

**Visual Perspective Taking as investigated by fast
periodic visual stimulation (FPVS)**

Visual Perspective Taking as investigated by fast periodic visual stimulation (FPVS)

Alexy-Assaf Beck

Thèse présentée en vue de l'obtention du grade de Docteur
en sciences psychologiques et de l'éducation

Promoteurs

Dana Samson (UCL)
Bruno Rossion (UCL)

Comité d'accompagnement

Martin Edwards (UCL)
Philippe Lefèvre (UCL)

Président

Vassilis Saroglou (UCL)

Comité d'accompagnement et jury

Martin Edwards (UCL)

Philippe Lefèvre (UCL)

Heather Ferguson (U of Kent, UK)

Olivier Collignon (UCL)

Abstract

Visual perspective taking (VPT) corresponds to our ability to see the world from another person's perspective. We investigated VPT using fast periodic visual stimulation (FPVS) approach while measuring electroencephalography (EEG). This thesis had two main objectives, the first was to extend the FPVS approach to study high-level socio-cognitive functions in the human brain. The second objective was to investigate the cognitive basis of mind-reading, and discriminate whether implicit VPT involves perspective taking mechanisms (mentalizing hypothesis) or only spatial cuing mechanisms (submentalizing hypothesis). Our findings lead to support another theoretical framework, the continuum scale theory which consider that submentalizing and implicit mentalizing are not dichotomous processes but reflect a continuum of processing of increased depth. Overall, this thesis provides a promising new approach for exploring the neural basis in human social cognition and their inter-individual difference.

Acknowledgments

First of all, I would like to thank my supervisors Dana Samson & Bruno Rossion for giving me the opportunity to conduct my doctoral research on a theme which confirm my deep interest in neuroscience and psychology. I am very grateful for all the guidance, support, and encouragement. I would especially like to give warm thanks to Dana for her support in helping me to improve the quality of this thesis.

I am grateful to the members of my jury, Drs Martin Edwards, Philippe Lefèvre, Olivier Collignon, Heather Ferguson for their interest in my work. I thank them all for their comments and useful remarks which have undoubtedly helped to make this thesis better. Thank you as well to Dr. Vassilis Saroglou for agreeing to preside the defense of my thesis.

I also owe many thanks to my colleagues and friends, whose support, feedback and encouragement have been indispensable, but whose names there is not space to fully enumerate here. The contributions of all of them are sincerely appreciated and gratefully acknowledged. In particular, I would like to thank Jacques Jonas for supporting me in our office (closing the window but why?), my colleague Talia Retter, Joan Liu, Genevieve Quek, Corentin Jacques for their comments on my papers but also Xiaoqing Gao, Sabine Meurrens, Gaëlle Meert, Laëticia Leloup and many others for their positivity and kindness.

Thanks to the administrative staff of the faculty, Isabelle Legrain, Yolande Nivaille, Brigitte Champagne and Martine Janssens for their help.

I am grateful to my partner Marion Maynau and also to Elsa Mandement for their help in the design of the thesis cover (not to mention my mom). Further thanks to my brother Romain and Natalie Wall for proof reading.

Last but not the least, I would like to thank my family: my parents, my grandmother, my brothers for providing constant support and encouragement, especially my mom without whom this thesis would never have been realized. Finally, I would like to express my profound gratitude to my partner Marion Maynau for her practical and emotional support, she has always been there for me throughout the good and the bad.

Table of contents

Abstract	5
Aknowledgments	7
Abbreviations	13
Chapter 1: General Introduction	15
I. Theoretical context	17
1. Theory of mind	17
1.1 ToM development	17
1.2 How does ToM work?	21
1.3 Neural networks underlying ToM	25
2. Visual perspective taking	31
2.1 Is VPT a spontaneous process?	35
2.2 Review of the VPT investigations which oppose the mentalizing and the submentalizing hypotheses	37
2.3 The schema theory	47
II. Methodological framework	49
1. The Electroencephalography	49
1.1 Principles of the EEG	49
1.2 Basic principles of signal processing	51
2. Fast periodic visual stimulation	52
2.1 Basic principles/considerations	52
2.2 Overview of the FPVS approaches	55
2.3 Stimulus parameters	55
2.4 FPVS applications	57
III. Overview of the thesis	63

Chapter 2: Empirical Study 1	65
An objective neural signature of rapid perspective taking	67
Abstract	67
1. Introduction	68
2. Materials and methods	70
3. Results	76
3.1 Behavioral Data	76
3.2 EEG Data	76
4. Discussion	81
5. Acknowledgements	82
6. Supplementary material	83
Chapter 3: Empirical Study 2	89
Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow	91
Abstract	91
1. Introduction	92
2. Materials and methods	95
3. Results	100
3.1 Behavioral Data	100
3.2 EEG Data	101
4. Discussion	105
5. Acknowledgements	108
6. Supplementary material	108
Chapter 4: Empirical Study 3	115
Neural intermodulation responses when viewing an object in front of a person are affected by the presence of an obstacle in the person's line of sight	117
Abstract	117
1. Introduction	118

2.	Materials and methods	121
3.	Results	127
3.1	Behavioral Data	127
3.2	EEG Data	127
4.	Discussion	137
5.	Acknowledgements	141
6.	Supplementary material	141
Chapter 5: Discussion		145
1.	Reminder of the goals of the thesis and the hypotheses	147
2.	Summary of findings under the lenses of the mentalizing and the submentalizing hypotheses	148
3.	The alternative interpretations	150
4.	The continuum scale model	153
5.	Strengths and limitations of the FPVS approach	158
6.	Future studies	160
7.	Conclusion	163
References		165
Appendices		181
Appendix 1: Opportunities and limitations of fast periodic visual stimulation for studying covert social attention		183
	Abstract	183
I.	Study 1: Explicit perspective taking instructions	184
1.	Methods	184
2.	Results	190
3.	Discussion of Experiment 1	192
II.	Study 2: Explicit spatial perspective taking instruction	192
1.	Methods	193
2.	Results	195
3.	Discussion of Experiment 2	198

Appendix 2: Replication of the Oddball paradigm (Study 1 of Chapter 2)	
without VPT instruction	201
Abstract	201
1. Materials and methods	202
2. Results	203
2.1 Behavioral Data	203
2.2 EEG analysis	203
3. Discussion	209
Appendix 3: Pre-test studies	211
1. Attentional paradigm	211
2. Oddball paradigm	211
3. Distraction paradigm	212
4. Intermodulation paradigm	213
5. Discussion	214

Abbreviations

ASD: Autism Spectrum Disorder
DC: Direct Current
EEG: Electroencephalogram
ERP: Event-Related Potential
FB: False Belief
FFT: Fast Fourier Transform
fMRI: Functional Magnetic Resonance imaging
FPVS: Fast Periodic Visual Stimulation
Hz: Hertz
IM: Intermodulation
IRI: Interpersonal Reactivity Index
mPFC: medial Prefrontal Cortex
PET: Positron Emission Tomography
ROI: Region-of-Interest
RT: Reaction Time
Sbl: baseline-corrected amplitude
SNR: Signal to Noise Ratio
SP: Selection Process
SSVEP: Steady-State Visually Evoked Potential
STS: Superior Temporal Sulcus
TB: True Belief
TMS: Transcranial Magnetic Stimulation
ToBy: Theory of Body mechanism
ToM: Theory of Mind
ToMM: Theory of Mind Mechanisms
TP: Temporal Poles
TPJ: Temporo Parietal Junction
VPT: Visual Perspective Taking

Chapter 1

General Introduction

I. Theoretical context

1. Theory of mind

Whereas we have no direct access to someone else's mind, we are able to attribute mental states to others (such as knowledge, desires, beliefs, thoughts). People engaged in social life construct many thoughts about others' inner world. Indeed, in everyday life, we are led to interpret other people's actions, intentions or beliefs to understand the social world around us and adapt our behavior accordingly. Humans can be considered as social animals with outstanding social cognitive abilities which allow them to make inferences about others' inner worlds based on unseen mental states. In 1978, Premack & Woodruff were the first to refer to these abilities as « Theory of Mind » (ToM) and to open a new area of research. Since, these abilities have never ceased to interest researchers (Frith & Frith, 2007; Samson, 2013; Tomasello & Call, 1997) and received other names such as « perspective taking » (PT; Hodges, 2015), mentalizing (Frith & Frith, 2006) or mind reading (Apperly, 2011; see Whiten, 1994 for a more complete list of names).

1.1 ToM development

ToM abilities are fundamental for humans. This is why researchers are interested in its development: how and when do humans develop ToM? Previous studies highlighted that human infants appear to start life with limited ToM abilities. For example, to investigate ToM development in children, Wimmer & Perner (1983) have developed the False-Belief task, which evaluates the ability to attribute beliefs to others, beliefs that can be divergent from our own. The false-belief task remains the most used task to investigate ToM, and comparatively few studies have examined other ToM abilities. Today, numerous versions of the false-belief task have been proposed. They consist of brief stories involving two characters, for example in its most common version, Sally and Anne (Baron-Cohen, Leslie, Frith, 1985; Figure 1.1). Sally and Anne are in the same room. Sally puts her marble in a basket. Sally then leaves the room and while she is away Anne takes the marble from Sally's basket, and puts it in a different location, in a box. Finally, Sally returns and the child is then asked where Sally will look first for her marble. The children know where the marble is, but

Sally has not seen her marble being moved. The child has a true belief (TB) regarding the location of the marble (he or she knows where the marble actually is) whereas, Sally has a false belief (FB), because for her, the marble is where she left it, i.e. in her basket. Thus, in order to pass the task, children must be able to understand that another's mental representation of the situation is different from their own, and they must be able to predict the other person's behavior based on that understanding.

There is evidence that infants already have some ToM abilities, for example they seem to represent another person's visual perspective (e.g., Sodian, Thoermer, Metz, 2007) or follow the gaze of other agents (Tomasello, 1999; Bräuer et al., 2005; Tremblay-Leveau, 1999). In addition, it seems that infant can recognize knowledge (Baron-Cohen et al., 1985), intention (Astington & Gopnik, 1991; Meltzoff, 1995), emotions (Harris, 1989; Avis & Harris, 1991) or can attribute goals and intentionality (Csibra, 2008; Senju & Csibra, 2008) and visual experiences (Kampis, Parise, Csibra, & Kovács, 2015). However, interestingly, before 4 years of age, children do not pass false-belief tasks (at least the one described above). Children do not seem to understand that others can have different knowledge of the surrounding world from the one they have. For them, Sally will look for her marble in Anne's box, i.e., where they know the marble currently is. Indeed, why would one seek the object that one wishes to find where it is not? For some authors (e.g., Perner, 1991; Perner & Ruffman, 1995; for a meta-analysis see Wellman, Cross, Watson, 2001), to have a ToM involves being able to understand others' beliefs, which is not the case with young children, under 4 years of age. For these authors, children are not born with innate ToM abilities; it is only after 4 years old that a radical shift in children's thought processes would happen which allows them to understand that another person can have a different mental state than their own. Thereby, it has been argued that ToM refers to complex abilities which emerge relatively late in development. However, different studies have found results which support the idea that children are endowed well before the age of 4 with more sophisticated ToM abilities.

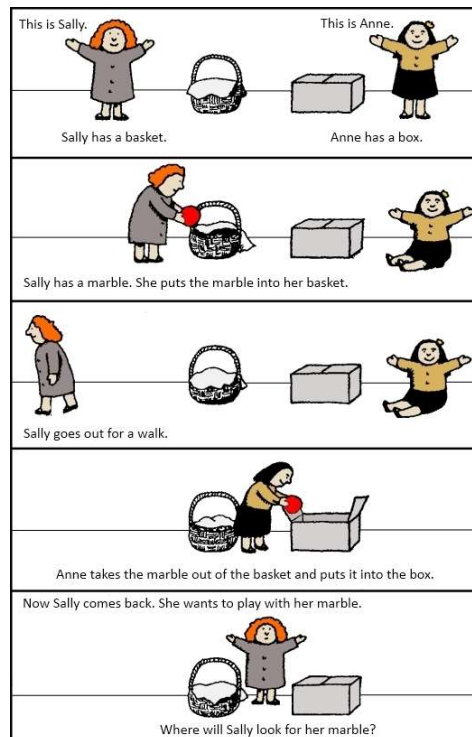


Figure 1.1. Illustration of the Sally–Anne task to test children’s ability to infer false beliefs.

Indeed, based on a simplified version of a ToM task (e.g., a non-verbal task coupled with eye tracking to measure a nonverbal response), researchers have found some evidence for an early development of certain aspects of belief reasoning. Onishi & Baillargeon (2005) modified the classic Sally–Anne task and measured gaze duration in 15-months-old infants with no linguistic capacities. In the first part of the experiment, infants were familiarized with an actor who placed an object into one of 2 boxes (Figure 1.2A). After placing the object into the box, different scenarios were possible. The object moved by itself to the other box either after the actor left the scene (and thus the actor knows the object new location, i.e. he has a true belief, TB) or the actor left the scene before the object moved (thereby, he does not know the accurate object location because he did not see that the object changed location and he has a false belief, FB). Then the actor comes back in both scenarios. In one FB scenario, the actor looks for the object in the box where he left it (Figure 1.2C), hence, his behavior is consistent with his FB (consistent condition). In another FB scenario, the actor looks where the object really is, and so his behavior is inconsistent with his FB

(inconsistent condition). Likewise, in the TB scenarios, the actor could look in the box either consistently or inconsistently with his belief. The results showed different patterns of gaze duration between the consistent and the inconsistent conditions. In fact, infants were surprised and looked longer when they observed the inconsistent condition than the consistent condition i.e., when the actor looked in the box where he is not supposed to. With a similar experiment Kovács et al. (2010) tested 7-month-old infants and highlighted that even at such an early age, infants also seem capable of processing others' belief. These studies support the idea that young infants have some understanding about what other people experience and have expectation about other's behaviors. However, as explained later such interpretation is highly controversial.

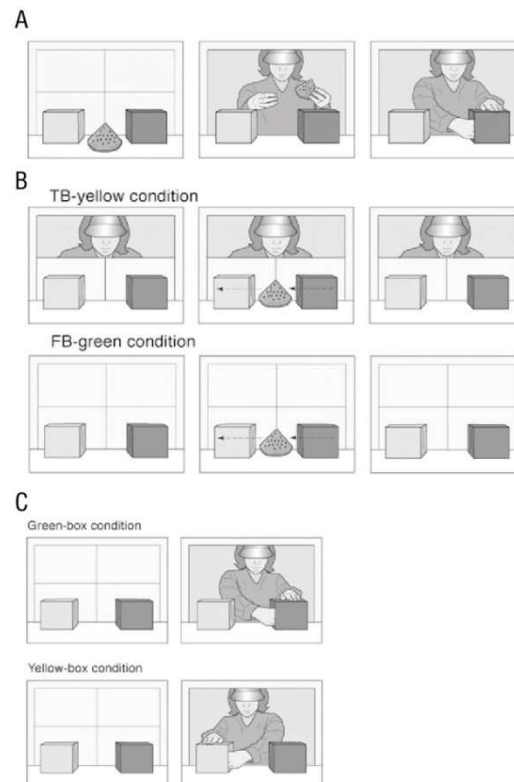


Figure 1.2. The procedure of Onishi and Baillargeon's (2005) (A) Familiarization (B) Belief induction (C) test trial

With age, children start to successfully infer more complex mental states which are not a direct reflection of the external world. Hence, several authors suggested that the abilities to understand others' mental states would be linked to the developmental paths of the underlying neural structures of the brain (ontogenetic development; e.g., Singer, 2006; Brüne & Brüne-Cohrs, 2006; Grosse Wiesmann et al., 2017). Moreover, Premack & Woodruff, have shown that other animals (obviously without linguistic capacities) such as chimpanzees share some aspects of ToM abilities with humans. For example, Call and Tomasello (2008) suggested that chimpanzees can at least understand goal, intention, and others' perception and this may emphasize a phylogenetic ToM emergence.

1.2 How does ToM work?

The fact that infants can pass simplified versions of ToM tasks supports an innate neural mechanism for ToM which could be the precursors to the "mature" ToM (Gopnik, Slaughter, & Meltzoff, 1994). However, there is an ongoing controversy about whether ToM development is driven by an innate sophisticated mechanism specifically involved in mental state reasoning and which would need to mature during childhood and puberty. According to Leslie (1984), the world can be naturally decomposed into three classes of stimuli and each one is processed by different cognitive components. A first component called "ToBy" ("Theory of Body Mechanism") allows the child to be aware of the mechanical properties of an agent or an object and to consider that an agent is self-ruling in his movements. The two other components are "ToM mechanisms" (ToMM) that focus on the "intentional" properties of agents. The ToMM1 ("Theory of Mind Mechanism system 1") would be responsible of the processing of the agent's goals. Whereas, the ToMM2 ("Theory of Mind Mechanism system 2") would be responsible of representing others as possessing mental states. During development, these three modules are distinctly and hierarchically organized: ToBy processes very basic information which are refined by ToMM1 and finally used by ToMM2 to infer mental states. However, at the infant ToM development stage, the infant is attracted to social stimuli and thus infers different mental states based on the most salient information which is assumed to be his own mental state. This would explain that children are less successful than adults in understanding others' mental states and cannot

pass a certain level of false-belief task because they cannot select someone else's perspective, and thus, project their own knowledge or belief to the other person.

Why are children unable to successfully perform false-belief tasks until they are around 4 years old? If we consider for example, a normal Sally and Anne task, the children deal with two conflicting beliefs: Sally's belief as she thinks that her marble is where she left it (FB), and children's own belief, as they know where Anne moved the marble (TB). Therefore, in order to predict where Sally is going to search for her marble, they have to select the relevant belief (i.e., Sally's belief) and inhibit the irrelevant belief (i.e., their own belief). Hence, the model proposes an additional mechanism, another ToMM's module linked with executive functions (i.e. a set of cognitive processes which refer to a family of top-down mental processes that are necessary for the cognitive control of behavior), called the Selection process (SP) which allows at a later stage (when their own belief and the other's belief is constructed) to inhibit the irrelevant belief (e.g., the child's belief) in order to adopt the relevant belief (e.g., Sally's belief).

In fact, inhibitory abilities are indeed important to non-egocentrically ascribe mental states to others. For example, a longitudinal study (Carlson, Mandell, & Williams, 2004) has showed that children's inhibitory abilities influence their later performances on ToM tasks. Furthermore, Samson et al. (2005)'s study emphasized the importance of inhibitory abilities in mature ToM. The authors observed in an adult patient with an impaired inhibitory control (caused by brain damage) a specific deficit in resisting interference from one's own perspective. Thereby, self-perspective inhibition seems to play a crucial role at all ages, when attributing mental states to others (Apperly, 2012; German & Hehman, 2006; Carlson, Moses, Claxton, 2004). According to the ToMM model, children before 4 years old possess the ToMM but not yet the inhibitory SP. Thus they miss social experience and the skill to use this experience, and this is why they fail standard false-belief tasks (see Leslie & Thaiss, 1992; Leslie & Polizzi, 1998).

At a more general level imputing mental states to others seems effortful and demanding on cognitive resources which are not involved in the core executive function skills (e.g., working memory, inhibition, planning). For example, Gagne & Coppola (2017), have found that deaf children show delays in false belief understanding, which supports a

strong link between language and ToM. Therefore, other brain functions (e.g., language) which at first sight do not seem directly related to ToM, play a valuable role in some aspects of it (e.g., to impute false belief to others). However, there is evidence that some ToM abilities are independent of language, as we will discuss below, infant and non-human species (both with limited language abilities) can nevertheless pass some ToM tasks.

The apparent discrepancy between a large body of evidence showing that children do not pass false-belief tasks before the age of 4 and accumulating evidence showing that infants as young as 7-months are sensitive to false beliefs led some authors to propose a “Two-system” model of ToM (Apperly, 2011; Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). This model posits the existence of two distinct processes, one would be a “low-level” and implicit mechanism, and another one would be a “high-level” and explicit mechanism. The implicit mechanism would be an early-developing system available in infancy and shared by humans and non-human animals. It allows us to track others’ mental states in a relatively spontaneous manner which is not flexible albeit, fast and efficient. The second system on the other hand would be more cognitively effortful and link to executive function. This explicit mechanism which may be specific to the human species would develop later in childhood (Grosse Wiesmann et al., 2016; Wellman et al., 2001) and remains effortful even for adults. According to this model, the implicit mechanism of ToM guides social interactions, whereas the higher mechanism allows us to reason about others’ mental states. This dichotomous distinction of ToM processing between implicit and explicit processes was supported by several empirical studies. For example, in the study by Schneider et al. (2013), the authors evaluated these two aspects of ToM in one group of people with symptoms of autism spectrum disorder (ASD) known to have impairment in social cognition and another group of control participants. In one task, participants had to read stories and answer questions about the mental states of the people from the stories. Thus, this task evaluated explicit ToM abilities, and results showed similar performances between the two tested groups. However, in the second task, the authors evaluated implicit ToM abilities. To do so, participants watched movies of TB, FB and control movies while eye movements were recorded. The results showed, that control participants looked at regions of interest that suggest that they implicitly tracked the mental states of the person in the movie, whereas such looking patterns were not observed in the ASD group. Such evidence

has been taken as support for the idea that the low-level and implicit ToM system is specifically impaired in cases of autism and that the two systems rely on different cognitive processes.

However, the idea of an implicit mentalizing system has generated considerable debate. In particular, the reliability of the various experimental paradigms used to demonstrate its existence has been questioned. For Heyes (2014), results showing an implicit mentalizing in infants can be explained otherwise. This author developed the *submentalizing* hypothesis, which proposes that the evidence of an implicit ToM system actually reflects the activation of domain general cognitive processes such as memory and attention (shared with others species) which do not process others' mental states per se. Such general mechanisms would process quickly and efficiently basic aspects of social input and simulate ToM or mentalizing in social contexts. Several studies which have shown implicit mentalizing could indeed be explained by Heyes's alternative explanation. For example, the implicit mentalizing interpretation suggests that, in the false-belief tasks used by Onishi & Baillargeon (2005) and Kovács et al. (2010), participants looked sooner and for longer at the location where the agent had last seen the target because they anticipated his behavior and consider that the agent was going to look in that location (due to his false belief about the target's location). The submentalizing interpretation of these results on the other hand, considers that participants did not attribute mental states to the agent and argue that the results (e.g., longer looking time when the agent has a FB and nevertheless seeks at the accurate location) are in fact due to the response to the novelty of low-level features¹ and retroactive interference whereby newly salient information (e.g., reappearance of the agent) inhibits the ability to recall previously acquired information (e.g., the new location of the target) and therefore the new information remains perceptually novel. Hence, domain-general processes would explain infant false-belief findings. However, according to Scott et al. (2014) perceptual novelty fails to explain all the

¹ e.g., during familiarization trial the agent always moved toward the green box whereas during the test the agent moved toward the yellow box which may have by itself created the "novelty/surprise" and increased gaze duration.

demonstrated evidence accumulated over the past 20 years that infants have an implicit ToM.

Thus, so far, it is still unclear whether there is actually an implicit ToM mechanism or whether there is only a more general mechanism which does not involve mentalizing but simulates the effect of mentalizing in social contexts. We will get back to this issue later in this chapter.

1.3 Neural networks underlying ToM

ToM is a complex construct which merges several skills needed to understand others. Neuroimaging studies of contrasting brain responses while participants perform ToM and non-ToM tasks highlighted that ToM involves a large network of heterogeneous brain regions (supporting the view that the ToM is not a unitary cognitive ability). Nevertheless, some specific regions seem to play a preponderant role when attributing mental states to others (Frith & Frith, 2003; Saxe, 2006; Van Overwalle, 2009).

Several functional magnetic resonance imaging (fMRI) or positron-emission tomography (PET) scan experiments have been designed to focus participant's attention on or away from thinking about mental states. For example, in Fletcher et al. (1995) study (PET experiment), participants read different kinds of stories and were asked to silently answer action-explanation questions. However, they were previously informed if the coming story was a "mental" story which involved thinking about thoughts and feelings in order to answer the question or a "physical" story which was used as a control story and did not involve thinking about thoughts and feelings to answer the question. The neural activity was increased during the "mental" stories in the medial prefrontal cortex (mPFC), the temporo-parietal junctions (TPJ), the temporal poles (TP) and the precuneus/posterior cingulate cortex (Figure 1.3). Furthermore, other studies highlighted similar brain activation while using different stimuli or methods. Gallagher et al. (2000), for example, found the same group of brain regions involved for verbal and non-verbal ToM stimuli (cartoon; see also Kobayashi, Glover, Temple, 2007; Sommer et al., 2007).

In fact, several studies investigated different aspects of ToM (e.g., beliefs, knowledge, desire) across a wide range different methods (e.g., PET, fMRI, transcranial magnetic

stimulation (TMS) using verbal or non-verbal stimuli) and revealed similar brain activation (even in congenitally blind adults; Bedny et al., 2009). A large number of reviews of these studies (e.g., Saxe, 2006; Frith & Frith, 2007) as well as meta-analyses (e.g., Arora, Schurz, Perner, 2017; Schurz et al., 2014; Mar, 2011; Frith & Frith, 2003; Gallagher & Frith, 2003; Van Overwalle, 2009) exist which show the deep interest of researchers in understanding the neural basis of ToM. Actually, there is large heterogeneity across the neuroimaging studies, related to the behavioral manipulations used (e.g., explicit or implicit ToM tasks) but also for imaging analysis (Schaafsma, Pfaff, Spunt, Adolphs, 2015) and fMRI resolutions. However, the literature highlighted a similar network of brain areas associated with ToM reasoning which is now commonly called the “ToM network” (Frith & Frith, 2006) or the “social brain” or the “mindreading network” (Figure 1.3B). However, note that there are various forms of ToM we can engage in according to the type of mental states we infer (e.g., beliefs, motivations, intentions, desires, visual experiences). Indeed, different patterns of activation within the “ToM network” have been found according to the type of mental state processed (e.g., false belief, intention, visual experience; Carrington & Bailey, 2009; Schurz et al., 2013, 2014). For example, the posterior superior temporal sulcus (STS) has been proposed to be involved while processing information about a person’s action (Frith & Frith, 1999). This possible specificity of processing for each type of mental states cannot be ignored and should be kept in mind.

The main regions generally included in the “ToM network” are the mPFC, the left and right TPJ, the TP and the posterior cingulate cortex (for a review, see Carrington & Bailey, 2009; Frith & Frith, 2003; Schurz et al., 2014; Van Overwalle, 2009), but there are also other regions such as the STS or the amygdala which are often found to be activated during ToM task. If the role of each of these regions is relatively blurred (possibly due to the subprocesses involved), there are nevertheless several hypotheses regarding their respective role.

It has been proposed that the mPFC is involved in the integration of enduring information such as personality traits (Van Overwalle, 2009) or to process social and emotional relevant information about others (Saxe & Powell, 2006) or even involved in distinguishing mental state representations from physical state representations (Frith &

Frith, 2003; Döhnelt et al., 2012). However, an alternative mPFC function has been proposed by Schuwerk et al. (2014a). They developed an fMRI study in which participants performed a belief reasoning task (based on a variation of the Sally and Anne paradigm) and highlighted that the mPFC and bilateral TPJ were connected to each other and involved in the representation of diverging mental states. TPJ would process relevant environmental information whereas the mPFC in case of FB would inhibit the TPJ in order to make the distinction between our own perspective and the other's different perspective. This mPFC function was also supported by a TMS study (Schuwerk, et al., 2014b) which found that the mPFC plays a role in establishing perspective differences (i.e., distinguishing between the other's and one's own perspective)².

Activation of the right inferior frontal gyrus was observed when there was a conflict between participants' own and someone else's perspective (Abraham et al., 2010; Van der Meer et al., 2011). For example, in a brain lesion study Samson et al. (2005) highlighted in patient with right fronto-temporal damage, a selective deficit in inhibiting self-perspective during false-belief tasks. In fact, the patient's pattern of results was similar to those of children under 4 years old, with a strong egocentric bias when judging Sally's belief. However, he was quite capable of inferring the mental states of others when the demands of inhibition from his own perspective were reduced (e.g., when he did not know where the object actually was). This evidence supports the models proposed earlier with the key notion of one system involved in the calculation of the perspectives and another system involved in perspective selection (see also Samson, Houthuys, Humphreys, 2015). It has therefore been suggested that the inhibition of one's own perspective could be underpinned, at least in part, by the right inferior frontal gyrus (Samson et al., 2005, 2015; Le Bouc et al., 2012; see also Schurz, et al., 2014; Vogeley et al., 2001; Amodio & Frith, 2006; Bahnemann et al., 2010; Mar, 2011).

² Authors applied TMS on the mPFC while participants performed a false belief task. The RTs showed a decrease in the discrepancy between trials in which participants considered another's or their own belief, irrespective of whether beliefs were consistent or inconsistent.

It has been proposed that the TPJ plays an important role during conscious attribution of mental states (e.g., Frith & Frith, 1999; Van Overwalle, 2009 but see Schneider et al., 2014), for example when participants are explicitly asked to take another person's perspective (e.g., participants describe a scene from another viewer's perspective; Saxe & Kanwisher, 2003; Saxe, 2006; Zacks et al., 2003; Aichhorn et al., 2006; but see also Vogeley et al., 2004 who did not find such activation). In addition, it has been proposed that the right and left TPJ play different roles. Unfortunately, the literature is inconsistent regarding the different involvement of the left and right TPJ during ToM tasks. More specifically, several studies suggested a critical role of the right TPJ compared to the left TPJ in mental state attribution (Saxe, 2006; Aichhorn et al., 2008). Indeed, left TPJ activation, contrarily to the right TPJ, has been also observed for non-social representation of the world (Aichhorn et al., 2008). However, recent investigations (e.g., Biervoye et al., 2016; Le Bouc et al., 2012; Samson et al., 2004, 2007) on brain-damaged patients have showed that left TPJ is of primary importance to spontaneously process the perspective of the other and there is growing evidence that the left TPJ is also necessary for representing others' beliefs (Samson et al., 2004) and thinking about perspectives (e.g., processing alternative perspectives Schurz et al., 2013). Thus, it seems that the left TPJ and the right TPJ are both of primary importance for efficient ToM processing.

In addition, as already mentioned, several studies have found other brain regions involved while mentalizing as the TP or the STS. More specifically, the role of the TP is still debated in the literature. It has been proposed that it is involved in the retrieval of social knowledge (Frith & Frith, 2003). However, for example Michel et al. (2013)'s study on brain damage patient with massive atrophy of the left TP observed spared performance on tasks which investigated different patient's ToM abilities (intention, knowledge, and beliefs). Furthermore, the STS was also proposed as another important brain region for ToM reasoning (Castelli et al., 2000; Fletcher et al., 1995; Gallagher et al., 2000). The STS has been found to be involved in detection of agency and implicit processing of biological movements that signal intentions (e.g., watching eye movements, facial expressions; Allison, Puce, McCarthy, 2000). In addition, there are other brain regions which have been found to be involved for ToM reasoning as the cingulate cortex (Northoff & Bermpohl, 2004;

David et al., 2006), the precuneus as well as the right insula (Gobbini, Koralek, Bryan, Montgomery, Haxby, 2007; Lissek et al., 2008; Sebastian et al., 2012).

Only few investigations examined brain region distinctions according to the implicit (i.e., participants are not asked to take another person's perspective) or explicit (i.e., participants have to actively and hence consciously take another person perspective) aspect of the imputed mental states. Schneider et al. (2014) in an fMRI study, showed Sally and Anne movies to participants and found the same region activated for both the implicit and the explicit conditions (see also Young & Saxe, 2009; Bardi, Desmet, Nijhof, Wiersema, Brass, 2016). Furthermore, in the explicit condition, the ToM network showed larger responses for FB than TB (Aichhorn et al., 2008; Sommer et al., 2007). During the implicit condition, this modulation of the brain activation according to the belief of the agent was also observed in different sub-regions of the ToM network (the left STS and the precuneus). Hence, because they found activity in the ToM network in the implicit belief condition and in addition highlighted sub-regions of the ToM network which were sensitive to whether the agent has a TB or a FB (see also Kovács et al., 2014), their findings support the theoretical model of Apperly & Butterfill, (2009) with an existing implicit system which spontaneously tracks others' mental states. Furthermore, Biervoye et al. (2016)'s study on brain damage patients highlighted different brain region involvement as a function of the explicit or implicit nature of the mentalizing instructions. The authors' observations were in accordance with the existence of an implicit system. Indeed, they highlighted the left (posterior) TPJ as a key region to spontaneously process other beliefs. Indeed, these patients had common lesions in the left (posterior) TPJ and both showed the same ToM deficit i.e., they were only able to take into account someone else's belief when they were explicitly asked to do so.

