted version of a continuous recognition task composed of two runs with the same picture set but arranged in different order. The first run measures storage and recognition capacity (item memory), the second run measures reality filtering. We found that accuracy and reaction times in response to all stimulus types of the task improved in parallel across ages. Importantly, at no age was there a notable performance drop in the second run. This means that reality filtering was already efficacious at age 7 and then steadily improved as item memory became stronger. At the age of 11 years, reality filtering dissociated from item memory, similar to the pattern observed in adults. However, performance in 11-year-olds was still inferior as compared to adults. The study shows that reality filtering develops early in childhood and becomes more efficacious as memory capacity increases. For the time being, it remains unresolved, however, whether this function already depends on the orbitofrontal cortex, as it does in adults, or on different brain structures in the developing brains of children.
3. DISCUSSION

The three studies presented in this thesis aimed to broaden our comprehension of orbitofrontal reality filtering. They allowed concluding that ORFi is independent from the capacity to verify the temporal order under which memories have been acquired, that it is not influenced by positive emotions, and that it develops early in childhood and in parallel with other memory capacities.

In the following, I will discuss these results more in detail, concluding with an overview of the limitations of the studies, and proposing future perspectives to further explore this memory mechanism.

3.1 ORFi and Temporal order judgment

In the first study (Liverani, Manuel, Bouzerda-Wahlen, et al., 2015), ORFi and temporal order judgment were directly compared in a unique experimental paradigm and within the same temporal frame, around the experienced present.

Behaviorally, the two mechanisms clearly dissociated: to select a memory that is pertinent to the ongoing reality is easier and more rapid than to place a memory in its correct temporal frame in the past. In the latter case, we need a cognitive effort to correctly perform the task.

This dissociation is further supported by the confidence rating that we asked participants to provide: responses in the ORFi condition were given with significantly more confidence than responses to temporal order judgment.

The difference in the accuracy between the two memory mechanisms mirrors memory challenges in everyday life. Failures in the discrimination of the origin of a memory or thought are quite frequent, even in healthy subjects (Mitchell & Johnson, 2000). The source monitoring framework assumes that there are evaluative processes by which a memory or a
thought is attributed to a specific source (Johnson et al., 1993; Mitchell & Johnson, 2000). People can succeed or fail in the discrimination of the origin of a mental experience, consequently generating accurate or distorted source attributions (Johnson et al., 1993).

Conversely, ORFi is a very efficient and reliable mechanism in healthy subjects. Even if we are continuously “stimulated” by a multitude of memories and thoughts, we are perfectly capable to select which memories are pertinent for our current tasks or duties, and to reject those that are not, maintaining our behavior in phase with the present situation. This is supported by other behavioral studies that tested ORFi with variations of the continuous recognition task presented in this thesis. Even in a much more difficult and rapid version of the task used with reality confusing patients, control subjects had a very high performance and produced few false positives (Schnider, von Daniken, et al., 1996b), indicating that ORFi is an automatic process that does not require a notable cognitive effort.

Electrophysiologically, we found a similar response between the two mechanisms at a very early stage of stimulus processing, followed by a dissociation from 310 ms on. Our interpretation is that at an early stage all stimuli are subjected to ORFi, irrespective of whether we are required to simply select pertinent memories or to make a temporal order judgment. After this first step, the two mechanisms dissociated, with different ERP amplitudes and topographies.

These results are a further demonstration of the distinctiveness of ORFi in comparison to other memory monitoring mechanisms, as already indicated by previous studies (Bouzerda-Wahlen et al., 2014; Bouzerda-Wahlen et al., 2013; Wahlen et al., 2011).

This conclusion is not surprising if we directly compare the core of these two theories: Source monitoring theories aim to verify the source of a memory in the past and its veracity. In contrast, ORFi verifies a thought's (memory's) relation with the present (“the now”). It is not thought to verify whether a memory about the past is correct or not or whether its context is
accurately retrieved or not. The only aspect that ORFi takes into account is the relevance of the activated memory for present behavior.

As already considered in the introduction of this thesis, the dissociation between the two mechanisms is supported by clinical data. Deficits in temporal order judgment have been documented in amnesics (Kopelman et al., 1997) and in patients with frontal lesions without amnesia (Milner, Petrides, & Smith, 1985; Shimamura et al., 1990), indicating that temporal order deficits do not by themselves induce reality confusion. Conversely, reality confusion is the hallmark of deficient orbitofrontal reality filtering (Bouzerda-Wahlen et al., 2013; Nahum et al., 2012a; Nahum et al., 2010; Schnider, Bonvallat, Emond, & Leemann, 2005; Schnider & Ptak, 1999).

The difficulty of temporal order judgments is also exemplified by a clinical observation. Our plan was to compare behavioral performances in orbitofrontal reality filtering and temporal order judgment in brain injured patients with amnesia. Similarly to a previous clinical study demonstrating dissociation between ORFi and content monitoring (Bouzerda-Wahlen et al., 2013), we expected to find dissociation from temporal order judgment. It rapidly turned out that the task combining the two memory mechanisms was too difficult for the patients. In particular, patients were incapable to perform the part of the task testing the temporal order judgment, which was, as for healthy subjects, perceived as more difficult than the part testing ORFi.

### 3.2 ORFi and positive emotions

In the second study presented in this thesis, we investigated the impact of positive emotions on ORFi (Liverani et al., 2016). We found that ORFi is not modulated by emotion. This result is in line with our clinical observation of reality confusing confabulating patients, who mainly enact daily habits rather than emotional memories.
Concerning the electrophysiological signature of ORFi, our results replicated previous studies (Bouzerda-Wahlen et al., 2014; Liverani, Manuel, Bouzerda-Wahlen, et al., 2015; Schnider et al., 2002; Wahlen et al., 2011), with a more positive potential for Distracters compared to Targets at about 250 - 380 ms.

Stimuli’s emotionality was processed differently by participants both from a behavioral and an electrophysiological point of view, confirming appropriate selection of images. Behaviorally, accuracy for neutral images was higher than for positive pictures. While astonishing at first, this result may be explained by the fact that the enhancing effect of emotional stimuli on memory is evident only after a delay of at least some hours (Javadi, Walsh, & Lewis, 2011; Maratos, Allan, & Rugg, 2000; Montagrin, Brosch, & Sander, 2013; Windmann & Chmielewski, 2008).

Electrophysiologically, the valence effect was evident at about 250 - 400 ms in both runs. This time window corresponds to the period in which the electrophysiological correlate of ORFi is expressed. However, the positivity associated with Distracters 2 was present separately for the positive or neutral valence of the stimuli, confirming the dissociation between ORFi and items emotionality.

We used only positive and neutral images, but not negative ones, on the basis of the suggestion that false memories of confabulating patients are characterized by a positive emotional bias (Fotopoulou, 2010; Fotopoulou, Conway, & Solms, 2007; Fotopoulou et al., 2008).

Future studies should also take into account negative emotions, in order to extend conclusions to emotions in general.

There is a broad literature about the effects of emotion on memory, especially regarding negative and arousing stimuli. Emotions can influence recognition, subjective experience and objective retrieval of the characteristics associated to a memory (Mather, 2007; Phelps &
Sharot, 2008). In comparison to neutral information, emotional ones can increase attention (Knight et al., 2007) and perception, with higher activation of posterior visual regions during the encoding of emotional stimuli (Phan, Wager, Taylor, & Liberzon, 2002; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004). This higher perceptual processing allows a better recognition of emotional compared to neutral items in the long term (Kensinger & Schacter, 2007; Mather, 2007). Amygdala is modulated by the emotionality of stimuli as well, and its activity correlates with memory accuracy (Canli, Zhao, Brewer, Gabrieli, & Cahill, 2000). Moreover, Kensinger et al. (2004) showed that amygdala activity correlates with hippocampal activity during encoding of arousing information, while the processing of negative non-arousing stimuli are associated to the activation of a prefrontal cortex - hippocampal network (Kensinger & Corkin, 2004).

Despite these well-documented effects of emotions on memory, our hypothesis is that ORFi would not be modulated by negative emotions. This assumption is based on diverse studies on extinction capacity, the ancestral mechanism that has been proposed to be at the basis of ORFi (Nahum, Bouzerda-Wahlen, Guggisberg, Ptak, & Schnider, 2012b; Nahum, Ptak, et al., 2009; Schnider, 2008a). In an EEG (Nahum, Morand, Barcellona-Lehmann, & Schnider, 2009) and an fMRI (Nahum, Simon, et al., 2011) study, Nahum et al. used a reversal learning task which combined extinction capacity and the response to threatening stimuli. In this task, two faces were presented to participants, changing randomly the right or left position. For each pair, they were asked to predict in which face a target stimulus would appear. The target stimulus was either a spider (threatening stimulus) or a disk (neutral stimulus). Participants were told that the target would normally appear on the same face (repetition trials), and would only occasionally switch to the other face (extinction trails). Thus, they had to change their response after the appearance of the non-target stimulus (extinction capacity). They found a dissociation between the processing of threatening stimuli and the ability to extinguish a
behavior that is no longer correct. Electrophysiologically, the processing of threatening vs. neutral stimuli occurred at about 100 - 200 ms and involved medial temporal regions. The processing of non-target stimuli instead appeared after 200 ms, and was mediated by ventromedial prefrontal areas (Nahum, Morand, et al., 2009). The fMRI study confirmed this dissociation, with higher activation of visual area for the processing of threatening stimuli and activation of the orbitofrontal cortex in extinction trials (Nahum, Simon, et al., 2011). These findings demonstrated that the non-occurrence of a previously correctly anticipated event is expressed later and activates different brain regions than the processing of threatening versus neutral stimuli.

As ORFi is based on extinction capacity, these results lead us to believe that threatening - and more in general, negative - stimuli would not have an impact on the filtration of memories pertinent to the present reality.

This is not the case for another memory monitoring mechanism that I discussed in this thesis, source monitoring. Accuracy in source memory is influenced by emotional information (Mitchell & Johnson, 2009). For example, in a working memory task in which participants were asked to retain pictures and corresponding locations, emotional items were associated to a lower accuracy compared to neutral ones (Mitchell, Mather, Johnson, Raye, & Greene, 2006). By contrast, evidences that source memory for emotional information is more accurate that for neutral ones are also present in the literature. Kensinger et al. (2005) showed that the accuracy in discriminating if an item had been seen or imagined at encoding was higher for emotional compared to non-emotional stimuli (Kensinger & Schacter, 2005).

Even if contradictory results are present in the literature, the fact that source memory is affected - either negatively or positively - by memory emotionality is another evidence of the dissociation between this memory mechanism from orbitofrontal reality filtering.
3.3 The early development of ORFi

The last study presented in this thesis concerned the development of ORFi in children. Even if the response accuracy in the continuous recognition task increases linearly between the three age groups (and reaction time diminishes accordingly), 7-years-old children are already capable to perform the task above chance level. Thus, at this age children are already capable to discriminate between memories that are pertinent with the present situation and memories that do not. Reality filtering improves in parallel with the development of item memory until the age of 9. From 11-years-old on, correlation analysis showed that ORFi continues to develop independently from item memory, as already found in other studies with adults (Schnider, 2008a; Schnider, von Daniken, et al., 1996a).

In a preliminary pilot study, we asked younger children (4 and 5 years old) to perform the same version of the paradigm. They answered by chance, making impossible to include children of this age in the study. Different factors could explain this difficulty: memory capacities are not developed enough to allow the correct completion of the task, instructions are too difficult, or the task is much demanding in terms of attention and concentration. To better explore this point, it could be useful to design an easier version of the experimental paradigm.

At 11 years of age, reality filtering is significantly better than in younger children, but there are still more false positive responses compared to adults. Therefore, we can argue that this mechanism continues to develop later on, probably also during adolescence, before reaching adult performance. Further investigations should test the paradigm with older children and adolescents, in order to determine at what age ORFi is completely functional.

While ORFi develops in parallel with the ability to store and recognize new information in 7 and 9-years-old children, the two processes become independent from 11 years old on. This
dissociation could rely on the different speed of maturation of the brain regions on which these two mechanisms are thought to depend.

As the association between ORFi and the orbitofrontal cortex is well documented in adults (Schnider, Treyer, et al., 2000; Treyer et al., 2003; Schnider, 2008 #5), we could speculate that the same brain region is involved in this mechanism also in children.

There is general consensus that OFC - and, more generally, the whole prefrontal cortex – is one of the brain regions that develops more slowly, reaching complete maturity only after 20 years or more in humans (Diamond, 2002). Nevertheless, studies about this development are contrasting. In a longitudinal study using MRI, Giedd et al. (1999) showed that the quantity of white matter (thus, of myelinated axons) increased linearly from 4 to 13 years in the prefrontal cortex. Gray matter attained its maximal volume between 11 and 12 year (Giedd et al., 1999). By contrast, in cross-sectional volumetric studies, Jernigan et al. (1991) and Sowell et al. (1999) documented a decrease in the gray matter volume from childhood to adolescence, with a significant correlation with source memory capacity (Jernigan, Trauner, Hesselink, & Tallal, 1991; Sowell et al., 1999).

The age groups included in our study ranged from 7 - to 11 - years - old. During this period, the development of the PFC parallels improvement in the speed of processing, in the capacity to use strategies and in the ability to hold information in mind and exert inhibition (Diamond, 2002).

The brain region associated to item memory is the temporal lobe, which reaches complete maturation earlier than the PFC (Buchsbaum et al., 1992). Thus, if we assume that reality filtering depends on the OFC also in children, the possible explanation of the dissociation between the two memory mechanisms from 11 - years - old on could rely on the difference in maturation speed of the subjacent brain regions.
Further investigations are needed to ascertain the association between reality filtering and OFC in children. To do that, a functional neuroimaging study using the same paradigm could be helpful. A limitation to this type of investigation is that functional MRI usually has artifacts with signal dropout in the region of the ORF, resulting in a weak signal (Ojemann et al., 1997). Hopefully, the advance in neuroimaging techniques will allow solving this problem and permitting to verify if the reality filtering mechanism resides on the same brain regions both in children and adults.