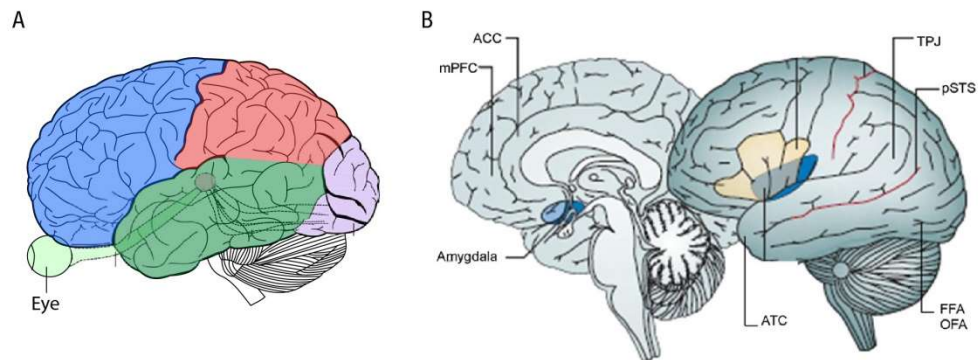


Figure 1.3. Human brain. (A) The blue region corresponds to the frontal lobe, the red to the parietal lobe, the purple to the occipital lobe, and the green to the temporal lobe. (B) The ToM network (taken from Heyes et al., 2010 and adapted from Blakemore, 2008). Regions shown: medial prefrontal cortex (MPFC), anterior cingulate cortex (ACC), temporo-parietal junction (TPJ), posterior superior temporal sulcus (pSTS), fusiform face area (FFA), occipital face area (OFA), anterior temporal cortex (ATC) and amygdala.

Regions of the “ToM network” can also be involved in cognitive functions that are not social per se. This underlines the complexity of the nature of the multiple neurocognitive mechanisms involved while mentalizing (e.g., general perceptual and cognitive processing of the scene, working memory, inhibition etc.). Indeed, as reviewed by Legrand and Ruby (2009), regions of the social network are also involved in other functions, which suggests that these regions serve as integration of the processed information and so, ToM would belong to domain general processing. However, the presence of a human character in the stimuli is not sufficient to elicit the activation of the “social brain” (Saxe & Powell, 2006) if it does not involve the attribution of mental states. For Frith (2007), only some regions of the social network serve as domain-general processing such as the mPFC, which might be involved in general (not only ToM) conscious experience. Furthermore, different studies suggest that the rTPJ is not specifically involved in ToM processing but also in attentional reorienting (Bzdok et al., 2013). However, Scholz et al. (2009)’s fMRI study investigated the rTPJ activation and highlighted that it is a distinct but neighboring region of the rTPJ which is in fact involved in attention orienting. Thus, several regions seem to be involved for ToM processing but they remain difficult to identify. The functional role of these regions in ToM

is difficult to clearly distinguish from the neural structures which are involved in cognitive functions that are not social as such.

2. Visual perspective taking

ToM referred to a set of abilities which allow us making inferences (e.g., beliefs, knowledge, visual, etc.) about others' inner world. Thereby, processing the visual experience of others (i.e., what is seen or how it is seen) is part of ToM and in this thesis, we will focus on this particular aspect of ToM. We will refer to it as visual perspective taking (VPT) which is the ability to make inferences about what is seen by another person, and allows us to reduce the complexity of ToM to some of its basic foundations. Studying VPT is particularly interesting since what other people look at can be used as a gateway to access their inner world. Indeed, what someone is looking at can inform us about what this person likes (or dislikes), intends to do, thinks, etc. To examine how we read other's minds, researchers have often looked at what we understand about others's visual perception.

According to Flavell et al. (1981) and Michelon & Zacks (2006), two levels of VPT can be differentiated. Level-1 VPT which is the focus of this thesis corresponds to the basic ability used to infer which objects can or cannot be seen by others. Empirical findings have shown that the larger the distance between the other person and a target object, the slower participants are to say if another person can see the target object. This suggests that the relationship between another person and the object (seen or not seen) is built through tracing a "line of sight" (Surtees, Apperly, Samson, 2013). Level-2 VPT goes a step further and refers to the understanding that an object viewed by different people (i.e., they have an unobstructed view) may nevertheless be seen differently, thus level-2 VPT establishes "how" a stimulus is perceived for another person from a different position (e.g., a '6' from another visual perspective can sometimes correspond to a "9"). Level-2 VPT is more cognitively effortful because it seems to involve a mental rotation process (i.e., a representation of ourselves in someone-else position) to reason about how an object is seen from another person's visual perspective (Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Surtees et al., 2013, 2016). Indeed, it has been found that participants' response time to judge how an avatar could see an object has been found to be dependent on the angular disparity between the participant and the avatar (Figure 1.4) which was not

the case when participants merely judged if the object was seen or not by the avatar (level-1 VPT). Since this mental rotation seems specific to level-2 VPT, the processes involved to adopt someone else's visual perspective could differ, at least in part, depending on the type of VPT in which we are engaged.

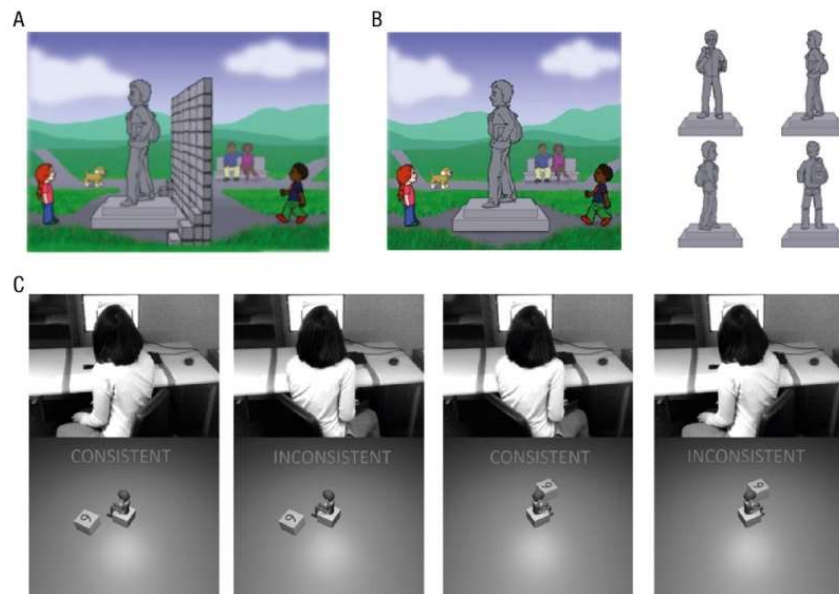


Figure 1.4. Top panel: Illustration of level-1 and level-2 VPT. **(A)** The girl but not the boy can see the statue (level-1 VPT). **(B)** The girl sees the front of the statue and the boy sees the back of the statue (level-2 VPT). Taken from Hutchins & Prelock, 2016. Bottom panel: **(C)** Different conditions showing possible angular disparity between the participant and the avatar. Participants were longer to judge whether the avatar sees a “6” or a “9” when angular disparity between the participant and the avatar was higher. Taken from Surtees et al. (2013).

Infants, from their first year, start to follow other people's gaze direction (e.g., Hood, Willen, Driver, 1998; Corkum & Moore, 1998; Scaife & Bruner, 1975) and, growing older, they become more sensitive (more or less susceptible to follow eye gaze) to whether another person can or cannot see an object (Brooks & Meltzoff, 2002; Butler, Caron, Brooks, 2000). Thus, level-1 VPT is one of the earliest forms of perspective taking found in human infants (Flavell et al., 1981; Flavell, Flavell, Green, Wilcox, 1980; Lempers, Flavell, Flavell, 1977; Moll & Tomasello, 2004, 2006) whereas level-2 VPT is not found in children under 4 years of age (Flavell et al., 1981; Gzesh & Surber, 1985; Hamilton, Brindley, Frith, 2009). Moreover, level-1 VPT is not only present in young infants, but humans share it with

different animal species, contrarily to level-2 VPT which seems to be a human-specific ability. For example, Hare et al., (2000) designed an experiment in which two chimpanzees, one subordinate and one dominant, competed for food. There were two pieces of food, and in one condition, both pieces were visible to the subordinate but one piece of food could not be seen by the dominant chimpanzee (Figure 1.5). In that condition, the subordinate chimpanzee went for the food that was hidden to the dominant chimpanzee suggesting that the subordinate chimpanzee took into account the visual experience of the dominant chimpanzee to choose the best course of action. Thus, authors have claimed that chimpanzees know what conspecifics do and do not see. In the literature, more examples of level-1 VPT ability in different species can be found in non-human primates (Tomasello et al., 1998; Bräuer et al., 2007; Shillito et al., 2005) but also mammals such as dogs or goats (Bräuer et al., 2004, 2006) or even among birds (Emery & Clayton, 2008).

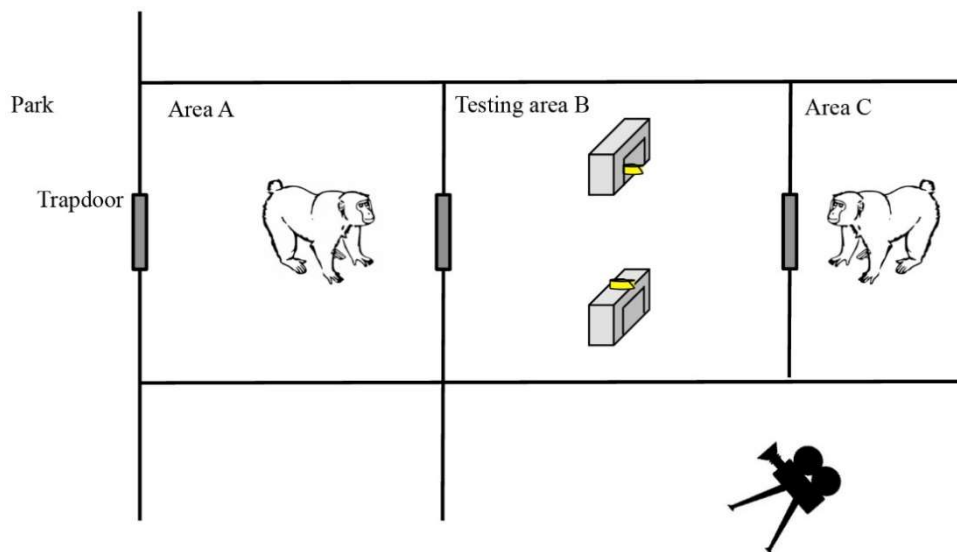


Figure 1.5. Illustration of the experimental apparatus from Canteloup et al. (2016) to test macaques. One piece of food was visible to the two macaques whereas the other piece was only visible to the macaque in area C. Taken from Canteloup et al. (2016).

In recent years, it has been mostly level-1 VPT that has received attention in the literature as some data have suggested that humans (adults) process spontaneously what is seen by someone else. In fact, it all started with the investigation of Samson and

collaborators (2010) in which participants were asked to count the number of targets (red discs) they could see on the walls of a virtual room from their own perspective (self-perspective) or from an avatar's perspective (other-perspective) standing in the middle of the room. The avatar and the participant could see the same number of targets (consistent condition) or they could see a different number of targets (inconsistent condition; Figure 1.6). The study consisted of three experiments. In the first experiment, the trials in which participants had to judge according to their own or to the avatar perspective were randomly mixed. The two other experiments in self and other perspective trials were more clearly distinguished to give participants a clearer opportunity to ignore the irrelevant perspective. Thus, in the second experiment, self- and other-perspective trials were presented in separate blocks of trials and, in the third experiment, participants always judged from their own perspective but in half of the trials a rectangle was used to replace the avatar in the room. The results of the three experiments showed that participants were slower to respond and more error prone when they could see a different number of targets than the avatar, and this occurred even in the condition in which participants had to only take their own visual perspective (i.e. the third experiment). Thus, even when participants were given the opportunity to ignore the irrelevant perspective, they were still influenced by this irrelevant perspective. This impact of someone else's perspective, referred to as altercentric intrusion, only occurred when the avatar was present in the room (not with the rectangle). These results have been replicated numerous times (e.g., Nielsen et al., 2015; Qureshi et al., 2010; Surtees & Apperly, 2012; Ferguson, Apperly, Cane, 2016).

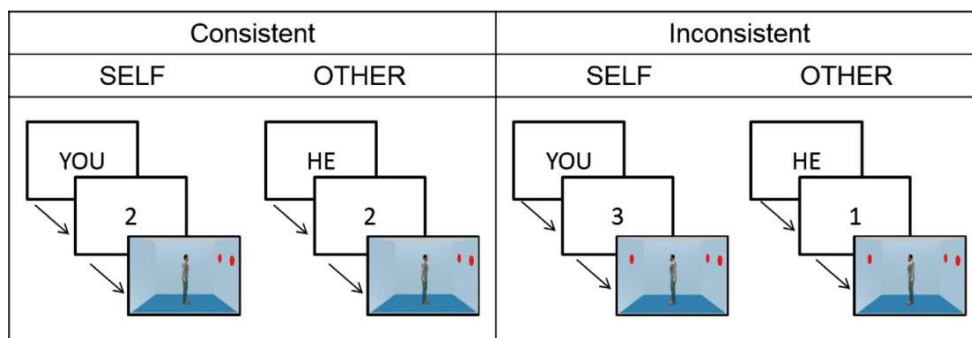


Figure 1.6. Example of trials used on the level-1 VPT task. Taken from Todd et al. (2017).

2.1 Is VPT a spontaneous process?

The three experiments of Samson et al. (2010) study suggest that we spontaneously process what someone else can or cannot see, because participants did it even when it was not necessary for the task, and it happened even when it had detrimental effects on the task performance (altercentric intrusion). In a follow-up study, Qureshi et al. (2010) used the same experimental materials as Samson et al. (2010), in a dual task situation. The authors observed that in dual task situation, the altercentric effect was accentuated in trials in which participants merely judged their own perspective. If the calculation of what the avatar sees was taxing on cognitive resources, the secondary task should have disrupted the processing of the avatar's perspective and the altercentric intrusion should have been at least reduced. However, results showed that the altercentric intrusion increased when subjects performed a secondary task. Therefore, the authors interpreted this as evidence that the calculation of what someone else sees is spontaneous and independent of cognitive load. Furthermore, participants made more egocentric errors on trials where they had to judge the avatar's perspective and thus, the secondary task increased both the egocentric and the altercentric interferences. The increased interference effects when cognitive resources were taken away suggest that the calculation of the irrelevant perspective was still taking place and hence occurred spontaneously. In other words, the calculation of the avatar's perspective as well as the calculation of participant's perspective are both spontaneous and not effortful. However, the increased interferences suggest that it is the processes involved in the selection of the task-relevant perspective which was disturbed by the secondary task and thus relies on effortful processes (see also Todd, et al., 2017). Thus, VPT processes seem to involve at least two subprocesses: on the one hand, the calculation of both the participant and the avatar's perspectives which seems to be spontaneous and less taxing on resources, and on the other hand, the selection of the relevant perspective which is a more effortful process. Note that in the literature (and in the present thesis) when referring to spontaneous VPT, we only consider the calculation process not the selection.

The assumed spontaneous aspect of the VPT ability (i.e., the calculation) has been questioned by different studies. Indeed, different investigations lead to assume that VPT might be less ubiquitous than first thought. For example, on the one hand, Ferguson et al.

(2016) have found that altercentric intrusion effect can be reduced when participants adopted their own perspective on consecutive trials (i.e., the strength of the altercentric intrusion effect was modulated by whether the previous trial was a self or other perspective trial), suggesting that the other perspective processing might not necessarily be spontaneous.

On the other hand, Bukowski et al. (2016)'s study also provides interesting observation concerning the spontaneous aspect of VPT ability. The authors investigated the distinction between a gaze-cuing paradigm (i.e., a paradigm in which a gaze induces a reflexive attentional orientation towards the location consistent with the gaze direction) and the VPT paradigm. Bukowski et al. (2016) showed a higher sensitivity to what someone is looking at in the context of the VPT paradigm than in the context of the gaze-cuing paradigm. The authors concluded that the type of paradigm (i.e., gaze cuing or VPT paradigms) changes the way that participants deploy their attention on the displayed stimulus in order to respond to the task. The "social mind set" induced in a VPT paradigm modifies the value of the other person's gaze. In fact, several studies suggest that the priority could be given to the agent by the context as is the case during a VPT task. In some cases, the mere fact of judging our own visual perspective (generally, experiments in which a perspective cue e.g., "YOU", primed the participant's perspective), can generate a social mind set which would be sufficient to naturally draw attention to the other person, which seems to increase the salience of the gazer and induce the processing of what the agent is looking at. Thus, attention towards the agent would prevent competition from other potential sources of attentional capture or orientation and induce the processing of the agent's perspective (see also Todd, Hanks, Galinsky, Mussweiler, 2011). Similarly, asking participants to take their own or the avatar's perspective increases the participants' social mind set, and thereby accentuate the value of the other person which would facilitate the processing of the avatar VPT. Indeed, if participants have to adopt the avatar's VPT on some trials, it might lead them to adopt the strategy of paying attention to the avatar's VPT even when they are merely asked to take their own perspective (Ferguson et al., 2016). Thereby, when participants were asked to switch perspectives between trials, the spontaneous processing of the avatar's VPT could be due to a task carry-over effect (Conway et al., 2016). Furthermore, concerning the spontaneous aspect of the VPT ability, different studies (Bukowski et al.,

2016; Gardner et al., 2017, Experiment 1) showed that the mere presence of an agent is not sufficient to induce the spontaneous processing of what he was seeing. This point emphasizes once again that the assumed spontaneous aspect of the VPT (i.e., the calculation of the perspectives) might not be always spontaneous and would be dependent on the attention deployment in response to the task instruction (context dependent).

However, note that the conclusions of the above paradigms are based on the assumption that VPT task requires the representation of others' views. However, to successfully perform a VPT task, it may not be required to process the other's perspective as discussed below. Indeed, according to the submentalizing hypothesis, the agent acts as an attentional cue which moves participant's attention to the targets in front of the agent in a reflexive manner and it is not the agent's mental state (i.e. his visual experience) which is processed per se.

2.2 Review of the VPT investigations which oppose the mentalizing and the submentalizing hypotheses

When participants are explicitly asked to take the avatar's perspective, authors agree that participants really process the avatar's visual perspective (i.e., what the avatar sees). However, there is a debate in the literature over the processes involved in the situation in which the participants are not explicitly asked to take the perspective of the other person. In those cases, the interpretation of the reasons why an altercentric intrusion effect is observed are debated. According to the submentalizing hypothesis, when participants adopt their own perspective, the avatar directs the participants' attention to one side of the room and it is not the agent's seeing which is processed as such. For example, in inconsistent trials (i.e., when participants do not see the same numbers of targets as the avatar), the targets' location are in conflict with the avatar's gaze direction. This would thus require a re-orientation of attention and causes a delay of the detection of the targets. In contrast, in the consistent trials (i.e., where participants see the same number of targets as the avatar), the avatar's gaze direction acts as a valid cue and facilitates the detection of the targets. This alone would suffice to explain the presence of the altercentric intrusion effect which would be caused by generic attentional process. However, according to the mentalizing

hypothesis, the processing goes beyond a simple shift of attention and involves the representation of other person's view. In that case, it is the conflict with the avatar's visual perspective which explains the presence of the altercentric intrusion effect.

Different types of evidence in the literature support the mentalizing hypothesis indirectly. For example, empathy, the ability to understand and share the feelings of another has been found to be associated with better VPT ability (Mattan, Rotshtein, Quinn, 2016). Furthermore, different studies highlighted that the emotional state of the participants, such as anxiety (Todd & Simpson, 2016), guilt or anger (Bukowski & Samson, 2016) can influence their VPT ability. For example, anger tends to make participants more self-centered (Bukowski & Samson, 2016) and anxiety has also been shown to have detrimental effects on perspective taking which seem specific to social stimuli (note however, that the non-social entity control used was a dual-colored stick without directional information; Todd & Simpson, 2016). In addition, Nielsen et al. (2015) have found that the altercentric intrusion effect observed with social stimuli was positively correlated with self-reported levels of everyday life empathic and perspective taking abilities (as measured by the Interpersonal Reactivity Index; IRI: Davis, 1980) but such correlation was not observed with other types of stimuli (semi-social and non-social stimuli; but see Conway et al., 2016). It could be argued that if empathy and emotional states influence VPT, it is more likely that the processes influenced are social in nature.

Ferguson et al. (2016) results also support the mentalizing hypothesis indirectly. Participants were eye-tracked while performing a VPT task. According to the submentalizing hypothesis the avatar acts as a directional cue which induces an automatic shift of attention to the avatar's line of sight direction. Therefore, the first participant's gaze fixation should be always the wall in front of the avatar. However, according to the mentalizing hypothesis, participants' eye gaze should be modulated whether participants were cued to take their own or the avatar's visual perspective i.e., they should not necessarily make their first fixation on the wall in front of the avatar. The eye tracking measures revealed that participants were most biased by the avatar's gaze direction when adopting the avatar's perspective and consequently made their first fixation on the wall in front of the avatar for those trials. However, when judging their own perspective, participants were equally likely

to make their first gaze fixation in front or behind the avatar i.e., they divided their attention between the possible target locations. Therefore, Ferguson et al. (2016)'s results support the mentalizing hypothesis and goes against the submentalizing hypothesis, for which the participant's first gaze fixation should be always the wall in front of the avatar.

On the other hand, we can also find in the literature different arguments which support the submentalizing hypothesis indirectly. Indeed, in a parallel line of research, it has been shown that attention is directed towards what someone is looking at. Indeed, a person's head and body orientation serve as facilitative spatial cues. In the literature, we can find that gaze direction (Friesen & Kingstone, 2003; George, Driver, Dolan, 2001; Hietanen, 1999; Nummenmaa & Calder, 2009; Schuller & Rossion, 2004) and head orientation (Laube, Kamphuis, Dicke, Thier, 2011; Nummenmaa & Calder, 2009) trigger attentional shifts. The gaze-cuing paradigm has shown that participants are quicker to detect (or identify) a target if it was preceded by a face which looked towards the target (gaze congruency effect) than when it appears in the opposite direction. Participants continue to be quicker even when the cue is not predictive (involuntary process) or when participants are under dual task conditions (effortless process; Law, Langton, Logie, 2010). In addition, other types of stimuli which are not faces (e.g., object with eyes, arrow) can also trigger a fast shift of attention which show that this effect is not confined to social stimuli and hence, supports the interpretation of an automatic attentional orienting caused by generic attentional processes (Driver et al., 1999).

However, gaze congruency effects have been positively correlated with the degree of agency of the gazer (Wiese, Wykowska, Zwicker, & Müller, 2012) and in addition, some studies suggested that mentalizing also mediates the gaze cuing effect (Teufel et al., 2009, 2010a, 2010b). For example, Teufel et al. (2010a) have developed a gaze-cuing paradigm in which the gazer wore opaque or non-opaque goggles, and have found that participants' processing of a subsequent target was modulated by the possibility of the agent to see or not the target. Participants were quicker to identify the target in the condition where the gazer could see it (non-opaque goggles). This study supports the involvement of more sophisticated processes which may compute the agent's mental state rather than merely

directing the participant's attention toward what is in front of the agent (see also Nuku & Bekkering, 2008 or Cole et al., 2015, 2017, Experiment 3 for contradictory results).

In addition, the study of Bukowski and collaborators (2015), also provides interesting findings concerning the gaze congruency effects. In this study, the authors investigated the distinction between the gaze-cuing paradigm and the VPT paradigm (presented above). They highlighted a higher sensitivity to what someone is looking at in the context of VPT paradigm than in the context of the gaze-cuing paradigm. Such distinctions are in line with the mentalizing hypothesis, and supports a particular processing during VPT task which is not merely caused by the shift of attention induced by the avatar. Furthermore, Hietanen et al. (2008)'s study emphasizes the distinction between the gaze congruency effects and attentional cuing. They provided evidence that both gaze and arrow cues trigger attentional shift, but brain measures revealed that they are based on different neural mechanisms. Their finding is also supported by a brain lesion (to the superior temporal gyrus) study, in which the patient showed intact cuing effect by arrows but not by gaze cues (Akiyama et al., 2006).

In the literature, it is possible to identify three different strategies which were used to more directly differentiate between the mentalizing and the submentalizing hypotheses. However, results are not consistent across studies.

A first strategy has been to replicate the Samson et al. (2010) paradigm, and add a condition with an arrow as central cue instead of an avatar to investigate whether it was the directional rather than the agentive feature of the agent which was important. Indeed, both an avatar and an arrow carry directional information, but mental states are assumed to be only attributed to the avatar because arrows have no agentive feature (Figure 1.7). According to the submentalizing hypothesis, participants' performances should show no significant difference between the arrow and avatar conditions (i.e., similar consistency effect). And indeed, different studies (Gardner et al., 2017; Conway et al., 2016; Santiesteban et al., 2014, 2017, Experiment 1) have found an altercentric intrusion of the same amplitude, whether it was an avatar or an arrow which was presented. These results support the assumption that the avatars as the arrows simply trigger a directional shift of attention and the authors concluded that the Samson et al. (2010) results were caused by

a shift of attention induced by the directional feature of the avatar (submentalizing hypothesis) and were not caused by a mentalizing process. However, in their interpretation of the results, the authors considered that it was not necessarily the same mechanism involved whether it was an avatar or an arrow presented. Indeed, it is possible (and probable) that general attentional processes are involved in both cases, but social stimuli could lead to a deeper processing, which involves mentalizing (Furlanetto, Becchio, Samson, Apperly, 2016). Thus, although different studies have showed similar patterns of behavioral response whether the central cue was an avatar or an arrow, if we consider that general attentional processes are also involved while adopting someone else's perspective, these studies do not allow us to discard the mentalizing hypothesis.

In direct line with the mentalizing hypothesis, some studies have found qualitative differences when the central stimulus was a person compared to an arrow. These results suggest that, in addition to provide a directional cue (common feature to a person and an arrow), the presence of a person leads to process his relationship with the objects in the room as "seen" or "not seen". For example, Nielsen et al. (2015) highlighted a significant altercentric intrusion effect stronger for stimuli involving a person than an arrow. Schurz et al. (2015) further showed that social stimuli (such as a person) triggered a particular brain processing which did not prevail with non-social stimuli (such as arrow). Indeed, the authors conducted an fMRI study and compared the implication of the neural structures during a VPT task while participants had to judge the number of targets they could see from their own perspective or from a central cue perspective (explicit vs implicit conditions). In addition, there were different types of central cues such as an avatar, an arrow or a lamp. For example when the participants had to evaluate the perspective of the lamp, they were asked to judge how many targets the lamp was lightening if it was turned on. The authors compared the neural responses whether the central cue was an avatar (a social stimulus) or a non-social stimulus. They also investigated whether spontaneous VPT recruits areas of the ToM network and whether implicit and explicit VPT recruit domain-specific processes. They found that VPT engaged ToM areas (in the rTPJ, the mPFC and the ventral precuneus) when spontaneously processing information linked to the other's perspective (during self-perspective judgments). These ToM areas of activation were found only when an avatar was used as central cue and were stronger when the participant and the avatar's visual

perspective were different from each other (i.e., in inconsistent trials). Therefore, their results support the mentalizing hypothesis because it highlights specific activation of ToM areas only for the VPT task which involved an avatar, i.e., a social agent (but see Catmur, Santiesteban, Conway, Heyes, Bird, 2016). These studies challenge the submentalizing hypothesis (which should have led to a similar response whether the central cue was an avatar or an arrow) because it strongly suggests that VPT paradigm leads to a processing which goes beyond a reflexive orienting effect.

However, in another study (Santiesteban et al., 2017) also based on the strategy of comparing responses when the central stimulus was a person compared to an arrow found results which this time support the submentalizing hypothesis and challenge the mentalizing hypothesis. In this brain stimulation study (Santiesteban et al., 2017), the authors used a disruptive repetitive TMS on a key brain region (rTPJ) which is assumed to be strongly involved during mentalizing. At the same time, participants performed a VPT task in which the central cue was, depending on the trial, a person or an arrow. According to the mentalizing hypothesis, we mentalize for social stimuli, therefore the disturbance of the rTPJ should affect the participant's performance more during the trials involving a person as central cue than an arrow. However, in line with the submentalizing hypothesis, the authors did not find significant difference in the performed VPT task whether the central cue was a person or an arrow.

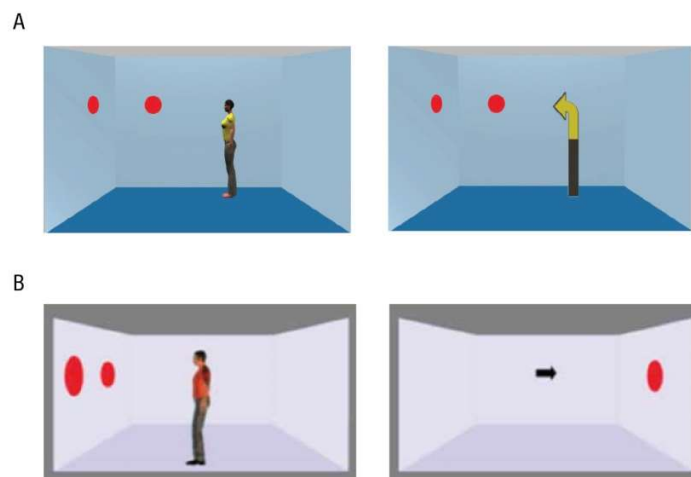


Figure 1.7. Stimuli used to test whether implicit VPT involved mentalizing or submentalizing mechanism. (A) Stimuli from Santiesteban et al. (2014). (B) Stimuli from Nielsen et al. (2015).

A second strategy used to more directly differentiate between the mentalizing and the submentalizing hypotheses has been to examine the effect of presenting an avatar facing different targets but not necessarily seeing all of them because some were hidden by an obstacle (e.g., a wall obstructing the view or the avatar wearing opaque goggles; Figure 1.8). For example, in the experiment of Baker et al. (2015; Figure 1.8A), the participant's and the avatar's point of view were either consistent (a barrier was displayed but not placed in the line of sight of the avatar or windows were added to the barrier), or inconsistent (a barrier obstructed the avatar's view preventing him of seeing the objects in front of him). Note that all the objects were always displayed in front of the avatar, i.e., the avatar's direction always acted as a valid cue. According to the submentalizing hypothesis, the presence or absence of an obstruction in the line of sight of the person should not have any effect on the speed or accuracy of participants' judgment of what they themselves see. Indeed, if participants do not compute what is seen by the other person but only what is in front of the person, their performance should be unaffected by the presence of an obstacle. In contrast, the mentalizing hypothesis predicts that the presence of an obstacle in the line of sight should affect participants' performance, because it affects what is seen by the person (e.g., longer reaction time (RT) to respond in case of avatar's conflicting visual perspective). Baker et al. (2015)'s results showed that the altercentric intrusion appeared only when the participant and the avatar's visual perspective were different from each other due to the presence of obstacle (inconsistent trials). This intrusion reflected the calculation of the perspective of the avatar and not a simple attentional orienting effect, since the avatar was always oriented towards the targets. However, Cole et al. (2016) conducted a similar study in which barriers were placed between the avatar and the walls of the room (Figure 1.8B) and did not replicate Baker et al. (2015)'s findings. In one of the conditions, the avatar did not see the targets because the barriers were completely closed, while in the other condition, he saw the targets through windows in the barriers. The researchers found that the altercentric intrusion appeared both when the barriers were open and when they were closed. Their results are in line with the submentalizing hypothesis (the presence or absence of an obstacle has no effect on participants' performances). However, the stimuli used by Cole et al. (2015, 2016, 2017) could explain their absence of an obstacle effect on participants' performance. Unfortunately, the condition which was supposed to depict a person seeing

an object has perspective issues which makes the stimuli less realistic and not obvious to understand (e.g., the object can be perceived as unseen where it is supposed to be perceived as seen, Figure 1.9).