Our study raises the general question of false memories in children. A large amount of studies have tried to understand the reliability of children’s memory, especially in the setting of eyewitness testimony. The typical paradigm used to investigate the quality of children’s memory consists in showing to children an enactment of an event or telling them a story and then ask questions about the event that include erroneous information. In the case of preschool children, not only the recall of the event is poorer compared to older children and adults (Cohen & Harnick, 1980; Ornstein, Gordon, & Larus, 1992), but children’s memories are also more vulnerable to distortions following misleading post-event information (Gathercole, 1998). This is particularly evident in 3- and 4- years-old children. The amount of memory distortions decreased linearly until 10 - 12 years (Ceci, Ross, & Toglia, 1987). One of the possible interpretation of increased suggestibility is that post-event information erroneously updates the original memory trace (Loftus, Schooler, & Wagenaar, 1985). Another explanation could be that biased information provided by the experimenter are used to “fill the gaps” of incomplete memory traces (McCloskey & Zaragoza, 1985), as in certain types of confabulations in adults (Conway & Tacchi, 1996). Alternatively, a failure in source monitoring has been advanced as well (Ackil & Zaragoza, 1995).

A particular paradigm used to investigate false memories is the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995). Participants are asked to remember a list
of semantically related words. At test, they have to recall the studied words. In this phase, many participants not only report the presented words, but also semantically related words (“lures”) not presented. Ghetti et al. (2002) investigated the developmental trajectory of false memories by testing 5- and 7- year old children using this paradigm. They found no significant age difference in false memory formation: children did not differ from adults in the amount of generated false words (Ghetti, Qin, & Goodman, 2002).

Based on these results, we can conclude that children’s memory is particularly vulnerable and sensitive to distortions during preschool age, as tested with the DRM paradigm and biased questioning. Our interpretation is that from 6 - 7 years on, the ability to control false memories, as well as the ability to select memories pertinent with the ongoing reality thanks to ORFi, are functioning. This may explain why children at this age are less prone to memory errors and to act according to fantasies.

3.4 Is ORFi important for psychosis and schizophrenia?

The mechanism of ORFi has been mainly studied in the context of confabulating and reality confusing patients. Less is known about the possible implication of a dysfunctioning ORFi in other neurological or psychiatric diseases, like schizophrenia.

Schizophrenia is a chronic and severe mental disorder characterized by positive symptoms (hallucinations, delusions, thought disorders) and negative symptoms (reduced speaking, affective flattening) (Association, 2013).

Deficient source monitoring had been proposed as an explanation of schizophrenia: for example, this kind of patients is less accurate than controls in the discrimination of the source and of the temporal context associated to a memory (Waters, Maybery, Badcock, & Michie, 2004).
The presence of positive symptoms in schizophrenia suggests a potential link between this disease and the inability to control interferences from ancient memories, like in reality confusing patients.

This assumption is supported by the observation that schizophrenia is characterized by abnormalities in ventral prefrontal cortex and orbitofrontal cortex (Davatzikos et al., 2005; Pantelis et al., 2003; Waltz & Gold, 2007). In particular, Waltz et al. (2007) showed a particular impairment in this kind of patients for the reversal of learned associations, a type of learning that depends on the activation of the ventral prefrontal cortex (Waltz & Gold, 2007). Furthermore, schizophrenia has been linked to a hyperdopaminergic state of the brain, which seems to be the cause of psychosis (Howes & Kapur, 2009).

Since ORFi is associated with the activation of the OFC (Schnider, Treyer, et al., 2000), its functioning is based on the extinction capacity (Schnider, 2008a, 2013) and is under dopaminergic control (Schnider et al., 2010) we could argue that schizophrenic patients are perturbed in their ability to filter memories and thoughts that are not pertinent with the present reality.

Two studies have already use an adapted version of our continuous recognition task with schizophrenic patients, founding that the severity of auditory hallucinations were correlated with the false positive responses in the task (Badcock, Waters, Maybery, & Michie, 2005; Waters, Badcock, Maybery, & Michie, 2003). Further EEG and fMRI studies would be useful to better investigate the implication of ORFi in this kind of disease.
4. CONCLUSION

The three studies presented in this thesis allowed us to clarify specific aspects of orbitofrontal reality filtering.

In the first study (Liverani, Manuel, Bouzerda-Wahlen, et al., 2015), we demonstrated that the ability to distinguish between memories that are relevant or irrelevant to the present situation behaviourally and electrophysiologically dissociates from the ability to retrieve the temporal context of a memory. This is a further evidence supporting the independence of ORFi from other monitoring mechanisms.

In the second study (Liverani, Manuel, Nahum, et al., 2015), we showed that ORFi is not influenced by positive emotions, confirming our clinical observations with reality confusing patients who mainly confabulate on daily habits and not on emotional memories.

Finally, in the third study (Liverani, Manuel, Nahum, et al., 2015) we proved that by the age of 7 ORFi is already functioning and continues to develop in parallel with recognition capacity until the age of 9, becoming an independent mechanism later on.

4.1 Limitations and future perspectives

Regarding the study comparing ORFi with TOJ, we found that participants perceived the TOJ condition as significantly more difficult in comparison to the ORFi condition. On one hand, this observation was a confirmation, at a behavioural level, of the dissociation between the two mechanisms. On the other hand, it could be supposed that the found electrophysiological differences could rely more on the difference of difficulty than on a difference of the cognitive mechanisms required for the two tasks per se. To solve this problem it could be useful to conceive an easier version of the task testing TOJ. This would also allow testing
reality confusing and amnesic patients and prove the hypothesized clinical dissociation between the two mechanisms.

In the study in which the influence of emotions on ORFi was investigated, we only tested positive emotions. This because previous researches emphasized the existence of a positive bias in confabulations, observation that was in contrast with our experience with reality confusing patients. Further studies including also negative stimuli are necessary in order to generalize our results to memory emotionality in general.

In the developmental study, we speculated that, as in adults, the OFC is the brain region associated to ORFi in children. Only an imaging study could unequivocally confirm this hypothesis, since the engagement of brain regions in children is often very different than in adults.

An easier version of the task to test also younger children would be useful to determine if ORFi is already functioning also in pre-scholar children. Moreover, since 11 - years - old children still present more false positives in comparison to adults, we would need further investigations to understand at what age the development of ORFi is complete, testing older children and adolescent.

Finally, until now the experimental paradigm of ORFi has been utilized only with reality confusing patients. It would be of interest to test it in other psychiatric and neurologic diseases like schizophrenia, post-traumatic stress disorder, Parkinson’s disease and so on.

A better comprehension of the different facets of this vital memory mechanism and its implications on diverse pathologies could give interesting hints on the role of cognitive rehabilitation and medication.
5. REFERENCES


6. ARTICLES

Study 1: Liverani et al. (2015a)
Study 2: Liverani et al. (2016)
Study 3: Liverani et al. (2015b)
Study 1

MEMORY IN TIME: ELECTROPHYSIOLOGICAL COMPARISON BETWEEN REALITY FILTERING AND TEMPORAL ORDER JUDGMENT

M. C. LIVERANI, a A. L. MANUEL, a A. BOUZERDA-WAHLEN, b M. GENETTI, b A. G. GUGGISBERG, a,c L. NAHUM a AND A. SCHNIDER a,c

a Laboratory of Cognitive Neurorehabilitation, Medical School, University of Geneva, Switzerland
b Department of Fundamental Neurosciences, University of Geneva, Switzerland
b Division of Neurorehabilitation, Department of Clinical Neurosciences, University Hospital and University of Geneva, Switzerland

Abstract—Orbitofrontal reality filtering (ORF) denotes a little known but vital memory control mechanism, expressed at 200–300 ms after stimulus presentation, that allows one to sense whether evoked memories (thoughts) refer to present reality and can be acted upon, or not. Its failure induces reality confusion evident in confabulations that patients act upon and disorientation. In what way ORF differs from temporal order judgment (TOJ), that is, the conscious knowledge about when something happened in the past, has never been explored. Here we used evoked potential analysis to compare ORF and TOJ within a combined experimental task and within a comparable time frame, close to the experienced present. Seventeen healthy human subjects performed an experiment using continuous recognition tasks that combined the challenges of ORF and TOJ. We found that the two mechanisms dissociated behaviorally: subjects were markedly slower and less accurate in TOJ than ORF. Both mechanisms evoked similar potentials at 240–280 ms, when ORF normally occurs. However, they rapidly dissociated in terms of amplitude differences and electrical source from 310 to 360 ms and again from 530 to 560 ms. We conclude that the task of consciously ordering memories in the immediate past (TOJ) is effortful and slow in contrast to sensing memories’ relation with the present (ORF). Both functions invoke similar early electrocortical processes which then rapidly dissociate and engage different brain areas. The results are consistent with the different consequences of the two mechanisms’ dysfunction: while failure of ORF has a known clinical manifestation (reality confusion as evident in confabulation and disorientation), the failure of TOJ, as tested here, has no such known clinical correlate. © 2015 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: reality filtering, orbitofrontal cortex, source memory, temporal order judgment, evoked potentials, time.

INTRODUCTION

The experience of remembering is fundamental to human action and thinking. A crucial aspect of memory is time. Two capacities are clearly important: the ability to sense whether a memory (thought) pertains to present reality – and may be acted upon – and the ability to recollect the temporal context in which a memory was acquired.

An impairment of the former capacity is associated with reality confusion, as evident in disorientation and confabulations that patients act upon (Schnider, 2008, 2013). For example, a 58-year-old woman left the examination convinced that she had to feed her baby – who was 35 years old at the time (Schnider, 1996); a dentist hospitalized after rupture of an aneurysm of the anterior communicating artery repeatedly left the hospital in the conviction that patients were waiting for him at his clinic (Ptak and Schnider, 1998). Schnider and colleagues explained this kind of behavior by the failure of a specific mechanism they now call orbitofrontal reality filtering (ORF) (Schnider, 2008, 2013). They found that, when reality confusing patients made repeated runs of a continuous recognition task, all of them composed by the same picture set, they produced increasingly more false positive responses, in contrast to correctly oriented amnesics (Schnider et al., 1996a; Schnider and Ptak, 1999; Nahum et al., 2012). Healthy subjects performing such a task had activation of the posterior orbitofrontal cortex and subcortical structures (Schnider et al., 2000; Treyer et al., 2003), which corresponds to the main lesion site of reality confusing patients (Schnider et al., 1996a; Schnider and Ptak, 1999; Schnider, 2013). Successful completion of the task by healthy subjects was associated with a frontal positivity in evoked potential responses at 200–300 ms, indicating that reality filtering (RF) occurs before the content of an upcoming thought or memory is processed at 400–600 ms (Schnider et al., 2002; Wahlen et al., 2011; Schnider, 2013).

ORF bears superficial resemblance to well-known memory-monitoring models. Strategic retrieval describes
a hypothetical series of processes leading up the evocation and verification of memories. It has been claimed to encompass ORF and to explain confabulation (Moscovitch and Melo, 1997; Gilboa et al., 2006). Experimental testing, however, showed that monitoring, as described within this model, electrophysiologically dissociated from ORF (Wahlen et al., 2011). In amnesic patients, only deficient ORF, but not strategic monitoring, was associated with reality confusion, as evident in disorientation and inappropriate acts in agreement with confabulations (Bouzera-Wahlen et al., 2013). Failures of source monitoring, the ability to judge the precise source (context, temporal frame, etc.) of a memory, is also a generally accepted explanation for confabulation (Johnson et al., 1993). Experimental evidence has not supported this claim (Johnson et al., 1997), and an evoked potential study demonstrated a dissociation between context source monitoring and ORF (Bouzera-Wahlen et al., 2014). Clearly, the ability to sense a memory’s (thought’s) relation with the present, as reflected in ORF, is different from the ability to decide whether and how a memory refers to the past, as covered by traditional memory-monitoring models.

Behavioral correlates of a disturbed sense for the temporal context in which memories were acquired are less clear. The capacity is usually measured with recency or temporal order judgment (TOJ) paradigms in which two lists of items are presented and subjects are asked to recall whether a given item had appeared in the first or in the second list or which one had been presented more recently (e.g., Tendolkar and Rugg, 1998; Duarte et al., 2010). The ability to correctly place memories in their temporal sequence has been suggested to be a mechanism for confabulation (Van der Horst, 1932; Talland, 1961; Dalla Barba et al., 1997), previously also for disorientation (Kraepelin, 1909), which has been the topic of very few studies. Data were ambiguous: difficulties in TOJ did not appear specific to confabulation and disorientation (Schnider et al., 1996b; Johnson et al., 1997). Disordered temporal order memory has been observed after dorsolateral prefrontal (Milner et al., 1985, 1991; Shimamura et al., 1990), medial temporal lobe, and diencephalic lesions (Kopelman et al., 1997). Imaging studies in healthy individuals demonstrated activation of the dorsolateral prefrontal cortex (Zorrilla et al., 1996; Konishi et al., 2002), together with cuneus, precuneus and right posterior parietal regions (Cabeza et al., 1997). In evoked potential studies, recency judgment evoked bilateral frontal positivity from around 300 ms (Tendolkar et al., 2004).

These observations indicate a functional and electrophysiological dissociation between ORF and TOJ. However, the paradigms used to explore them not only differed in design but also in temporal frame: while healthy subjects need to perform continuous recognition tasks in rapid succession to challenge ORF (and induce orbitofrontal cortex activation) (Schnider, 2008, 2013), TOJ has traditionally been tested on items separated by minutes (e.g., Tendolkar et al., 2004; Grove and Wilding, 2008).

For the present study, we designed an experimental task exploring the two processes within a similar time frame around the extended present (Fraisse, 1984), with task components separated by short intervals (around 1 min). Given the suspected dissociation between the two mechanisms, we expected that ORF – an intuitive faculty – would behaviorally dissociate from explicit, typically effortful TOJ by more accurate and faster performance. On the electrophysiological level, we expected stimuli critical for ORF to induce a frontal positivity at 200–300 ms. In contrast, we expected the stimuli specific for TOJ to induce potential changes at a later stage of processing, beyond 300 ms.

**EXPERIMENTAL PROCEDURES**

**Participants**

Twenty-three right-handed healthy subjects gave written informed consent and were paid to participate in the study, which was approved by the Institutional Ethical Committee. They reported no history of neurological or psychiatric disorders or medication use. Five subjects were excluded from the study because of poor signal quality and one subject because of poor comprehension of the task. Seventeen subjects (11 women and six men, age 24 ± 3.3 years) were included in the analysis.

**Procedure and task**

We devised a task composed of four different blocks. Two blocks tested RF, two tested TOJ. The composition of the task is shown in Fig. 1. Blocks’ order was pseudo-randomized: half of the subjects started with the RF condition, the other half with the temporal judgment (TJ) condition. Blocks were separated by 5 min to avoid fatigue.

Each block was composed of a different set of images and was divided in 3 runs: the first two runs (Fig. 1A) corresponded to the learning phase, in which two different sets of 40 concrete line drawings (Snodgrass and Vanderwart, 1980) were shown for the first time (distracters, D; n = 40) and then repeated after 6–8 intervening stimuli (repetitions, targets, T; n = 40). Subjects had to answer the following question: “Have you already seen this image within this run?” by pressing the left button with the right index to indicate “yes” or the right button with the right middle finger to indicate “no”. These two runs were separated by a 3-min pause.

In the third run of each block, we tested alternatively RF or TOJ. RF (Fig. 1B) was tested according to the same logic as in previous studies (Schnider et al., 2002; Schnider, 2008, 2013): all images from the two previous runs were mixed together to compose a continuous recognition run with picture repetitions after 6–8 intervening images. Subjects had to answer the following question: “Have you already seen this image within this run?” by pressing the left button with the right index to indicate “yes” or the right button with the right middle finger to indicate “no”. In the following, picture recurrences within this run will be designated “targets within reality filtering task (TRF)”. Items appearing for the first time within the third run (previously presented either in the first or the second
run of the learning phase) will be named “distracters within the reality filtering task (DRF)”.

TOJ (Fig. 1C) was investigated by presenting all pictures of the two preceding runs and asking the participants the following question: “Have you seen this image in the just preceding (second) run (rather than the very first run)?” They pressed the left button to answer “yes” (in the following “target within the temporal judgment task, TTJ”), and the right button to answer “no” (in the following “distracters within the temporal judgment task, DTJ”).

In both tasks, participants judged the confidence in their responses. To do this, a black rectangle was displayed in the center of the screen after disappearance of each image. Subjects then had to press the left button if they were sure of their answer or the right button if they were not sure.

Images were presented on a computer screen for 2000 ms. Black rectangles for the confidence judgment were displayed for 1500 ms after each item. Interstimulus interval with a fixation cross in the center of the screen was 700 ms. Stimulus presentation was controlled using E-prime (Psychology Software Tools, Inc., Pittsburgh, USA; www.pstnet.com/eprime).

In the first and in the second run of each block there were 80 stimuli, 40 distracters (D) and 40 targets (T). In the third run of the blocks, in which RF was tested, there were 160 stimuli: 80 targets within the TRF and 80 distracters (40 from the first learning phase, 40 from the second) within the reality filtering task (DRF).

In the third run of the blocks, in which TOJ was tested, there were 80 stimuli, of which 40 DTJ and 40 TTJ.

As the study aimed to compare the mechanisms underlying RF and TOJ, only the stimuli of the third runs were retained for the analysis. Therefore, data from the third runs of the two blocks testing RF and the third runs of the two blocks testing TOJ were pooled for the analyses. Moreover, only one half of the DRF were randomly retained for the analysis (20 belonging to the first learning phase and 20 to the second learning phase) in order to have the same number of trials per condition.
EEG acquisition and raw data preprocessing

The electroencephalogram was continuously recorded during the task with a 128-channel Active-Two Biosemi EEG system (Biosemi V.O.F. Amsterdam, Netherlands). Signal was sampled at 512 Hz in a passband of 0–104 Hz. Electrodes impedance was kept below 20 KΩ.

Analyses were performed using Cartool Software by Denis Brunet (https://sites.google.com/site/fbmlab/cartool/).

Epochs from 100 ms before to 800 ms after the onset of the stimulus were averaged for each condition (i.e. DRF, DTJ, TRF, and TTJ) and for each subject to calculate the event-related potentials (ERPs). ERPs were band-pass filtered to 1–30 Hz and recalculated against the average reference. Baseline correction was applied on the 100-ms pre-stimulus period. Only trials with correct answers entered the analysis. Traces were visually inspected to detect and reject epochs with eye movements or other sources of transient noise. In addition, epochs with artifacts higher than 100 μV were automatically excluded. Channels showing substantial noise were interpolated using a spherical spline interpolation (Perrin et al., 1987). The epoch from 0 to 800 ms after the stimulus presentation was retained for the analysis. The mean number ± SD of epochs across subjects per condition was: DRF = 48.3 ± 14.6; TRF = 48.6 ± 14.4; DTJ = 40 ± 14; TTJ = 36.5 ± 12.1.

Behavioral data analysis

2x2 repeated measures analysis of variances (ANOVAs) on accuracy, reaction times and confidence judgment were performed with the within-subjects factors Stimulus Type (distracters, D, targets, T) and Condition (reality filtering, RF, and temporal judgment, TJ). Post-hoc analyses were performed with Fisher’s least significant test.

In addition, an index of confidence judgment (CJ) was calculated, which should reflect how often a subject gave a correct response and was convinced about it. It was defined as follows:

\[ CJ = \frac{CR}{TI} \times \frac{PJ}{TI} = \frac{CR}{TI} \times \frac{PJ}{TI^2} \]

CR ("correct responses") denotes the sum of the correct answers of the subject, TI ("total items") the total number of items, and PJ ("positive judgment") the sum of the items for which the subject was sure about his response.

We performed a 2 × 2 repeated ANOVA on this index, with the two within-subject factors Stimulus Type and Condition, and a post-hoc Fisher’s test.

Topographic pattern analysis (TANOVA)

To determine periods with significant differences in the topography of the electric field across conditions, we performed a time point by time point statistical comparison of scalp field maps over 128 electrodes using a non-parametric randomization test based on global dissimilarities between electric fields (TANOVA, topographic analysis of variance) implemented in the RAGU software (Koenig et al., 2011). Analysis was performed using a within-subject design with two factors (Stimulus Type and Condition) and with 5000 randomization runs. The output revealed statistically significant periods of interest over 0–800 ms as defined by \( p < 0.05 \). To account for temporal autocorrelation we considered reliable only periods which remained significant (\( p < 0.05 \)) for more than 20 ms (Guthrie and Buchwald, 1991). Secondly, we averaged the scalp field data across participants and periods of interest in order to have one time point per interval. We then computed statistics on this average to avoid false positives due to multiple testing across time.

Waveforms analysis

Five regions of interest (ROIs) were defined: frontal, central, posterior, lateral right and lateral left. All were composed of 15 electrode sites, except for the frontal and the posterior ones that comprised 16 electrodes (Dien and Santuzzi, 2005). ERPs from electrodes pertaining to a region were averaged for each subject and each condition. According to the results obtained with the Topographic ANOVA, the statistical analysis of amplitude differences was performed in three time periods: 240–280 ms, 310–360 ms, and 530–560 ms.

We assessed amplitude differences by performing a 2 × 2 repeated-measures ANOVA for each ROI and across the four time windows with the factors Stimulus Type (D or T) and Condition (RF or TJ). Then, we used post-hoc Fisher’s tests with a significance level of \( p < 0.05 \).

Source localization

As a final step we used a distributed linear inverse solution based on a Local Auto-Regressive Average (LAURA) (Grave de Peralta Menendez et al., 2004) in order to estimate the generators of the different scalp topographies. LAURA sorts out the source configuration that best simulates the biophysical behavior of electric fields (i.e., activity at one point in the brain depends on the activity at neighboring points according to electromagnetic laws).

The model comprises a 3D realistic head model with 3005 solution points (Grave de Peralta Menendez et al., 2001) distributed within the gray matter of the average brain provided by the Montreal Neurological Institute (MNI). We first calculated LAURA inverse solutions for each subject’s ERP at each time point. Then, with paired t-tests, we statistically compared the estimated electric current for each solution point (current source density) between D (mean of DRF and DTJ) and T (mean of TRF and TTJ), DRF and TRF, DRF and DTJ, RF (mean of DRF and TRF) and TJ (mean of DTJ and TTJ) within the periods in which the topographic ANOVA revealed main effects or interactions, namely at 240–280 ms (main effect of Stimulus Type), 310–360 ms (interaction) and 530–560 ms (main effect of Condition). Only nodes with \( p < 0.05 \) for at least 20 ms were retained.
RESULTS

Behavioral results

Behavioral results are summarized in Table 1. To compare performance in the decisive runs 3 (TOJ vs. ORF), 2 × 2 repeated-measures ANOVA was done. It revealed a significant effect of Task type on accuracy (F(1,16) = 55.1, p < 0.001). Post hoc tests showed that participants were significantly less accurate in response to both distracters and targets in the TJ condition compared to the RF condition.

Reaction times had a significant main effect of Task type (F(1,16) = 99.7; p < 0.001) and Stimulus Type (F(1,16) = 13.4; p = 0.002). Fisher’s post-hoc test revealed that subjects were slower in TJ than RF, and in response to distracters than targets in both conditions.

Regarding confidence judgment, there was a significant main effect of Task type (F(1,16) = 36.0; p < 0.001) and a significant interaction between Task type × Stimulus Type (F(1,16) = 6.1; p = 0.026). Subjects were significantly more confident about their answers in RF than TJ. Regarding the interaction, the feeling of confidence was significantly higher for the targets during RF than to the distracters during TJ.

As the RF task had the same instructions as the two learning runs (“Have you already seen this picture within this run?”), we tested for training effects within the ORF task. A 2 × 3 repeated measures ANOVA involving the three runs (Learning runs 1 and 2, ORF task) and the two stimulus types revealed a significant interaction regarding accuracy (F(2,32) = 9.6; p = 0.01), which was due to lower accuracy in run 3 (ORF) than the two learning runs. Reaction times were slower for distracters than targets overall (F(1,16) = 10.8; p = 0.004). A significant interaction between Run and Stimulus (F(2,32) = 11.62; p < 0.001) was explained by slower reactions to distracters than targets in run 2 and run 3 (ORF). Thus, there did not appear to be a training effect between the three runs.

Topographic pattern analysis (TANOVA)

There was a main effect of Stimulus Type at 240–280-ms post stimulus onset (F(1,16) > 4.5, p < 0.001), a main effect of Condition (F(1,16) > 4.5, p = 0.024) and an interaction (F(1,16) > 4.5, p = 0.021) between 530 and 560 ms and an interaction between the two factors at 310–360 ms (F(1,16) > 4.5, p = 0.010). These results indicate different topographies in different time periods and conditions, implying distinct activity generators (Michel et al., 2004).

Waveform analysis and source estimation

In the following, we describe separately the waveform analysis and the source estimation for these three periods.

Period 240–280 ms. Repeated measures ANOVAs on the amplitude differences revealed a main effect of Stimulus Type on frontal (F(1,16) = 8.9; p = 0.01), central (F(1,16) = 10.4; p = 0.005), lateral right (F(1,16) = 10.0; p = 0.006) and lateral left (F(1,16) = 6.1; p = 0.025) ROIs. Fig. 2 shows that distracters of both conditions induced a greater positivity than targets over frontal and central electrodes (post-hoc test, p < 0.05). In the lateral right ROI, DTJ was more positive than TRF and TTJ, while over lateral left electrodes, DRF was more positive than TTJ. In addition, a significant interaction between Stimulus Type and Condition over posterior electrodes was found (F(1,16) = 10.9; p = 0.004), with a higher positivity of TRF compared to all the other stimuli.

In order to estimate the generators of the main effect of Stimulus Type reported in the TANOVA, we performed t-tests comparing distracters and targets in source space, which showed significant activation differences (T(16) > 1.74; p < 0.05). As visible in Fig. 3A, targets (in blue) more strongly activated the left middle temporal gyrus than distracters.

Thus, in this early period there is no dissociation between the two mechanisms, but between the two types of stimuli (D and T).

Period 310–360 ms. There was a significant interaction Condition × Stimulus Type over frontal (F(1,16) = 5.31, p = 0.03), lateral right (F(1,16) = 4.52, p = 0.05) and lateral left (F(1,16) = 11.9, p = 0.00) electrodes, and a main effect of Condition over posterior electrodes (F(1,16) = 12, p < 0.001). Post-hoc tests revealed a trend for DRF to produce a higher positivity than DTJ over frontal electrodes (p = 0.06; Fig. 2). Over posterior electrodes, TRF tended to induce a stronger negativity than DTJ (p = 0.07). Over lateral left electrodes instead, DRF induced a greater positivity than TRF and DTJ (p < 0.05).

The interaction between Stimulus Type and Condition observed in the TANOVA was explained by a significant difference between DRF and TRF and between DRF and DTJ.

Table 1. Behavioral results. The second column indicates the percentage of correct responses ± standard deviation (SD) for each type of stimulus. The third column indicates mean of reaction times ± SD in milliseconds for each type of stimulus. The third column indicates mean of reaction times ± SD in milliseconds for each type of stimulus.
Source estimation revealed that DRF (red in Fig. 3B) differed from TRF by stronger activation of the right middle frontal gyrus, the left superior temporal gyrus, and the bilateral postcentral gyrus and right insula, while TRF (blue in Fig. 3B) induced greater activity in left precuneus, insula, thalamus and in right middle temporal gyrus.

Compared to DRF, DTJ (blue in Fig. 3C) more strongly activated bilateral middle frontal gyrus and the left anterior cingulum, left caudate nucleus, and right middle temporal gyrus.

Thus, 310 ms after stimulus onset, RF (DRF) started to dissociate from TOJ (DTJ).

**Period 530–560 ms.** Waveform analysis revealed a significant interaction between Stimulus Type and Condition over posterior ($F_{(1,16)} = 13.9$, $p = 0.002$), lateral right ($F_{(1,16)} = 5.0$, $p = 0.040$) and lateral left ($F_{(1,16)} = 5.3$, $p = 0.035$) electrodes. Post-hoc tests ($p < 0.05$) showed that DRF was more positive than TRF and DTJ over posterior electrodes, and that TRF was more negative than TTJ over posterior and lateral right electrodes. In the lateral left ROI, the response to TRF was significantly more negative than DTJ and TTJ. No significant differences were found over frontal and central electrodes.

To estimate the generators of the main effect of Condition reported in the TANOVA, we performed a t-test on the current density between RF (DRF, TRF) and TJ (DTJ, TTJ) (Fig. 3D). During RF (in red), there was greater activation of the left middle occipital gyrus, the left caudate and bilateral thalamus. By contrast, TJ (in blue) more strongly activated the right middle frontal gyrus.