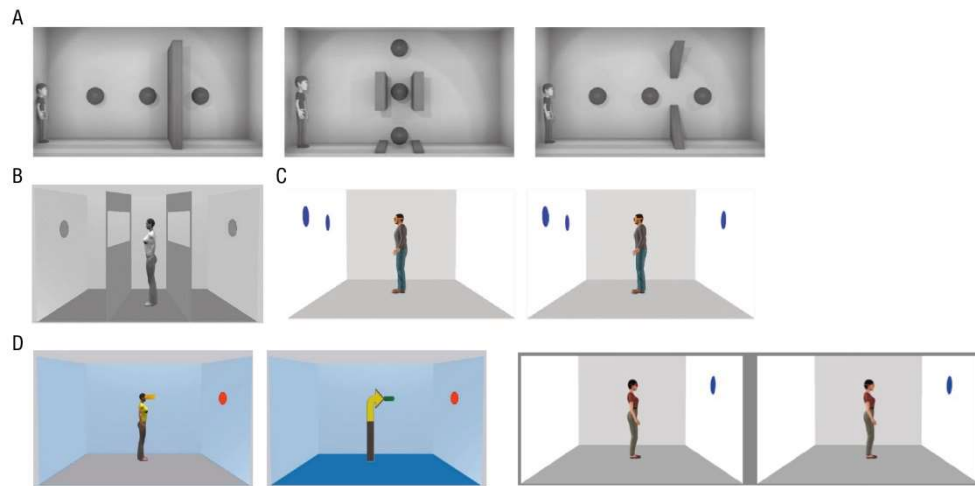


Figure 1.8. Stimuli used by (A) Baker et al. (2015), (B) Cole et al. (2016), (C) Furlanetto et al. (2016), (D) Conway et al. (2016).

Still based on the strategy of placing an obstacle in the line of sight of the avatar, Furlanetto et al. (2016), developed an experiment in which the avatar displayed in the center of the room wore goggles or no goggles (Figure 1.8C). Participants, prior to the experiment, were informed that one type of goggles was transparent (e.g., the red goggles) or opaque (e.g., the orange goggles). Thus, when participants performed the task the avatar's visual experience depended on the goggles it wore. Results showed altercentric intrusion effect only when participants believed that the avatar could see (i.e., when the avatar wore transparent goggles or no goggles at all). Their findings are in line with the mentalizing hypothesis. However, Conway et al. (2016) in a similar study (Figure 1.8D) did not replicate Furlanetto et al. (2016) results. The magnitude of the altercentric intrusion effect was similarly modulated whether the avatar wore opaque or transparent goggles. Therefore, participants did not seem to compute the avatar as "seeing" the targets (ascription of mental states) and thus, the results support the submentalizing hypothesis. Conway et al. (2016), proposed that the altercentric intrusion effect observed by Furlanetto et al. (2016) (when participants believed that the avatar could see) was in fact, a false

positive. Nevertheless, Freundlieb et al. (2016) by using a similar strategy in a study in which they investigated the Level-2 VPT, have found evidence in accordance with the mentalizing hypothesis. They showed that participants spontaneously took into account how somebody else perceives the environment only when that person could actually see. If the person wore opaque glasses, the participants seemed not to spontaneously adopt the person's visual perspective.

Conway et al. (2016) combined the two strategies presented and placed an avatar or an arrow in the center of a room with different telescopes which masked or did not mask the targets on the wall (Figure 1.8D). Before the experiment, the participants manipulated the two telescopes (one was yellow whereas the second was green) to discriminate between the one which allowed them to see the targets and the one which was masking them. In accordance with the submentalizing hypothesis, the authors found similar altercentric intrusion effect, for the avatar and the arrow irrespective of whether the telescope showed or masked the targets.

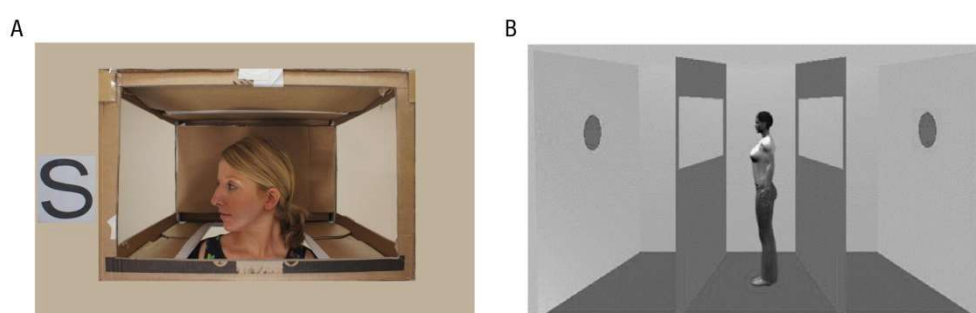


Figure 1.9. (A) Stimuli used in Experiment 1 by Cole et al. (2015). Note that the “S” placed at the front of the edge of the box is not necessarily seen by the person in the box whereas stimuli are assumed to depict the seeing condition. (B) Stimuli used by Cole et al. (2016). Note that the walls are not realistic, they have no thickness.

Finally, a recent study proposed a third strategy to test the mentalizing versus the submentalizing hypotheses. Based on the literature it is known that spatial cueing tasks facilitate the detection of a target (1 cue preceded the appearance of the target; Posner, 1980). Similarly, double cuing versions of this task (2 cues preceded the appearance of the target) also facilitate the detection of the target, and so, attention can be facilitated at two

locations. Michael et al. (2017)'s study used similar stimuli to those used by Samson et al. (2010) but there were always targets on both, the right and the left wall. Depending on the trial 0, 1 or 2 avatars were displayed in the center of the room (Figure 1.10A). Note that when 2 avatars were displayed, each had a perspective which was inconsistent with the perspective of the participant (targets were always displayed on both walls). Therefore, if the avatar acts as a directional cue, the detection of the targets (on both walls) should be facilitated if there are 2 avatars (similarly to the double spatial cuing task which facilitates the participant's detection of the target). In contrast, if the avatar triggers perspective taking mechanisms, trials in which there are 2 avatars should lead to an increase in the conflict between the avatars' and the participant perspective, and delay the detection of the targets. Participants always judged according to their own perspective only, and authors observed that participants were slower to detect the target when 2 avatars were displayed on the stimuli (Figure 1.10B). Thereby, the pattern of results suggests that the avatar triggered perspective taking mechanisms, thereby, supporting the mentalizing hypothesis.

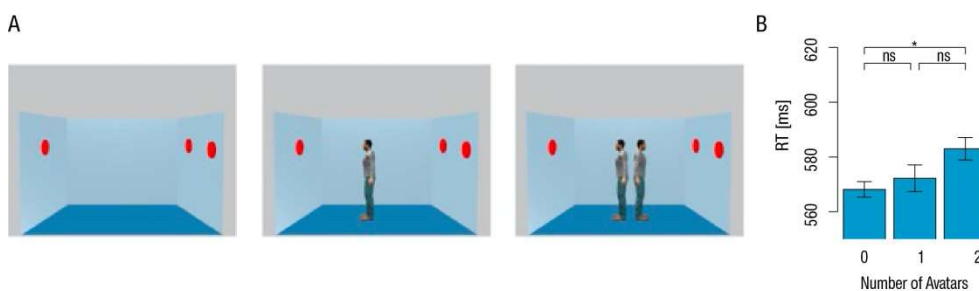


Figure 1.10. (A) Stimuli used by Michael et al. (2017). (B) Results (Experiment 2) from Michael et al. (2017).

Thus, despite the different studies which ingeniously investigated VPT, it is still unclear whether we process what someone else "sees" per se (mentalizing hypothesis) or if the avatar drives the attention (as an arrow) to what is in front (submentalizing hypothesis). Indeed, we observed difficulties in replicating studies (e.g., Furlanetto et al. (2016) vs. Conway et al. (2016); Nielsen et al. (2015) vs. Conway et al. (2016); Cole et al. (2016) vs. Baker et al. (2015); Santiesteban et al. (2014, 2017) vs. Schurz et al. (2015)). In such a context, some authors have tried to develop a new theoretical framework, the schema theory, to try to accommodate the different results.

2.3 The schema theory

The schema theory goes beyond the submentalizing hypothesis in the sense that gaze cuing is thought to generate a “schema” which is not specific to social cues. These schemas are abstract mental representations that can be summarized in structures of preconceived ideas (knowledge about concepts), which are then used to categorize different types of knowledge of the environment (e.g., word meaning/representation, relationships between gaze direction and relevant stimuli) in order to understand the world. The schemas are learned from repeated sequences of actions required to achieve a particular goal, and the human mind uses schemas to organize, retrieve, and encode chunks of important information. For example, the repetition of the link between a gaze direction and relevant stimuli generates the formation of a schema. A schema, once activated, is executed automatically in a bottom-up fashion. If the gaze direction schema is activated, the gaze will trigger a shift of spatial attention towards the location gazed at. However, schemas are generally thought to have a level of activation, which can be modulated by top-down processing such as attentional processing or priming. Top-down processing will give more value to a specific schema and increase its probability to be selected, when multiple conflicting schemas can be applied to the same information. Thereby, Gardner et al. (2017) and Cole et al. (2015) proposed that the probability of activating a particular schema will be modulated by the strength of the top-down processes that regulate the schema activation threshold. So, activation of the gaze cuing schema will depend on the one hand on the power of the directional cue (e.g., how clear is the eyes movement?) and on the other hand, on the top-down processes which influence the gaze cuing schema threshold. For example, viewing the eye movement of a person known to be blind might be less efficient (than a seeing person) to activate the gaze cuing schema because of the top-down processes which might inhibit its activation (why follow what is certainly a non-informative gaze?). Similarly, if the gaze cue is powerful (e.g., clear eye movements), the gaze cuing schema can be rapidly activated and trigger a rapid and involuntary shift of attention (even when the gaze is irrelevant for the task). However, if the gaze cue is not powerful enough to activate the gaze cuing schema (e.g., the eyes are hardly visible), or if the participant was primed with a social mind set, the gaze cuing effects can be modulated by contextual information and

participants being more or less susceptible to consider the gaze cue (see Gardner et al., 2017 for the schema theory interpretation of Samson et al. (2010)'s results).

To simplify, the schema theory is a variant of the submentalizing hypothesis, with no specific processing engaged for social cue. Implicit VPT and gaze congruency effect are the consequence of learned association, and do not involve mental state attribution. In other words, any overlearned directional cue (e.g., arrow) can induce a schema which will generate the same pattern of results as gaze cuing. According to the schema theory, it is the strength of the directional cue, (i.e., if the directional cue is or not ambiguous, e.g., avatar wearing opaque goggles) and the participant's social mind set, which can lead the participant to actually inferring mental states (e.g., the avatar can or cannot see).

In the light of the highly debated and conflicting results of studies trying to disentangle the mentalizing and the submentalizing hypotheses, the main goal of this thesis was to examine the extent to which a specific EEG technique, fast periodic visual stimulation (FPVS) could offer a new empirical approach to study VPT and advance in the mentalizing/submentalizing debate. In the next section, the EEG technique and the FPVS approach will be presented.

II. Methodological framework

1. The Electroencephalography

An electroencephalogram (EEG) allows us to record the electrical brain activity, and the majority of the analyzed data in this thesis were based on EEG recordings during visual stimulation.

1.1 Principles of the EEG

Richard Caton was a British scientist who used a galvanometer (instrument for detecting and measuring electric current) in 1875 to observe for the first-time electrical impulses on the surface of the brain of living animals. Some years later, in 1929, Hans Berger, a German scientist developed this technique and recorded with an EEG, the electrical activity on the human brain from the surface of the head. It was the first evidence of oscillatory patterns of neural activity. The electrical current of the brain occurs when neurons communicate. The neurons' discharge is generated by currents of potassium (K^+) and sodium (Na^+) which cross the cell membrane, and cause the depolarization of the membrane to some threshold which then causes them to "fire". This event is called action potential, and tracking these discharges over time reveals the brain activity.

In the brain a particular type of cell, the pyramidal neurons, are assumed to be the main generators of the EEG response. These neurons are radially oriented in the cortex (orthogonal to the surface; Lopes da Silva, 2002) and thanks to the difference of ion charge between the superior and the inferior part of the cells, these neurons act as electromagnetic dipoles which co-align and co-activate, add up and generate an electric field. If the neurons were located in the brain without any symmetric aligned structure, the electric fields would cancel each other out, and they would not be detectable on the scalp. Also, not all electrical fields generated by the brain are strong enough to spread all the way through tissue, bone and skull towards the scalp surface. More precisely, EEG is most sensitive to electrical current generated in superficial layers of the cortex; this means that deep brain structure has a limited contribution to the EEG signal (as well as the electrical current which are tangential to the skull). Furthermore, EEG can take hundreds to thousands of snapshots of electrical activity within a single second on several points on the scalp but can only detect

a large-scale of neurons which fire near-synchrony. Indeed, according to Nunez and Srinivasan (2006), each EEG's sensor (called electrode) records the voltage (ionic electrical activity) of approximately 100 million to one billion of neurons. However, if the EEG is silent and relatively tolerant of subject movement (unlike most of other neuroimaging techniques; e.g., fMRI) it remains sensitive to ocular movement and electrical environment which are a source of noise and cause perturbations on the EEG recording.

To sum up, EEG is a non-invasive method with a low spatial resolution but an excellent time resolution of the order of milliseconds.

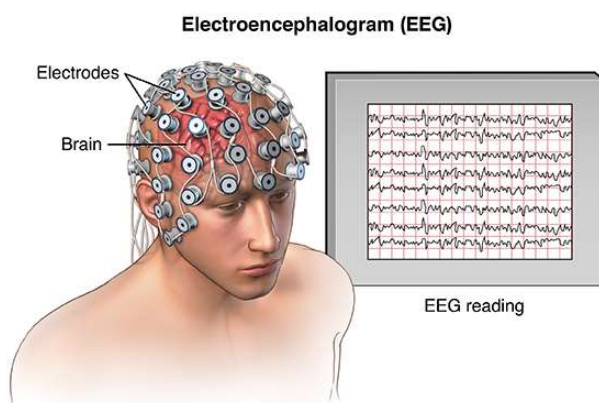


Figure 1.11. Electrodes placed on the head of a person with the EEG recording. EEG's electrodes are placed on the scalp on standardized location. The EEG system uses certain anatomical landmarks which ensure similar spatial positioning for each participant.

EEG measures the brain activity and it is generally admitted that cognition is linked to brain functions which are themselves underpinned by the neuronal network activity. Since EEG gives access to the variation of the electric potential among time, it is an ideal technique to investigate the precise time course of brain processes involved in various cognitive and perceptual functions (Figure 1.11). Therefore, EEG can be considered as a gateway to understanding neurocognitive processes underlying human behavior. EEG is also used as diagnostic tool in several clinical conditions such as epilepsy, or to locate areas of damage

in the brain following head injuries, (stroke, tumor) or to differentiate between psychiatric and organic diseases (Libenson, 2010).

1.2 Basic principles of signal processing

An event, which can be either external (e.g., perception of stimulus) or internal (e.g., thought) generates a brain response called Event-Related Potential (ERP). This electrical potential can be recorded on the skull by EEG in the form of a trace, the EEG's spectrum (Figure 1.11). Usually, ERPs are recorded in response to an isolated stimulus event (transient ERP). However, the brain is never "silent" (there are always neurons which are firing) and the relationship between the stimulus and the brain response is variable. Multiple presentations of the same stimulus over several trials, which allow to average the EEG activity specific to the stimulus. Indeed, averaging several trials allows on the one hand, to decrease the neurons' responses which are not related to the stimulus processing (the noise, which is stochastic) and on the other hand, to increase the neurons' response specifically linked to the event. In other words, by averaging, we decrease the noise and we increase the signal that is systematic and gives access to the resulting components over time (based on the assumption that there is a systematic signal that occurs as a consequence of the event). For this reason, it is crucial to know the exact time of the delivery of the stimulus, in order to extract the precise brain response of the stimulus (Luck, 2005). The EEG spectrum is then assumed to represent the link between stimuli and the brain response and it is largely used to investigate human cognition.

To illustrate EEG findings, I will briefly present one study which thanks to ERP measure explored the neural basis of ToM. More precisely McCleery et al. (2011)'s study investigated the different neural mechanisms associated with VPT. In the context of VPT, ERP measures can help to explore the neuronal time course associated with different brain regions in the calculation and the selection of one's own perspective or the perspective of another person. McCleery et al. (2011) used a similar paradigm to Samson et al. (2010) coupled with EEG and observed that the posterior cortex was the first involved and seemed to process the representation of our own versus another perspective. A late ERP component was also observed (slow wave 600-800 ms post-stimulus) over lateral frontal scalp regions and has been proposed to reflect the neural mechanisms involved to differentiate whether the

avatar and the participants saw the same or different numbers of targets. However, if the ERPs technique allows us to study the neuronal response along time for a whole

complex stimulus (e.g., avatar in a context, seeing or not objects), it is not possible to separate the specific neuronal responses for the different parts which constitute the complex stimulus (e.g., the neural response for the avatar and the object separately). Therefore, this technique remains blind to the neuronal response of the different elements that constitute the stimulus and cannot investigate interactions among individual elements because the ERPs response corresponds to all stimuli presented in the visual field.

2. Fast periodic visual stimulation

2.1 Basic principles/considerations

In contrast to the usual EEG studies which record brain responses to isolated stimulus events (transient ERP), in this thesis, our experiments involved the measure of brain responses to fast periodic visual stimulations (FPVS). The FPVS approach consists in presenting to a participant a train of periodic stimuli (for example a sequence of flashes) which will generate the same periodic pattern in the brain's activity (Figure 1.12). This periodic brain response has been referred to as the "steady-state" visually evoked potential (SSVEP; Regan, 1966, 1977) and follows the same periodicity as the displayed stimulus. To record a specific neural response to the presentation of a stimulus, transient ERPs are analyzed in the time domain, whereas with the FPVS method we analyze the brain response in the frequency domain. Indeed, the repeated stimulation of the flickering stimulus generates ERPs at the exact same frequency of stimulation. Thus, presenting the stimulus at a fixed frequency (i.e., periodically) will generate brain responses at the exact same frequency. Thus, it makes sense to analyze the brain's response in the frequency domain instead of the time domain.

The FPVS approach offers lower temporal resolution, but in return, very high frequency resolution. Indeed, this method is able to precisely isolate the neural activity induced by the stimulus itself as peak in the frequency domain at the frequency of stimulation. However, besides the EEG response at the frequency of stimulation, additional cerebral responses can also be observed at integer multiples of the stimulation frequency

(e.g., 6 Hz stimulation will also lead to responses at 12 Hz, 18 Hz, etc.) because of the non-linear nature of the visual response. The harmonics can provide additional insights into the underlying neural encoding mechanisms. However, these responses tend to decrease in amplitude (lesser peak amplitude in the frequency domain) for harmonics of the fundamental frequency.

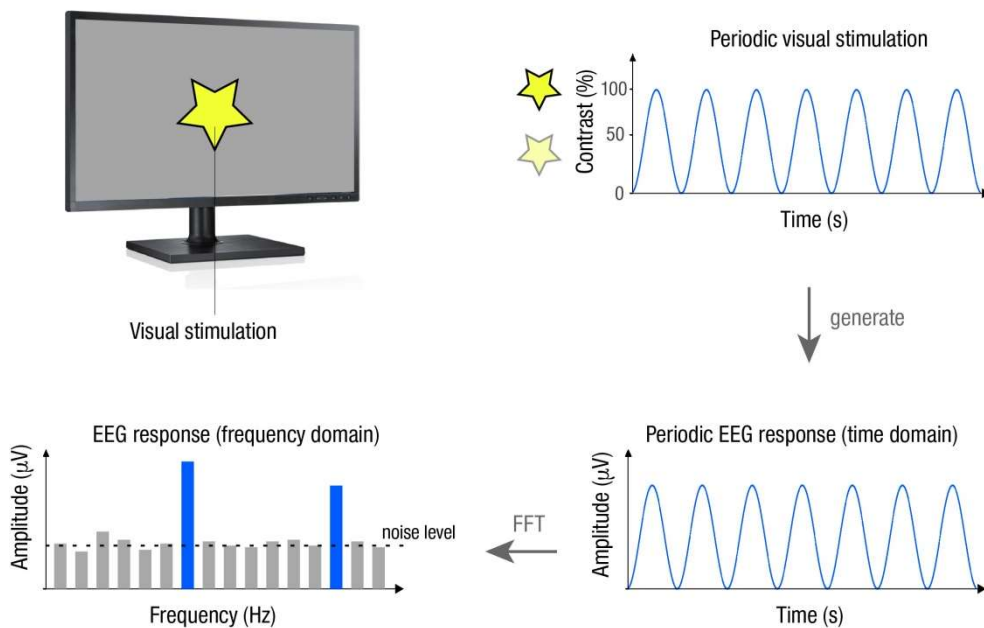


Figure 1.12. The principle of the FPVS approach. A stimulus is shown at a fixed frequency (with a periodic variation of its contrast). The stimulation generates a periodic EEG response at the exact same frequency. This response can be quantified into the frequency domain at the stimulation frequency. Note that depending on the actual form of the time domain response, the response can be distributed across several harmonics (integer multiples of the frequency of the visual stimulation).

To precisely isolate the neural activity induced by the periodic stimulation, we applied on the EEG recording a fast fourier transform (FFT): it is an algorithm which converts the EEG signal (which is at this stage in the time domain) into the frequency domain (and vice versa; Figure 1.13). To sum up, the FFT decomposes any waveform (as in EEG, the electrical signal over time) into a set of sine waves of different frequencies and amplitudes. The sum of these sine waves corresponds to the original wave form (Luck, 2005). In the frequency domain, the interval between frequencies corresponds to the frequency resolution, which is inversely proportional to the duration of the EEG segment. To illustrate, if the EEG

segment lasts 50 seconds in the time domain, the frequency resolution would be equal to 0.02 Hz ($1/50$) in the frequency domain. Thus, each bin (frequency component) of the frequency spectrum would be a multiple of 0.2 Hz.

The main advantage of the FPVS approach is that the FPVS responses can be easily identified. Indeed, the signal of interest (response for the visual stimulation) is concentrated within a particular frequency (and its harmonics) whereas the noise is stochastically distributed across broad frequencies. Therefore, the measure with this approach provides a high signal to noise ratio (SNR). The SNR allows us to compare the signal of interest with the surrounding noise, in order to indicate the reliability of the measure of interest. Thus, the FPVS response can be precisely (depending on the frequency resolution) and objectively identified at a predetermined frequency i.e., at the frequency of stimulation used.

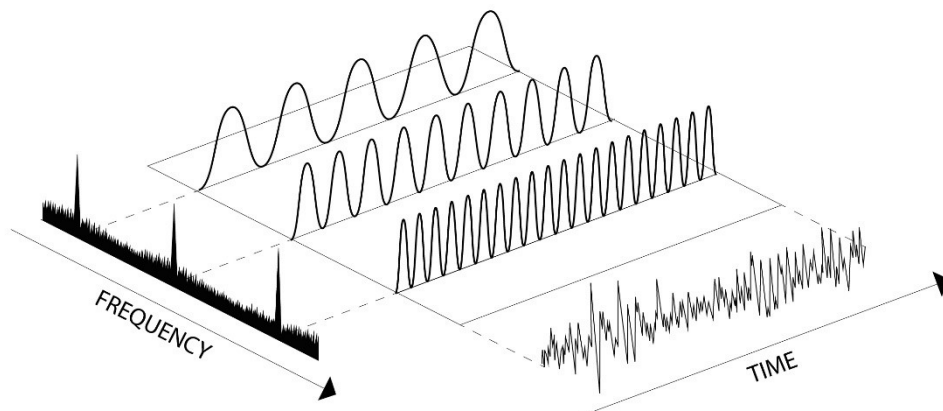


Figure 1.13. View of the same signal in the time and frequency domain. FFT converts the EEG signal from the time domain to the frequency domain (and vice versa).

Several studies localized the major generators of the FPVS response in the visual cortex (e.g., Singla & Haseena, 2014; Andersen et al., 2008, 2009, 2010). Accordingly, the highest FPVS amplitude responses were recorded at occipital and parietal scalp region. However, fast periodic stimulation method is not restricted to the visual system and can be adapted to study other sensory modalities such as the auditory (e.g., Giani et al., 2012) and the somatosensory systems (e.g., Vlaar, van der Helm, Schouten, 2015). Currently it is mainly used to study neural activity involved in low-level processing such as perception or

attention. The technique found large applications with brain computer interfaces. More recently, the method has been adapted to study higher-level visual processing, like face perception and reading (Liu-Shuang et al., 2014; Lochy et al., 2015; Rossion & Boremanse, 2011; for a review Norcia, Appelbaum, Ales, Cottureau, Rossion, 2015).

2.2 Overview of the FPVS approaches

The FPVS approach has been successfully adapted to various stimulation paradigms to investigate different brain mechanisms such as for examples, visual attention or rapid categorization of stimuli (e.g., face or object, word or non-word). The development of our paradigms requires a review of different FPVS paradigms in the literature. Note that there is large variability between experimental setups, e.g., frequencies used, stimuli (light flash, letters, faces, dots), analyses (one harmonic or several harmonics included), which makes it difficult to directly compare the results.

2.3 Stimulus parameters

Several parameters are known to play a role in the amplitude of the FPVS response, but one of the main factors which will interest us, is the frequency of stimulation.

It is possible to obtain FPVS responses for a large range of frequencies, as shown by Herrmann (2001), which presented a flickering light at frequencies from 1 to 100 Hz and found FPVS responses in the EEG spectrum at all the frequencies tested up to at least 90 Hz. However, the amplitude of the FPVS response is linked to the stimulation frequency (Herrmann, 2001; Pastor et al., 2003; Figure 1.14) and particular caution must be taken with the frequency used (Ding et al., 2006) which can be a critical point in investigating some mechanisms. This is why studies often used different frequencies depending on the investigated mechanisms. On the one hand, we can note that studies which investigated the effect of attention on low-level processes use relatively high frequency rates, mostly in the range of 7 Hz to 20 Hz (e.g., Andersen et al., 2008, 2012; Kim et al., 2007; Morgan et al., 1996; Müller & Hubner, 2002; Walter et al., 2012). Studies which inspired this thesis have targeted higher-levels of processing (e.g., face identity coding) and relied on slower frequency rates, e.g., 6 Hz, which has been determined as the optimal frequency rate for this function (Alonso-Prieto et al., 2013; Gentile & Rossion, 2012). In these studies, it was

shown that there is a range of frequencies giving effects, but that using too high frequency rates (e.g., > 10 Hz) does not work because there was not enough time to process each stimulus fully.

In fact, an optimized paradigm relies largely on using optimal stimulation frequency rates for the investigated mechanism. However, finding the “good frequency” for a particular brain mechanism can be one of the main difficulties in the development of a FPVS paradigm. In addition, computer screen limitations must also be taken into account in the choice of the frequency used when developing the experimental paradigm. Indeed, computer screens are limited by refresh frames, and the frequencies used should be an integer multiple of screen refresh rates to display the stimuli properly.

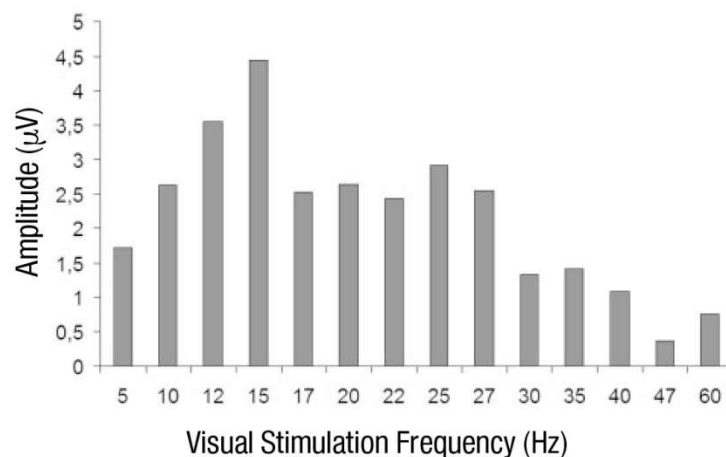


Figure 1.14. EEG responses to flickering stimuli in the frequency domain. Note that the optimal stimulation frequency observed here was 15 Hz. Taken from Pastor et al. (2003).

If the choice of the frequency is a crucial parameter, there are in addition other parameters such as color, luminosity, contrast, size, stimulus type (i.e., light devices, checkerboards, graphical representation) which can modulate the amplitude of the FPVS response (which could be themselves not independent of frequency). Not controlling for these factors may induce a bias in favor of the largest stimulus or the one with the highest contrast for example. Unfortunately, the current state of knowledge in the field does not yet give a comprehensive list of optimal FPVS parameters.

2.4 FPVS applications

For a long time, and even today, the potentials of the FPVS approach have been underestimated and restricted to the study low-level visual processing. For example, Regan (1977) first wrongly considered that the FPVS response was not particularly affected by attention and were therefore useless in studying this mechanism. However, it has now been successfully used to investigate attentional processing (Müller et al., 1998; Müller & Hillyard, 2000; Fuchs et al., 2008), and more recently face processing (Rossion et al., 2012; Dzhelyova & Rossion, 2014; Liu-Shuang et al., 2014), facial expressions (Dzhelyova, Jacques, Rossion, 2015) and lexical processing (Lochy, Belle, Rossion, 2015; Lochy, Van Reybroeck, Rossion, 2016).

In fact, the FPVS approach is more largely used since Silberstein et al. (1990) and Morgan et al. (1996), have first showed that FPVS responses were in fact attention sensitive. Morgan et al. (1996) frequency-tagged two squares, each one at a particular frequency (8.6 Hz and 12 Hz; Figure 1.15A top). Participants were cued to attend to a sequence of stimuli in one visual hemifield and to ignore the opposite hemifield. The stimuli were superimposed on both flickering squares positioned on the two-horizontal sides of the screen. The authors observed an increase in the FPVS response over the occipito-temporal scalp area elicited by the square positioned within the attended visual hemifield (Figure 1.15A bottom). Therefore, the authors demonstrated that FPVS responses are attention sensitive, with the increase of the FPVS response reflecting the increase of the neuronal response to all stimuli within the attentional “spotlight”. In another study, Muller and Hubner (2002) have used two superimposed stimuli (i.e., a small letter was embedded in the center of a larger letter), which were both frequency-tagged at different frequencies (Figure 1.15B top). Participants were cued to attend to one of these stimuli and an enhancement of the FPVS response was observed only for the frequency of the attended letter (Figure 1.15B bottom). The small letter amplitude response strongly increased when it was attended compared to when it was ignored. Therefore, the authors concluded that the spotlight of attention could be shaped as a doughnut. Indeed, if the spotlight of attention had been a unitary beam, the small letter amplitude response should have remained unchanged whether it was attended or ignored (the small letter was embedded in the center of a larger letter).

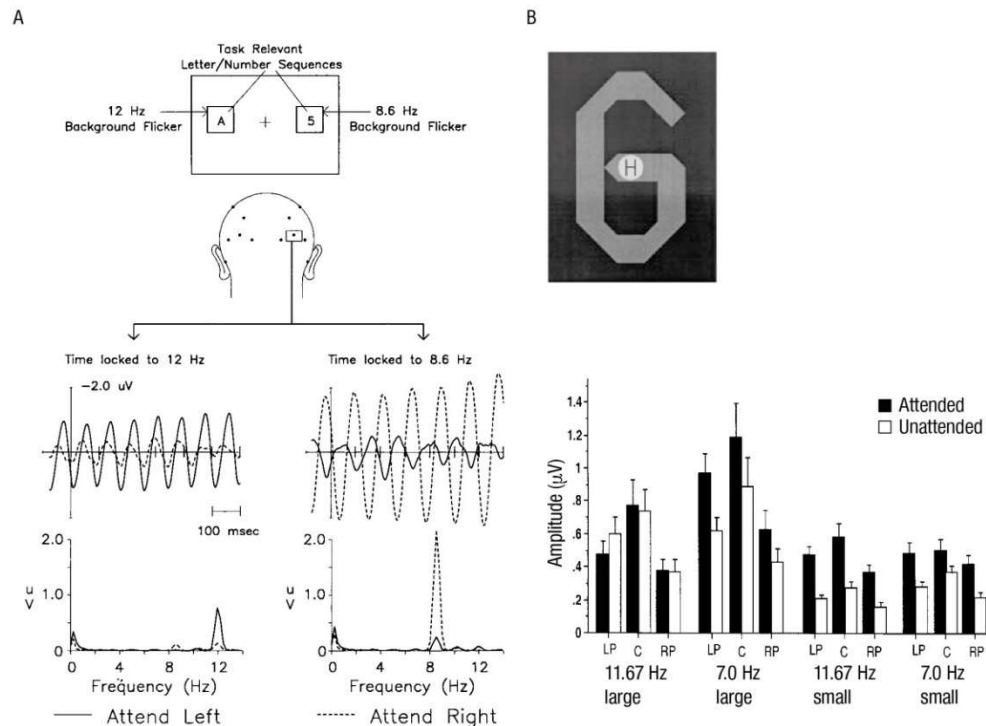


Figure 1.15. (A) Illustration of the stimuli displayed and the EEG response in the time and frequency domains. The neuronal response was maximal for stimuli within the attentional “spotlight”. Taken from Morgan et al. (1996). (B) Stimuli used by Muller and Hubner, (2002). The two letters were frequency tagged at two different frequencies, one was presented at 7 Hz and the other one at 11.67 Hz (and vice versa). The responses in the frequency domain for the different conditions were maximal for the attended letter over three location clusters: LP (left parietal), C (central), and RP (right parietal). Taken from Muller & Hubner, (2002).

More recently, the FPVS approach has been used to study face processing. For example, Liu-Shuang et al. (2014, 2015) have developed an oddball paradigm in which a sequence of different image categories (e.g., object, flower, etc.) were flickering at 5.88 Hz. Inside the sequence, at a fixed interval (and thus periodically and at a certain frequency; e.g., every 4 stimuli which corresponded to $5.88/5 = 1.18$ Hz), a specific category of stimuli was introduced (e.g., faces, a determined face identity; Figure 1.16A). The authors observed in the EEG frequency spectrum large peaks at the fundamental frequency (5.88 Hz) but also at the embedded frequency (1.18 Hz). The specific response at the frequency which corresponds to the periodic presentation of the embedded face frequency (i.e., 1.18 Hz) showed that the FPVS approach can be used to objectively characterize the discrimination

of visual patterns such as faces among objects (individual face discrimination). Furthermore, a similar oddball paradigm was presented to infants and provided evidence for an early selectivity to face category processing (de Heering & Rossion, 2015). Furthermore, Lochy et al. (2015, 2016) successfully used the oddball paradigm in a study examining reading, with evidence of automatic discrimination between words and non-words.

Other studies have also used the FPVS approach to investigate more social aspects of human cognition (Figure 1.16B, C). For example, Wieser & Keil (2014), have investigated facial expressions (fearful, neutral, happy) presented within a visual context (pleasant, neutral, threat) both flickering at a particular frequency (12 vs. 15 Hz; Figure 1.16B). Thus, their paradigm allowed them to study the brain activity engaged in the processing of the faces and the context separately. They showed that arousing background scenes generally drove larger FPVS response amplitudes than neutral scenes. Furthermore, their results showed that background scenes and facial expressions interact. Fearful facial expressions affected the brain processing of the visual background i.e., the FPVS response for the background was higher when viewed with a fearful facial expression. Thus, the authors suggested that fearful faces elicit vigilance for detecting potential threat in the environment. We can find in the literature other FPVS studies investigating facial expressions and how they might attract attentional resources (e.g., Bekhtereva, Craddock, Müller, 2015; Müller, Andersen, Keil, 2008; Keil et al., 2005). This indicates that the perceived emotional content of the stimulus can modulate the FPVS response.

Interestingly, the FPVS method can be sensitive enough to measure inter-individual differences. For example, it has been shown that each individual has his own sensitivity profile for different frequencies (Zhu et al., 2010; Herrmann, 2001). Furthermore, Belmonte (1998) emphasized a link between the FPVS response amplitude and behavioral performance. Indeed, he observed a relationship between the behavioral performances and the amplitude of the FPVS response. Participants who were fast to shift their attention to detect a target also had faster amplitude changes in their FPVS response (see also Andersen & Müller, 2010; Andersen et al., 2009; Stormer, Winther, Li, Andersen, 2013).

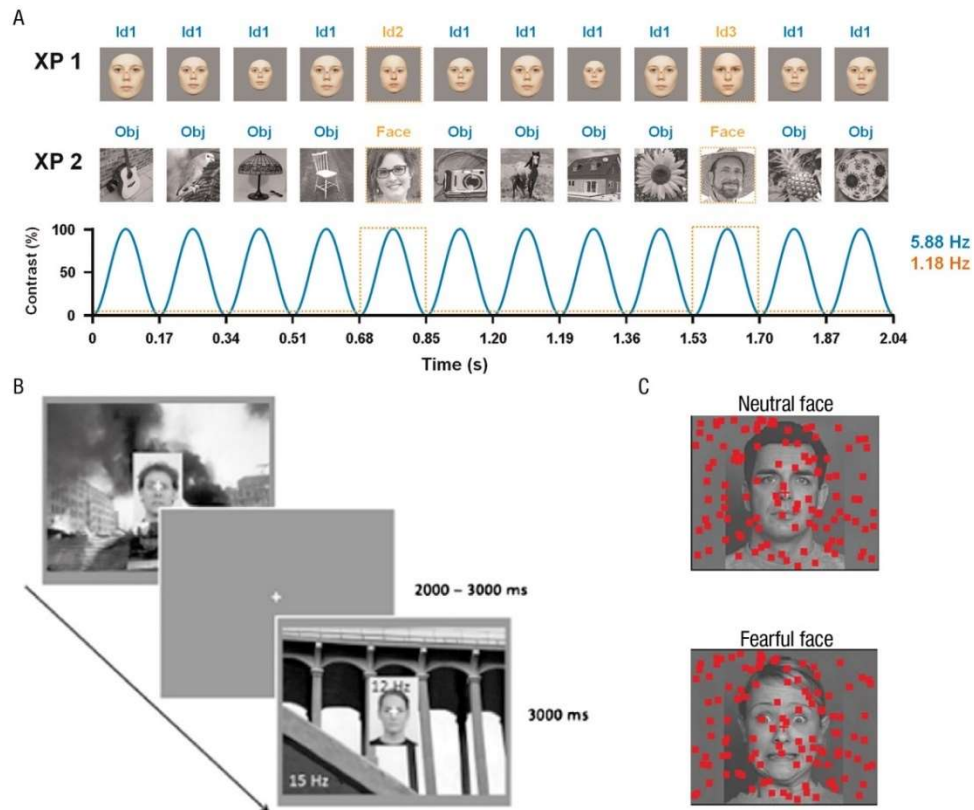


Figure 1.16. (A) Stimuli of the oddball paradigm from Liu-Shuang et al. (2015). Oddball images were presented at regular interval within the sequence of images (B) Illustration of two trials from Wieser and Keil (2014). The face was flickering at 12 Hz and the background at 15 Hz. (C) Example of stimuli used by Bekhtereva et al. (2015). Superimposed red squares flickered at 15 Hz.

Some studies reported more complex responses when at least two different frequencies were used. These additional responses were called intermodulation components (IMs) and were observed at frequencies equal to the combinations of the frequencies used. The response at such frequencies which were not present in the stimulus can be observed because of the non-linear nature of the visual response. The responses of the different frequencies which are present in the stimulus interact which cause the appearance of IMs (Figure 1.17).

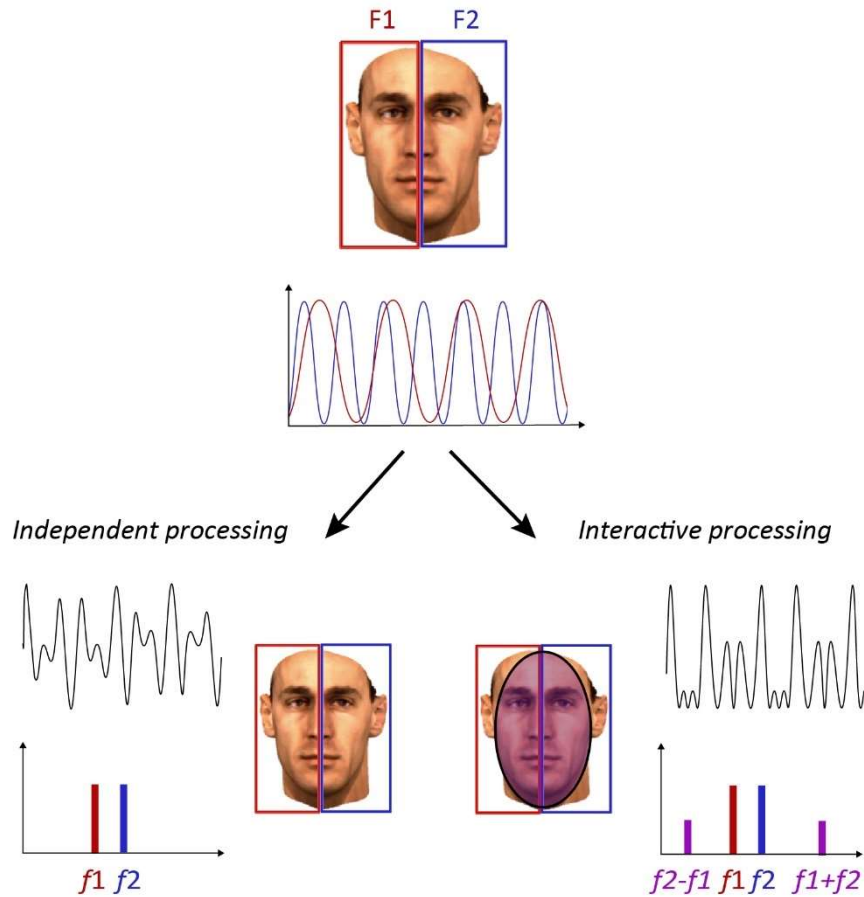


Figure 1.17. Responses to the inputs presented at two different frequencies (F1 & F2). Bottom left: Independent processing: The only observable responses in the frequency domain are the responses to the different frequencies which are present in the stimuli (f_1 and f_2). Bottom right: Interactive processing: Because of the non-linear nature of the visual response, the frequencies of the input are processed interactively which produce additional responses appearing at frequencies equal to the combinations of the frequencies of the input (e.g., $f_1 \pm f_2$). Adapted from Boremanse et al. (2013).

For example, Boremanse et al. (2013) frequency-tagged two halves of a face at two different frequencies (5.88 Hz and 7.14 Hz), and have found large IM responses only when the two halves formed a whole face (peaks were observed at 1.26 Hz = 7.14 Hz – 5.88 Hz and at 13.02 Hz = 7.14 Hz + 5.88 Hz), whereas the IM responses were substantially reduced if there was a gap or if the two halves were misaligned (Figure 1.18) although, the responses at the two fundamental frequencies remained similar between conditions. Such responses

could only emerge from a population of neurons that receive information from the two flickering stimuli and so the IM response represents the integration of the 2 stimuli. The authors concluded that IM responses were a signature of the neural representation of the whole face that was distinct from the neuronal response of the two parts themselves. In addition, in a similar study, Boremanse et al. (2014) found higher IM responses when the two halves of a face belonged to the same person compared to when they corresponded to different identities (see also Gundlach & Müller, 2013; Alp et al., 2016, 2017 for other examples of IM response sensitivity). These observations suggest IMs reflect high-level representations, that is the integration of the different parts of the stimulus.

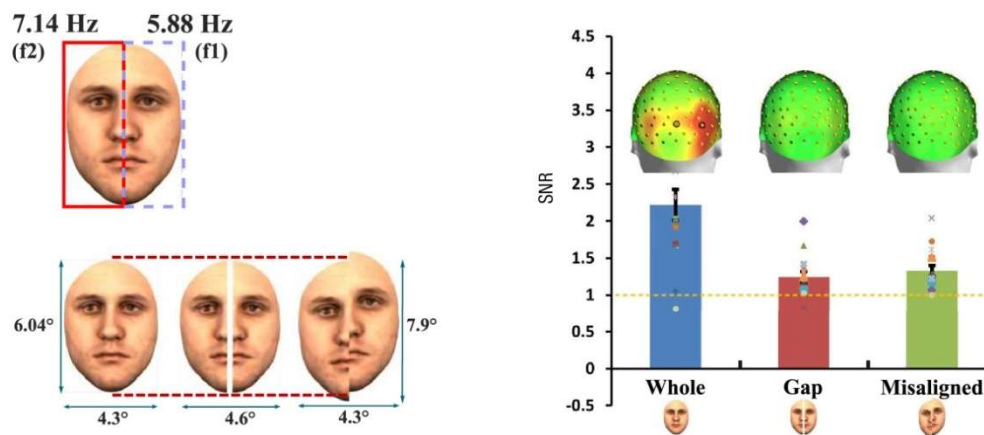


Figure 1.18. Stimuli and conditions from Boremanse et al. (2013) and the quantification of the IM response (SNR) and the related topographical maps per condition. The IM response was maximal for the whole face condition.

III. Overview of the thesis

This thesis has had two main objectives. The first was to extend the FPVS approach to study high-level socio-cognitive functions in the human brain. This would challenge the widespread view that FPVS approach is confined to study “low-level” cognition. The approach has already been successfully adapted to investigate higher-level brain functions such as individual face perception or lexical processing; for a majority of the scientific community, the FPVS approach remains only adapted to study “low-level” brain functions. The second objective was to investigate the cognitive basis of mind-reading, especially, to discriminate whether implicit VPT involves perspective taking mechanisms (mentalizing hypothesis) or only spatial cuing mechanisms (submentalizing hypothesis). Discriminating between these two hypotheses has far reaching implications in understanding the nature of the processes underlying humans’ social interactions.

The aim of the first empirical studies was to explore whether the FPVS approach could be extended to capture VPT processing. Indeed, as for now, many important questions regarding this remarkable VPT ability remain largely unanswered. Does perspective-taking occur spontaneously? What are the underlying processes? Are these processes shared across species? Are these processes available from infancy? Progress in finding answers is limited by the absence of reliable perspective-taking signatures that are transferrable to different populations and easily amenable to experimental manipulations. A first study is reported in appendix 1 which did not lead to strong enough results. However, in Study 1 of the Chapter 2, we report a study which showed that FPVS can provide a reliable signature of explicit VPT. In addition, we also explored if our paradigm can be used to investigate implicit VPT (see Appendix 2).

In Study 2 of Chapter 3, we examined the potential of the FPVS approach to measure implicit VPT processing and discriminate between the mentalizing and the submentalizing hypotheses. More specifically, we measured the specific brain responses elicited while a frequency-tagged object was presented in the vicinity of a person versus an arrow to examine whether this would yield similar brain responses (as predicted by the submentalizing hypothesis) or different brain responses (as predicted by the mentalizing hypothesis).

In Study 3 of Chapter 4, we tested whether we could isolate specific neural responses for the integration of the avatar and the object presented in the scene. If IM responses could be found, we wanted to examine further whether the IM responses were affected by the fact that the avatar can or cannot see the object. In one condition a barrier obstructed the line of sight of the avatar (the object was not visible to the avatar) and in another condition the barrier did not obstruct the avatar's line of sight. According to the submentalizing hypothesis and contrarily to the mentalizing hypothesis, any IM response should remain similar whether the object was seen or not by the avatar.

Finally, in Chapter 5, I summarized and discussed our main findings and their theoretical implications as well as the potential of FPVS in relation to a lot of parameter-optimization work (see also Appendix 3).

Chapter 2

Empirical Study 1

An objective neural signature of rapid perspective taking

Beck, Alexy-Assaf¹, Rossion, Bruno¹ & Samson, Dana¹

¹ Psychological Sciences Research Institute (IPSY), Institute of Neuroscience (IoNS), University of Louvain, Belgium.

<https://doi.org/10.1093/scan/nsx135>

Social Cognitive and Affective Neuroscience

Abstract

The frequency-tagging approach has generally been confined to study low-level sensory processes and always found related activation over the occipital region. Here for the first time, we investigated with it, high-level socio-cognitive functions i.e. the processing of what other people are looking at which is referred to as level 1 visual perspective taking (VPT). Sixteen participants were presented with visual scenes alternating at 2.5 Hz which were depicting a person and an object in a room, while recording electrophysiological brain activity. The person orientation and object position changed at every stimulus but the person in the room always faced the object, except on every fifth stimulus. We found responses in the EEG spectrum exactly at the frequency corresponding to the presentation of the scenes where the person could not see the object, i.e. 0.5 Hz. While the 2.5 Hz stimulation rate response focused on typical medial occipital sites, the specific 0.5 Hz response was found mainly over a centro-parietal region. Besides a robust group effect, these responses were significant and quantifiable for most individual participants. Overall, these observations reveal a clear measure of level 1-VPT representation, highlighting the potential of EEG frequency-tagging to capture high-level socio-cognitive functions in the brain.

1. Introduction

Consider the visual scene displays shown in Figure 2.1A: which ones are the odd ones out? Answering this question requires considering not only the position of the person in the scene and the left/right location of the object in the scene, but the specific *relationship* between the two: in three of the visual scenes only, the person looks away from the object. The present study uses an original electrophysiological frequency-tagging approach to provide neural evidence for the rapid representation of this relationship between another person and what he or she is looking at, i.e. a basic visual perspective-taking (VPT) ability, without direct contamination by overt responses.

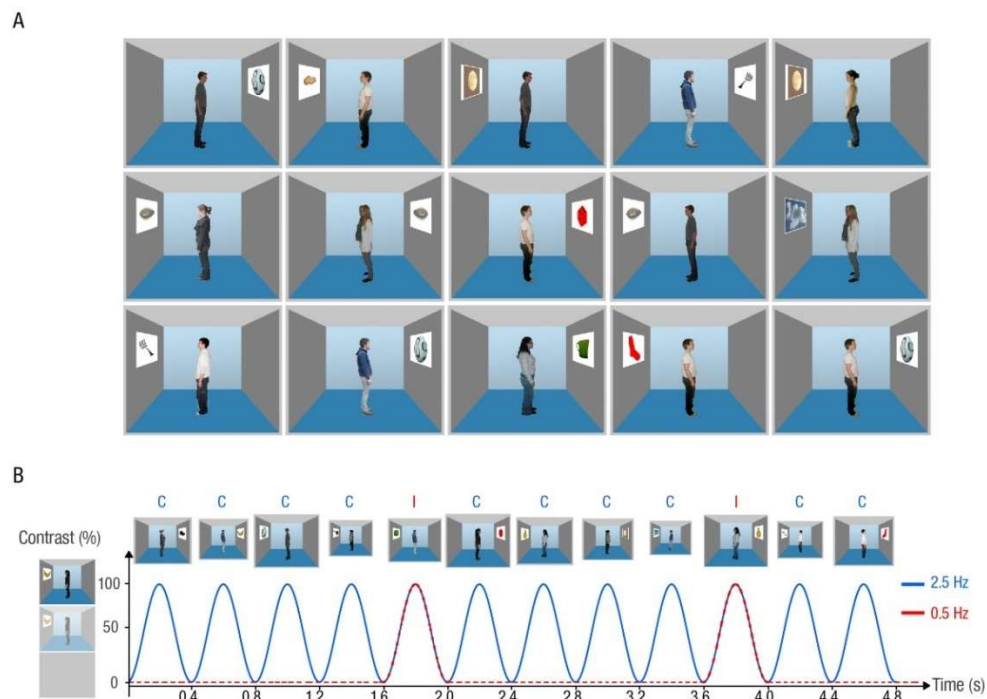


Figure 2.1. Stimuli and paradigm. **(A)** Examples of stimuli used in the experiment. **(B)** Experimental paradigm. Different stimuli appear at each stimulation cycle (no repetition within a stimulation sequence), with a randomly changed size to prevent pixel-based adaptation. Stimuli are presented by sinusoidal contrast modulation at the rate of 2.5 Hz (i.e., stimuli reached full contrast after 0.2 s). The embedded periodic response of interest was 0.5 Hz. Note that the side of the object location and the direction faced by the person could vary randomly at every cycle, and only the specific relationship between the two changed at 0.5 Hz. A full stimulation sequence lasted 94 s (84 s stimulation plus 5 s of fade in and fade out; see the movie S1 in supplementary material).

Considering another person's viewpoint plays a fundamental role in our everyday life social interactions. Amongst the various forms of perspective-taking we can engage in, processing what the other person is looking at is particularly important as it provides valuable information to make more complex social inferences, such as inferring what the other person likes, intends to do, or talks about, for example. This basic VPT ability is usually referred to as level 1 VPT (contrary to level 2 VPT which consists in establishing "how" a stimulus is perceived by someone else (Flavell et al., 1981)). Level 1 VPT is one of the earliest developing form of explicit perspective taking in children (Sodian et al., 2007) and it is a perspective-taking ability that higher primates such as chimpanzees might share with humans (Hare et al., 2000).

Human adults easily represent what another person can see when explicitly instructed to do so. Under certain circumstances, they also do it involuntarily. For example, Samson and collaborators (2010) asked participants to count how many dots they could see in a visual scene. A human avatar similar to those depicted in Figure 2.1 was present in the scene and could either see the same amount of dots visible to participants or could only see a subset of the dots. Participants were slowed down and more error prone in judging their own visual perspective when the avatar saw a different amount of dots than themselves (Baker et al., 2015; Furlanetto et al., 2016; Todd et al., 2017). This observation indicates that participants involuntarily represented what the avatar was seeing. However, in this original paradigm, the signature of level 1 VPT is concomitant to the overt motor response. Isolating a neural marker of VPT that is independent from any motor response would not only be important for understanding better the information processing stages when making unconscious or conscious decisions about what others see, but may open new perspectives for measuring Level 1 VPT in populations with immature or deficient decisional processes such as young children, individuals with cognitive deficits, or non-human species without extensive training. The present study addresses this issue by developing a new marker of perspective taking. We measured Level 1 VPT with scalp electrophysiology, taking advantage of a technique called "frequency-tagging" or "Fast Periodic Visual Stimulation" (FPVS), in which the process of interest is captured objectively (i.e., at the experimentally defined frequency) in the electroencephalography (EEG) frequency domain by Fourier Transform. Since its original demonstration by Adrian & Matthews (1934), this approach

has essentially been confined to study low-level sensory processes ("Steady State Visual Evoked Potentials", SSVEPs, Regan, 1966), including their modulation by selective and spatial attention (e.g., Keil et al., 2005; Morgan et al., 1996; Muller & Hubner, 2002). However, in recent years, changing high level visual properties at a periodic frequency rate and adapting this rate to the time-constant of the process of interest has shown that this highly sensitive and objective approach can shed light on higher level brain functions, such as individual face perception (Rossion & Boremanse, 2011; Liu-Shuang et al., 2014) or lexical processing (Lochy et al., 2015). Here we extend this FPVS-EEG paradigm a step further to capture visual perspective-taking (VPT) processes in the human brain.

More specifically, we designed a paradigm based on Figure 2.1's display, in which we successively, rapidly (2.5 Hz) and continuously (i.e., for 84 seconds without interruption) presented visual scenes where the person faced an object. Every fifth stimulus, the person could not see the object. Participants were unaware of the ratio and periodic rate of the two types of visual scenes. We hypothesized to observe a differential EEG signal depending on whether the avatar could or could not see the object, i.e., a signal in the EEG spectrum corresponding exactly to the specific frequency at which the stimulus depicting a person who could not see the object was presented (i.e. 0.5 Hz - 2.5 Hz/5). Furthermore, until now the scalp topography observed in studies using FVPS found activation on the occipital scalp region. However here, since previous studies highlighted the implication of parietal and frontal areas during VPT (e.g., Schurz et al., 2015; McCleery et al., 2011), we expected to find a more anterior neural response than the one usually found with FPVS. Such unique scalp activation would support the great interest of extending the use of the FPVS technique to investigate high-level cognitive functions.

2. Materials and methods

Participants

Sixteen healthy volunteers (female: 14, mean age: 21 ± 1.67 , all right-handed) took part in the experiment in exchange of a small honorarium. The number of participants was decided on the basis of the number of participants recruited in the original study by Samson et al. (2010) and on the basis of a previous EEG study investigating explicit visual perspective

taking (McCleery et al., 2011). All participants gave written informed consent prior to the experiment and the experiment received the approval from the ethics committee of the Psychological Sciences Research Institute. All experiments were performed in accordance with relevant guidelines and regulations.

Stimuli

The stimuli were created with Adobe Photoshop CS6. Sixteen photographs of the profile view of student volunteers (8 females) were taken, which all gave informed consent for publication of identifying images. The picture of the person was cropped and placed in the middle of a room, with the left, back, and right walls visible. The vanishing point of the room was localized at the level of the person's head. Half of the stimuli depicted the person looking at the left wall and the other half of the stimuli depicted the person looking at the right wall (Figure 2.1A).

Twenty objects with high familiarity and low visual complexity ratings were selected from the Moreno-Martinez & Montotro (2012) data base and were displayed on either the left or right side wall. Two different sets of stimuli were made. For the first set, the objects were always placed in front of the person either on the right or the left wall (Figure 2.1A). For those stimuli, the person and the participant saw the object (i.e., *consistent perspective stimuli*). For the second set of stimuli, the objects were always placed on the wall behind the person. Here, the person and the participant had a different visual experience because only the participant could see the object (i.e., *inconsistent perspective stimuli*). Each person and each object were presented a similar amount of time. The color of two objects (the dice and the sock) was changed to red. These two objects were the only red objects, (the other colors of the objects were yellow, green, black, gray, etc.).

Without any object and person displayed in the room, the luminance of the right and left walls was 127 cd/m², the luminance of the background wall was 199.55 cd/m² and the luminance of the floor was 117.37 cd/m². The size of the room was approximatively 16.1° horizontal by 12.9° vertical, all pictures were displayed on a light gray background (198/255, 198/255, 198/255), resulting in 641 x 480 pixel images (307680 pixels). Note, however, that every displayed stimulus had a randomly varied image size (min: 227683 pixels and max:

369216 pixels) (see Figure S2.1 of the Supplemental Material for additional information about the size of the image).

Procedure

We selected as frequency of picture presentation 2.5 Hz through a pilot experiment (see Appendix 3). This means that every 400 ms a new picture was presented. Note that this presentation rate matches the finding of McCleery et al. (2011) who found an event-related potential (ERP) peaking around 450 ms that reflected perspective taking processes. 2.5 Hz is a slower frequency rate than generally used in previous studies. However, since we investigated high-level cognitive function, we assumed that the stimuli needed a longer presentation to be processed. A stimulation sequence (or trial) consisted of a series of 235 pictures (total trial duration: 94 s). In the periodic stimulation condition, a trial consisted of a sequence composed of regular trains of 4 consistent perspective stimuli (C) followed by an inconsistent perspective stimulus (I) (i.e., CCCCICCCICCCI...). This second frequency rate (0.5 Hz) was based on the successful use of this 1/5 ratio in previous studies (e.g., Liu-Shuang et al., 2014; see Retter & Rossion, 2016 for manipulation of ratios between oddball stimuli). A higher rate of presentation of inconsistent stimuli may have made the periodicity more obvious to participants. A lower rate of presentation, on the other hand, would have resulted in an even smaller frequency of interest responses (i.e., below 0.5 Hz) and it is known that the lowest part of the EEG spectrum is characterized by a higher level of noise. Hence, every fifth stimuli (i.e., 0.5 Hz), there was a change in perspective consistency. Critically, the side of the object (Right or Left, R or L) – and thus of the body orientation – was fully randomized, so that the inconsistent-consistent relationship concerned only the *relationship* between the person and the object (object seen versus unseen) (e.g., LC-RC-RC-LC-LI- RC-RC-RC-LC-RI- RC, etc.; Figure 2.1). In the non-periodic stimulation condition, a trial consisted of a sequence composed of the exact same stimuli but shown in an entirely random order. In total, the same amount of consistent/inconsistent perspective stimuli were shown as in the periodic sequence (188 consistent perspective stimuli, 47 inconsistent perspective stimuli per trial). This control and non-periodic sequence was used to ensure that any response of interest (0.5 Hz) was above noise level only due to the periodicity of the perspective change.

There were 8 different person's identity (4 females) and 10 different objects plus the two red objects (sock and dice) within one stimulation sequence. Furthermore, there were always the same amount of images with a male and a female person and the same amount of persons looking at the left or to the right wall equally spread across the two sexes.

During EEG recording, participants were seated in a light- and sound-attenuated room, at a viewing distance of 80 cm from an LED monitor (BenQ XL2420T) with a 1920 x 1080 resolution and a 120 Hz refresh rate. Stimuli were periodically presented through sinusoidal contrast modulation (Rossion & Boremanse, 2011) at a base frequency of 2.5 Hz (Figure 2.1).

Each stimulation sequence started with the onset of a fixation cross. After a random time period of 2 to 5 s, the stimulation gradually (i.e., linearly) appeared and reached full contrast after 5 s (contrast increase; fade-in 5 s). After 84 s, the scene gradually disappeared (contrast decrease; fade-out 5 s). The fade-in and fade-out periods (e.g., Liu-Shuang et al., 2014) were used to prevent ocular artefacts and were not included in the analyses. In addition, to minimize low-level adaptation effects, every newly displayed stimulus had a randomly varied image size (min: 227683 pixels and max: 369216 pixels). The experiment started with one trial of periodic sequence and one trial of non-periodic sequence (presented in counterbalanced order across participants) to practice the task before the EEG recording. Then, participants performed 16 test trials (8 periodic trials). Half of the participants started with a periodic trial and the other half started with a non-periodic trial, and then we alternated trial by trial between these two conditions.

Participants received written instructions prior to the EEG experiment. They had to maintain fixation on the cross placed in the middle of the screen and had to respond by pressing the space bar as quickly and accurately as possible. Their task was to detect the person viewing the red dice (12 targets per trial). They did not have to respond if the person saw any other object (including the red sock) or if the person could not see the red dice (because it was in his/her back). Thus, accurate performance in this task required detecting a specific visual experience of the person and the goal of this behavioral task was to ensure that participants maintained their attention to both stimuli throughout the trials. However, while participants had to monitor the other person's visual perspective to do the task,

crucially, this task was not to respond whether the perspective was consistent or not. In both conditions, (periodic and non-periodic) stimuli depicting the person viewing the red dice were randomly inserted in the image sequences. Thus, correct behavioral responses (detecting the person viewing the red dice) were not periodic. Thereby, the EEG measure of interest was not concomitant with the motor response to the task instruction. A response was considered as correct if it was given within three standard deviations of the participant's mean response time.

EEG acquisition

EEG data were recorded using a 128 scalps channel at a sampling rate of 512 Hz with an ActiveTwoBiosemi electrode system (Biosemi, Amsterdam, Netherlands). Two additional electrodes, CMS (Common Mode Sense) and DRL (Driven Right Leg), were used as reference and ground. Each individual's electrode direct current (DC) offset were adjusted below ± 50 μ V prior to recording by injecting the electrode with a conductive gel (Signa). Eye-movements were recorded by four additional electro-oculograms around the subjects' eyes. These channels were used for monitoring vertical and horizontal eye movements to identify trials contaminated by blinks.

EEG pre-processing

Off-line analyses were performed with Letswave 5 (<http://nocions.webnode.com/letswave>) and Matlab 2012 (Math-Works, Natick, MA), similarly to previous FPVS studies that used this type of paradigm (e.g., Liu-Shuang et al., 2014)). Individual EEG data were first bandpass filtered between 0.1 and 120 Hz (zero phase shift Butterworth filter, order 4). Then, we reduced the sampling rate down to 256 Hz to save space and processing time. Eyes blinks were identified based on the recording of the electro-oculograms. Noisy channels containing other artefacts than blinks or excessive noise (larger deflections than 200 μ V) were interpolated by the nearest neighboring electrodes (a maximum of 5 % of channels were recreated per participant). In addition, one subject had only two ocular electrodes instead of four due to a technical problem during the setting up. EEG data were subsequently segmented for each trial, 2 s prior to the start of the trial plus 2 s after the end (-2 s to 96 s). Since the approach provides a response of

interest confined to a small frequency bin, whereas the noise is stochastically distributed across broad frequencies, the approach is highly resistant to artifacts. Thus, in the present study, there was no need to reject EEG trials. Afterwards, a common average reference computation was applied to all channels excluding ocular channels for each subject independently.

Frequency-domain analyses

Frequency domain analyses were performed to quantify the magnitude of the response of the overall FPVS responses in each stimulation condition (periodic and non-periodic). For that, each epoch was segmented a second time, right after the end of the fade-in to the end of the trial just before the beginning of the fade-out (21504 time bins in total = 84 s). Then, the resulting segments were averaged separately for each participant and each periodic/non-periodic trial. In addition, we also computed the grand average across participants and conditions. A Fast Fourier Transformation (FFT) was applied on the averaged segments to extract from the EEG signal, the amplitude spectra for each electrode. Thanks to the long time windows (84 s), the frequency analysis yielded spectra with a high frequency resolution ($1/84$ s, i.e., 0.0119 Hz) and allowed the unambiguous identification of the response at the exact frequencies of interest (i.e., 2.5 Hz for the base stimulation rate and 0.5 Hz for the perspective consistency discrimination response).