Thus, in this third period of interest, there is a dissociation between the two tasks, both regarding waveforms (in the posterior region) and brain generators.

**DISCUSSION**

The aim of this study was to compare two memory control mechanisms that potentially share partially common...
Mechanisms but that have never been directly compared: ORF and TOJ. We found that these processes, when tested within a unique task and within a similar time frame around the experienced present, induce a similar electrophysiological response at an early stage of processing, but then rapidly dissociate, from 310 ms on, on the behavioral and electrophysiological levels.

In our task, the critical stimuli to investigate RF were the distracters of the third run (DRF). The inability to control this type of interference characterizes patients who confuse reality as evident in disorientation (Schnider et al., 1996a; Nahum et al., 2012) and confabulations that patients act upon (Schnider and Ptak, 1999; Nahum et al., 2012). Healthy subjects performing a task similar to the present one had activation of the posterior medial OFC (Schnider et al., 2000; Treyer et al., 2003, 2006), in agreement with the lesion site of reality confusing patients (Schnider and Ptak, 1999; Schnider, 2013).

No such strict stimulus-behavior association is known for TOJ, which was solicited to treat both distractors and targets of the third run (DTJ and TTJ) of the TJ condition. In our analysis, we concentrated on the comparison of distracters of the temporal judgment condition (DTJ) with the distracters in the reality filtering condition (DRF) because they shared all qualities except the cognitive demand specific to the third run: both types of distracters had been seen twice before presentation in the third run, had a similar temporal distance to their previous occurrence (in the first or second run) and required the same type of answer (rejection, “no”). In addition, effects specific to the two task conditions, irrespective of the type of answer (“yes or no”) were explored.

Behaviorally, the two control processes dissociated. Responses to stimuli requiring RF (DRF of run 3) were significantly more accurate, rapid, and given with more confidence than TOJ (responses to DTJ of run 3). These differences were not due to fatigue since the experimental blocks were pseudo-randomized between subjects. Thus, deciding the correct temporal context of a memory (TOJ) is more difficult and time consuming than discriminating whether a memory belongs to the present reality or not (RF), even when both refer to a similar frame of time around the present.

Electrophysiologically, the two processes shared the same configuration at an early stage: between 240 and 280 ms, distracters of both tasks differed from targets. This was visible in the topography, the waveforms (Fig. 2), and the source estimation (Fig. 3A). In the RF task, responses to distracters (DRF, the stimuli that reality confusing patients cannot handle) induced the same frontal positivity as observed in previous studies (Schnider et al., 2002; Wahlen et al., 2011; Bouzerda-Wahlen et al., 2014). We had expected a distinct signal in this period for RF, based on earlier studies (Schnider et al., 2002; Wahlen et al., 2011) but not necessarily for TOJ, on which no such data are available. The finding suggests that, at this early stage, stimuli that have just been seen undergo RF, irrespective of whether the task requires a conscious decision about previous occurrence (TOJ) or an intuitive sense about a stimulus’ relation with the present (RF).

From 310 ms on, the two mechanisms dissociated, as reflected in different topographies and ERP amplitudes. Waveform analysis revealed a trend toward greater positivity of DRF compared to DTJ over frontal electrodes and significantly stronger positivity over left lateral electrodes between 310 and 360 ms (Fig. 2). Source analysis indicated distinct anatomical correlates (Fig. 3B): TJ (processing of DTJ) activated a distributed network encompassing frontal and temporal regions. Previous functional magnetic imaging studies revealed partially similar regions in tasks of source monitoring: medial temporal lobe (Peters et al., 2007; Ross and Slotnick, 2008), precuneus (Lundstrom et al., 2005), and prefrontal cortex (in particular middle frontal gyrus) (Mitchell and Johnson, 2009). Among these regions, the prefrontal cortex has been considered particularly critical for the recollection of temporal context (Zornilla et al., 1996; Dobbins et al., 2003; Duarte et al., 2010).
Within the RF task, strong differences in cortical activation appeared between distracters and targets (Fig. 3B). Indeed, ORF is considered to dissociate memories that relate to reality (targets) from memories that do not (distracters) by modulating the extent of their cortical representation (see review in Schneider, 2008, 2013).

Our task demanded the ability to "exclude" (reject) specific stimuli, i.e., to say that a stimulus has not been seen in the current run (DRF) or has not been presented in the just preceding run (DTJ). To this extent, the task resembles the exclusion paradigm by Jacoby (1991). Wilding and colleagues (1997) examined the electrophysiological signature of a similar exclusion paradigm. At study, subjects were presented with a list of words spoken either by a male or a female voice. At test, previously heard and new words were presented. Participants were asked to indicate the old words that they had heard only from one of the two voices (i.e. male or female). The ERP component associated with successful source memory retrieval was a right frontal wave starting at 400-ms post-stimulus, which we did not observe in our study. This finding underscores that our results reflect temporal processing in memory rather than the subjects’ negative ("no") response.

Tendolkar and Rugg (1998) explored the evoked potentials induced by a more specific recency judgment task. At study, subjects saw two lists of written words. At the test, 5 min later, word pairs composed either by two words from the two study lists (Old + Old condition), by one word from a study list and a new word (Old + New condition) or by two new words (New + New condition) were presented. Subjects were asked to indicate the word which had been presented in the more recent list. The authors found a positive left parietal effect in response to Old + Old and Old + New pairs from 400 ms on, suggested to reflect the left parietal old/new effect specific to recollection (Rugg et al., 1996). In addition, they found a bilateral frontal positive wave from 300 ms on in response to Old + Old pairs that were correctly recollected. Scalp topography confirmed this frontal positivity specific to Old + Old items compared to the other type of items. The authors interpreted these results as indicative of a prefrontal contribution to recency judgment, a process dissociating from item recollection (characterized by the left parietal old/new effect) and from source memory retrieval, as studied by Wilding (characterized by the right/frontal effect) (Wilding and Rugg, 1997).

Despite using a markedly different paradigm, our results essentially concur with Tendolkar and Rugg’s (1998): we also found a frontal positivity specific to recency judgment, which started at a comparable point in time (> 300 ms). The resemblance is astonishing in light of the differences in paradigm: first, while we used images, Tendolkar and Rugg used words. Secondly, in contrast to their study, we did not include new items in the test phase. Thus, our results cannot be attributed to novelty detection. Third, the timing of the experiments differed: in the study phase, the pause between the two runs was only 3 min in our experiment, while it was 5 min in Tendolkar and Rugg’s experiment; before the test phase, there was no break interval in our task, while there was a five-minute pause in Tendolkar and Rugg’s experiment. Finally, during the test phase, Tendolkar and Rugg’s subjects had to choose the correct item from a pair, while only one item was displayed in our task. The electrophysiological similarities despite different paradigms support the idea that the frontal positivity observed both in Tendolkar and Rugg’s (1998) and our present study truly reflects processing specific to recency judgment and is independent of novelty detection.

Our results support a further conclusion from Tendolkar and Rugg (1998), namely, the dependence of recency judgment on the prefrontal cortex. Their conclusion was based on observation of the frontal potential in the waveform analysis. The type of source estimation used in the present study, which takes the static electromagnetic inverse problem into account (Michel et al., 2004), also revealed stronger bilateral activity in frontal regions for the TOJ condition. In addition, there was greater activity in temporal and parietal areas.

These results indicate that from around 300 ms after stimulus presentation the ability to explicitly place memories in their correct temporal context in the past has a specific electrophysiological correlate, which is distinct from the ability to sense memories’ relation with present reality, that is, RF. Electroencephalographic responses again dissociated between 530 and 560 ms. Inverse solutions revealed the activation of different areas, with more extended activation in subcortical (caudate, thalamus) and occipital regions during RF, and more extended activation in right frontal regions during TOJ. These late differences, in particular the frontal activation during TOJ, might reflect the cognitive effort required by this process, as revealed by the lower accuracy and longer reaction times, and by lower confidence in the judgment. However, it cannot be excluded that the differences at 530–560 ms reflect, or are at least are influenced by, varying degrees of motor preparation. Indeed, it is likely that motor preparation for the fastest responses in RF (from around 600 ms on, Table 1) has already started in this period. A relative argument against this interpretation is that there was more intense activation of right middle frontal gyrus in response during TJ, thus, for the task with slower reaction times and on the “wrong” side: the frontal cortex ipsilateral to the ensuing motor response. Nonetheless, while the reaction time differences in this study are an essential result in their right, they make it impossible to decide whether the evoked potential differences observed at 530–560 ms reflect differences in motor preparation, temporal processing, or both.

Another potential confound lies in the fact that the targets in the two tasks differed by the frequency of previous occurrence, and, therefore, potentially by familiarity: those of the TOJ task had been seen only twice before, those of RF three times. This confound appears unlikely: While familiarity effects in previous evoked potential studies on continuous recognition tasks have indeed been visible at 400–600 ms, they only appeared when comparing first and second presentations
of the stimuli (first or second run of the present task), but not when comparing the first and second appearance within a subsequent run (third run of the present task) (Schnider et al., 2002; Wahlen et al., 2011).

Of note, these caveats do not apply to the differences observed at 240–280 ms and at 310–360 ms: First, these periods largely precede reaction times and are thus unlikely to be influenced by motor preparation. Secondly, the main results from these periods referred to the responses to the distracters of the two tasks (similar response at 240–280 ms; different response at 310–360 ms), which shared all qualities (frequency of and distance to previous encounter, response) except the cognitive demands specific to the third run.

This study contributes to the growing evidence that there are multiple, dissociable memory control processes. It shows that ORF dissociates from TOJ, despite sharing early common processing, when tested within a similar temporal frame. An earlier study demonstrated that it is different from strategic content monitoring (Wahlen et al., 2011). Others studies indicated that it also dissociates from source monitoring (Johnson et al., 1993; Bouzerda-Wahlen et al., 2014), from reality monitoring (Johnson and Raye, 1981) and from the ability to discriminate between real and imagined events in the past: distinguishing between imagined and seen stimuli (Leynes, 2012) or between performed and watched actions (Leynes and Kakadia, 2013), seems to be expressed 1000 ms after stimulus onset, that is, even later than TJ or content monitoring.

In conclusion, our study demonstrates that the capacity to select memories that pertain to present reality (ORF) dissociates from the capacity to consciously place memories in their recent temporal context (TOJ), even when they refer to a similar temporal frame, the extended present. This dissociation is most evident at the behavioral level, with faster and more accurate performance during RF than TOJ. Evoked potential analysis indicates that they share the same mechanism at the initial stage, during which a memory’s relation with the present might be processed. Then, they rapidly dissociate in two following steps, with particular signals during TOJ. One step might reflect processing of the recency of a memory. The subsequent step might reflect the cognitive effort required to make a conscious TOJ or, alternatively, the earlier motor preparation in RF, mirroring the comparative ease of this capacity.

Acknowledgment—This work was supported by the Swiss National Science Foundation, Switzerland, grant no. 32003B-155947 to AS.

REFERENCES


GLOSSARY

ORF: orbitofrontal reality filtering, the ability to sense whether a memory pertains to present reality
TOJ: temporal order judgment, the ability to recollect the temporal context in which a memory was acquired

(Accepted 23 December 2014)
(Available online 14 January 2015)
Study 2

The influence of emotion on orbitofrontal reality filtering. *Submitted.*
The Influence Of Emotion On Orbitofrontal Reality Filtering

Maria Chiara Liverani¹, Aurélie L. Manuel¹, Adrian G. Guggisberg¹, Louis Nahum¹, Armin Schnider*¹

¹Department of Clinical Neuroscience, University of Geneva, Switzerland

Submitted to Journal:
Frontiers in Behavioral Neuroscience

Article type:
Original Research Article

Manuscript ID:
190546

Received on:
09 Feb 2016

Frontiers website link:
www.frontiersin.org
Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

MCL: conception of the research, design of the experiment, subjects' recruitment, data acquisition, data analysis, writing of the paper.
AM: data acquisition, data analysis, writing of the paper, revision of the paper.
LN: conception of the research, design of the experiment, data analysis, revision of the paper.
AG: conception of the research, design of the experiment, data acquisition, revision of the paper.
AS: conception of the research, design of the experiment, data analysis, writing of the paper, revision of the paper.

Keywords

Memory, reality, Emotions, orbitofrontal cortex, EEG

Abstract

Word count: 247
Orbitofrontal reality filtering is a mechanism that allows us to keep thought and behavior in phase with reality. Its failure induces reality confusion with confabulations and disorientation. Confabulations have been claimed to have a positive emotional bias, suggesting that they emanate from a tendency to embellish the situation of a handicap. Here we tested the influence of positive emotion on orbitofrontal reality filtering using a paradigm validated in reality confusing patients and with a known electrophysiological signature, a frontal positivity at 200-400 ms after memory evocation. High-density evoked potentials were recorded from nineteen healthy subjects as they performed two runs of a continuous recognition task, both composed of the same set of pictures, which were either neutral or positive. The first run measures learning and recognition. The second run, when all items are familiar, is a measure of reality filtering. Performance was more accurate and faster on neutral than positive pictures in both runs and all conditions. Evoked potential correlates of emotion and reality filtering occurred at 260-350 ms but dissociated in terms of amplitudes and topography. While positive stimuli evoked a more negative frontal potential than neutral ones, reality filtering induced a frontal positivity superimposed on the potential difference induced by stimuli's emotional valence. We conclude that the electrophysiological expression of reality filtering is not influenced by emotional positivity of the processed material. The result is consistent with the observation that reality-confusing patients with defective reality filtering usually enact common habits rather than emotionally salient activities.

Funding statement

This work was supported by the Swiss National Science Foundation, grant no. 32003B-155947 to AS.

Ethics statement

(Author are required to state the ethical considerations of their study in the manuscript including for cases where the study was exempt from ethical approval procedures.)

Did the study presented in the manuscript involve human or animal subjects: Yes

Please state the full name of the ethics committee that approved the study. If the study was exempt from this requirement please state the reason below.
Ethics Committee of the Canton of Geneva

Please detail the consent procedure used for human participants or for animal owners. If not applicable, please state this.
Participants gave written informed consent before their participation in the study. They were informed about the aim of the study, the advantages, disadvantages and potential risks regarding the technique used. They were informed that they were covered in case of potential damages occurring during the experiment, and that they could stop the experiment at any time. Finally, they were informed about the confidentiality of the data.
Please detail any additional considerations of the study in cases where vulnerable populations were involved, for example minors, persons with disabilities or endangered animal species. If not applicable, please state this.