In order to correct for noise level for each participant's spectrum, at each frequency, we either subtracted (for baseline-corrected amplitude used for statistical analysis) or divided (for signal-to-noise ratio (SNR) used for the data visualization), the average voltage amplitude of the 10 surrounding bins (5 on each side, excluding the 2 immediately adjacent bins and the two extreme bins; e.g., Dzhelyova & Rossion, 2014; Rossion et al., 2012). Here we used only 10 surrounding bins instead of 20 in previous studies since our periodic response of interest at 0.5 Hz is surrounded by frequency bins in the lowest part of the EEG spectrum, characterized by a high level of noise. Z-scores were also calculated to assess the significance of the response at specific frequencies (e.g., 0.5 Hz, 1 Hz, 1.5 Hz etc.) by computing the difference between amplitude at the frequencies of interest and the mean amplitude of the 10 surrounding bins (excluding the 2 immediately adjacent bins and the

two extremes bins) and dividing this difference by the standard deviation of the 10 surrounding bins (Rossion et al., 2012). For the group analyses, individual baseline corrected amplitude and SNR spectra were averaged separately for each periodic/non-periodic trial per condition.

For further analysis, we determined a range of relevant harmonics (based on significant Z-scores) for each frequency based on the group-level data. Different regions-of-interest (ROIs) were also determined based on electrodes that showed a particularly large SNR on the average between periodic and non-periodic trials.

3. Results

3.1 Behavioral Data

All participants successfully performed the task (mean = 89% $\pm$ 3.54; false alarm = 14.08% $\pm$ 6.77; response time (RT) = 0.60 s $\pm$ 0.044). There was no significant difference of RT (periodic: 0.60 s $\pm$ 0.046; non-periodic: 0.60 s $\pm$ 0.043; $t(15) = -0.09$, $p = 0.926$, Cohen's $d = -0.024$) and accuracy (periodic: 88.77 % $\pm$ 4.80; non-periodic: 89.17 % $\pm$ 4.18; $t(15) = -0.29$, $p = 0.773$, Cohen's $d = -0.073$) between the periodic and non-periodic trials, suggesting that the level of difficulty did not differ across the periodic and non-periodic stimulation conditions.

3.2 EEG Data

Discrimination between consistent and inconsistent perspectives

Inconsistent perspective stimuli correspond to images where the object was placed behind the person, so that the person could not see that object. In the periodic stimulation condition, these images were presented every fifth stimuli. Hence, neural discrimination between consistent and inconsistent perspectives should lead to a response at the exact periodic frequency of interest, which corresponds to 0.5 Hz (2.5 Hz/5) and to the harmonics in the EEG spectrum.

In order to determine the number of harmonics to take into account for further analysis, we pooled the 128 electrodes for each participant. We then computed the Z-score

(computed as described above) for each harmonic on the grand averaged data across stimulation conditions (periodic and non-periodic trials together). The highest significant harmonic was found at 3 Hz (6th harmonic; threshold of significance placed at a Z-score of 1.65; $p < 0.05$, one-tailed, signal > noise).

We then cropped the grand average FFT spectrum, centered at the periodic frequency of interest (0.5 Hz, 1 Hz, 1.5 Hz, 2 Hz, 3 Hz) with the 10 surrounded neighboring bins on each side. We excluded the fifth harmonic (2.5 Hz) because it corresponds to the base frequency rate. We summed the spectrum of the harmonics and their neighboring bins (Retter & Rossion, 2016), and computed the SNR (Figure 2.2A).

Based on the topographical map of the grand average of the combined (i.e., unbiased) periodic and non-periodic conditions (Figure 2.2A), we selected two groups of electrodes which show the maximal summed-harmonic response at the periodic frequency of interest. These groups of electrodes were used as ROIs for further analysis. The main ROI was defined by 15 contiguous channels in the centro-parietal region (FCC1, FCC2, FCC1h, FCC2h, FCz, C1, C1h, C2, C2h, CCP1h, CCP2h, Cz, CCPz, CPz, CPPz), which were associated with the highest SNR on grand averaged data. Another ROI was defined by 3 channels in dorsolateral prefrontal region (AF8, F8, FT8) associated with the highest SNR outside of the centro-parietal region.

As hypothesized, EEG spectra (Figure 2.2B) showed large and significant responses (based on the Z-score) at the frequency of interest and at its harmonics for each ROI, only in the periodic condition (see Table S2.1 in the Supplemental Material). Note that the results of the quantification analysis were independent of the number of harmonics included in the analysis (Figure S2.2 of the Supplemental Material).

On one hand, we summed the spectrum of the harmonics and their neighboring bins and computed the SNR for the periodic and non-periodic trials over the ROIs (see Figure S2.3 of the Supplemental Material). The resulting Z-score were significant in the periodic trials only (one-tailed Z-score threshold set at 1.65, $p < 0.05$; Frontal ROI: $Z = 7.59$ versus $Z = -2.54$ and Centro-parietal ROI: $Z = 11.03$ versus $Z = 0.03$ for the periodic and non-periodic trials, respectively).

On the other hand, we computed the baseline corrected amplitude then summed the harmonics, and then we compared with one-sample t-test against 0 (i.e noise level). T-test showed a significant response only for the periodic condition (Frontal ROI: $t(15) = 5.35$, $p = 0.001$, Cohen's $d = 1.34$ and $t(15) = -1.63$, $p = 0.125$, Cohen's $d = -0.41$, Centro-parietal ROI: $t(15) = 4.63$, $p = 0.000$, Cohen's $d = 1.16$ and ROI: $t(15) = 0.27$, $p = 0.792$, Cohen's $d = 0.07$ for the periodic and non-periodic trials, respectively). In summary, these results reveal an EEG signature of the detection of the change of perspective consistency. Note however that there was no correlation between the behavioral response time and the baseline corrected amplitude of the summed harmonics of the frequency of interest for the 2 ROIs (Frontal ROI: $r = 0.115$, $p = 0.672$, Centro-parietal ROI: $r = 0.125$, $p = 0.644$).

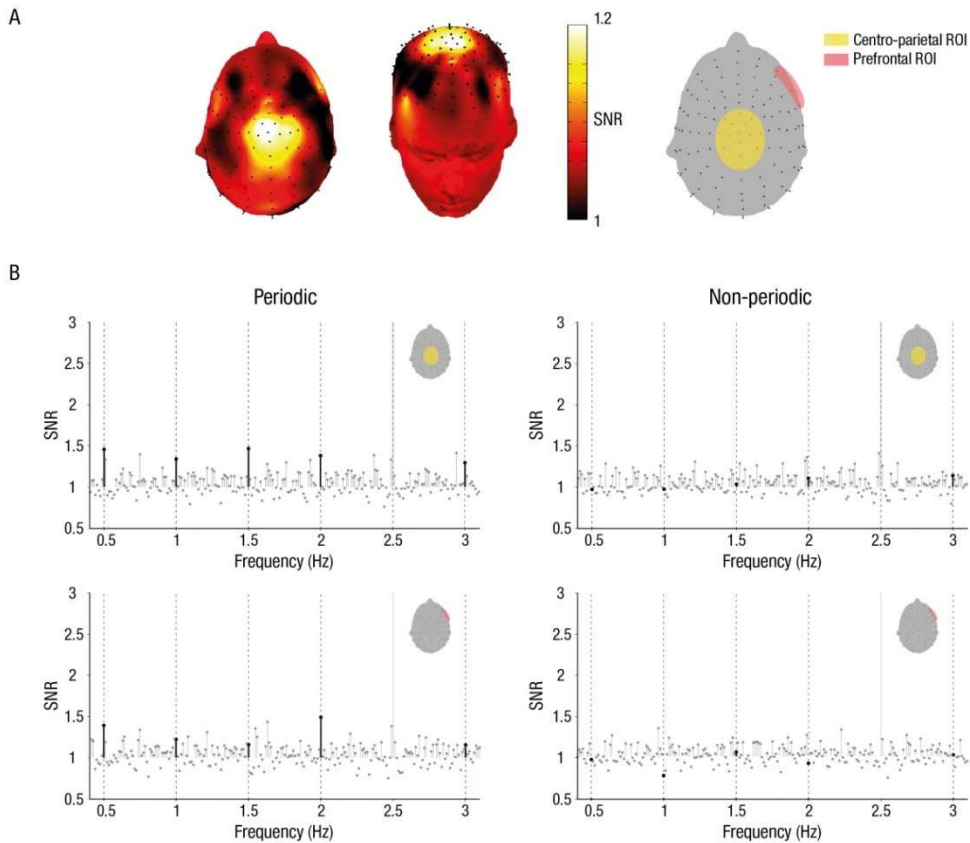


Figure 2.2. Responses at the frequencies of interest. (A) Scalp topography (SNR) of the grand average of the periodic and non-periodic trials of the sum of the 5th first harmonic (excluding 2.5 Hz) with the selected ROIs for further analysis. (B) SNR spectrum over the centro-parietal ROI and the frontal ROI of the periodic trials (left) and the non-periodic trials (right).

Individual analysis

To evaluate the sensitivity of our paradigm, we tested whether each individual participant showed a discrimination response between the consistent and inconsistent perspective. To do so, for each participant and over each ROI, we cropped the FFT spectrum, centered at the periodic frequency of interest (0.5 Hz, 1 Hz, 1.5 Hz, 2 Hz, 3 Hz) with the 10 surrounded neighboring bins on each side. We summed the spectrum of the harmonics and their neighboring bins and computed the Z-scores and the SNR. The resulting Z-scores were significant in at least 1 of the 2 ROI for 14 participants out of 16 (see Table S2.2 in the Supplemental Material). Figure 2.3 shows an illustration of the individual results.

Base frequency analysis

As expected, we also found a response to the base frequency rate (i.e., the rate at which a new image was presented), merely reflecting the synchronization of the visual system to the visual stimulation. This response to the base frequency rate focused on medial occipital electrodes for both the periodic and non-periodic trials (Figure S2.4 in the Supplementary material).

There was no significant difference in the baseline corrected amplitude across the two stimulation conditions (periodic and non-periodic; Figure S2.4 in Supplementary material). Since frequency-tagged EEG responses are highly sensitive to spatial and selective attention (Andersen et al., 2008; Morgan et al., 1996; Müller & Hillyard, 2000; Walter et al., 2012) the absence of difference coupled with the equal behavioral performance suggest that participants paid equally attention to the stimuli in the two stimulation conditions.

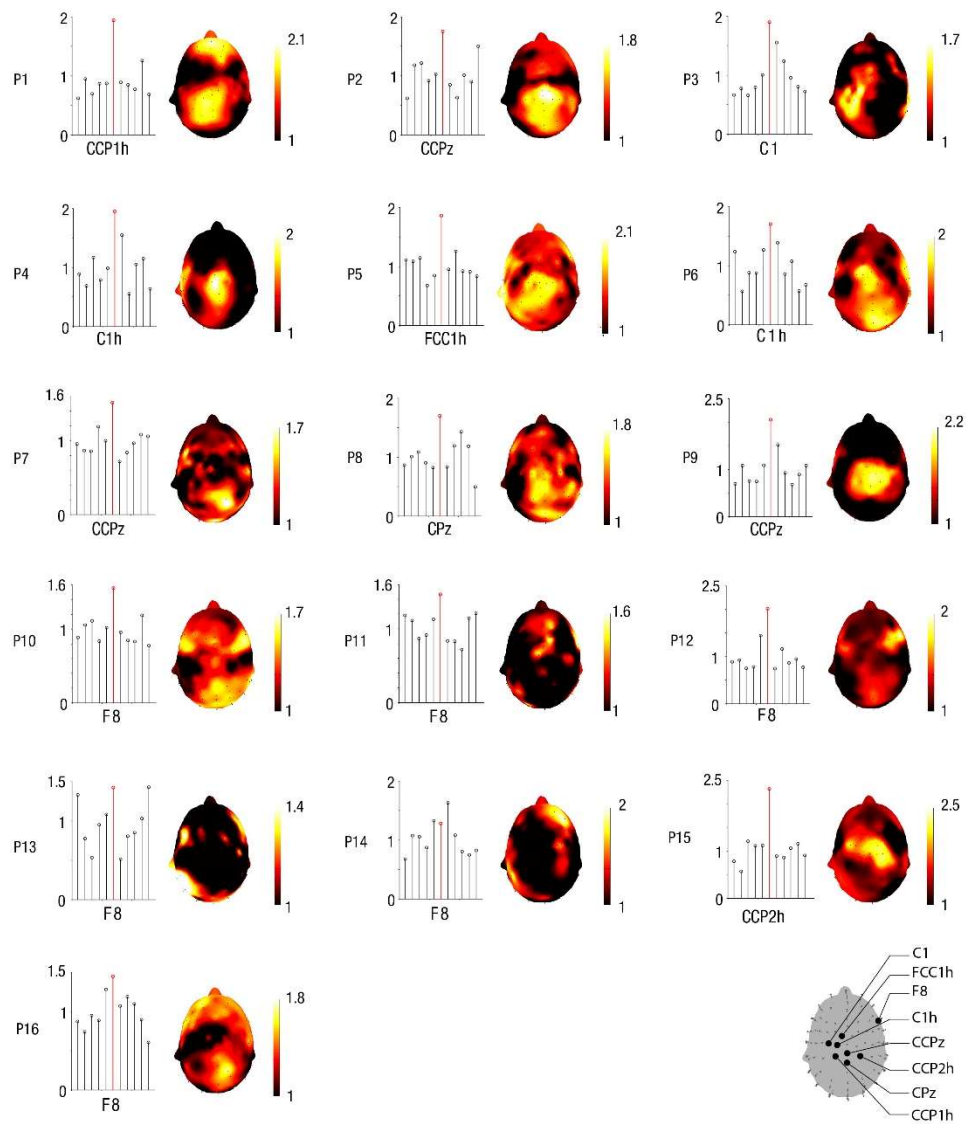


Figure 2.3. Illustration of individual participants' SNR response to changes of perspective consistency over one electrode belonging to one ROI. The spectrum is centered at the response of the summed of the 5 significant harmonics, surrounded by the summed of the neighboring bins, which indicate noise levels (for visibility only the 5 neighboring bins on each side are shown).

4. Discussion

We found a neural signature at the exact frequency at which participants could witness the perspective of another person changing from being consistent with their perspective (the other person saw the object on the wall that participants could also see) to becoming inconsistent (the other person could not see the object that participants could see). This EEG response was clear (i.e., well above noise level) and objective (i.e., occurring exactly at a frequency defined by the experimental design), and was significant in virtually all individual participants tested in the study. This process was measured at a relatively fast rate (i.e., every 2 seconds, with 400ms/image) and, crucially, was not contaminated by a motor response. Since participants were explicitly asked to judge the avatar visual perspective throughout the experiment, the EEG measure reflects explicit perspective taking processes.

Besides a robust group effect, the responses were significant and easily quantifiable for most individual participants. The strength of the FPVS response may however vary according to factors (such as personality traits or clinical features) known to affect interindividual variability in perspective taking. This is an avenue for further investigation as FPVS may provide a more ecological and sensitive measure to capture such variability.

Virtually all EEG responses recorded so far in this type of fast periodic “oddball” paradigm were localized over occipital and ventral occipito-temporal brain regions, reflecting the differential process of objects, faces, letters or words (e.g., Dzhelyova et al., 2015; Liu-Shuang et al., 2014; Lochy et al., 2015). In contrast, the VPT neural response was found here over central and right prefrontal electrode sites. These scalp localizations are fully compatible with the known neurofunctional basis of VPT processing. The central topography may be due to processes generated in temporo-parietal brain areas, involved in the representation of the other person’s perspective (Aichhorn et al., 2006; McCleery et al., 2011; Van Overwalle, 2009). Previous studies have also pointed to a role of the right prefrontal cortex in inhibiting our own point of view to allow the selection of the other person’s perspective when both perspectives are in conflict (Hartwright et al., 2015; McCleery et al., 2011; Samson et al., 2005; Shibata & Inui, 2011; Vogeley et al., 2001). However, a strong word of caution is required here as EEG alone is not suitable to make

strong brain localization claims. Nevertheless, it is particularly interesting that the extension of the FPVS-EEG approach to higher-level cognitive processes reveals relatively anteriorly located responses on the scalp.

The current results do not inform about the nature of the information processed, and more specifically at which stage the information processed conforms to someone's mental state (e.g., the person can or cannot see the object). It is possible that at early stages of processing, only the spatial link between the other person and the object is computed and triggers a shift of attention to the object in front of the person. Future studies could address this issue with the FPVS-EEG approach by using obstacles in the other person's line of sight, similarly to what has been done in previous behavioral studies (Baker et al., 2015; Cole et al., 2016; Furlanetto et al., 2016).

More generally, our study opens new ways to explore the neural basis of perspective taking in social cognition (Frith & Frith, 2007). For example, whether implicit and explicit perspective taking share the same type of processes is a highly debated question (Apperly & Butterfill, 2009; Heyes, 2014; Grosse Wiesmann et al., 2016; Biervoye et al., 2016; Bardi et al., 2016). FPVS responses could be compared in the presence *versus* absence of explicit perspective taking instructions (without the need to change the nature of the stimuli) to provide further insights into this debate: will there be differences in terms of the strength of the EEG signal or is the topography on the scalp different? Furthermore, should future results show the presence of a FPVS response even in the absence of explicit instructions, this could provide a powerful measure to investigate perspective taking in infants (see for example FPVS responses to faces in infants, de Heering & Rossion, 2015), non-humans, perhaps even allow cross-ages or cross-species comparisons. Indeed, the strength of the technique relies in the fact that it only requires watching a train of images.

5. Acknowledgements

We thank Andrea Conte for having created the application used to test the subjects. We also thank Talia Retter and Corentin Jacques for helping with the data analyses.

6. Supplementary material

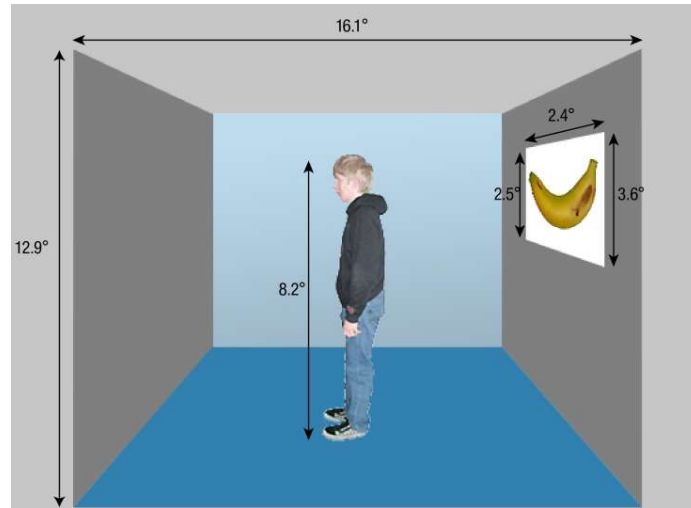


Figure S2.1. Original size of the images in visual angle at a viewing distance of 80 cm. Note that each displayed stimulus had a randomly varied image size.

Table S2.1. Grand average response for the periodic and non-periodic trials for different harmonics of the periodic response of interest. Data are reported for the 2 ROIs. *p*-Values were calculated on the basis of Z-scores (one-tailed Z-score threshold set at 1.65, $p < 0.05$).

Frequency (Hz)	Periodic Z-score		Non-periodic Z-score	
	Centro-parietal ROI	Prefrontal ROI	Centro-parietal ROI	Prefrontal ROI
1F = 0.5	5,21***	13,21***	-0,51	-0,88
2F = 1	5,06***	0,88	-0,83	-4,69
3F = 1.5	6,16***	1,60	-0,14	0,28
4F = 2	9,96***	7,07***	1,58	-1,16
6F = 3	3,94***	1,79*	1,53	0,54

Note: * $p < .05$. ** $p < .01$. *** $p < .001$.

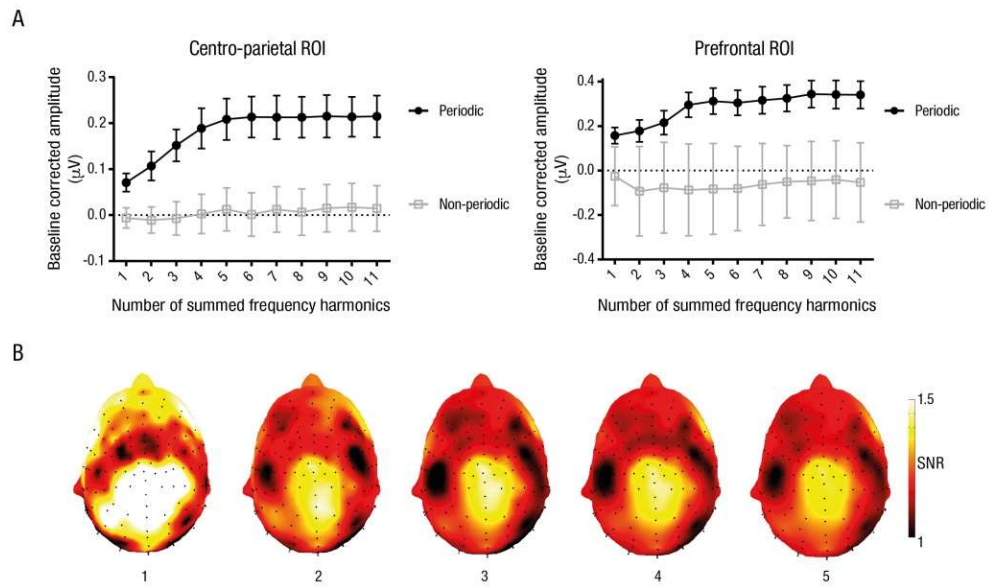


Figure S2.2. The pattern of response for the VPT inconsistency detection is independent from the number of harmonics included in the quantification of the EEG response. **(A)** Baseline corrected amplitude response as a function of the number of summed harmonics. **(B)** Topographical maps (SNR) of the periodic condition as a function of the number of summed harmonics.

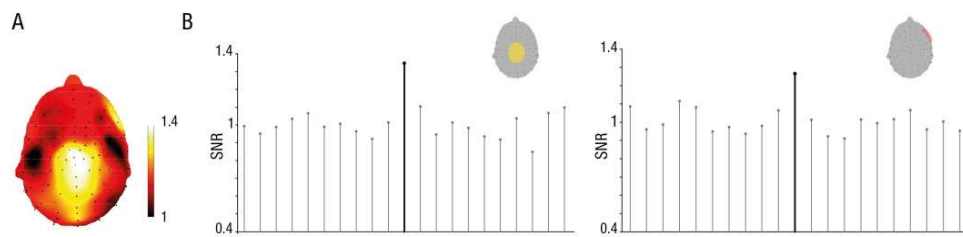


Figure S2.3. Sum of harmonics of the periodic frequency of interest for periodic trials. **(A)** Scalp topography (SNR). **(B)** Grand average SNR over the 2 ROIs (centro-parietal ROI on the left panel and dorsolateral prefrontal ROI on the right panel). The spectrums are centered at the response of the sum of the 5 harmonics, surrounded by the summed of the neighboring bins, indicating noise levels.

Table S2.2. Individual participants' Z-score response computed on the sum of the harmonics of the periodic response of interest over the two ROI independently. *p*-Values were calculated on the basis of Z-scores (one-tailed Z-score threshold placed at 1.65, $p < 0.05$).

Participant	Z-score	
	Centro-parietal ROI	Prefrontal ROI
P1	8,08***	2,42**
P2	3,02**	2,11*
P3	1,13	0,77
P4	4,11***	-1,24
P5	7,41***	7,31***
P6	4,53***	2,4**
P7	0,47	0,74
P8	4,89***	0,29
P9	7,67***	0,18
P10	4,11***	2,83**
P11	0.85	4,62***
P12	1,13	3,34***
P13	-0,71	3,24***
P14	-1,83	4,83***
P15	10,95***	2,18*
P16	1,74*	7,35***

Note: * $p < .05$. ** $p < .01$. *** $p < .001$.

Base frequency analysis

A response at the base frequency rate merely reflects the synchronization of the visual system to the visual stimulation. Thus, the brain response to the base frequency rate is expected to be very similar in the periodic and the non-periodic trials. To illustrate a typical base rate response in an FPVS paradigm, we show a SNR spectrum of a single electrode (POOz, medial occipital) with a clear peak at the base rate and its harmonics in the EEG spectrum (Figure 2.2A). The SNR spectra showed a virtually identical response at 2.5 Hz and harmonics for periodic and non-periodic trials.

In order to determine the number of harmonics to take into account for further base rate analysis, we grand averaged the FFT spectrum (periodic and non-periodic trials) of all participants. We first pooled the 128 electrodes, and then we computed the Z-score (threshold of significance placed at a Z-score of 3.30, $p < 0.0005$, one-tailed tests, i.e., signal above noise). There were significant responses up to the 15th harmonic (37.5 Hz). We cropped the individual FFT spectrum, centered at the periodic frequency of the base rate up to the 15th harmonic, with the 10 surrounded neighboring bins on each side. On the one hand, we summed the spectrum of the harmonics and their neighboring bins and computed the SNR. On the other hand, we computed the baseline corrected amplitude then summed the harmonics and computed the grand average of the periodic and non-periodic trials independently and together.

The topographical maps showed an almost identical response for the periodic and the non-periodic trials (Figure S2.4B). Based on the SNR grand average (non-periodic and periodic trials), the highest SNR was found on electrode POOz (snr = 9.01). We then selected an occipital ROI localized in this particular region (POOz, POz, POO5, POO6, Oz). There was no significant difference in the baseline corrected amplitude across the periodic and non-periodic trials ($t(15) = -0.917$, $p = 0.374$, Cohen's $d = 0.23$; Figure S2.4C).

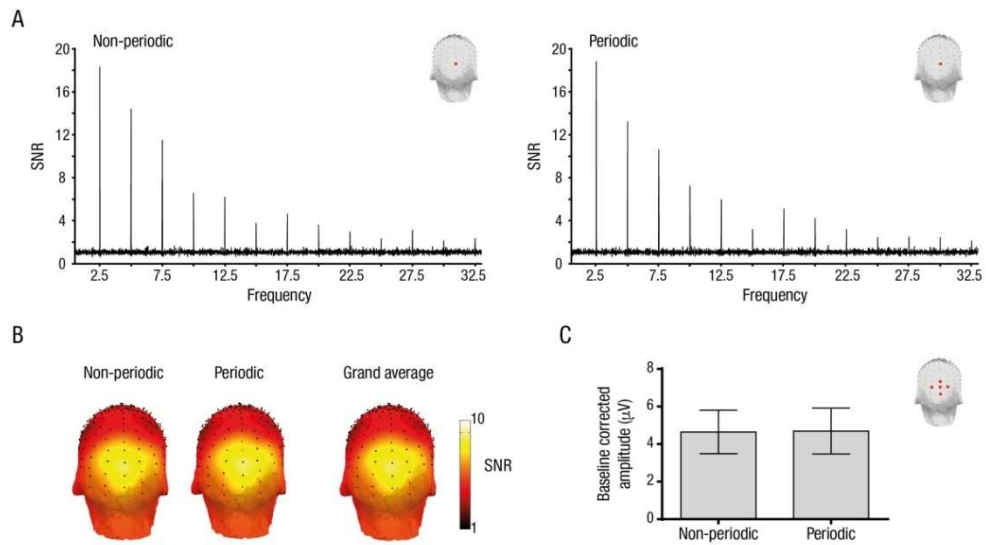


Figure S2.4. Base frequency response (2.5 Hz and harmonics). **(A)** Grand-average EEG spectra (SNR) over POOz showing the first 13 base frequency harmonics for the non-periodic and periodic trials separately. **(B)** Scalp topography of the sum of the significant harmonics of the base frequency for the non-periodic and periodic trials and the grand average (SNR). **(C)** Histogram of the sum of the baseline corrected amplitude of the base frequency over the occipital ROI for the non-periodic and the periodic trials. Error bars represent mean with 95% CI.

Chapter 3

Empirical Study 2

Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow

Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow

Beck, Alexy-Assaf¹, Rossion, Bruno¹ & Samson, Dana¹

¹ Psychological Sciences Research Institute (IPSY), Institute of Neuroscience (IoNS), University of Louvain, Belgium.

Abstract

There is accumulating evidence that adults process what is in another person's line of sight both rapidly and involuntarily. However, it remains unclear whether in the absence of explicit perspective-taking instructions, the processing reflects mentalizing or simply an attentional orienting phenomenon to the objects in front of the other person (submentalizing hypothesis). The present study used frequency-tagging with electroencephalography (EEG) to measure brain responses elicited by an object (frequency-tagged at 2.4 Hz) was presented in the vicinity of a person versus an arrow, flickering at a different frequency. Even though participants were not asked to take the other person's perspective there was a stronger EEG response specifically at the object frequency when it was presented in the vicinity of a person than an arrow. This observation shows that even in the absence of explicit perspective-taking instructions, adults give a special status to objects presented in a social context.

1. Introduction

Visual perspective taking (VPT) refers to the ability to process someone else's visual experience. This ability is fundamental in our social interactions. For instance, viewing a person gazing at a particular location can help us to predict that person's upcoming behavior. There has been an increasing interest in understanding the mechanisms that sustain VPT with one of the main debated issues arising from the observation by Samson and collaborators (2010) that adults rapidly and involuntarily process what is in another person's line of sight. In that study, participants were asked to judge how many discs they could see in a room in which a centrally positioned person could either see all the discs in the room or none/some of the discs. Participants were slower and more error prone in judging how many discs they could see in the room when the other person in the scene saw a different number of discs to what they themselves could see. While the findings were replicated several times (e.g., Bukowski & Samson, 2015, 2017; Baker, Levin, & Saylor, 2015; Nielsen, Slade, Levy, Holmes, 2015; Furlanetto et al., 2016; Todd, Cameron, Simpson, 2017; Todd & Simpson, 2016), the interpretation of the effect is still highly controversial: does it mean that participants spontaneously processed what the person "saw" (mental connection between the other person and what he visually perceives; mentalizing hypothesis) or does it simply mean that participants shifted their attention to the discs in front of the person without construing the object as being *seen* by the person (submentalizing hypothesis)? Discriminating these two hypotheses has far reaching implications for understanding the nature of the processes that humans utilize during social interactions and to further explore how these processes develop in humans and evolved phylogenetically. Indeed, evidence also points to efficient processing of what is in someone else's line of sight in infants (e.g., Sodian, Thoermer, Metz, 2007; Poulin-Dubois, Sodian, Metz, Tilden, Schoeppner, 2007; Brooks & Meltzoff, 2009) and nonhuman species (e.g., Hare, Call, Agnetta, Tomasello, 2000; Tomasello, Call, Hare, 1998; Clayton, Dally, Emery, 2007), raising the question whether infants and nonhuman species are also able to mentalize. Solving the mentalizing/submentalizing debate in adults may provide useful cues about how to examine and understand the processes used by infants and nonhumans.

Two main strategies have been followed so far to discriminate between the submentalizing and mentalizing hypotheses in adults. One strategy has been to examine the effect of placing an obstacle in the line of sight of the person presented in the scene which obstructs the view of the object in front of the person. According to the submentalizing hypothesis, the presence of such an obstacle should not affect what participants process since they do not process what is seen by the person but only what is in front of that person. In contrast, the mentalizing hypothesis posits that participants process the mentalistic link between the person and the object. What participants process should thus be different in the presence compared to the absence of an obstacle in the other person's line of sight. While some evidence suggests that participants perform the same way irrespective of the presence or absence of an obstacle, supporting thus the submentalizing hypothesis (Cole, Atkinson, Le, Smith, 2016; Conway, Danna, Mobin, Catmur, Bird, 2016), other evidence suggests that the performance was different, supporting thus the mentalizing hypothesis (Baker et al., 2015; Furlanetto et al., 2016).

Another strategy has been to compare the effects of having a person versus an arrow in the scene. The logic behind this approach is that both persons and arrows offer a directional cue that can shift participants' attention to a certain part of space. According to the submentalizing hypothesis, this alone can explain the effects observed in Samson et al.'s VPT paradigm. Santiesteban et al. (2014) used this paradigm and added a condition with an arrow instead of a person. They found the exact same pattern of performance irrespective as to whether the central stimulus was an arrow or a person (see also Conway et al., 2016; Santiesteban, Catmur, et al., 2014; Santiesteban, Kaur, Bird, Catmur, 2017). A more recent study (Santiesteban et al., 2017) showed that disrupting the functioning of the right temporo-parietal junction (rTPJ) which is believed to be a key brain region for visual perspective taking (Arora et al., 2017), had similar detrimental effects on task performance, here also, irrespective as to whether an arrow or a person was presented in the VPT task. However, others studies found differences when the central stimulus was a person compared to an arrow, in accordance with the mentalizing hypothesis which considers that in addition to providing a direction cue (common feature to a person and an arrow), the presence of a person leads to the objects in the scene to be construed as "seen" or "not seen". For example, Nielsen et al. (2015) showed that participants were less efficient in their

judgements of the number of discs they could see in a room when the number of discs in front of another person in the scene did not match the total number of discs in the room than when the number of discs in front of an arrow present in the scene did not match the total number of discs in the room. Furthermore, Schurz et al. (2015) showed differential activation of the rTPJ in a VPT task depending on whether an arrow or a person was present in the scene.

Thus, the mentalizing/submentalizing debate is still an open question. In the present study, we developed an original experiment to discriminate between the submentalizing and the mentalizing hypotheses. We used electroencephalography (EEG) together with “Fast periodic visual stimulation” (FPVS; e.g., Rossion et al., 2015, Retter & Rossion, 2016, Quek & Rossion, 2017). The FPVS technique consists in presenting visual stimuli at a fast periodic (i.e. fixed) rate generating an electrical brain response at the same frequency as the stimulus, recorded with an EEG device (Adrian & Matthews, 1934) and analyzed in the frequency domain through Fourier decomposition (Regan, 1966; Norcia, Appelbaum, Ales, Cottureau, & Rossion, 2015). For example, a perceived continuous stimulation at 8 Hz (8 images are presented per second) elicits an electrophysiological response at exactly 8 Hz. While being able to provide high signal-to-noise ratio (SNR) responses in a short amount of time and being a highly objective approach (i.e., signals are measured only at frequencies determined by the experimenter), the FPVS approach is also advantageous in being able to separate signals coming from different spatial sources (i.e., frequency-tagging; Regan & Heron, 1969). Hence, it has been used successfully to measure effects of spatial and selective attention in visual displays (e.g., Morgan, Hansen, Hillyard, 1996; Müller et al., 1998; Painter, Dux, Travis, Mattingley, 2014).

In our study, we were interested in the neural responses to visually presented objects depending on whether these objects are displayed in the presence of another person versus an arrow. We separated the processing of the objects from the processing of the person/arrow by tagging them at different frequencies. According to the submentalizing hypothesis, objects should be processed similarly (i.e., neural response that does not differ in amplitude) irrespective as to whether they are presented simultaneous with a person or an arrow, because in both cases, participants simply process the objects in their spatial

Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow

relations to the directional stimulus. In contrast, according to the mentalizing hypothesis, the objects should be processed differently in the presence of a person compared to an arrow, because only in the presence of a person, can these objects be processed as “seen” or “not seen” in addition to their spatial relation to the directional cue. Furthermore, according to the mentalizing hypothesis, if participants processed the object as “seen” or “not seen” by the avatar, we could expect to observe an interaction between the nature of the central stimuli (avatar or arrow) and the object being in front or behind the central stimuli, with a stronger neuronal response for an object in front of an avatar (and hence seen by the avatar) than an object in front of an arrow. Thus, compared to previous studies, we were able to focus on the (neural) processing of the objects, independently of the observers/arrows, to disentangle the mentalizing and submentalizing views. Importantly while viewing the stimuli, participants had to perform a task which was unrelated to the directional cue provided by the person/arrow, and they did not have to provide any overt behavioral response on any relationship between the person/arrow and the object. The recorded EEG response of interest was thus independent from any behavioral response and not contaminated by decisional processes.

2. Materials and methods

Participants

Thirty-two healthy volunteers (mean age: 22.4 ± 3.16 , all right-handed) took part in the experiment in exchange for monetary compensation. Half of the participants were assigned to the person condition while the other half was assigned to the arrow condition (8 females per condition). The number of participants was decided on the basis of the number of participants recruited in the original study by Samson et al. (2010) and on the basis of a previous EEG study investigating visual perspective taking (McCleery et al., 2011). All participants gave written informed consent prior to the experiment and the experiment received the approval from the ethics committee of the Psychological Sciences Research Institute.

Stimuli

Stimuli consisted of a room with the left, back, and right walls visible. We superimposed an image of an object (e.g., a vegetable) on either the left or right wall, and an image of either a person or an arrow in the center of the room. The person stimuli were full-body profile view photographs of 8 females and 8 males. The arrow stimuli were created by phase-scrambling the persons' photographs and shuffling the image pixels. More specifically, each head of the 16 persons was scrambled to form the color scheme of the directional part of one of the arrows and each body of the 16 persons was scrambled to form the color scheme of the base and non-directional part of one of the arrows (Figure S3.1 of the Supplemental Material). Thus, each person image had a corresponding arrow image in which low level properties (i.e., amplitude spectrum) were roughly preserved. In addition, both the person and the arrow had always a colored contour of 1 pixel which could be green, pink, yellow or orange. This color contour was related to one of the features participants were asked to attend to (participants' task is explained further below). Finally, 24 objects with high familiarity and low visual complexity ratings were selected from the Moreno-Martínez & Montoro (2012) data base and were displayed on either the left or right side wall depending on the stimulation sequence.

Without any object and central cue displayed in the room, the luminance of the right and left walls was 127 cd/m², the luminance of the background wall was 199.55 cd/m² and the luminance of the floor was 117.37 cd/m². All pictures were displayed on a light grey background (198/255, 198/255, 198/255), resulting in 641 x 480 pixel images. The size of the room was approximatively 16.1° horizontal by 12.9° vertical (see Figure S3.2 of the Supplemental Material for additional information about the size of the images).

Procedure

The frequencies used have been chosen to be in a similar range as the frequency used in Study 1 of Chapter 2 (i.e., 2.5 Hz). These frequencies were slower than the frequency rates generally used in the literature because here, we investigated higher level cognitive functions and assumed that stimuli needed a longer presentation to be processed. A stimulation trial consisted of a sequence of 160 person or arrow pictures presented one

Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow

after the other at a rate of 2 Hz (500 ms/image). Within a trial, there were always 10 different person or arrow identities, always with a monochrome colored contour (either orange, yellow or pink). Among them, 2 person or arrow identities had the green monochrome target contour (the specific task instructions are detailed below). The 10 image identities appeared in random order (5 repetitions of each image), with no identical repetitions between consecutive images. In addition, during a stimulation trial, the object displayed on one wall was changing at a rate of 2.4 Hz (416 ms/image), with each of the 24 different objects being presented 8 times (see Figure 3.1 for stimulus illustrations). The orientation of the person/arrow was always fixed within a stimulation trial and so was the position (right/left wall) of the object. In total, there were 4 types of stimulation trials corresponding to the various combinations of the person/arrow stimulus facing the left versus right wall and with the object being placed on the wall in front or behind the person/arrow stimulus.

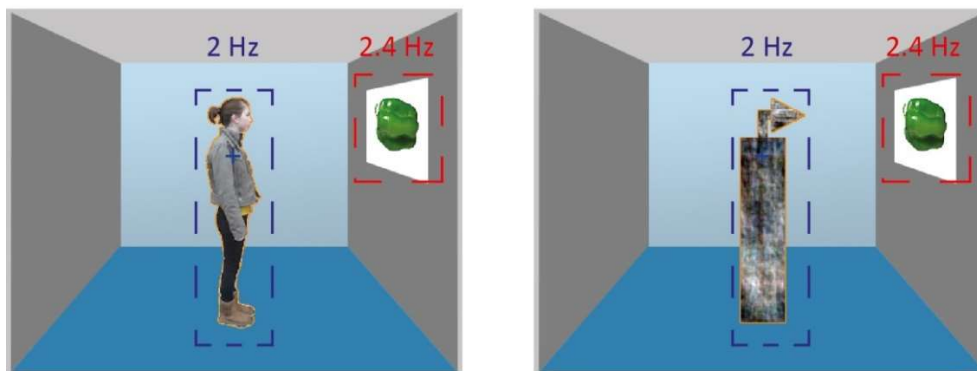


Figure 3.1. Examples of stimuli used in the person (left panel) and the arrow (right panel) condition with their corresponding presentation rates. The dotted rectangle and the text are for explanatory purpose only and were not visible during the experiment.

During EEG recording, participants were seated in a light- and sound-attenuated room at a viewing distance of 80 cm from an LED monitor (BenQ XL2420T) with a 120 Hz refresh rate screen. Stimuli (persons/arrows and objects) were periodically presented through sinusoidal contrast modulation (Rossion & Boremanse, 2011) at their corresponding frequency (2 Hz for persons/arrows, 2.4 Hz for objects).

At the beginning of each trial, participants only viewed the fixation cross during 2 to 4 s. Then, the background appeared and the contrast of the stimuli gradually (i.e. linearly) increased and reached full contrast after 5 s (contrast increase; fade-in 5 s). After 73 s, the scene gradually disappeared (contrast decrease; fade-out 2 s). The fade-in and fade-out periods were used to prevent ocular artefacts and were not included in the analyses. In addition, to minimize low-level adaptation effects, every newly displayed person/arrow had a randomly varied image size (around min: 9315 pixels and max: 11902 pixels). The experiment started with one practice trial in which the person/arrow was facing the object and one practice trial in which the object was behind the person/arrow (presented in counterbalanced order across participants) to familiarize with the task before the EEG recording. Then participants performed 32 test trials (8 per type of stimulation trials). Half of the participants started with the person/arrow facing the object, the other half with the object behind the person/arrow.

Participants received written instructions prior to the EEG experiment, and were unaware that they were taking part in a VPT experiment (there was no mention of “perspective taking”, “seen” or “unseen”, “front” or “back” in the instructions). Their task was to fixate the cross placed in the middle of the screen and to detect as quickly and accurately as possible when the displayed person/arrow had a green contour (10 targets/trial) and each time that the object *soccer ball* was shown on the wall (8 targets/trial). The two events had not to occur concomitantly. Thus, importantly, the task was unrelated to any relationship between the person/arrow and the object. The task was simply used to ensure that participants maintained their attention to both stimuli throughout the trials. A response was considered correct if it was given within three standard deviations of the participant’s mean response time.

EEG acquisition

EEG was recorded with an ActiveTwoBiosemi (Biosemi, Amsterdam, Netherlands) system using a 128 scalps channel based on the 10/20 system. Two additional electrodes, CMS (Common Mode Sense) and DRL (Driven Right Leg) were used as reference and ground, respectively. For each participant, the magnitude of the direct current (DC) offset of all

Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow

electrodes was held below 50 μ V. The vertical eye-movements were recorded using two electrodes above and below the right eye, and for the horizontal eye-movements, we used an electrode next each temple, which were used for monitoring vertical and horizontal eye movements to identify trials contaminated by blinks. The recordings were sampled at 512 Hz.

EEG pre-processing

Off-line analyses of the recorded EEG were carried out using Letswave 5 (<http://nocions.webnode.com/letswave>) running over MATLAB R2012b (Math-Works, Natick, MA). Individual EEG data were bandpass filtered at a high-pass cut-off of 0.1 Hz and a low-pass cut-off of 120 Hz (zero phase shift Butterworth filter, order 4). Then, data files were downsampled to 256 Hz to reduce processing time and reduce file size. A first segmentation from the start to the end of each trial sequence was performed (0 s to 80 s). Then, noisy channels containing other artefact than blinks (which were identified based on the electro-oculograms) were interpolated by the nearest neighboring electrodes (a maximum of 5 % of channels were recreated per participant). In addition, epochs with large artifacts (other than cause by blinks) were excluded from analysis. In total, trial rejection did not exceed 0.68 %. Afterwards, all EEG channels were referenced to a common average (excluding ocular channels).

Frequency-domain analysis

Data were segmented a second time to contain an integer number of 2 Hz and 2.4 Hz cycles, starting after the fade in, until 70 s (i.e., 17920 time bins) to allow the identification of the response at the exact frequencies of interest. The trials were averaged separately for each condition (person or arrow) and types of stimulation sequences (object in front or behind the person/arrow). Finally, the averaged time domain epochs were transformed into the frequency domain using a fast Fourier Transformation (FFT) which extracts the amplitude of the EEG signal for each electrode. The resulting frequency resolution was approximately 0.0143 Hz (1/70 s). We computed the SNR across the whole spectra by dividing the response at each frequency bin by the average voltage amplitude of the 20 surrounding bins (10 on each side, excluding the 2 immediately adjacent bins and the two

extreme bins; e.g., Rossion et al., 2012). However, to quantify a response, we computed the baseline corrected amplitude by subtracting the response at each frequency by the average voltage amplitude of the 20 surrounding bins (10 on each side, excluding the 2 immediately adjacent bins and the two extreme bins; e.g., Retter & Rossion, 2016). In addition, Z-scores were calculated by computing the difference between amplitude at the frequencies of interest and the mean amplitude of the 20 surrounding bins (excluding the 2 immediately adjacent bins and the two extremes bins) and dividing this difference by the standard deviation of the 20 surrounding bins (Rossion et al., 2012). We used the Z-score with a threshold of significance placed at 1.65 ($p < 0.05$, one-tailed, signal > noise) to assess the significance of the response at specific frequencies (e.g., 2 Hz, 2.4 Hz, 4 Hz, 4.8 Hz etc.; computed on the pooled 128 electrodes) and to determine a range of significant harmonics. Thus, we determined the range of significant harmonics for the person, the arrow and the object separately which were then summed to get an aggregate response. Based on the topographical maps we define regions-of-interest (ROIs; for the central cue and the object separately) for statistical analyses. Separate statistical analyses on these ROIs were performed based on individual participants' values (baseline-corrected amplitude) using repeated measures analyses of variance (ANOVAs) with the factors *Condition* (Person vs. Arrow), *Central stimulus-object link* (person/arrow oriented towards vs. away from the object) and *Object position* (Left vs. Right wall) for the object analyses or *Central stimulus orientation* (Left vs. Right) for the central cue analyses.

3. Results

3.1 Behavioral Data

Accuracy and response times for the detection of the two independent targets (green person/arrow contour and object *soccer ball*) were collected during the experiment. Accuracy (Person: $93.6\% \pm 4.46\%$; Arrow: $92.96\% \pm 6.89\%$; $t(29) = 0.305$, $p = 0.763$), false alarms (Person: $2.41\% \pm 0.75\%$; Arrow: $3.66\% \pm 3.08\%$; $t(29) = -1.528$, $p = 0.147$) and response times (Person: 0.57 ± 0.03 ; Arrow: 0.58 ± 0.03 ; $t(29) = -1.207$, $p = 0.237$) did not differ across conditions, suggesting that the level of difficulty was similar across conditions.

Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow

3.2 EEG Data

Object response

The analysis of the EEG response to the object was directly related to our aim to discriminate the submentalizing hypothesis from the mentalizing hypothesis. As a reminder, according to the submentalizing hypothesis, participants who are not directly asked to take someone's perspective simply compute what is in front or behind the person but not what the person "sees". Accordingly, the object in our paradigm should be processed in the same way when it is displayed in the presence of a person as when it is displayed in the presence of an arrow. According to the mentalizing hypothesis however, the object should be processed differently across the two conditions since only in the presence of a person can the object be construed as "seen" or "not seen". The critical effect to discriminate both hypotheses is thus the EEG signal related to the object in the presence of a person versus an arrow: there should be no significant difference according to the submentalizing hypothesis while we should expect a significant difference according to the mentalizing hypothesis.

In all conditions, we observed robust responses at 2.4 Hz and harmonics (4.8 Hz, etc.) in the EEG spectrum (Figure 3.2), which reflect the representation of the flickering objects (Figure 3.2), independently of the representation of the person/arrow, tagged at a different frequency. Z-scores were significant up to the 7th harmonic (16.8 Hz), which were summed for estimating the global SNR and amplitude responses (see methods). We excluded the 12 Hz (5th harmonic) for further analysis because it is a harmonic of both tagged stimuli.

The objects being presented on one of the lateral walls (i.e., in the left or right visual field), they led to a strongly contralateralized response (Figure 3.3A). Thus, we selected 2 ROIs of 4 electrodes which showed the strongest responses, each based on the grand average topographical maps depending on the objects being presented on the left wall (POO6, PPO6, O2, PO8) or right wall (PPO5, POO5, PO7, O1; Figure 3.3B; see Figure S3.3 of the Supplemental Material for the electrodes positions). The analysis of the baseline corrected amplitude of the ROI was submitted to an ANOVA with *Condition* (Person vs. Arrow), *Object position* (Left vs. Right wall), and *Central stimulus-object link* (object in front

vs. behind the central stimulus) as independent variables. The main effect of *Condition* was significant, $F(1, 30) = 6.512$, $p = 0.016$, $\eta^2 = 0.178$, with a larger response for the objects presented in the person condition compared to the objects presented in the arrow condition. There was a marginal but not significant effect of *Object position*, $F(1, 30) = 3.480$, $p = 0.072$, $\eta^2 = 0.104$, with a trend for a larger response for objects displayed on the left than the right wall. There was no main effect of *Central stimulus-object link*, $F(1, 30) = 1.105$, $p = 0.302$, $\eta^2 = 0.036$. None of the interactions were significant ($F_s < 1.560$ $p_s > 0.221$).

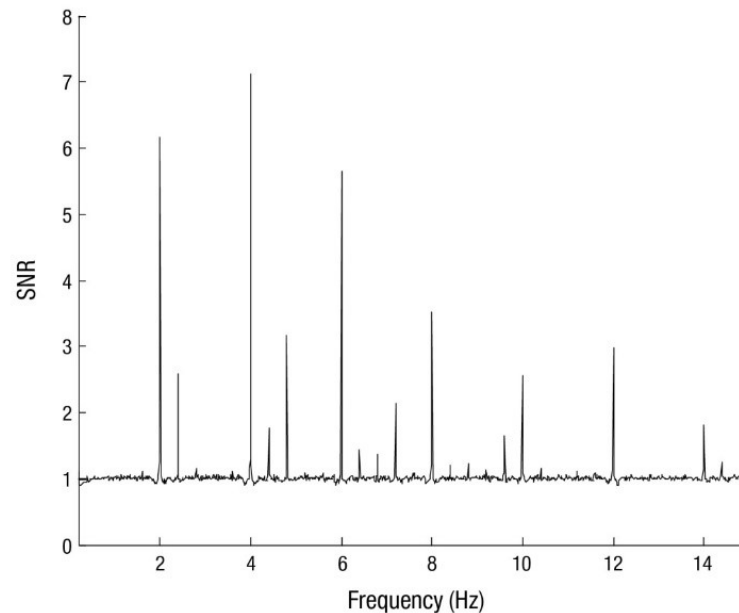


Figure 3.2. Frequency spectrum of the Grand average (all participants and conditions) over the 128 pooled channels.

In summary, there were high contralateralized responses at 2.4 Hz (and harmonics). In line with the prediction stemming from the mentalizing hypothesis, responses to the objects were significantly stronger in the person compared to the arrow condition (Figure 3.3C), but irrespective of whether the object/person pointed in the direction of the objects or not.

Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow

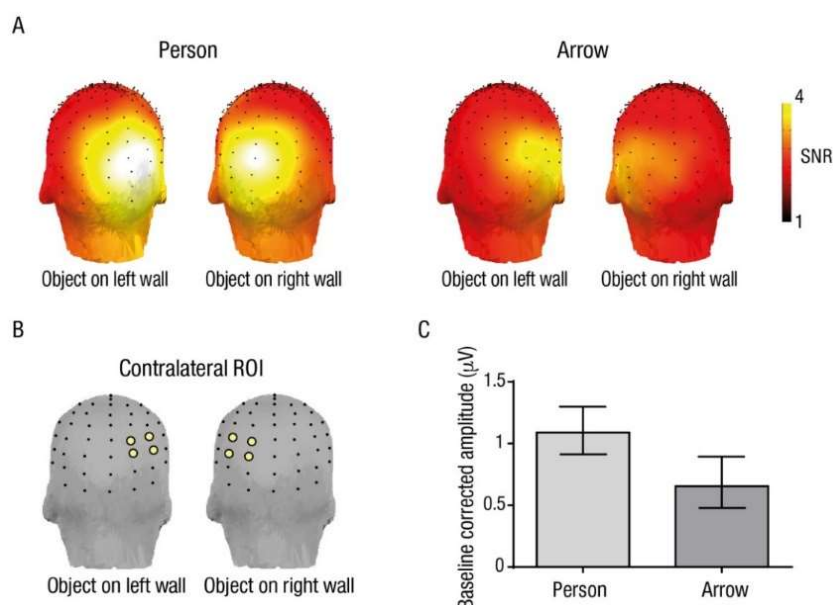


Figure 3.3. (A) Topographical maps of the sum of the harmonics of the object frequency per condition and depending of the object position. Note that scales are different between conditions. (B) Contralateral to the object's visual field position and the selected ROIs for further analysis. (C) Bar graph of the sum of the baseline corrected amplitude for each condition (person vs. arrow) over the occipital ROI. Error bars represent mean with 95% CI.

Central stimulus response

The significantly larger response for the objects in presence of the person than the arrows could simply be that participants in the group stimulated with the person have an overall larger EEG response. Here, thanks to frequency-tagging, we were able to test this possibility by measuring the response at 2 Hz, the frequency of stimulation of the central item. Figure 3.2 show that there were large peaks at 2 Hz and its specific harmonics (4 Hz, 6 Hz, etc.) on posterior channels (Figure 3.4).

There were significant responses up to the 15th harmonic (30 Hz). For further analysis, we excluded the 6th (12Hz) and the 12th (24 Hz) harmonics, which were not specific of the person and arrow responses because they corresponded also to the 5th and the 10th harmonics of the object response.

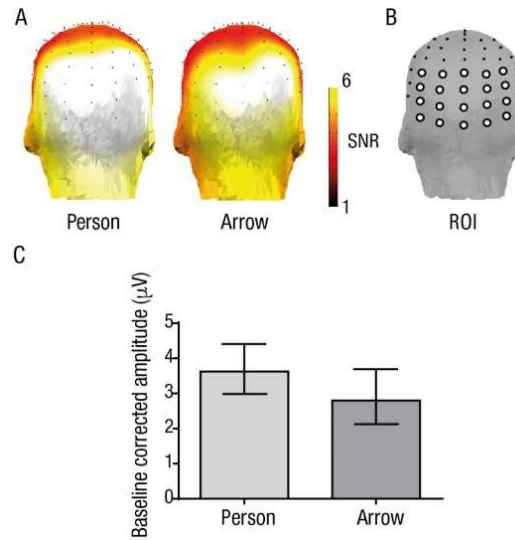


Figure 3.4. (A) Scalp topography (SNR) of the sum of the 13th first harmonic (excluding the 6th and the 12th harmonic) of each condition (person vs. arrow) and their grand average. (B) Selected ROIs for further analysis. (C) Bar graph of the sum of the baseline corrected amplitude for each condition (person vs. arrow) and type of sequence (Left vs. Right orientation of the person/arrow and orientation towards or away from the object displayed on the wall) over the occipital ROI. Error bars represent mean with 95% CI.

Based on the topographical map of the SNR grand average (person and arrow; Figure 3.4A), we identified 20 electrodes which showed the strongest summed-harmonic response for the central item (PPO5, PO7, PO9, PO11, POO5, O1, POI1, I1, POOz, Oz, Olz, Iz, POO6, O2, POI2, I2, PPO6, PO8, PO10, PO12; Figure 3.2B) and selected them to define the occipital ROI which was used for further analyses on the central stimulus. The analysis of the baseline corrected amplitude of the ROI was submitted to an ANOVA with *Condition* (Person vs. Arrow), *Central stimulus orientation* (Left vs. Right), and *Central stimulus-object link* (person/arrow oriented towards vs. away from the object) as independent variables. The ROI's analysis revealed no main effect of *Condition*, $F(1, 30) = 2.080$, $p = 0.160$, $\eta^2 = 0.06$. Although numerically the response was larger in the person than the arrow condition. The effect of *Central stimulus orientation* was not statistically significant, $F(1, 30) = 2.947$, $p = 0.096$, $\eta^2 = 0.089$. There was main effect of *Central stimulus-object link*, $F(1, 30) = 1.001$, $p = 0.325$, $\eta^2 = 0.032$. None of the interactions were significant ($F_s < 1.619$ $p_s > 0.213$).

4. Discussion

We presented frequency-tagged objects at 2.4 Hz in two conditions, either in the presence of a person or an arrow themselves both frequency-tagged at 2 Hz. Participants were asked to make judgments unrelated to perspective taking. We observed large electrical responses exactly at the frequencies of interest and their harmonics over the posterior scalp region providing a robust neural marker of the processing of the central directional cue (person/arrow) and the object in front or behind the central stimulus.

More important for the purpose of our study were the neural responses to the objects. According to the submentalizing hypothesis, in the absence of explicit perspective-taking instructions, participants simply process what is in front of the other person without construing these objects as being “seen” by the other person. Accordingly, an object in the scene should be processed similarly in the presence of any other type of directional stimulus (i.e., an arrow) since what matters is the spatial relation between the orientation of the directional stimulus and the position of the object. According to the mentalizing hypothesis, despite the absence of explicit perspective-taking instruction, participants construe objects as “seen” or “not seen” and go thus beyond simply shifting their attention to the object in the location prompted by the directional information conveyed by the other person’s body orientation. We would thus expect a deeper processing of the object in the presence of a person compared to a nonhuman directional cue.

Our results unambiguously showed a stronger neural response to objects in the presence of a person than in the presence of an arrow. Both the person and arrow stimuli shared directional information thus the difference across conditions must be attributed to another factor. We posit that the difference is linked to the mentalizing hypothesis and that participants were construing objects in the person condition as objects that were “seen” or “not seen”. Before discussing further the implications of these results, it is important to consider whether there are any other explanations for our pattern of results. For example, one objection to the interpretation that our results support the mentalizing hypothesis is that our FPVS paradigm did not allow us to capture anything related to the processing of what is in the other person’s line of sight. Whereas, a behavioral post-test clearly showed that participants extracted the directional information more efficiently from the arrow

stimuli than the person stimuli (Figure S3.4B of the Supplemental Material). Indeed, based on our FPVS responses there is no direct evidence that participants engaged in perspective taking or had their attention shifted to the object in front of the directional cue, because we had no overt behavioral response that relates to these processing. The neural response was furthermore similar when the object appeared in front compared to behind the central directional cue. The absence of difference for objects in front or behind the central stimulus could be explained by the fact that participants had to focus on the object to succeed in the behavioral task. Thereby, the expected stronger activation of the neural population related to the attention shift (and/or perspective taking mechanism) could be masked or cancelled out, especially since the frequency-tagging procedure could have acted as bottom-up attention attractor and could have prevented the measurement of any shift of attention related to the orientation of the central stimulus. Thus, our results do not mean that there was no shift of attention and/or perspective taking mechanism involved; such shift of attention may simply not have been measurable when examining the neural response to the objects presented in the scene, which could also explain that we did not observe the interaction between the nature of the central stimuli (avatar or arrow) and the object being in front or behind the central stimulus which would support the mentalizing hypothesis. Furthermore, the observed stronger response for the object presented in the vicinity of an avatar compared to an arrow could be due to the objects which become more engaging when put in a social context with a human agent because a person might use it (spectator to an action without involving perspective taking processes), compared to the arrow that is just another inanimate object. Finally, according to the mentalizing hypothesis, objects in the presence of another person are construed as “seen” or “not seen”. Such construal applies thus both to objects within the line of sight of another person (processed as “seen”) and objects outside their line of sight (processed as “not seen”). Knowing what is not seen by someone else is equally important in everyday life. For example, this is what allows us to stop someone from crossing the road if we notice that that person did not see a car coming. Knowing what someone cannot see plays also a critical role in deceiving or in inferring someone’s false belief. Thus, our adaptive behaviors in social situations result from the fact that we tag both, objects that are seen and objects that are not seen by someone else, both tagging must therefore generate a corresponding brain response. We can also hypothesize

that the distinction between an object seen or not seen by someone else depends on different neurons which are sensitive to the relationship between the person/arrow and the object and not necessarily on the neurons which respond to the persons/arrows or the objects themselves. This is an interesting avenue to explore further in the future. For the purpose of the current study, however, the absence of difference in the neural response to objects in front or behind the person remains compatible with the mentalizing hypothesis. Furthermore, one could argue that the observed stronger neural response to objects in the presence of a person than in the presence of an arrow could be due to a difference between the responses to the person versus the arrow. Note however, that this difference was not significant. Nevertheless, to ensure that our effect (i.e., stronger neural response to objects in the presence of a person than in the presence of an arrow) was not due to a difference between the responses to the person versus the arrow, we computed the same analyses by excluding the participant which showed the largest response to the person and the one which showed the lowest response to the arrow (see the Additional analyses of the Supplemental Material). In that case, the analyses showed no trend for a larger response for the person than the arrow, whereas the observed strongest neural response to objects in the presence of a person than in the presence of an arrow remains significant.

Overall, it thus seems that our pattern of results refutes the submentalizing hypothesis formulated in terms of a purely attentional orienting phenomenon. Participants do more than just paying attention to the object in front of a directional cue when they view a person surrounded by objects. Our pattern of results does, however, not provide a definitive answer concerning the exact nature of what is processed. Between the extreme versions of the submentalizing and mentalizing hypotheses, they may be a continuum of processing of increased depth with the final layer of processing corresponding to a conscious construal of the mental experience of the other person. The neural response we measured in the current study may be somewhere along this continuum but is further away than the most extreme version of the submentalizing hypothesis. This is particularly remarkable since compared to previous studies (e.g., Cole et al., 2016; Conway, Lee, Ojaghi, Catmur, Bird, 2017; Santiesteban et al., 2014; Santiesteban et al., 2017), participants in our study were not prompted to take any perspective, since their sole task was to detect the color of the central stimulus (person or arrow) contour and to identify the object displayed on a wall. This opens

the way to consider that mentalizing processes could be much more ubiquitous than sometimes acknowledged (Santiesteban, Shah, White, Bird, Heyes, 2014; Heyes, 2014).

In conclusion, the present observations provide evidence that adults process objects differently in their environment if these objects are presented alongside another person compared to a control directional object (i.e. an arrow). This indicates that specific social processes may be at play despite the absence of explicit perspective taking instructions. This study adds new empirical evidence in favor of the hypothesis that adults engage in more than attentional orienting when viewing someone else even when they are not instructed to mentalize.

5. Acknowledgements

We thank Andrea Conte for having created the application used to test the subjects. We also thank Joan Liu-Shuang and Genevieve Quek for their comments on this paper.

6. Supplementary material

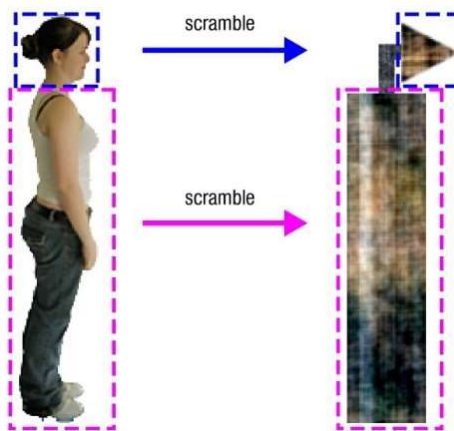


Figure S3.1. Creation of the arrows. Low level properties were roughly preserved between the person picture and its corresponding arrow.

Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow

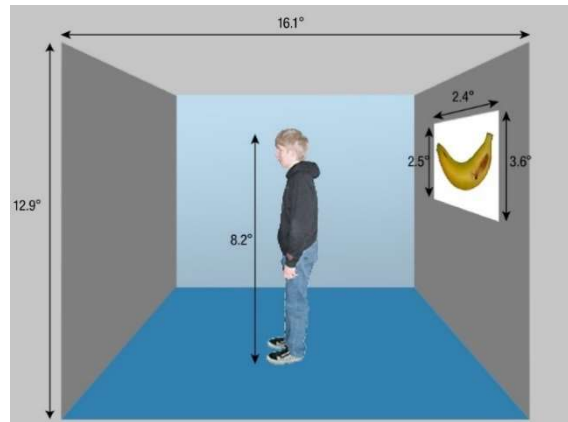


Figure S3.2. Original size of the images in visual angle at a viewing distance of 80 cm. Note that each displayed person/arrow had a randomly varied image size.