No vulnerable populations were involved in this study.
The Influence Of Emotion On Orbitofrontal Reality Filtering

Maria Chiara Liverani¹, Aurélie L. Manuel¹, Adrian G. Guggisberg¹, Louis Nahum¹, Armin Schnider¹*

¹ Laboratory of Cognitive Neurorehabilitation, Division of Neurorehabilitation, Department of Clinical Neuroscience, University Hospital and University of Geneva, Geneva, Switzerland

Running title:
Orbitofrontal Reality Filtering and Emotions

Correspondance:
Prof. Armin Schnider
armin.schnider@hcuge.ch

Keywords: memory, reality, emotions, orbitofrontal cortex, EEG

Number of figures: 2
Word count: 3711
Abstract

Orbitofrontal reality filtering is a mechanism that allows us to keep thought and behavior in phase with reality. Its failure induces reality confusion with confabulations and disorientation. Confabulations have been claimed to have a positive emotional bias, suggesting that they emanate from a tendency to embellish the situation of a handicap. Here we tested the influence of positive emotion on orbitofrontal reality filtering using a paradigm validated in reality confusing patients and with a known electrophysiological signature, a frontal positivity at 200-400 ms after memory evocation. High-density evoked potentials were recorded from nineteen healthy subjects as they performed two runs of a continuous recognition task, both composed of the same set of pictures, which were either neutral or positive. The first run measures learning and recognition. The second run, when all items are familiar, is a measure of reality filtering. Performance was more accurate and faster on neutral than positive pictures in both runs and all conditions. Evoked potential correlates of emotion and reality filtering occurred at 260-350 ms but dissociated in terms of amplitudes and topography. While positive stimuli evoked a more negative frontal potential than neutral ones, reality filtering induced a frontal positivity superimposed on the potential difference induced by stimuli’s emotional valence. We conclude that the electrophysiological expression of reality filtering is not influenced by emotional positivity of the processed material. The result is consistent with the observation that reality-confusing patients with defective reality filtering usually enact common habits rather than emotionally salient activities.
1 Introduction

To sense whether a thought constitutes a souvenir from the past, a potential plan for the
future, or refers to current reality is crucial for meaningful behavior. Failure to make this
distinction is associated with reality confusion, as expressed in disorientation and
confabulations, which patients act upon. We have called this syndrome, which corresponds to
the original description of the Korsakoff syndrome (Korsakoff, 1891; Jolly, 1897),
“behaviorally spontaneous confabulation” (Schnider, 2003; 2008).

Confabulations have received multiple interpretations as a combination of amnesia with
executive dysfunction (Kapur and Coughlan, 1980; Papagno and Baddeley, 1997; Nys et al.,
2004), a failure to “monitor” memory output (Burgess and Shallice, 1996; Moscovitch and
Melo, 1997; Johnson and Raye, 1998; Gilboa et al., 2006), a disturbed sense of time (Talland,
1961; Dalla Barba, 1993), or a desire to fill gaps in memory to avoid embarrassment (Flament,
1957) and maintain self-coherence (Conway and Tacchi, 1996). A related proposal holds that
motivational factors may induce a positively biased memory recall in amnesia, which would
lead to confabulations (Fotopoulou et al., 2007; Fotopoulou et al., 2008, Fotopoulou, 2010
#354). Fotopoulou et al. (2008) based this proposal on the observation of a positive emotional
bias in the content of confabulations. Similarly, Alkathiri et al. (2015) described a positive
bias in the false recall and recognition of confabulating patients.

These hypotheses tried to explain confabulations as a purely verbal phenomenon, irrespective
of the presence of reality confusion. We found that patients who confabulate in the context of
reality confusion, as reflected in acts according to the confabulations and disorientation,
specifically failed in an experimental task. They underwent repeated runs of a continuous
recognition task, each composed of the same set of pictures but arranged in different order,
and had to indicate whether they had seen the picture in the current run. Healthy subjects
usually have no difficulty in distinguishing between pictures seen in previous or in the current
run. In contrast, patients with behaviorally spontaneous confabulations had a specific increase
of false positives from the second run on (Schnider et al., 1996a;b; Schnider and Ptak,
1999; Nahum et al., 2012). Recovery from reality confusion was accompanied by
normalization of the false positive rate (Schnider et al., 2000a). Lesions involved the posterior
medial orbitofrontal cortex or directly connected structures (Schnider et al., 1996a; Schnider
and Ptak, 1999); healthy subjects performing the described task activated this area (Schnider
et al., 2000b). We have called the ability measured with this task orbitofrontal reality filtering
(ORFi) (Schnider, 2013). Evoked potentials (ERPs) in healthy subjects performing the task
showed a distinct frontal positivity at about 200-300 ms in response to the stimuli on which
the patients had failed, namely, first appearances (“distracters”) within the second run
(Schnider et al., 2002; Wahlen et al., 2011; Liverani et al., 2015).

Patients with behaviorally spontaneous confabulation typically enact common habits: they
think they have business meetings or familial obligations. There is no obvious positive
emotional bias in their perception of reality, which is at odds with the claim that confabulations
do have such a bias. If Fotopoulou et al.’s (2008; Fotopoulou) hypothesis also applied to
behaviorally spontaneous confabulation, then orbitofrontal reality filtering would be expected
to be influenced by the emotional valence of stimuli, too. In the present study we investigated
whether orbitofrontal reality filtering is indeed influenced by positive emotion.

We composed two runs of a continuous recognition task, similar to previous clinical and
imaging studies (Schnider, 2003; 2008), but including both neutral and emotionally positive
pictures. We expected that there would be a valence effect in the ERP’s from about 200-300
ms on (Palomba et al., 1997; Cuthbert et al., 2000; Maratos et al., 2000; Bradley et al.,
2007; Van Strien et al., 2009). Critical for the present study, however, we expected that
stimuli’s emotion would have no influence on reality filtering: there would be the typical
frontal positivity in response to first appearances of pictures in the second run around 200-300 ms irrespective of or superimposed on emotional valence (Schnider et al., 2002; Wahlen et al., 2011; Liverani et al., 2015).

2 Materials and Methods

2.1 Participants

Twenty right-handed participants gave written informed consent and were paid to participate in the study. They reported no history of psychiatric or neurological disorders or medication use. One subject was excluded from the study because of poor signal quality of electrophysiological data. Finally 19 subjects (10 females, age 27 ± 5 years) were included in the analysis. The research was approved by the local Ethics Committee and conducted according to the Declaration of Helsinki.

2.2 Stimuli

The set of positive and neutral stimuli consisted of 240 realistic, high-quality photographs from the Nencki Affective Picture System (NAPS) (Marchewka et al., 2013), divided into five categories: landscapes, objects, animals, people and faces. The NAPS was chosen as it allows one to match pictures according to visual features, such as complexity, luminosity and contrast, which are critical for an ERP study. Pictures were selected on the basis of their original valence rating, which was collected using a bipolar scale ranging from 1 to 9 (with 1 = very negative, 5 = neutral, 9 = very positive) (Marchewka et al., 2013). Half of the pictures had a neutral valence (4.7 - 5.3; mean = 5 ± 0.2), the other half had a positive valence (7.3 - 8.5; mean = 7.7 ± 0.3; t(119) = 273, p < .001). Positive pictures were also more relaxing: they had lower arousal ratings (2 - 4.5; mean = 3.8 ± 0.9) than neutral pictures (4.1 - 7.4; mean = 5.2 ± 0.6; t(119) = 14.65, p < .001). Stimuli were matched with respect to their luminance, contrast and entropy (t(119) > 172, p > .65).

2.3 Task composition

To avoid fatigue, subjects performed three independent blocks of the task, separated by a 10-minutes' break. Each block was constituted of a different set of 80 images (40 positive, 40 neutral) used to compose two runs of a continuous recognition task (Figure 1). The two runs of a block were composed of the same pictures. Picture repetitions occurred after 8 to 16 intervening images. Participants were asked to indicate picture repetitions within, and only within, the currently ongoing run by pressing the right button with the right middle finger to indicate first presentations (distracters; “No, not yet seen”) within the run and the left button with the right index for repeated images within the ongoing run (targets; “Yes, already seen.”). The first run measures learning and recognition and can be solved on the basis of familiarity alone. In the present study, this run thus contained the following stimulus types: positive distracters, PDrun1; neutral distracters, NDrun1; positive targets, PTrun1; neutral targets, NTrun1.

The second run was used to test reality filtering (Schnider, 2008; 2013). The same set of images was presented, but rearranged in a different order. As subjects are already familiar with all stimuli, this run requires the ability to sense whether familiarity emanates from previous occurrence within the ongoing run (“ongoing reality”) or from the previous run. This run contains the following stimulus types: positive distracters, PDrun2; neutral distracters, NDrun2; positive targets; PTrun2; neutral targets, NTrun2.
Amnesic subjects typically have difficulty with this task’s first run (Schnider et al., 1996a; Schnider and Ptak, 1999); correct performance by healthy subjects activates the medial temporal lobe (Schnider et al., 2000b). Reality-confusing patients with confabulations and disorientation fail in this second run: they typically have a specific increase of false positives (Schnider and Ptak, 1999; Nahum et al., 2012). Performing this task’s second run activates the posterior orbitofrontal cortex in healthy subjects (Schnider et al., 2000b). Thus, the processing of distracters (PDrun2; NDrun2) is the critical capacity in the second run.

The two runs within each block were made in immediate succession. Images were presented on a 17 inch computer screen for 2000 ms followed by an interstimulus interval with a fixation cross in the center of the screen presented for 700 ms. The task was performed using E-prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, USA; www.pstnet.com/eprime). Before the task, participants performed a training session with geometrical figures (N = 10) instead of the experimental pictures, in order to familiarize with task instructions.

2.4 EEG acquisition and raw data processing
Electroencephalography (EEG) was continuously recorded with a 128-channel Active-Two Biosemi EEG system (Biosemi V.O.F. Amsterdam, Netherlands). Signal was sampled at 512 Hz and filtered at bandwidth of 0-104 Hz. Electrodes impedance was kept below 20 KΩ.

EEG data preprocessing and analyses were performed with the Cartool software (https://sites.google.com/site/fbmlab/cartool) developed by Denis Brunet (Brunet et al., 2011). Epochs from 100 ms before to 800 ms after the stimulus onset were averaged for each subject and each condition to calculate the event-related potentials (ERPs). ERPs were band-pass filtered to 1-30 Hz and recalculated against the average reference. Baseline correction was applied on the 100 ms pre-stimulus period. Only correct trials were retained in the analysis and epochs with artifacts higher than 100 µV were automatically excluded. In addition, data were visually inspected and EEG epochs with eye blinks, movements or other sources of transient noise were rejected. Before the group averaging, channels with substantial noise were interpolated using a spherical spline interpolation (mean 4.2% of interpolated electrodes) (Perrin et al., 1987). The epoch from 0 to 800 ms after stimulus onset was retained for analysis. The mean number (± SD) of accepted epochs was 73 ± 20.8 for PDrun1, 75.5 ± 21 for NDrun1, 74.7 ± 20.4 for PTrun1, 74.3 ± 21.2 NTrun1, 72.9 ± 23 for PDrun2, 71.4 ± 25.1 for NDrun2, 68.8 ± 25.7 for PTrun2 and 70.7 ± 27.4 for NTrun2.

2 x 2 x 2 repeated measures ANOVA with factors Run (1, 2), Stimulus (Distracter, Target) and Emotion (Positive, Neutral) revealed no statistical differences in number of epochs (p > 0.05).

2.5 Behavioral data analyses
2 x 2 x 2 repeated measures analysis of variances (ANOVAs) on percentage of correct responses and reaction time were performed with the within-subjects factors Run (1, 2), Stimulus (D, T) and Emotion (P, N). When appropriate, post-hoc Fisher’s tests were performed, with a significance level of p < 0.05.

2.6 Topographic patterns analyses (TANOVA)
To identify periods with significant statistical differences in the topography of the electric field across conditions, a time-wise 2 x 2 topographic ANOVA (TANOVA) was performed for each run. This is a non-parametric randomization test (5000 randomizations per time point) based on global dissimilarities between electric fields implemented in the RAGU software (Randomization Graphical User Interface; (Koenig et al., 2011)). Analysis was
performed using a within-subject design with the factors Stimulus (D, T) and Emotion (P, N). The statistically significant periods of interest were defined by a p < 0.05 for ≥ 20 ms (Guthrie and Buchwald, 1991) (See Murray et al., (2006) and Toepel et al., (2014), for a similar procedure).

2.7 Global Waveform Analyses
Electrode- and time-wise 2 x 2 repeated measures ANOVAs on the ERP waveform for each of the 128 electrodes were conducted with the factors Stimulus (D, T) and Emotion (P, N) for each run. This analysis was performed with the Statistical Toolbox for Electrical Neuroimaging (STEN) developed by Jean-François Knebel (http://www.unil.ch/line/home/menuinst/about-the-line/software--analysis-tools.html). To account for temporal autocorrelation only periods that remained significant (p < 0.05) for ≥ 20 ms were considered reliable. We also applied a spatial extent criterion of at least 5 electrodes (i.e. 4% of the electrode montage) for each time sample considered statistically significant.

2.8 Regional Frontal Waveform Analysis
Since the electrophysiological correlate of orbitofrontal reality filtering is typically expressed in the frontal region (Schnider et al., 2002), we defined a region of interest (ROI) composed of 16 electrodes in the frontal part of the scalp (Dien and Santuzzi, 2005). ERPs of electrodes in this region were averaged for each condition and each subject. Apparent differences were statistically tested using repeated-measures ANOVAs with the factors Stimulus (D, T) and Emotion (P, N) for each run and subjected to post-hoc tests.

3 Results

3.1 Behavioral results
Behavioral results are summarized in Table 1. Accuracy was better for neutral than positive pictures (F (1,18) = 14.56, p = 0.001, η^2_p = 0.45). There was a significant interaction Run x Stimulus (F (1,18) = 9.93, p = 0.006, η^2_p = 0.36). Post-hoc tests revealed that accuracy was better in response to Distracters than Targets in the first, but not the second run (t (18) = 1.73, p = 0.01). Reaction times were longer in run 2 than run 1 (F (1,18) = 14.60, p = 0.001, η^2_p = 0.45), in response to Distracters than Targets (F (1,18) = 17.56, p = 0.001, η^2_p = 0.49) and in response to positive than neutral pictures (F (1,18) = 17.49, p = 0.001, η^2_p = 0.49).