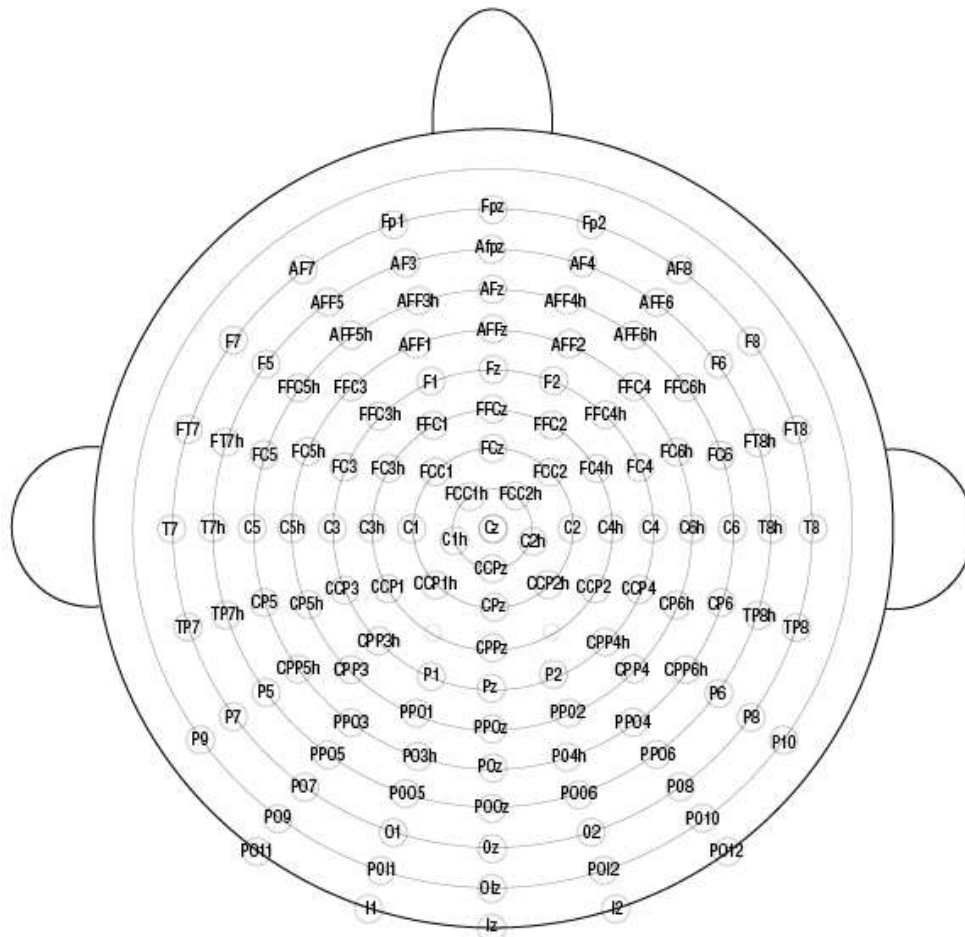


Figure S3.3. Electrodes names and position of the 128-channel Biosemi system.

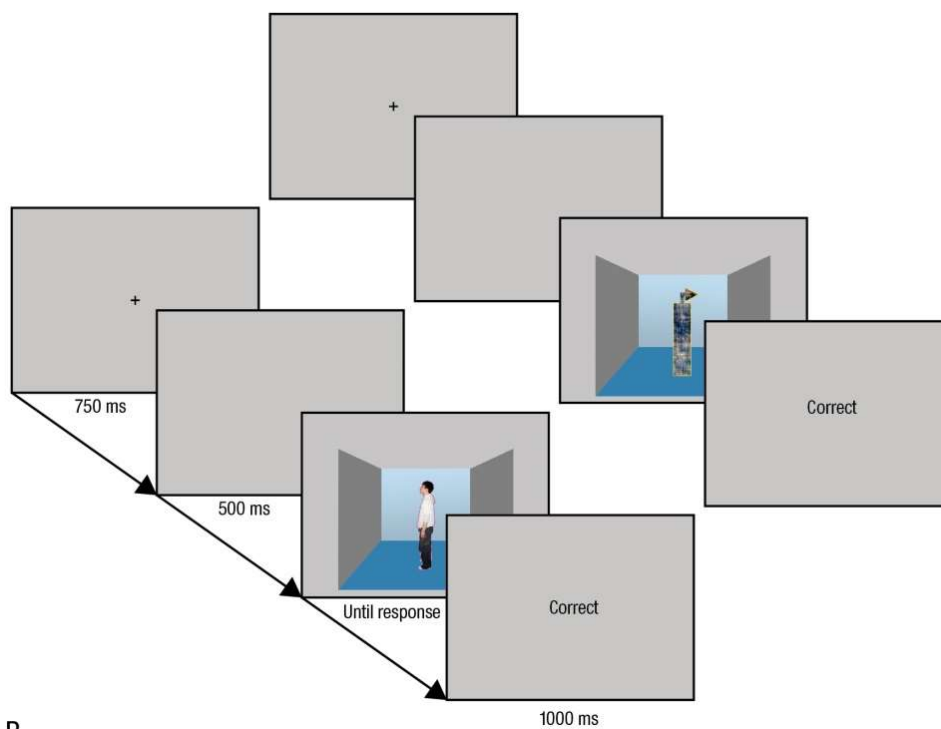
Post-test

E-prime 2.0 software (Psychology Software Tools) was used to program the experiment. Each trial started with a fixation cross presented for 750 ms. Then, the fixation cross disappeared and 500 ms after, the image of the room and the central cue appeared until participants responded or until 2000 ms had elapsed. Finally, the last screen displayed the feedback about the accuracy of the response for 1000 ms. Participants received written instructions prior to the experiment, and had to judge as quickly and accurately as possible if the central cue (i.e., person or arrow) was oriented to the left or right side by pressing the corresponding keyboard arrow. Participants responded by pressing the left keyboard arrow if the central cue was oriented to the left or the right keyboard arrow if the central cue was oriented to the right. Sixteen participants (11 female, 14 right-handed, mean age: 31.3 ± 4.46) which were different from the EEG experiment, performed 2 blocks of 128 trials each (64 trials with a person and 64 trials with an arrow in the center of the room and the same number of trials for all colors contour, i.e. 16 trials for each color contour per condition and block; person and arrow trials appeared in a random order within a block). Furthermore, there was the same number of trials for each cue orientation (left or right). Note that the room was presented without object.

Based on the nature of the central cue, the results (Response Time: Person: 409.23 ± 50.08 , Arrow: 397.52 ± 51.34 ; $t(15) = 6.107$, $p < 0.000$; Accuracy: Person: $98.93 \% \pm 0.010 \%$; Arrow: $99.22 \% \pm 0.014 \%$; $t(15) = -1.065$, $p = 0.304$) showed that the directional information could be extracted quicker from the arrow than the person stimuli while accuracy was the same (Figure S3.4B).

Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow

A



B

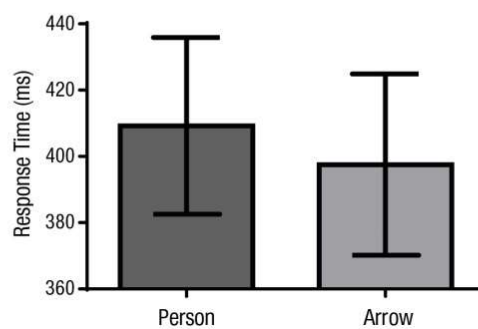


Figure S3.4. Information about the post-test. (A) Illustration of two trials (one with arrow stimulus and another one with person stimulus), in which participants merely judged whether the central stimulus (i.e., person or arrow) was oriented to the left or the right side. (B) Response time in the task for the person and the arrow stimuli separately.

Additional analyses

We excluded both the participant which showed the largest response to the person and the one which showed the lowest response to the arrow based on the average of all the electrodes and on the sum of the significant harmonics.

Central stimulus response: The analysis of the baseline corrected amplitude of the ROI was submitted to an ANOVA with *Condition* (Person vs. Arrow), *Central stimulus orientation* (Left vs. Right), and *Central stimulus-object link* (person/arrow oriented towards vs. away from the object) as independent variables. There was no main effect of *Condition*, $F(1, 28) = 0.742$, $p = 0.396$, $\eta^2 = 0.026$. The effect of Central stimulus orientation was not statistically significant, $F(1, 28) = 3.674$, $p = 0.066$, $\eta^2 = 0.116$, but showed a trend for a larger response for central cue oriented to the right wall. There was no main effect of *Central stimulus-object link*, $F(1, 28) = 0.94$, $p = 0.761$, $\eta^2 = 0.003$. None of the interactions were significant ($F_s < 3.433$ $p_s > 0.074$).

Object response: The analysis of the baseline corrected amplitude of the ROI was submitted to an ANOVA with *Condition* (Person vs. Arrow), *Object position* (Left vs. Right wall), and *Central stimulus-object link* (object in front vs. behind the central stimulus) as independent variables. The main effect of *Condition* was significant, $F(1, 28) = 4.337$, $p = 0.047$, $\eta^2 = 0.134$, with a larger response for the objects presented in the person condition compared to the objects presented in the arrow condition. There was a significant effect of *Object position*, $F(1, 28) = 5.119$, $p = 0.032$, $\eta^2 = 0.155$, with a larger response for objects displayed on the left than the right wall. There was no main effect of *Central stimulus-object link*, $F(1, 28) = 0.798$, $p = 0.379$, $\eta^2 = 0.028$. None of the interactions were significant ($F_s < 2.867$ $p_s > 0.102$).

This analyses support that there is a stronger neural response to objects in the presence of a person than in the presence of an arrow which seems not due to the difference between the responses to the person versus the arrow.

Mentalizing or submentalizing during implicit visual perspective taking? Frequency-tagging shows that objects are processed differently in the presence of another person compared to an arrow

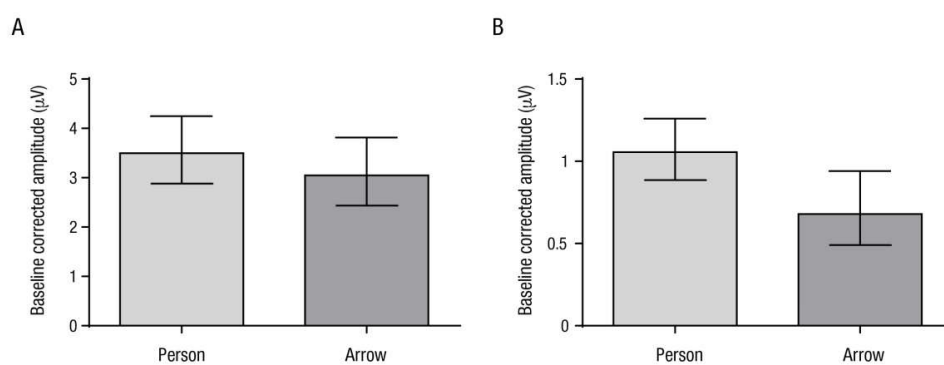


Figure S3.4. Bar graph of the sum of the baseline corrected amplitude for each condition (person vs. arrow) over the occipital ROI. (a) Central cue response. (b) Object response. Error bars represent mean with 95% CI.

Chapter 4

Empirical Study 3

Neural intermodulation responses when viewing an object in front of a person are affected by the presence of an obstacle in the person's line of sight

Beck, Alexy-Assaf¹, Rossion, Bruno¹ & Samson, Dana¹

¹ Psychological Sciences Research Institute (IPSY), Institute of Neuroscience (IoNS), University of Louvain, Belgium.

Abstract

Adults rapidly and involuntarily process what is in another person's line of sight, but the underlying processes remain highly controversial. It is still unclear whether, in the absence of explicit perspective-taking instructions, we spontaneously process what another person "sees" (mentalizing hypothesis) or if we merely shift our attention to what is in front of that person (submentalizing hypothesis). Electroencephalogram (EEG) was recorded while we displayed stimuli in which a person (frequency tagged at 2.07 Hz) faced an object (frequency tagged at 3 Hz). We found in the EEG spectrum responses exactly at these frequencies, but most importantly clear intermodulation components (IMs), which appeared at frequencies equal to the combinations of the object and person frequencies (e.g., 2.07 Hz + 3 Hz = 5.07 Hz). Such responses emerge from population of neurons which received information from the two flickering stimuli and correspond to the objective neural signature of the processes involved in their integration. IMs responses were increased when the person could see the object compared to when the person could not see the object due to an obstacle in his line of sight. These responses were significant and quantifiable for most individual participants even though participants were not asked to pay attention to the person's visual experience. These results discard the submentalizing hypothesis formulated in terms of a purely attentional orienting phenomenon and show that EEG frequency tagging is a promising avenue for clarifying the processes involved during visual perspective taking.

1. Introduction

There is increasing evidence suggesting that in many situations, adults spontaneously take into account other people's perspective (Samson et al., 2010; Nielsen et al., 2015; Schurz et al., 2015; Baker et al., 2015; Furlanetto et al., 2016; Qureshi et al., 2010; Michael et al., 2017). Whether this is result of cognitive processes that genuinely allow to infer the other person's mental state is currently highly debated. The current study addresses this debate in the context of visual perspective taking (VPT), i.e. the ability to infer what another person is seeing.

One finding in particular sparked a debate about the mechanisms involved when adults observe another person in the vicinity of objects. Samson et al. (2010) presented to participants images of a person standing in the middle of room with discs displayed on the walls. Results revealed that when participants were simply asked to judge the number of discs they could see in the room, they were slower and more error prone when some of the discs were not visible to the other person (because some of the discs were behind the person) compared to when the other person could see all the discs (because all the discs were in front of the person). This was observed even when participants were never explicitly asked to take into account the other person's perspective. The difference of performance was taken as reflecting an altercentric interference effect. Still today, the interpretation of this effect remains controversial. Is it due to the fact that participants computed the other person's visual experience and that this is harder when the self and the other have different visual experiences (mentalizing hypothesis)? Or is it due to the fact that the other person acted as a spatial cue orienting attention to the objects in front of the other person and that on some occasions this cue was not valid because some discs were located somewhere else (submentalizing hypothesis)?

To disentangle the mentalizing and submentalizing hypotheses, authors used two main strategies. One strategy has been to directly compare the presence of a person with the presence of an arrow in the scene. Both stimuli carry indeed directional information and act as spatial cues to orient attention (e.g., Driver et al., 1999; Posner, 1980) but only the person has mental states. If the altercentric intrusion effect merely results from an attentional orienting mechanism (submentalizing hypothesis), the effect should be similar

when a person or an arrow is presented in the scene. This is precisely what several authors found (Conway et al., 2016; Gardner et al., 2017; Santiesteban, Catmur, Hopkins, Bird, Heyes, 2014; Santiesteban, Kaur, Bird, Catmur, 2017). Unfortunately, this approach remains blind as to whether it is really the same underlying mechanism which causes the same pattern of results. For example, some studies have shown that the attentional orienting neural networks triggered by an arrow or gaze direction are not fully overlapping (Hietanen et al., 2008; Akiyama et al., 2006). Furthermore, in other studies, the altercentric interference effect was found to be stronger when a person compared to an arrow was presented in the scene (Nielsen, Slade, Levy, Holmes, 2015) and was accompanied by the activation of the temporo-parietal junction (a brain region thought to be involved in mentalizing, e.g., Frith & Frith, 1999; Van Overwalle, 2009 but see Schneider et al., 2014) only in the presence of the person (Schurz et al., 2015; but see Catmur et al., 2016). In addition, based on this strategy, in a previous study (Study 2 reported in Chapter 3) we compared the neural response elicited while an object was presented in the vicinity of a person versus an arrow. We observed a stronger brain response for the object when it was presented in the vicinity of a person than an arrow which suggests that stimuli presented in a social context have a special status.

The other strategy to disentangling the mentalizing and submentalizing hypotheses consisted in placing an obstacle in the other person's line of sight. For example, a barrier was placed between the person and the object in front of her or the person wore a pair of opaque goggles. Thus, despite the fact that the person was facing an object, she could not necessarily see that object (e.g., Figure 4.1). According to the submentalizing hypothesis, participants only process the spatial relationship between the person and the target (e.g., in front, behind) thereby responses should be independent of whether the person can or cannot see the target. Thus, the altercentric interference effect should not be modulated by the presence versus absence of an obstacle in the other person's line of sight. This is precisely what some authors found (Cole, Atkinson, Le, Smith, 2016; Conway, Danna, Mobin, Catmur, Bird, 2016). However, some authors found, on the contrary, that the presence of an obstacle in the line of sight modulated participants' performance (Baker et al., 2015; Furlanetto, Becchio, Samson, Apperly, 2016). These latter results suggest that

participants go further than computing the mere spatial relationship between the person and the objects in front of her.

Thus, despite the different strategies which investigated implicit VPT, the controversial result observed did not allow to disentangle between the mentalizing and submentalizing hypotheses. The current study aimed to propose an original approach which would help to differentiate between these two hypotheses. To do so, we used the “Fast periodic visual stimulation” (FPVS) technique (e.g., Rossion et al., 2015; Retter & Rossion, 2016; Quek & Rossion, 2017) coupled with electroencephalogram (EEG) in order to bring a new perspective in the current debate. This method consists in presenting stimuli periodically and at a fixed rate. The train of a periodically perceived stimuli generates the same periodic pattern of activity in the brain (e.g., 12 Hz presentation rate elicits an electrophysiological response at exactly 12 Hz; Regan, 1966). The brain’s response is then analyzed in the frequency domain which gives access to the specific neuronal response to the flickering items. It is possible to tag at different frequencies different parts of the stimuli to generate responses at each flickering frequency to readily quantify aspects of processing associated with each stimulus in the display. The brain’s response for each frequency-tagged element can thus be objectively identified in the EEG frequency spectrum at the exact predefined frequencies (Regan & Heron, 1969). Previous studies which frequency-tagged multiple items (at different frequencies) found in some cases the emergence in the EEG spectrum of frequencies which were not present during the visual stimulation. Such responses appeared at frequencies which were equal to the combinations of the frequencies $F1$ and $F2$ ¹ (e.g., $f1 \pm f2$; Baitch & Levi, 1988; Zemon & Ratliff, 1982; Alp, Kogo, Van Belle, Wagemans, Rossion, 2016; Boremanse, Norcia, Rossion, 2013). These responses which cannot be explained by the independent process of the individual frequencies are referred to as “intermodulation components” (IMs, Zemon & Ratliff, 1982, 1984). Such responses can only emerge from

¹ e.g., the person at the frequency $F1$ and the object at the frequency $F2$, capital letter refers to the frequency used to tag a stimulus and small letter are used to refer to the neuronal response at the corresponding frequency (e.g., $f1$ or $f2$).

Neural intermodulation responses when viewing an object in front of a person are affected by the presence of an obstacle in the person's line of sight

population of neurons which received information from the two flickering stimuli and are assumed to represent integration processes of these two stimuli.

The present study investigated IMs resulting from the common integration of an object and a person (each frequency tagged at a particular frequency) which, crucially, should generate different patterns of results according to the submentalizing or the mentalizing hypothesis. Indeed, according to the submentalizing hypothesis, participants simply process the spatial relationship between the directional stimulus (i.e., the person) and the target. This spatial relationship remains unchanged whether an obstacle obstructs the other person's line of sight or not. Therefore, IMs which represent the integration of the person and the object should be unaffected by the presence of an obstacle. In contrast, according to the mentalizing hypothesis, participants go further and process the target as "seen" or "not seen" by the person. Thus, the presence of an obstacle in the person's line of sight should modify whether the other person sees or not the object in front of her and this should influence the presence or strength of the IM response.

2. Materials and methods

Participants

We tested sixteen healthy volunteers (female: 8, mean age: 22.68 ± 2.24 , all right-handed) in exchange of a small honorarium. The number of participants was decided on the basis of the number of participants recruited in the original study by Samson et al. (2010) and on the basis of a previous EEG study investigating visual perspective taking (McCleery et al., 2011, Studies 1 and 2 reported in Chapter 2 and 3 respectively). All participants gave written informed consent prior to the experiment and the experiment received the approval from the ethics committee of the Psychological Sciences Research Institute.

Stimuli

Stimuli were created with Poser 9 (version 1.0.1, Smith Micro) and consisted of 3D avatar (16 different identities, 8 females) placed in the middle of a room (with the left, back and right wall visible). The avatar could face either the left or the right wall. In addition,

each avatar had a colored contour of 1 pixel which could be green, blue, red or pink. This color contour was related to one of the features participants were asked to attend to (participants' task is explained further below). Twenty-four objects (e.g., a vegetable) with high familiarity and low visual complexity ratings were selected from the Moreno-Martínez & Montoro (2012) database and were always displayed on the wall in front of the avatar. Moreover, a board was placed in front of the avatar, which could prevent or not the avatar from seeing an object displayed on the wall depending on the height of the board.

The room with only the board displayed had a luminance of 119 cd/m². The size were approximately 11.60° horizontal by 9.67° vertical for the room, 7.16° for the avatar and 1.72° horizontal by 2.15° for the object (see Figure S4.1 of the Supplemental Material for additional information about the size of the image).

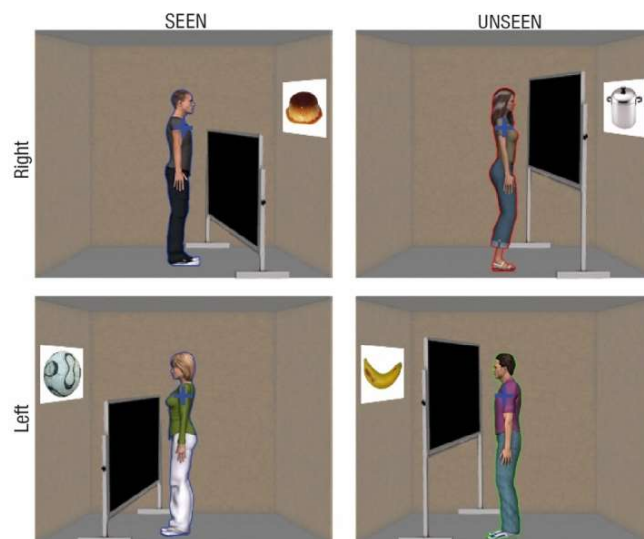


Figure 4.1. Stimuli of the Seen (facing left or facing right) and Unseen (facing left or facing right) conditions. The avatar was frequency-tagged at 2.07 Hz and the object at 3 Hz (see the movie S1 in supplementary material).

Procedure

Each trial lasted 80 s. It started with a fixation cross presented for 2 to 4 s followed by the appearance of the room and the stimuli (avatar and object). The contrast of the avatar and the object gradually (i.e. linearly) increased and reached full contrast after 5 s (contrast

increase; fade-in 5 s). At the end of the trial the avatar and the object gradually disappeared (contrast decrease; fade-out 2 s). The fade-in and fade-out periods were used to reduce eye-movements or blinks due to abrupt stimulation onset or offset and were not included in the analyses. A trial always displayed a room which does not flicker (it appears stable). Superimposed to it were a flickering avatar and a flickering object which critically were frequency tagged at different periodic rates through sinusoidal contrast modulation (Rossion & Boremanse, 2011; Figure 4.2). More specifically, the avatar was presented at 2.07 Hz (more precisely at 2.068966 Hz) such that each avatar appears for 483 ms and the object at 3 Hz (i.e., 333 ms/image). The frequencies used to tag the avatars and the objects were adapted from previous studies (Study 1 and 2 reported in Chapter 2 and 3 respectively). However, we did not choose the exact same frequencies than in Study 2 of Chapter 3 in which the responses to the fundamentals, their harmonics, and their combinations sometimes overlapped². In addition, the frequencies were constrained to contain an integer number of frames on a monitor with a refresh rate of 120 Hz. Thus, each avatar's presentation took 58 frames and each object's presentation took 40 frames.

The frequencies chosen correspond for each trial to 166 avatar images and 240 object images presented in a train. Within each trial the participant saw 10 different avatar identities (5 females) always displayed with a monochrome colored contour (blue, pink and red). Two avatars (1 female) had the green monochrome target contour (the specific task instructions are detailed below). Per trial, each avatar image was presented 5 or 6 times and each object passed 10 times. The avatars and the objects were presented in a randomize order without repetition between consecutive images.

Both the orientation (left/right) of the avatar, the position of the object (left/right wall) and the position and the height (up or down) of the board were fixed during the course of a single trial. Thus, during a trial the avatar always saw or not the object placed in front of

² To clearly identify IMs it is important that the combinations of fundamental and harmonic frequencies do not correspond to the frequencies of the fundamentals and their harmonics themselves.

him or her, depending on whether the board obstructed the avatar's line of sight or not. The different types of stimulation sequences were Seen-Right (the avatar faces the right wall and the board does not obstruct the line of sight), Seen-Left (the avatar faces the left wall and the board does not obstruct the line of sight), Unseen-Right (the avatar faces the right wall and the board obstructs the line of sight), and Unseen-Left (the avatar faces the left wall and the board obstructs the line of sight). Either condition was repeated 8 times, making for a total of 32 trials.

To familiarize with the task (see below), participants performed two practice trials, one of the Seen condition and one of the Unseen condition. Then, participants performed 32 test trials (8 trials of each condition: Seen-Right, Seen-Left, Unseen-Right, and Unseen-Left). Half of the participants started with the Seen condition and the other half with the Unseen condition, and then we alternated trial by trial between these two conditions.

Participants received written instructions prior to the EEG experiment, and were ignorant that they were taking part in a VPT experiment. We avoided words which could engage the participant to create a link between the avatar and the object (e.g., "seen", "front"). They had to maintain fixation on the cross placed in the middle of the screen throughout all trials and they had to respond by pressing the spacebar as quickly and accurately as possible. Their task was to detect each time that an avatar with a *green contour* was presented (10 to 12 avatars targets per trial) and each time that the object *soccer ball* was presented (10 objects targets per trial). In total there were 20 to 22 targets to respond to per trial. Note that, accurate performance in this task does not require to process the avatar's visual experience (i.e., the task was not a VPT task). The goal of the behavioral task was to ensure that participants maintained their attention to both stimuli throughout the trials. A response was considered as correct if it was given within two standard deviations of the participant's mean response time.

During the EEG recording, participants were seated in a comfortable chair in a moderately darkened and sound-attenuated room, at 80 cm from an LED monitor (BenQ XL2420T) with a 1920 x 1080 resolution.

Neural intermodulation responses when viewing an object in front of a person are affected by the presence of an obstacle in the person's line of sight

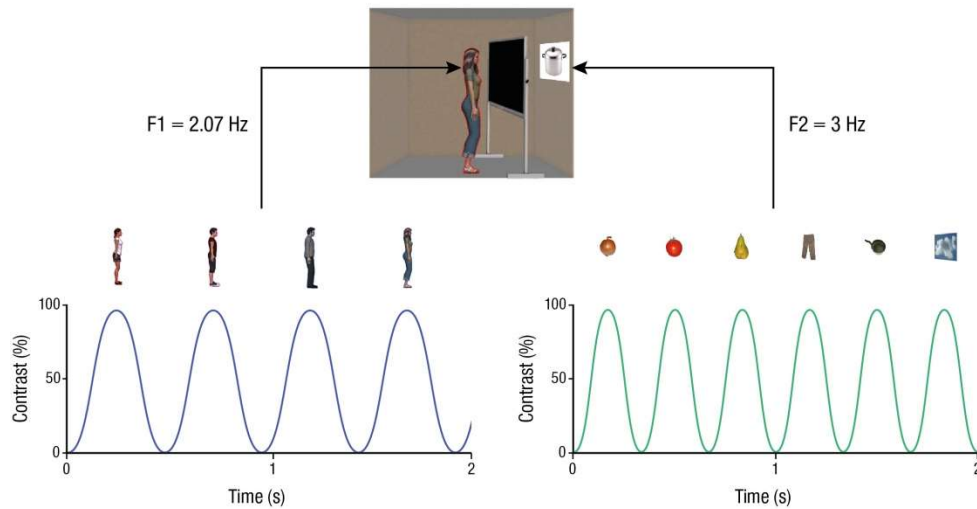


Figure 4.2. Different stimuli (avatar and object) appear at each stimulation cycle (no repetition between consecutive images). Stimuli are presented by sinusoidal contrast modulation at the rate of 2.07 Hz for the avatars and 3 Hz for the objects.

EEG acquisition

The EEG was recorded with a BioSemi ActiveTwo (Biosemi, Amsterdam, Netherlands) system using 128 electrodes at a 512 Hz sampling rate. All recordings were performed relative to two additional electrodes, CMS (Common Mode Sense) and DRL (Driven Right Leg), which were used as reference and ground respectively, and the magnitude of the direct current (DC) offset of all electrodes was held below 50 μ V. Vertical and horizontal eye-movements were recorded by four additional channels, two were placed above and below the right eye, and the two others were placed at the corner of each eye. These electro-oculograms electrodes were used for monitoring vertical and horizontal eye movements to identify trials contaminated by blinks.

EEG pre-processing

Off-line analyses were performed using Letswave 5 (<http://nocions.webnode.com/letswave>) and Matlab 2012 (Math-Works, Natick, MA). Butterworth band-pass filter between 0.1 and 120 Hz (zero phase shift Butterworth filter, order 4) was applied to the continuously recorded individual participant data. Then, after

downsampling to 256 Hz, noisy channels containing other artefact than blinks (which were identified based on the recording of the electro-oculograms) were interpolated by the nearest neighboring channels (a maximum of 6 channels were interpolated per participant). In addition, epochs with large artifacts (other than cause by blinks) were excluded from analysis. In total, trial rejection did not exceed 0.59 %. Segmentation was performed, 2 s prior the start, plus 2 s after the end (-2 s to 84 s) of each trial. Afterwards, all channels were re-referenced to the common average reference, excluding the ocular channels for each subject independently.

Frequency-domain analyses

Each trial was segmented a second time, in such manner which excluded the fade-in, fade-out of all trial, and contained an integer multiple of the two frequencies, F1 and F2 (18688 time bins in total = 73 s). Then, the resulting epochs were averaged separately for each participant and each condition to decrease the stochastic noise and increase the signal that is systematic. The resulting waveforms were passed in the frequency domain by a Fast Fourier Transformation (FFT) which extracts from the EEG signal, the amplitude spectra for each electrode. The resulting frequency resolution is the inverse of the time sequence (73 s), thus, the frequency analysis yielded spectra with a very high frequency resolution (1/73 s, i.e., 0.0137 Hz between each bins) and allowed the unambiguous identification of the response at the exact frequencies of interest. We took account of variations in noise across the frequency spectrum by computing the baseline-corrected amplitude (Sbl) for each frequency bin. Specifically, we subtracted the response at each frequency by the average voltage amplitude of the 10 surrounding bins (5 on each side, excluding the 2 immediately adjacent bins and the two extreme bins; e.g., Rossion et al., 2012; Dzhelyova & Rossion, 2014). Here we used only 10 surrounding bins to avoid possible overlap between harmonics responses (e.g., responses at 10.34 Hz i.e., $5f_1$: 5×2.07 and 10.24 Hz i.e., $11f_2 - 11f_1$: $11 \times 3 - 11 \times 2.07$). Z-scores were also calculated from the amplitude spectrum to assess the significance of the response for each discrete frequency bin (e.g., 2.07 Hz, 3 Hz, 5.07 Hz etc.) by computing the difference between the amplitude at the frequencies of interest and the mean amplitude of the 10 surrounding bins (excluding the 2 immediately adjacent bins and the two extremes bins) and dividing this difference by the standard deviation of the 10

surrounding bins (Rossion et al., 2012). Frequency harmonics with a Z-score greater than 3.11 ($p < 0.001$; 1-tailed, i.e., signal > noise) were included for further analysis. A conservative threshold was used because we wanted to focus our analyses on the most robust harmonics. Based on this criterion, we determined the range of significant harmonics for the avatar, the object and their IMs separately and then we summed the baseline-corrected amplitudes of the significant harmonic responses for the avatar, the object and their IMs separately. Topographical maps were used to define regions-of-interest (ROIs; for the avatar, the object and their IMs separately) for statistical analyses. Separate statistical analyses on these ROIs were performed based on individual participants' values (baseline-corrected amplitude) using repeated measures analyses of variance (ANOVAs) with the factors Visibility (Seen vs. Unseen) x Avatar-object side (Left vs. Right). To compare the scalp distribution between the Seen and Unseen conditions, the amplitudes values were normalized (McCarthy & Wood, 1985) and finally we subtracted the Unseen from the Seen condition (Seen-Unseen).

3. Results

3.1 Behavioral Data

We compared the response times (RT) and accuracy of the detection of the two targets together (*green contour* avatar and object *soccer ball*) across the Seen and Unseen conditions. There were no significant differences of RT (Seen: $0.55 \text{ s} \pm 0.06 \text{ s}$; Unseen: $0.55 \text{ s} \pm 0.05 \text{ s}$; $t(15) = -0.965$, $p = 0.350$) and accuracy (Seen: $89.83 \% \pm 3.89 \%$; Unseen: $91 \% \pm 3.72 \%$; $t(15) = -1.521$, $p = 0.149$). The absence of significant difference provided no evidence that the level of difficulty differed across the two conditions.

3.2 EEG Data

The frequency spectrum (Figure 4.3) reveals several peaks corresponding to harmonics of the avatar and object presentation frequencies, as well as IMs (e.g. combinations of these harmonics).