3.2 Topographic pattern analysis (TANOVA)
The TANOVA revealed significant effects of Emotion in both runs, but more temporally consistent in run 1, in a relatively early period, between about 270 and 390 ms (Figure 2A). The effects of Stimulus type (Figure 2B) were more extended in both runs but with distinct differences. Specifically, significant topographic differences were observed in the following periods:

3.2.1 Run 1: There was an effect of Emotion at about 270 - 390 ms post stimulus onset (Figure 2A, left column). Stimulus types induced an early main effect at about 70 - 170 ms, and a later, prolonged effect from 280 ms on (Figure 2B, left column).
3.2.2 Run 2: There was a main effect of Emotion at 210 - 240 ms and a brief effect at 350 - 370 ms (Figure 2A, right column). A main effect of Stimulus was present at 170 - 230 ms, and from 250 ms on (Figure 2B, right column).

3.3 Global Waveform analysis

Global waveform analysis using ANOVAs over all electrodes yielded results consistent with the TANOVA: Emotion was expressed in both run, but more distinctly in run 1, mainly at about 280-400 ms, with brief effects before and after. Stimulus type had extended main effects, from about 350 to 800 ms in the first run. In the second run, extended effects were present at an earlier stage, at around 200 and from about 280 to 350 ms. In details, there were the following effects:

3.3.1 Run1: As visible in Figure 2C (left column), ANOVAs over all electrodes showed an effect of emotion at 45 - 80 ms, 105 - 130 ms, 260 - 400 ms, 720 - 740 ms over frontal and lateral right scalp regions. There was a main effect of Stimulus over extended periods and extended scalp regions at 50 - 160 ms, 270 - 390 ms and 400 - 800 ms (Figure 2B, left column).

3.3.2 Run2: There was a significant main effect of Emotion at 120 - 140 ms, 200 - 255 ms and 290 - 390 ms (Fig. 2C, right column), and a main effect of Stimulus at 165 - 220 ms, 250 - 380 ms, 390 - 460 ms, 490 - 670 ms, and 700 - 790 ms (Figure 2D, right column).

3.4 Regional Frontal Waveform Analysis

As Figure 2E shows, Emotion was expressed both runs at around 270-400 ms. In the first run, this effect prevailed over the effect of Stimulus type. In the second run, the effect of Stimulus type was superimposed on the effect of Emotion, in both cases with Distracters being less negative than targets. This corresponds to the expected signature of ORFi. After 400 ms, there was only an extended effect of stimulus type, but no effect of Emotion, consistent with the TANOVA and the Global Waveform Analysis. In detail, the following findings were obtained:

3.4.1 Run 1: A directed 2 x 2 repeated measures ANOVA on amplitude difference in the frontal ROI during the period showing the main effect of Emotion and of Stimulus (270-390 ms) in the TANOVA and waveforms results showed that neutral stimuli (both distracters and targets) evoked more positive responses than positive stimuli (F (1,18) = 13.42, p = 0.002, ηp² = 0.43).

The late main effect of Stimulus present in both TANOVA and waveform analysis at about 400 - 800 ms corresponded to a higher positive potential for Distracters compared to Targets in the frontal region (F (1,18) = 4.71, p = 0.04, ηp² = 0.21).

3.4.2 Run 2: There was a main effect of Stimulus at 250 - 380 ms: neutral stimuli evoked more positive responses than positive stimuli (F (1,18) = 7.06, p = 0.02, ηp² = 0.28). In both cases, distracters evoked more positive responses than targets (F (1,18) = 8.41, p = 0.02, ηp² = 0.29), similar to earlier studies on ORFi (Schnider et al., 2000b; Wahlen et al., 2011; Bouzerda-Wahlen et al., 2014; Liverani et al., 2015). There was no significant interaction (p = .127).

4 Discussion

This study shows that orbitofrontal reality filtering (ORFi) is not dependent on or influenced by positive emotional valence of the processed information. The experiment used to test this is a confirmed surrogate marker of reality confusion in confabulating patients (Schnider et al., 1996a; b; Schnider and Ptak, 1999; Schnider et al., 2000a; Gilboa et al., 2006; Nahum et al., 2012). It activates the orbitofrontal cortex (area 13) in healthy subjects (Schnider et al., 2000b; Treyer et al., 2003; 2006) and has a known electrophysiological signature: a relative
frontal positivity (or absence of negativity) at about 200 - 300 ms in response to distracters of run 2 (Schnider et al., 2002; Wahlen et al., 2011; Bouzera-Wahlen et al., 2014; Liverani et al., 2015).

In the present study, positive stimuli clearly differed from neutral ones: they were less well recognized in both runs of our task and they induced a more negative evoked potential over the frontal electrode in both runs around 250 to 400 ms. The behavioral result, albeit in contradiction to our hypothesis, is compatible with previous observations: emotional stimuli are typically better recognized after a delay (Palomba et al., 1997; La Bar and Phelps, 1998; Ochsner, 2000; Versace et al., 2010), but they may have no advantage (Van Strien et al., 2009; Treese et al., 2010) or even be less well recognized, typically due to more false positive responses, with short delays (Maratos et al., 2000; Windmann and Chmielewski, 2008).

Another possible explanation lies in the lower arousal associated with the positive stimuli used – the price for truly “positive” pictures, which may have a relaxing effect. However, similar to valence, the behavioral effect of arousal is difficult to predict: while better retained at long intervals, high arousal stimuli may be less well recognized at short intervals (Kleinsmith and Kaplan, 1963). For the sake of the present study, it is important to note that the emotional stimuli did induce a different behavioral response than the neutral ones.

Similar considerations hold for the evoked potential response. Effects of emotionality have strongly varied across studies depending on the type and presentation of material and the specific task (Olofsson et al., 2008). In most studies, emotional stimuli (positive or negative) evoked more positive responses than neutral ones (Palomba et al., 1997; Cuthbert et al., 2000; Maratos et al., 2000; Bradley et al., 2007) at various time points and over different electrodes. Although results between studies differed considerably, it appears that valence effects (positive - negative) are typically expressed at an earlier stage, around 200 - 300 ms, than arousal effects, which occur at 300 - 1000 ms (Van Strien et al., 2009; Versace et al., 2010). The main potential differences in the present study occurred in the earlier period, around 250 - 400 ms, and may thus reflect the influence of the positive valence rather than the lower arousal of our emotional stimuli.

The present study was interested in the influence of positive emotion on ORFi and, therefore, in the time period between 200 - 400 ms. Thus, the crucial result is that positive stimuli did differ both behaviorally and electrophysiologically from neutral stimuli already when ORFi was not yet challenged, that is, in the first run. However, while the difference between positive and neutral stimuli also came out in the critical second run, the signature of reality filtering was present irrespective of the emotional valence of the stimuli. Thus, reality filtering worked on neutral and emotionally positive stimuli in the same way.

What does this result mean for the mechanism of reality confusion and confabulation? First, it indicates that the core mechanism of reality confusion, as it may result from damage to the posterior medial OFC or directly connected structures (Schnider et al., 1996a; b; Schnider and Ptak, 1999; Nahum et al., 2012), is not modulated by emotion. This is compatible with an additional line of evidence. As we have previously shown, the tendency of reality-confusing patients to enact daily routines appears to reflect an inability to abandon anticipations that are not currently valid. In other words, they have a specific defect of extinction capacity (Nahum et al., 2009; Nahum et al., 2012). The typical signature of this need to abandon a behavior - a frontal positivity also at about 200 - 300 ms (Schnider et al., 2007) - does not depend on whether the presented outcome is positive (reward) or neutral (no reward), but on the need to adapt behavior (Nahum et al., 2011). Thus, similar to the result of the present study, extinction capacity, too, does not appear to be dependent on the emotional connotation of the outcome.

Our result refers to a specific mechanism of reality confusion and confabulation - orbitofrontal reality filtering. It leaves open the possibility that motivational factors, such as
positive emotions, influence the content of confabulations and maybe also the desire to verbally express them. Such factors have long been speculated to influence confabulations occurring in the context of amnesia (Flament, 1957; Weinstein, 1987; Fotopoulou et al., 2008), but also in other, non-cerebral disease states, “to explain away manifestations of illness” (Weinstein and Kahn, 1955). This notion is difficult to apply to confabulations emanating from reality confusion, all the more so because profoundly negative enactments have been described, such as, the repeated urge to organize the funeral of (still alive) loved ones (Korsakoff, 1891; Nahum et al., 2012).

In conclusion, orbitofrontal reality filtering, whose failure causes reality confusion with confabulation and disorientation, does not appear to be modulated by positive emotional valence of the treated material. While positive emotional bias may still influence the content of false memories and their expression, it does not appear to influence the core mechanism of the reality confusion underlying specific confabulations and disorientation.

List of non-standard abbreviations:

ORFi: Orbitofrontal reality filtering
OFC: Orbitofrontal cortex
ERP: event related potential
TANOVA: topographic ANOVA
PD: positive distracter
ND: neutral distracter
PT: positive target
NT: neutral target

Acknowledgment:
We thank Prof. David Sander for advice in the choice of the stimuli.

Founding:
This work was supported by the Swiss National Science Foundation, grant no. 32003B-155947 to AS.

Conflict of interest:
The authors declare no competing financial interests.

Author contributions:
MCL: conception of the research, design of the experiment, subjects’ recruitment, data acquisition, data analysis, writing of the paper.
AM: data acquisition, data analysis, writing of the paper, revision of the paper.
LN: conception of the research, design of the experiment, data analysis, revision of the paper.
AG: conception of the research, design of the experiment, data acquisition, revision of the paper.
AS: conception of the research, design of the experiment, data analysis, writing of the paper, revision of the paper.
References


Table 1: Behavioral results of the reality filtering task.

The second column indicates the percentage of correct responses ± standard deviation (SD) for each type of stimulus. The third column indicates mean of reaction times ± SD in milliseconds for each type of stimulus. PDrun1 = positive distracters of run 1; NDrun1 = neutral distracters of run 1; PTrun1 = positive targets of run 1; NTrun1 = neutral targets of run 1; PDrun2 = positive distracters of run 2; NDrun2 = neutral distracters of run 2; PTrun2 = positive targets of run 2; NTrun2 = neutral targets of run 2;

<table>
<thead>
<tr>
<th>Stimulus type</th>
<th>% Correct Responses</th>
<th>Reaction times (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDrun1</td>
<td>98 ± 2.6</td>
<td>809.6 ± 222.3</td>
</tr>
<tr>
<td>NDrun1</td>
<td>99 ± 2.3</td>
<td>786.9 ± 203.8</td>
</tr>
<tr>
<td>PTrun1</td>
<td>94.2 ± 6.5</td>
<td>732.7 ± 85</td>
</tr>
<tr>
<td>NTrun1</td>
<td>96.3 ± 4</td>
<td>727.6 ± 85.9</td>
</tr>
<tr>
<td>PDrun2</td>
<td>96.2 ± 3.6</td>
<td>835.2 ± 249.2</td>
</tr>
<tr>
<td>NDrun2</td>
<td>96.7 ± 2.8</td>
<td>829.8 ± 237.5</td>
</tr>
<tr>
<td>PTrun2</td>
<td>95.2 ± 5.2</td>
<td>767.7 ± 89.3</td>
</tr>
<tr>
<td>NTrun2</td>
<td>96.2 ± 4.4</td>
<td>747.7 ± 83.3</td>
</tr>
</tbody>
</table>

Figure legends:

Figure 1: Task design
Each block was composed of two runs. Positive Distracters (PD) and Neutral Distracters (ND) are positive and neutral images that are presented for the first time within a run, respectively. Positive Targets (PT) and Neutral Targets (NT) are positive and neutral images that are repeated within the same run, respectively.

Figure 2: Electrical neuroimaging results.
(A), (B): Time-wise topographic analysis on run 1 and 2. Results of the 2 x 2 repeated measures ANOVAs on the topography with the factors Stimulus (Distracters, Targets) and Emotion (Positive, Neutral). The dotted line indicates p < 0.01. Periods with significant main effects (p < 0.01, ≥ 20 consecutive ms) are showed in black. (C), (D): Electrode- and time-wise ERP waveform analysis on run 1 and 2 and on different scalp regions (gray triangles). Results from the 2 x 2 repeated measures ANOVAs on the waveforms with the factors Stimulus (Distracters, Targets) and Emotion (Positive, Neutral). Black lines indicate significant effects with p < 0.01 and lasting at least 20 ms. (E): Grand average ERPs waveforms for the 16 -electrodes frontal ROI in response to Positive Distracters (PD, in magenta), Neutral Distracters (ND, in blue), Positive Targets (PT, in green) and Neutral Targets (NT, in black). Traces are displayed in microvolts (µV) as a function of time relative to stimulus onset. Repeated measures ANOVA were calculated in the run 2 across the time window derived from TANOVA and boxed in gray (260 - 350 ms).
Figure 2.

A. Main Effect Emotion run 1

B. Main Effect Stimulus run 1

C. Main Effect Emotion run 1

D. Main Effect Stimulus run 1

E. Frontal electrodes run 1

F. Frontal electrodes run 2

- PD
- ND
- PT
- NT
Study 3

Children’s sense of reality: The development of orbitofrontal reality filtering

Maria Chiara Liverani, Aurélie L. Manuel, Louis Nahum, Veronica Guardabassi, Carlo Tomasetto & Armin Schnider

To cite this article: Maria Chiara Liverani, Aurélie L. Manuel, Louis Nahum, Veronica Guardabassi, Carlo Tomasetto & Armin Schnider (2015): Children’s sense of reality: The development of orbitofrontal reality filtering, Child Neuropsychology, DOI: 10.1080/09297049.2015.1120861

To link to this article: http://dx.doi.org/10.1080/09297049.2015.1120861

Published online: 18 Dec 2015.

Submit your article to this journal

Article views: 40

View related articles

View Crossmark data
Children’s sense of reality: The development of orbitofrontal reality filtering

Maria Chiara Liverani, Aurélie L. Manuel, Louis Nahum, Veronica Guardabassi, Carlo Tomasetto and Armin Schnider

Laboratory of Cognitive Neurorehabilitation, Medical School, University of Geneva, Switzerland; Department of Psychology, University of Bologna, Bologna, Italy; Division of Neurorehabilitation, Department of Clinical Neurosciences, University Hospital and University of Geneva, Switzerland

ABSTRACT

Orbitofrontal reality filtering denotes a memory control mechanism necessary to keep thought and behavior in phase with reality. In adults, it is mediated by the orbitofrontal cortex and subcortical connections and its failure induces reality confusion, confabulations, and disorientation. Here we investigated for the first time the development of this mechanism in 83 children from ages 7 to 11 years and 20 adults. We used an adapted version of a continuous recognition task composed of two runs with the same picture set but arranged in different order. The first run measures storage and recognition capacity (item memory), the second run measures reality filtering. We found that accuracy and reaction times in response to all stimulus types of the task improved in parallel across ages. Importantly, at no age was there a notable performance drop in the second run. This means that reality filtering was already efficacious at age 7 and then steadily improved as item memory became stronger. At the age of 11 years, reality filtering dissociated from item memory, similar to the pattern observed in adults. However, performance in 11-year-olds was still inferior as compared to adults. The study shows that reality filtering develops early in childhood and becomes more efficacious as memory capacity increases. For the time being, it remains unresolved, however, whether this function already depends on the orbitofrontal cortex, as it does in adults, or on different brain structures in the developing brains of children.

Memory encompasses different processes that do not develop at the same time during childhood. The ability to store and recollect new information (item memory) is already present at the age of 4 years (Marshall, Drummey, Fox, & Newcombe, 2002), similar to the ability to place a memory in its correct temporal sequence (Friedman, 1991, 1992; McCormack & Russell, 1997). By comparison, source memory, i.e., the capacity to attribute memories to specific sources and contexts, develops only from 6 to 7 years of age onwards (Cycowicz, Friedman, Snodgrass, & Duff, 2001; Czernochowski, Mecklinger, & Johansson, 2009; Drummey & Newcombe, 2002).

CONTACT Armin Schnider armin.schnider@hcuge.ch Service de Neurorééducation, Hôpitaux Universitaires de Genève, Av. de Beau-Séjour 26, CH-1211 Geneva 14/Switzerland.

© 2015 Taylor & Francis
Orbitofrontal reality filtering refers to the capacity to discriminate between thoughts (memories) that relate to ongoing reality, and thoughts that do not (Schnider, 2008, 2013). The existence of this mechanism was derived from the observation of patients who, after brain damage, confuse reality: they are disoriented and act according to ideas that have no relation with true reality and justify their behavior with confabulations. For example, a 58-year-old mother who had suffered orbitofrontal damage following the rupture of an aneurysm of the anterior communicating artery repeatedly left the examination in the conviction that she had to feed her baby—but her child was now a 38-year-old adult (Schnider, von Däniken, & Gutbrod, 1996b). In comparison to amnesic patients, who do not exhibit such inappropriate behavior, these patients specifically failed in an experimental task: while their performance was similar to other amnesic patients when they carried out a continuous recognition task for the first time, they had a dramatic performance drop when they carried out the task for the second time, with the same pictures arranged in different order. They had a specific increase in false positives (Nahum, Bouzerda-Wahlen, Guggisberg, Ptak, & Schnider, 2012; Schnider & Ptak, 1999; Schnider et al., 1996b), which also strongly correlated with the level of disorientation (Nahum et al., 2012; Schnider, von Däniken, & Gutbrod, 1996a). The disappearance of false acts and confabulations were specifically accompanied by the normalization of the false positive rate on the second run (Schnider, Ptak, von Däniken, & Remonda, 2000). In adults, the mechanism depended on the orbitofrontal cortex: lesions always involved the posterior medial orbitofrontal cortex (area 13) or directly connected structures (Schnider, 2008; Schnider et al., 1996a; Schnider & Ptak, 1999). Functional imaging studies in healthy subjects showed activation in the orbitofrontal cortex area 13 in the second run of continuous recognition tasks, either when using pictures (Schnider, Treyer, & Buck, 2000; Treyer, Buck, & Schnider, 2003) or meaningful sounds (Treyer, Buck, & Schnider, 2006). The mechanism underlying orbitofrontal reality filtering appears to be similar to extinction capacity, the ability to abandon anticipations that are no longer valid (Nahum, Ptak, Leemann, & Schnider, 2009). Recent studies have shown that this process is rapid (200–300 ms after stimulus presentation; Schnider, Valenza, Morand, & Michel, 2002), and that it dissociates from other memory control processes like context source memory (Bouzerda-Wahlen, Nahum, Liverani, Guggisberg, & Schnider, 2014; Johnson, Kounios, & Nolde, 1997; Mitchell & Johnson, 2009), strategic retrieval monitoring (Bouzerda-Wahlen, Nahum, Ptak, & Schnider, 2013; Gilboa et al., 2006; Wahlen, Nahum, Gabriel, & Schnider, 2011) and temporal order memory (Liverani et al., 2015). These processes are laborious and have different signatures in evoked potential studies (Bouzerda-Wahlen et al., 2014; Liverani et al., 2015; Wahlen et al., 2011). Orbitofrontal reality filtering is also different from reality monitoring, the ability to distinguish between the memory of a thought and a memory of a perception experienced in the past (Johnson & Raye, 1981). At what age orbitofrontal reality filtering develops in children has never been examined.

Children’s behavior is at times reminiscent of reality confusion by confabulating patients. Children are more vulnerable to memory distortions than adults (Ceci & Bruck, 1993; Schacter, Kagan, & Leichtman, 1995), and they produce more false narratives when exposed to leading questions (Ceci & Bruck, 1993). Similar to
confabulating patients, these false memories often have an autobiographical content, are rich in details, and are firmly held by the children even when faced with disconfirming evidence. In addition, children have difficulty with source monitoring. Foley and Johnson (1985) asked children to distinguish between a task that they had really performed and a task that they had only imagined. Children aged 6 to 9 years failed to correctly recall this kind of information. Gopnik and Graf (1988) asked 3- to 5-year-old children to retrieve information about the content of a series of drawings that was acquired in three different ways: they actually saw the drawings, were told about them, or had to infer the contents from a clue. The youngest children had more difficulties in retrieving the source of their memories than the older children.

These observations suggest a possible link between children’s susceptibility to memory distortion and immaturity of reality filtering. In this study we developed an adapted version of the task used in adults (Nahum et al., 2012; Schnider et al., 1996b; Schnider & Ptak, 1999). It consists of repeated runs of a continuous recognition task, all composed of the same set of pictures, which are arranged in a different order on each run. The first run measures the ability to encode and subsequently recognize pictures (item memory). In the second run, all items are familiar, so that the correct recognition of the picture repetitions demands the ability to sense whether a picture is known from a previous occurrence in the present run (“ongoing reality”) or from an occurrence in the previous run (“the past”). This task thus allows comparison of the development of item memory as opposed to reality filtering. We tested three groups of children, aged 7, 9 and 11 years, and adults, aged around 30 years, to determine when reality filtering develops and whether it does so in parallel with or separately from item memory.

Method
Participants
A total of 91 children and 20 adults participated in the study. Children attended public primary schools in suburban areas in the North of Italy. Of the children, 5 were excluded from the analysis because they always pressed the same response button during the reality filtering task and 3 were excluded because of reaction times exceeding 4000 ms in the reality filtering task; thus, 83 children were included in the analyses. Of these remaining children, 20 attended the first grade in primary school (7-year-old group; \(M_{\text{age}} = 84.9\) months, age range = 77.4–92.4 months), 33 attended the third grade (9-year-old group; \(M_{\text{age}} = 107.7\), age range = 104.9–110.5 months) and 30 attended the fifth grade (11-year-old group; \(M_{\text{age}} = 130.6\), age range = 127.2–134 months). All participants were of typical age for their grade. A fourth group of 20 adults served as controls for the reality filtering task (\(M_{\text{age}} = 31.5\) years, age range = 28–34).

Materials and procedure
All subjects gave informed consent. For the children, parents provided informed consent. All children attended normal classes. Children in need of an individual education plan (Italian schools: special programs for children with cognitive disabilities) were excluded. Permission to conduct the study was granted by school principals and
parents. The study was approved by the Institutional Ethical Committee of the University of Bologna, Italy.

Children performed the tasks and filled out the questionnaire with the help of the experimenter during school hours, in a quiet room and individually.

**Reality filtering task**

Item recognition (IR) and reality filtering were tested with an adapted version of a continuous recognition task developed by Schnider and colleagues ([Figure 1; Schnider, 2003, 2013; Schnider et al., 1996a]).

The task was composed of two runs in which the same set of 30 cartoon images of animals was presented. Each cartoon was repeated once after 6 to 9 intervening pictures. In both runs, participants were asked to answer to the question “Have you already seen this image within the ongoing run?” by pressing the blue button for the images presented for the first time (distracters) and the red button for the images presented for the second time (targets). The first run of the task evaluates the capacity

![Figure 1. Reality filtering task design.](image)

**Note.** Each block was composed of two runs. The distracters (D) are images that are presented for the first time within a run; the targets (T) are images that are repeated within the same run.
to learn and recognize stimuli based on familiarity; this run activates the medial
temporal lobe (Schnider et al., 2000; Schnider & Ptak, 1999).

In the second run, all items are already familiar. Thus, the ability to sense whether or not the memory of a stimulus emanates from a previous occurrence in the ongoing run (“present reality”) is required. This distinction requires reality filtering (Schnider, 2003, 2013). Clinical and imaging studies with adults have shown that the critical capacity required in the second run is the control of the interference by the distracters (first appearances) in the second run (Nahum et al., 2012; Schnider & Ptak, 1999). Performance of the second run activates the posterior medial orbitofrontal cortex in adults (Schnider et al., 2000).

Stimuli remained on-screen until the subject answered. The interstimulus interval with a fixation cross was 1500 ms. In each run, 30 images were presented for the first time (distracters), and each of them subsequently repeated (targets).

Before the real task, a training block with 20 trials using different images was presented.

Stimulus presentation was controlled with E-Prime 2 (Psychology Software Tools, Pittsburgh, USA). The percentage of correct responses and the mean reaction time for each type of stimulus (distracters and targets from run 1, distracters and targets from run 2) were calculated for the analysis. Recognition performance in the first run (IR) was calculated in a similar way to that in earlier studies (Schnider et al., 1996a, 1996b): Hits1-FP1, where Hits1 describes the number of correctly-recognized items in the first run and FP1 describes the false positive responses in the first run.

Temporal context confusion (TCC), a measure of impaired reality filtering (Schnider et al., 1996a), was calculated as follows: TCC = (FP2/Hits2) – (FP1/Hits1), where FP2 and FP1 are the false positives in run 2 and run 1 (i.e., the incorrect “yes” responses to images that have not yet been presented in the ongoing run), and Hits2 and Hits1 are the correct recognitions of targets in run 2 and run 1 respectively (i.e., the correctly-recognized repetitions). In addition, the $d'$ index on the first run of the task for each participant was calculated. The $d'$ is the standardized difference between the means of the false alarm rates and the hit rates. The formula is as follows: $d' = z(FA) - z(H)$, where $FA$ and $H$ are the false alarm and hit rates, respectively (Swets, Tanner, & Birdsall, 1961).

As the reality filtering task had never been used with children, a paraphrasing technique was adopted to verify that the children understood the task. Specifically, children were asked to repeat the task instruction in their own words. The technique is commonly used for identifying sources of response error due to lack of comprehension of items in the process of developing and adapting survey and assessment tools for children (Willis, 2004).

**Control tasks**

**Working memory**

Working memory was assessed with the subtest of the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV; Wechsler, 2003), that is, forward and backward digit repetition.
Flanker task
To evaluate inhibition capacity, children performed the flanker task (Eriksen & Eriksen, 1974) on computer. In the version adapted for children (Christ, Kester, Bodner, & Miles, 2011), five fish were presented in one row and participants were instructed to pay attention to the orientation (right or left) of the central fish (target). The other fish were oriented either in the same direction (congruent stimulus) or in the opposite direction (incongruent stimulus). Children were asked to indicate the direction of the central target by pressing the left or the right button on the keyboard. The test session was preceded by a training block composed of 20 trials. The main test contained 60 trials. The maximal duration of the trials was 3000 ms. The task was displayed using Inquisit 4.0 (Millisecond Software, Seattle, WA, USA). The percentage of errors and the mean reaction time for congruent and incongruent stimuli were calculated for the analysis.

Orientation questionnaire
Orientation is a continuous measure of reality filtering (Nahum et al., 2012; Schnider et al., 1996a). It was evaluated using a child-adapted questionnaire composed of four questions in each of four domains of orientation: person (e.g., “What is your name?”), situation (e.g., “Where are we now?”), place (e.g., “Which city are we in?”), and time (e.g., “What day is it today?”). The complete list of items is reported in the Appendix.

Strengths and difficulties questionnaire
As a final step, parents were asked to complete the “Strengths and Difficulties Questionnaire” (Goodman, 1997) for their child, a 25-item questionnaire composed of five subscales assessing emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behavior.

Statistical analysis
Reality filtering task
All four age groups performed this task. A $2 \times 2 \times 4$ repeated measures analysis of variance (ANOVA) on accuracy and reaction time was performed with the within-subject factors of Run (1, 2) and Stimulus (Distractor, D, Target, T) and the between-subject factor of Age Group (7-year-olds, 9-year-olds, 11-year-olds, adults). A one-way ANOVA was performed on the $d'$ index, the IR and the TCC score with the between-subject factor of Age Group (7-year-olds, 9-year-olds, 11-year-olds, adults). The Pearson’s correlation was calculated between the measures of TCC and IR for each group. When needed, post hoc Fisher’s tests with a significance level of $p < .05$ were used.

Working memory
Working memory was tested only in the three groups of children. One-way ANOVA with the between-subject factor of Age Group (7-year-olds, 9-year-olds, 11-year-olds) was calculated on the two scores (forward and backward) of the WISC-IV. The Pearson’s correlation was calculated between the working memory scores and performance in reality filtering.
Flanker task
Only the three groups of children performed the flanker task. Two $2 \times 3$ repeated measures ANOVAs with the percentage of errors and the reaction time as dependent measures were performed, with the within-subject factor of Stimulus Type (congruent, incongruent) and the between-subject factor of Age Group (7-year-olds, 9-year-olds, 11-year-olds). The Pearson’s correlation between the performance in the flanker task and reality filtering task was calculated.

Orientation questionnaire
All groups responded to this questionnaire. One-way ANOVAs for each domain of orientation were calculated with the between-subject factor of Age Group (7-year-olds, 9-year-olds, 11-year-olds, adults). The Pearson’s correlation between the orientation scores and the TCC was also calculated.

Strengths and difficulties questionnaire
The parents of the three groups of children filled in the questionnaire. The Pearson’s correlations between the questionnaire scores and reality filtering were calculated.

Results
Comprehension of instructions
All but two children correctly paraphrased and understood the instructions of the reality filtering task. Two participants (one aged 7 and one aged 9) initially failed to paraphrase the instruction correctly but were able to complete the task after the experimenter repeated the instructions.

Reality filtering task
The behavioral results are summarized in Table 1. Both accuracy and reaction times in response to all stimulus types improved across all ages. The $2 \times 2 \times 4$ repeated measures ANOVA on the percentage of correct responses revealed a significant main effect of Age Group, $F(3, 99) = 3.63, p = .015, \eta_p^2 = .09$, and a significant interaction between Run and Stimulus, $F(1, 99) = 12.02, p = .001, \eta_p^2 = .10$. The post hoc tests show that participants were more accurate in their response to the distracters in run 1 compared to the distracters in run 2, $F(1, 99) = 7.55, p = .007, \eta_p^2 = .07$, whereas the difference in accuracy between the Targets in run 1 and the Targets in run 2 falls short of significance, $F(1, 99) = 3.48, p = .065, \eta_p^2 = .03$. The general accuracy, combining all stimuli, is significantly higher for adults compared to children aged 7 years, $t(38) = 2.93, p = .004$, and 9 years, $t(51) = 2.97, p = .009$, but does not differ significantly from that of children aged 11 years, $t(48) = 1.14, p = .165$.

The analysis of the reaction times revealed a main effect of Stimulus, $F(1, 99) = 6.88, p = .010, \eta_p^2 = .06$, with slower responses to distracters compared to targets, and a main effect of Run, $F(1, 99) = 5.49, p = .021, \eta_p^2 = .05$, with faster responses for the first compared to the second run. Furthermore, a main effect of Age Group was found, $F(3, 99) = 32.68, p < .001, \eta_p^2 = .49$. The post hoc tests show that all age groups differ from...
each other: the children aged 7 years were slower than the children aged 9 years, \( t(51) = 2.66, p = .009 \), the children aged 11 years, \( t(48) = 5.00, p < .001 \), and the adults, \( t(51) = 9.32, p < .001 \); the children aged 9 years were slower than the children aged 11 years, \( t(61) = 2.62, p = .008 \), and the adults, \( t(51) = 7.73, p < .001 \); and the children aged 11 years were slower than the adults, \( t(48) = 5.20, p < .001 \).

The TCC score, the IR score and the \( d' \) index are shown in Table 2. The one-way ANOVA revealed significant differences between age groups for the IR, \( F(1, 99) = 3.99, p = .009 \), \( \eta_{p}^2 = .11 \), and the \( d' \) index, \( F(1, 99) = 5.24, p = .002 \), \( \eta_{p}^2 = .13 \), but not for the TCC score, \( F(1, 99) = 1.18, p = .32 \), \( \eta_{p}^2 = .03 \). The IR score is significantly higher for adults compared to children aged 7 years, \( t(38) = 3.17, p = .002 \), and 9 years, \( t(51) = 2.72, p = .008 \), but not compared to children aged 11 years, \( t(48) = 1.43, p = .15 \). The \( d' \) index is higher for adults compared to children aged 7 years, \( t(38) = 3.67, p < .001 \), 9 years, \( t(51) = 3.18, p = .002 \), and 11 years, \( t(48) = 2.07, p = .041 \). The TCC and IR scores significantly correlate in the children aged 7 years, \( r = .616, p < .001 \), and 9 years, \( r = .448, p = .02 \), but not in the children aged 11 years or the adults, \( r < .07, p > .70 \).

### Control tasks

#### Digit span

The one-way ANOVAs revealed a significant effect of Age Group on both the forward, \( F(2, 82) = 7.83, p < .001 \), \( \eta_{p}^2 = .16 \), and backward, \( F(2, 82) = 26.94, p < .001 \), \( \eta_{p}^2 = .40 \), digit repetition scores. For the forward digit repetition, the post hoc tests denote a
significant difference between the children aged 7 and 11 years, \( t(48) = -3.90, p < .001 \), and between the children aged 9 and 11 years, \( t(61) = -2.05, p = .021 \). For the backward digit repetition there is a significant difference between the children aged 7 and 9 years, \( t(51) = -2.94, p < .001 \), the children aged 7 and 11 years, \( t(48) = -8.14, p < .001 \), and the children aged 9 and 11 years, \( t(61) = -4.83, p < .001 \). The correlation between working memory and reality filtering is not significant.

**Flanker task**

The 2 (Stimulus: congruent vs incongruent) \( \times \) 3 (Age Group) repeated measures ANOVA on the children’s percentage of errors shows a significant main effect of Stimulus, \( F(1, 80) = 401.09, p < .001, \eta_p^2 = .83 \), with a higher rate of errors in response to incongruent stimuli than to congruent stimuli. Neither the main effect of Age Group nor the interaction term yielded significant effects, all \( Fs < 2.20 \), all \( ps > .11 \).

The analysis of the reaction times revealed a main effect of Stimulus, \( F(1, 80) = 34.89, p < .001, \eta_p^2 = .30 \), with faster responses to congruent compared to incongruent stimuli, and a main effect of Age Group, \( F(2, 80) = 24.25, p < .001, \eta_p^2 = .37 \), whereas the Stimulus \( \times \) Age Group interaction is not significant, \( F < 1, p = .65 \). The post hoc test revealed that the children aged 7 years were slower than the children aged 9 years in response to congruent, \( t(51) = -3.00, p < .001 \), and incongruent, \( t(51) = 2.65, p = .01 \), stimuli, and slower than the children aged 11 years in response to congruent, \( t(48) = 7.18, p < .001 \), and incongruent, \( t(48) = 6.65, p < .001 \), stimuli. The children aged 9 years were slower than children aged 11 years in response to congruent, \( t(61) = 4.867, p < .001 \), and incongruent, \( t(61) = 4.31, p < .001 \), stimuli.

The Pearson’s correlations between the performance in the flanker task and the performance in response to the stimulus critical for the reality filtering (the distracters in run 2) are not significant.

**Orientation questionnaire**

The scores for the orientation questionnaire are summarized in Table 3. The one-way ANOVA revealed a significant difference between the four age groups for the total orientation score, \( F(3, 99) = 146.83, p < .001, \eta_p^2 = .81 \), and for the subscales of personal, \( F(3, 99) = 90.94, p < .001, \eta_p^2 = .73 \), circumstantial, \( F(3, 99) = 53.31, p < .001, \eta_p^2 = .61 \), spatial, \( F(3, 99) = 52.19, p < .001, \eta_p^2 = .61 \), and temporal, \( F(3, 99) = 52.28, p < .001, \eta_p^2 = .61 \), orientation. Adults’ performance was significantly better than all other groups in all subscales of the questionnaire. The children aged 7 years performed worse compared to the children aged 9 years for the domains of personal, \( t(51) = -4.176, p < .001 \), spatial, \( t(51) = -5.370, p < .001 \), and temporal, \( t(51) = -5.774, p < .001 \), orientation, and

<table>
<thead>
<tr>
<th>Table 3. Orientation questionnaire.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>7-year-olds</td>
</tr>
<tr>
<td>9-year-olds</td>
</tr>
<tr>
<td>11-year-olds</td>
</tr>
<tr>
<td>Adults</td>
</tr>
</tbody>
</table>

*Note:* Statistically significant differences are described in the Results section.
for the total orientation score, $t(51) = -7.574, p < .001$. Compared to the children aged 11 years, the children aged 7 years performed worse in the domain of personal, $t(48) = -4.443, p < .001$, spatial, $t(48) = -5.297, p < .001$, and temporal, $t(48) = -7.674, p < .001$, orientation, and for the total orientation score, $t(48) = -8.525, p < .001$. The children aged 9 years performed worse compared to the children aged 11 years only for the temporal domain, $t(61) = -2.417, p = .019$.

The Pearson’s correlation for each age group between the total orientation score and the TCC score revealed a significant correlation only for 7-year-old children, $r(20) = .58$, $p < .001$. The correlation between the total orientation score and the RI score is significant only for 7-year-old children as well, $r(20) = .54, p = .013$. The correlations between the total orientation score and the performance in the reality filtering task are significant for the Targets in run 1 in 7-year-old children, $r(20) = .59, p = .005$, and for the distracters in run 2 in the 11-year-old children, $r(30) = .46, p = .010$.

**Strengths and difficulties questionnaire**

The within-group Pearson’s correlations between the total parent-reported score of the questionnaire and the measures of the reality filtering task are not significant.

**Discussion**

This study investigated the development of orbitofrontal reality filtering in children. In adults, this memory control mechanism is critical for maintaining thought and behavior in phase with reality. Its failure induces a confusion of reality, evident in acts that are inappropriate for current reality, confabulations, and disorientation (Schnider, 2008, 2013).

The most important result of the present study is that at no age was there a performance drop in the second run of the task. As memory capacity (performance in the first run of the task) increased over the examined ages, so did the ability to control interference in the second run. Even the 7-year-old children were able to suppress the interference of items seen in a previous run but occurring for the first time within the ongoing run (the distracters in run 2). In other words, reality filtering is already functional at age 7, in proportion to the ability to store information. As item memory improves from 7 to 11 years of age, and then into adulthood, reality filtering also becomes more rapid and accurate. Even at the age of 11 years, however, performance is not yet at an adult level.

The reaction times significantly differed between all age groups and with regard to all stimulus types. In part, this observation can be explained by a maturation of cognitive functions in general, as reflected by the fact that reaction times in children steadily improve from 3 to 4 years of age through to young adulthood (Elliott, 1970), especially between the ages of 9 and 12 years (Williams et al., 2015). However, item memory (IR score) and reality filtering (TCC score) dissociated in the 11-year-old and adult groups. This corresponds to previous observations in healthy adults and amnesic patients (Schnider, 2008; Schnider et al., 1996a). This observation suggests that reality filtering develops in parallel with item memory until the age of 9 years, but then continues to develop as an independent memory function. Even if we cannot be sure that reality filtering in children depends on the same brain regions as in adults (the orbitofrontal
cortex), a possible explanation of this dissociation could lie in the later development of the brain structures critical for reality filtering relative to those critical for recognition memory. While the ability to store and recognize new information depends on the medial temporal lobes (Eichenbaum, Yonelinas, & Ranganath, 2007; Squire, Wixted, & Clark, 2007), reality filtering in adults critically depends on the posterior medial orbitofrontal cortex area 13 (Schnider, 2003; Schnider et al., 1996a, 2000; Schnider & Ptak, 1999). The temporal lobe matures until the second decade of life (Buchsbaum et al., 1992), while the prefrontal cortex is a region that reaches complete maturation later; its development continues into adulthood (Casey, Giedd, & Thomas, 2000; Casey, Tottenham, Liston, & Durston, 2005). The main difference between this region and the rest of the brain is the continuation of the pruning of synapses into young adulthood (Bourgeois, Goldman-Rakic, & Rakic, 1994), a process accompanied by the improvement of cognitive capacities like memory, inhibition and attention (Casey et al., 2000). It is also possible that at the ages of 7 and 9 years, structures other than the orbitofrontal cortex assume the task of reality filtering and that the dissociation between item memory and reality filtering from the age of 11 years onwards indicates that the orbitofrontal cortex has matured enough to assume this function. This interpretation, too, is compatible with the known difference in the speed of maturation of the medial temporal lobe and the prefrontal cortex.

In adults, reality filtering also does not correlate with inhibition capacity (Nahum et al., 2009) or working memory. The present study confirms this finding. While inhibition is one of the last high-order functions to develop, the ability to reject memories that do not pertain to ongoing reality (reality filtering) and to master a task requiring this form of “inhibition” is already significantly present at the age of 7 years.

In adult amnesic subjects, orientation about time, place and circumstances is a continuous measure of orbitofrontal reality filtering and highly correlates with its measure (healthy adults typically have perfect orientation; Nahum et al., 2009, 2012; Schnider et al., 1996a). In the present study, we found a trend towards improved orientation from the ages of 7 to 11 years, in particular in the temporal domain. This observation replicates previous research showing that information concerning the day, month and year starts to be correctly reported only from the ages of 8 to 9 years onwards, while the other domains of orientation are acquired earlier (Hotz, Plante, Helm-Estabrooks, & Nelson, 2014). However, the score for reality filtering (TCC score) correlates with orientation only in 7-year-old children. Overall, orientation as an abstract concept evolves between the ages of 7 and 11 years, but reality filtering is not more of a key determinant of its development than other memory and cognitive functions.

In conclusion, this study shows that orbitofrontal reality filtering, as tested with repeated runs of a continuous recognition task, is already functional at the age of 7 years and then improves its efficacy as item memory evolves. The study supports the claim (Schnider, 2013) that increasing storage capacity in memory requires the parallel development of an automatic, rapid mechanism that allows one to sense whether or not a thought relates to present reality, and thus, whether or not it can be acted upon. The results of this study may be part of the explanation for why at
the age of 7 years, the tendency to act in a fantasy world has already largely subsided.

Acknowledgements

We thank Claudia Ugolini and Dr Claudia Lietti for their help with data collection and data analysis.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Swiss National Science Foundation [grant no. 32003B-155947].

References


Schnider, A., Valenza, N., Morand, S., & Michel, C. M. (2002). Early cortical distinction between memories that pertain to ongoing reality and memories that don’t. *Cerebral Cortex*, 12(1), 54–61. doi:10.1093/cercor/12.1.54


**Appendix**

**Orientation questionnaire**

**Personal domain:**
- What is your name?
- How old are you?
- Which class are you attending?
- Where were you born?

**Spatial domain:**
- Which city are we in?
- What is the name of this school?
- Which floor are we on?
- Which region of Italy are we in?

**Situational domain:**
- Where are we now?
- What do you do here during the day?
- What are we doing at this moment?
- What are the names of your teachers?

**Temporal domain:**
- What day is it today?
- What month is it?
- What year is it?
- What time is it?