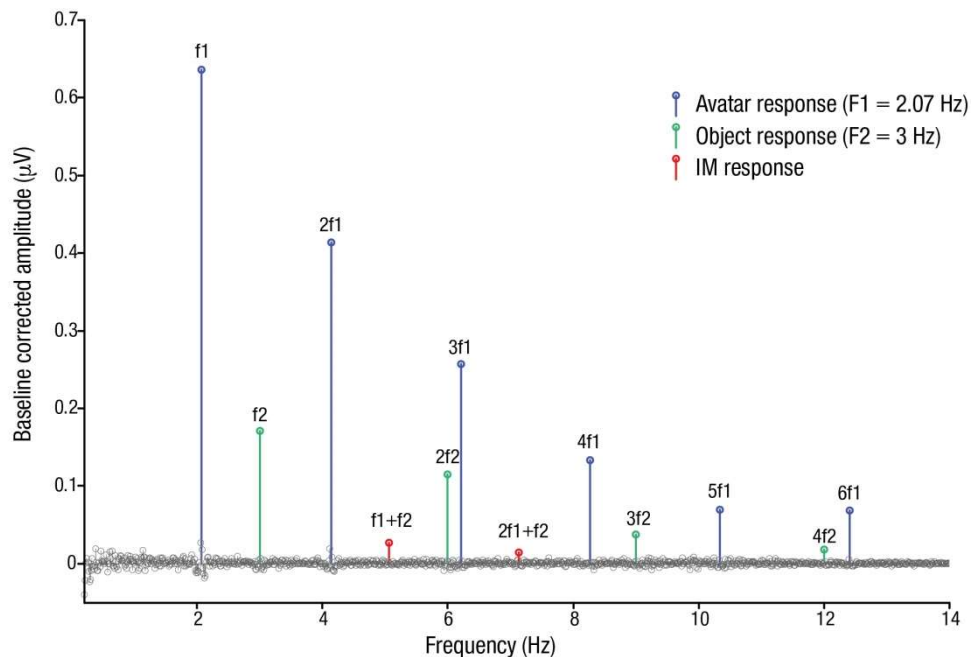


Figure 4.3. Frequency spectrum of the Grand average (all participants and conditions) over the 128 pooled channels ($f_1 = 2.07$ Hz, $f_2 = 3$ Hz).

Intermodulation components

IMs represent the common integration of the two flickering stimuli. We were particularly interested in comparing the IM responses between the Seen and Unseen conditions. According to our predictions, if only attentional cuing mechanisms are involved (submentalizing hypothesis) when participants watch a person in the vicinity of objects in the absence of explicit perspective taking instructions, then participants will simply process the spatial relationship between the avatar and the object. The IMs should therefore be unaffected (same amplitude response) by the presence of an obstacle in the avatar's line of sight. In contrast, if participants go beyond the mere processing of the spatial relation between the avatar and the object and process whether the avatar can or cannot see the object (mentalizing hypothesis), IMs should be affected (different amplitude response) by the presence of an obstacle in the avatar's line of sight. More precisely, we expected in that case stronger IM responses in the Seen than the Unseen condition.

Neural intermodulation responses when viewing an object in front of a person are affected by the presence of an obstacle in the person's line of sight

A large number of potential IM components could exist in our data. In order to select only the most robust IM components, Z-scores were calculated on the FFT grand-averaged data pooled across all electrodes and all conditions until the fifth-order IM components ($4f_1+f_2$; f_1+4f_2 ; $4f_2-f_1$; $4f_1-f_2$). We then selected for further analysis continuous significant harmonics. Ten significant IM components were significantly larger than noise level (see Table 4.1 for the IMs and their corresponding Z-scores values)

Table 4.1. The 10 significant IM with their corresponding Z-scores (one-tailed Z-score threshold set at 3.11, $p < 0.001$, i.e., signal > noise).

IM	Hz	Z-score
$f_1 + f_2$	5.07	18.76
$2f_1 + 2f_2$	10.14	4.60
$2f_1 + f_2$	7.14	10.64
$f_1 + 2f_2$	8.07	5.56
$f_1 + 4f_2$	14.07	3.56
$2f_1 + 5f_2$	19.14	6.07
$2f_2 - f_1$	3.93	9.20
$3f_2 - f_1$	6.93	3.45
$3f_1 - f_2$	3.21	5.34
$4f_1 - f_2$	5.28	3.27

The sum of the IM baseline-corrected amplitude response was contralateral to the Avatar-object side over the occipital scalp region. We selected electrodes which showed the strongest activation and defined 2 ROIs of 6 electrodes (Figure 4.4) each based on the separate grand average topographical maps of the Avatar-object side (left or right; Avatar-object left side: O2, PO8, POI2, PO10, I2, PO12; Avatar-object right side: PO7, O1, PO9, POI1, PO11, I1; see Figure S4.2 of the Supplemental Material for the electrodes positions). Based on a previous study in which we highlighted a neural signature of rapid visual perspective taking over the centro-parietal scalp region (Study 1 of Chapter 2), we choose an additional ROI (identical to this previous study; FCC1, FCC2, FCC1h, FCC2h, FCz, C1, C1h, C2, C2h, CCP1h, CCP2h, Cz, CCPz, CPz, CPPz).

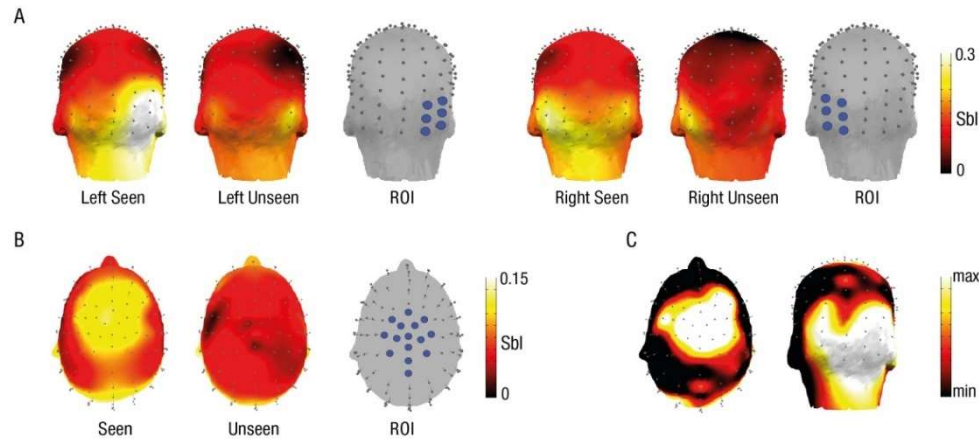


Figure 4.4. (A) Scalp topography (Sbl) of the sum of the 10 significant IM components of each condition (Visibility: Seen vs. Unseen x Avatar-object side: Left vs. Right) and the average of each condition separately with their corresponding ROI. (B) Top view of the topographical map of the grand average (all conditions) and the selected ROI. (C) Scalp distribution of the Seen and Unseen conditions (difference between the normalized topography of the Seen and Unseen condition).

A repeated ANOVA was conducted with the factors ROI x Avatar-object side x Visibility ($2 \times 2 \times 2$). A significant effect of ROI was found, with a larger response for the occipital ROI than the centro-parietal ROI ($F(1, 15) = 43.545$, $p < 0.001$, $\eta p^2 = 0.744$). There was no significant effect of the Avatar-object side ($F(1, 15) = 2.879$, $p = 0.110$, $\eta p^2 = 0.161$). A main effect of Visibility was found with a stronger response for the Seen than Unseen condition ($F(1, 15) = 15.017$, $p = 0.001$, $\eta p^2 = 0.500$). There was no significant two-way interaction ($F_s < 2.046$, $p_s > 0.173$) and a non-significant trend for a three-way interaction ($F(1, 15) = 3.865$, $p = 0.068$, $\eta p^2 = 0.205$). To explore this three-way interaction, a repeated ANOVA (Avatar-object side x Visibility) was conducted on each ROI. There was a significant two-way interaction only over the centro-parietal ROI (centro-parietal ROI: $F(1, 15) = 5.438$, $p = 0.034$, $\eta p^2 = 0.266$; occipital ROI: $F(1, 15) = 1.078$, $p = 0.316$, $\eta p^2 = 0.067$). Post-hoc t-tests revealed a trend for a larger response in the Unseen-Left than the Unseen-Right condition ($t(15) = -1.915$, $p = 0.075$; and no significant difference between the Seen-Left and the Seen-Right condition ($t(15) = 0.667$, $p = 0.515$).

The main effect of Visibility was essentially due to two of the IMs ($f1+f2 = 5.07$ Hz; $2f1+f2 = 7.14$ Hz; Figure 4.5). Therefore, we conducted the ANOVA for each of these two IM separately. For $f1+f2$, a significant effect of ROI was found, with a larger response for the

occipital ROI than the centro-parietal ROI ($F(1, 15) = 23.168, p < 0.001, \eta p^2 = 0.607$). There was no significant effect of the Avatar-object side ($F(1, 15) = 0.056, p = 0.816, \eta p^2 = 0.004$). A main effect of Visibility was found with a stronger response for Seen than Unseen ($F(1, 15) = 9.980, p = 0.006, \eta p^2 = 0.400$). There was no significant interaction ($F_s < 1.751, p_s > 0.206$). For 2f1+f2, a significant effect of ROI was found, with a larger response for the occipital ROI than the centro-parietal ROI ($F(1, 15) = 18.075, p = 0.001, \eta p^2 = 0.546$). There was no significant effect of the Avatar-object side ($F(1, 15) = 1.738, p = 0.207, \eta p^2 = 0.104$). A main effect of Visibility was found with a stronger response for Seen than Unseen ($F(1, 15) = 14.852, p = 0.002, \eta p^2 = 0.498$). There was a significant interaction ROI x Avatar-object side ($F(1, 15) = 11.728, p = 0.004, \eta p^2 = 0.439$) which was due to the fact that occipital ROI responded more for the left than the right side ($t(15) = 2.245, p = 0.040$). There was a trend for a significant ROI x Visibility interaction ($F(1, 15) = 3.121, p = 0.098, \eta p^2 = 0.172$). Post-hoc t-tests showed an overall stronger response over the occipital ROI compared with the centro-parietal ROI whether the object was seen or not seen ($t(15) = 3.952, p = 0.001$ and $t(15) = 3.012, p = 0.009$ respectively). All other interactions were not significant ($F_s < 0.232, p_s > 0.637$).

In summary, results show that the sum of the IM components were modulated by the visibility manipulation with a stronger response in the condition in which the avatar could see the object (Seen) than when the avatar could not see the object (Unseen). This effect was mostly due to 2 particular IMs (f1+f2; 2f1+f2) over 2 particular scalp regions (Figure 4.4C and Figure 4.5C). More specifically, the IM component f1+f2 over the occipital ROI was on average 25 % stronger (SD: ± 50 %) in the Seen than the Unseen condition and that same component over the cento-parietal ROI was 96% (SD: ± 228 %) stronger in the Seen than Unseen condition. The IM component 2f1+f2, on the other hand, was 3 % (SD: ± 174 %) stronger for the Seen than the Unseen condition over the occipital ROI and 171% (SD: ± 373 %) stronger for the Seen than the Unseen condition over the cento-parietal ROI.

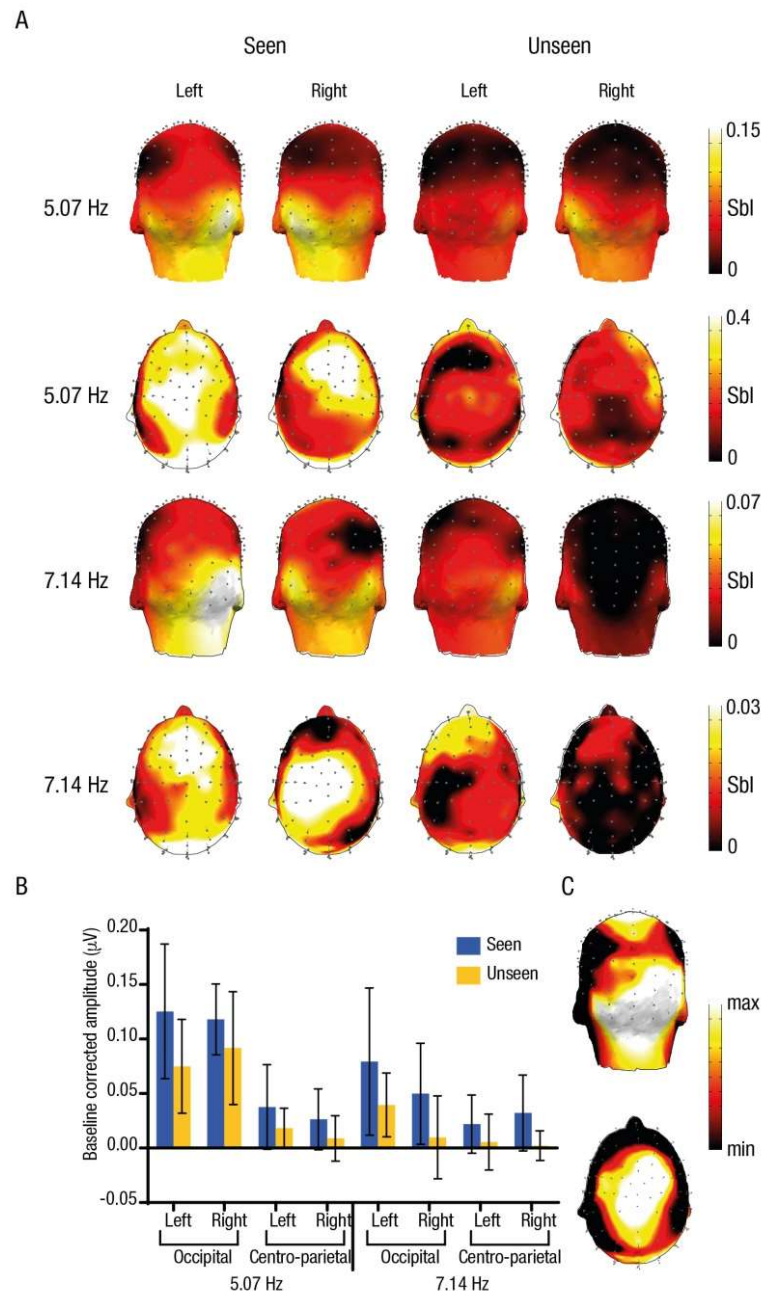


Figure 4.5. IM responses ($f_1 + f_2 = 5.07$ Hz and $2f_1 + f_2 = 7.14$ Hz). **(A)** Scalp topography (Sbl) of the IM (5.07 Hz and 7.14 Hz respectively). **(B)** Bar graph of the baseline-corrected amplitude of the two IM components over the two ROIs for the different conditions. Error bars represent mean with 95% CI. **(C)** Scalp distribution of the sum of the two IMs (5.07 Hz + 7.14 Hz) between the Seen and Unseen conditions (normalized topographie Seen condition - normalized topographie Unseen condition).

Neural intermodulation responses when viewing an object in front of a person are affected by the presence of an obstacle in the person's line of sight

Individual analysis

For the individual analysis, we cropped the individual FFT spectrum, centered at the frequencies of the 2 IMs components which were largely involved in the effect of visibility (i.e., $f_1 + f_2 = 5.07$ Hz and $2f_1 + f_2 = 7.14$ Hz) with the 10 surrounded neighboring bins on each side, and computed the baseline-corrected amplitude and summed them. Figure 4.6 depicts the baseline-corrected amplitude response over the occipital ROI. The bin of interest (the central bin) was quantifiable and above the noise (noise level is represented by the surrounding bins) in the Seen condition in virtually all the participants (14/16 participants with a significant Z-score) and in about half of the participants in the Unseen condition (9/16 participants with a significant Z-score). Furthermore, the response of interest (the central bin) was almost always stronger in the Seen than the Unseen condition.

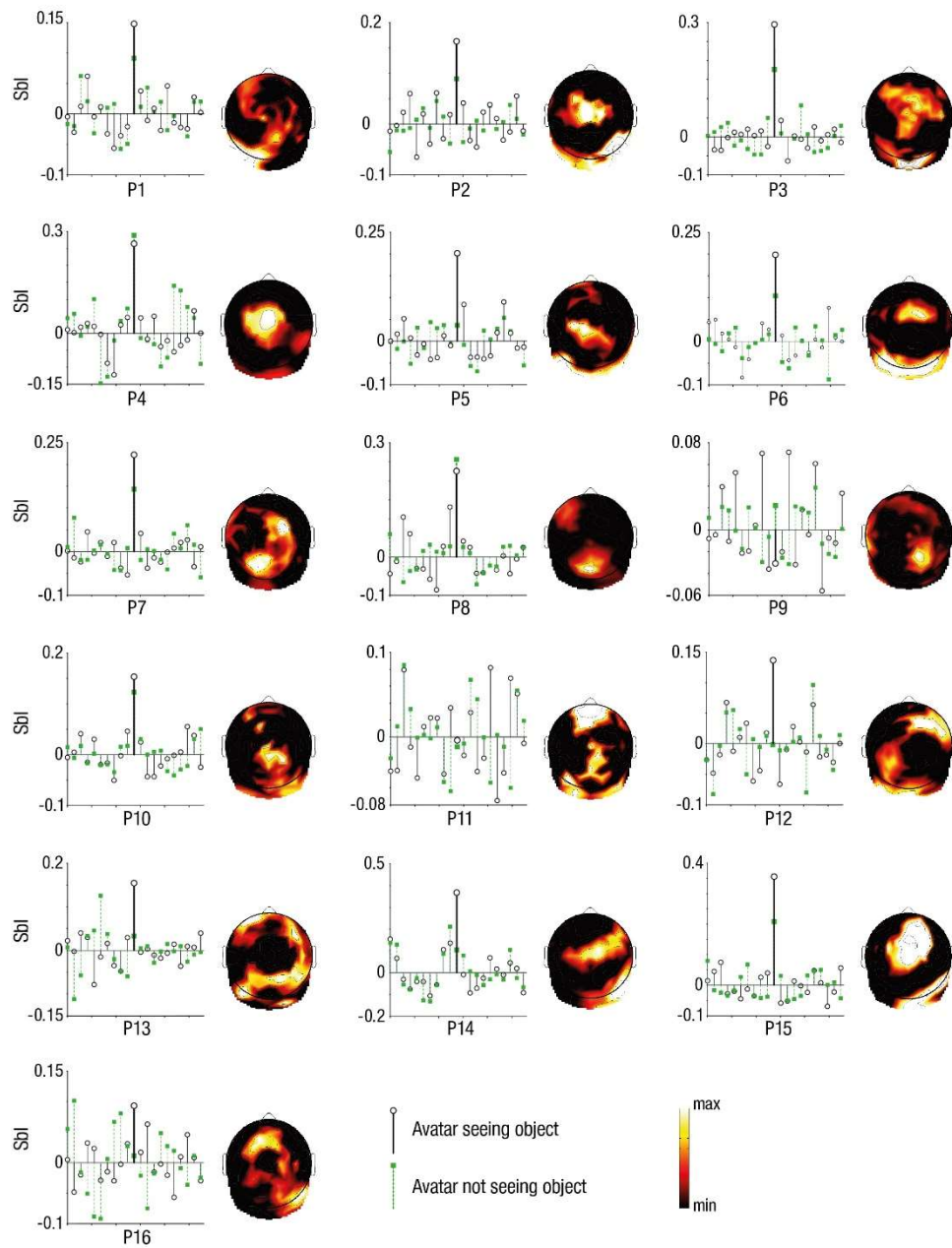


Figure 4.6. Sbl spectrum of the sum of the individual IMs responses (5.07 Hz + 7.14 Hz) over the occipital ROI in the Seen and Unseen conditions separately and the topographical map (of the difference of the normalized values of the Seen and Unseen conditions). The spectrum is centered at the response of interest surrounded by the summed of the neighboring bins, which indicate noise levels. The color scales of the topographical maps

Base responses

As visible on Figure 4.3, the avatar and object elicited clear responses at the fundamental frequencies and harmonics (e.g., avatar: $1f_1 = 2.07$ Hz; $2f_1 = 4.13$; $3f_1 = 6.20$ Hz; object: $1f_2 = 3$ Hz; $2f_2 = 6$ Hz; $3f_2 = 9$ Hz). To evaluate the number of significant harmonic responses to include in the analysis, we pooled together the 128 electrodes and then we grand averaged the FFT spectrum of all participants and all the conditions. Finally, we computed the Z-score for each frequency (as previously described).

Avatar-related responses

Avatar frequency responses were continuously significant up to the 10th harmonic ($10f_1 = 20.69$ Hz; threshold of significance placed at a Z-score of 3.11; $p < 0.001$, one-tailed, signal > noise; see Table S4.1 in the Supplemental Material). Based on this criterion, the avatar response was quantified as the sum of the significant harmonics. To do so, we cropped the individual FFT spectrum, centered at the fundamental frequency of interest (2.07 Hz) up to the highest significant harmonic ($10f_1 = 20.69$ Hz) with the 10 surrounding bins on each side, thus ending at 10 epochs. The baseline corrected amplitude was computed on each epoch separately, and then were summed together and grand averaged.

The grand average response for the avatar was spread towards the occipital region (Figure 4.7). We thus selected a large occipital ROI of 15 electrodes with the maximal summed-harmonic response (PO7, PO9, PO11, O1, PO11, I1, Oz, Olz, Iz, O2, PO12, I2, PO8, PO10, PO12).

A repeated ANOVA was conducted with the factors Avatar-object side x Visibility (2 x 2). There was no significant effect of the Avatar-object side ($F(1, 15) = 0.026$, $p = 0.873$, $\eta p^2 = 0.002$). No effect of Visibility ($F(1, 15) = 0.071$, $p = 0.794$, $\eta p^2 = 0.005$) and no significant interaction ($F(1, 15) = 0.125$, $p = 0.729$, $\eta p^2 = 0.008$).

The result showed that the avatar response was similar across conditions (Figure 4.7) and was not modulated by the object visibility to the avatar.

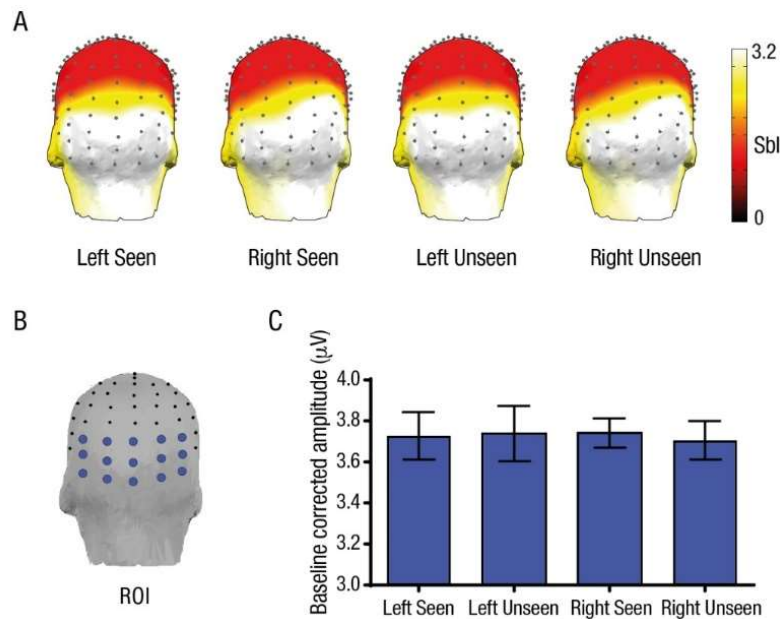


Figure 4.7. Response to the avatar frequency. (A) Scalp topography (Sbl) of the four stimulation sequences and their grand average (B) selected ROI for further analysis. (C) Bar graph of the sum of the baseline corrected amplitude of the avatar base frequency over the ROI for the different conditions. Error bars represent mean with 95% CI.

Object-related responses

For the object response, and based on the Z-score, the harmonics were continuously significant up to 6th harmonic ($6f_2 = 18$ Hz; see Table S4.2 in the Supplemental Material). Similarly to the avatar's response, the object's response was quantified as the sum of the significant harmonic. We cropped the individual FFT spectrum, centered at the fundamental frequency of interest (3 Hz) up to the highest significant harmonic ($6f_2 = 18$ Hz) with the 10 surrounded neighboring bins on each side, thus ending at 6 epochs. The baseline corrected amplitude was computed on each epoch separately, and then were summed together and were grand averaged.

As participants fixated the cross placed in the center of the screen while the object could be presented on the left or right visual field (depending on the condition), this lead to strong contralateralized responses. We selected 2 ROIs (the 2 same ROIs as for the IM response) of 6 electrodes (Figure 4.8) each based on the grand average topographical maps

Neural intermodulation responses when viewing an object in front of a person are affected by the presence of an obstacle in the person's line of sight

when the objects were presented on the right or left visual field (Avatar-object left side: O2, PO8, POI2, PO10, I2, PO12; Avatar-object right side: PO7, O1, PO9, POI1, PO11, I1).

A repeated ANOVA was conducted with the factors Avatar-object side x Visibility (2 x 2). There was no significant effect of the Avatar-object side ($F(1, 15) = 0.552, p = 0.469, \eta p^2 = 0.036$). No effect of Visibility ($F(1, 15) = 0.576, p = 0.459, \eta p^2 = 0.037$) and no significant interaction ($F(1, 15) = 0.567, p = 0.463, \eta p^2 = 0.036$).

The result showed that the response to objects was similar across conditions (Figure 4.8) and was not influenced by the object visibility to the avatar.

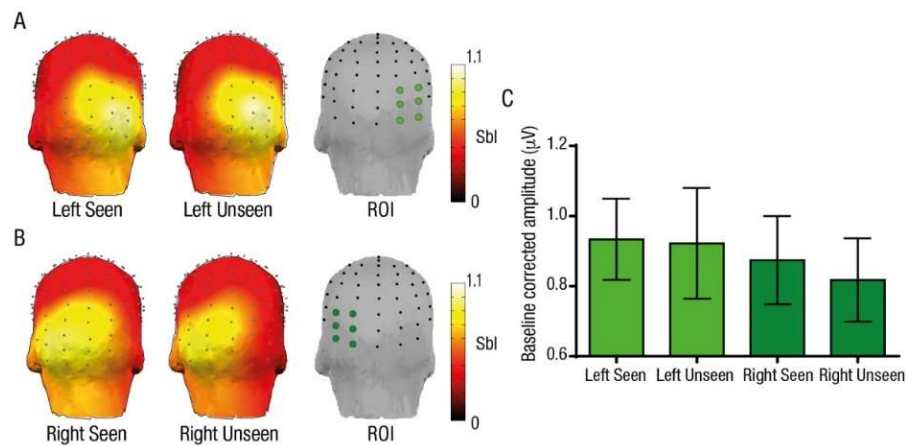


Figure 4.8. Response to the object frequency. (A) Scalp topography (Sbl) of Avatar-object left side condition and (B) right side condition with the selected ROIs for further analysis. (C) Bar graph of the sum of the baseline-corrected amplitude of the object base frequency over the contralateral ROI to the Avatar-object side for the different conditions. Error bars represent mean with 95% CI.

4. Discussion

Applying the FPVS approach has allowed us to disentangle in the EEG spectrum, the neural response underlying each frequency-tagged element (avatar and object) of the displayed stimulus. The result showed that both the avatar and the object responses remain similar whether the avatar faced the object on the right or left wall. In addition, the neural response for the avatar and the object was not affected by the presence of an obstacle in the avatar's line of sight and so, was not modulated by the object visibility to the avatar. However, our most important observation was the emergence of IM components. The IM

responses appeared at frequencies which were not present in the input stimulation and corresponded to the combinations of the frequencies $F1$ and $F2$ (e.g., $f1 + f2$). Such responses, strongly suggest that these responses came from population of neurons which received information from the two flickering stimuli. This is the first time that IM components reflect a quantifiable signature of the invisible link which associates a social stimulus (i.e., the avatar) and an object in his environment. The responses of interest were objectively identifiable in the EEG spectrum (i.e., occurring exactly at a frequency defined by the experimental design or to their combinations) and clear (i.e., well above noise level) for the majority of the tested participants. Importantly, these IMs were observed despite the absence of explicit perspective-taking instructions.

The central goal of the present experiment was to investigate whether the neural integration of the avatar and the object would be affected by the presence of an obstacle in the avatar's line of sight since different predictions can be drawn according to the submentalizing and the mentalizing hypothesis. Indeed, if in the absence of explicit perspective-taking instructions, only attentional orienting mechanisms are involved, participants should process the spatial relationship between the avatar and the object regardless of the presence or absence of an obstacle in the avatar's line of sight (submentalizing hypothesis). In that case, IM responses should be unaffected by the presence or absence of an obstacle (similar IMs amplitude responses for the Seen and Unseen conditions). In contrast, if participants go beyond the computation of the spatial relationship between the avatar and the object and process what the avatar can or cannot see (mentalizing hypothesis), then IM responses should be affected by the presence of the obstacle (different IMs amplitude responses for the Seen and Unseen conditions) because the avatar's visual experience changes in presence of an obstacle in his or her line of sight. We observed IMs whether the obstacle was present or not in the avatar's line of sight. This supports that an invisible link between the avatar and the object is created whether the object is visible or not from the avatar's perspective. However, as supported by the mentalizing hypothesis, the observed link between the avatar and the object was stronger when there was no obstacle in the avatar's line of sight. Among the different IM components observed, we identify 2 IM components ($f1 + f2 = 5.07$ Hz and $2f1 + f2 = 7.14$ Hz) which were particularly affected by the presence of the obstacle in the avatar's line of

sight. Indeed, there were both significantly stronger in the absence of the obstacle in the avatar's line of sight suggesting deeper neural integration processes when the object was not only in front of the avatar but also visible to him or her.

While our findings go against the submentalizing hypothesis formulated in terms of a purely attentional orienting phenomenon, they do not necessarily mean that the processing went as far as mentalizing. The pattern of the IM response reflects, at the very least, deeper perceptual integration in the absence of an obstacle between the avatar and the object. The IMs responses recorded reflect information association which may be only a part of the processes involved to finally get a construal of the mental experience of the avatar (e.g., he sees or not the object). We propose that instead of having dichotomous types of processing, one type depending on generic attentional processes and another type corresponding to mentalizing processes, there could be a continuum of processing of increased depth. In other words, implicit mentalizing might reflect processes somewhere in between generic attentional and mentalizing mechanisms and it is the amount of integrated information which would lead to effects compatible with the submentalizing or the mentalizing hypothesis. The bottom of the scale would depend on generic attention mechanisms and would process the mere spatial relationship between the person and the object, whereas the top of the scale would correspond to a conscious representation of the mental experience of the other person. In the setting of our experimental design, participants had to focus on the avatar and the object but they were not asked to take the avatar's perspective. Thus, our findings indicate that deeper integrated information processing than simply computing the spatial relation between the avatar and the object could be much more spontaneous than sometimes acknowledged (Santiesteban, Shah, White, Bird, Heyes, 2014; Heyes, 2014; Gardner, Hull, Taylor, Edmonds, 2017; Cole, Atkinson, D'Souza, Smith, 2017).

Furthermore, the IM response over the centro-parietal scalp region showed the highest sensitivity to the presence of an obstacle in the avatar's line of sight for both IM. This particular involvement of the centro-parietal scalp region is compatible with attentional processing of the stimuli (Ge, Matsuoka, Ueno, Iramina, 2007; Kalla, Muggleton, Juan, Cowey, Walsh, 2008) but it is also in line with the known neurofunctional basis of VPT

processing. Indeed, it may be due to processes generated in temporo-parietal brain areas, involved in the representation of the other person's perspective (Aichhorn, Perner, Kronbichler, Staffen, Ladurner, 2006; McCleery et al., 2011; Van Overwalle, 2009). It is remarkably interesting that the extension of the FPVS-EEG approach to study higher-level cognitive processing reveal more anterior located responses on the scalp than others studies which investigated "less high processing" (for a review see Norcia, Appelbaum, Ales, Cottureau, Rossion, 2015).

The pattern of results observed also provides useful cues concerning the nature of the FPVS response. First, the apparent sensitivity of the IM responses to a presence or absence of an obstacle in the avatar's line of sight while the responses for the avatar and object themselves remain similar indicates that the response to the common integration of the avatar and the object (avatar-object link) is different than the mere sum of the avatar and the object responses themselves. Indeed, otherwise the amplitude of the IMs responses observed should have been just equal to the combinations of the amplitudes responses of the object and avatar frequencies (e.g., the amplitude response at 5.07 Hz should have been equal to the sum of the amplitude response for the avatar (2.07 Hz) and the amplitude response of the object (3 Hz)). Thus, our findings support that the neural response at the fundamental frequencies reflects lower sensitivity to the stimuli (f_1 and f_2 remain similar whether the obstacle was placed or not in the avatar's line of sight), whereas IMs reflects long-range neural interaction with higher sensitivity to the stimuli (IM response were affected by the presence/absence of the obstacle). Secondly, our data also suggest that the order of the IM component can reflect different levels of the visual hierarchy processing. Indeed, when comparing the overall percentage of the response amplitude of the Seen and Unseen conditions, the second-order component (IM: $f_1 + f_2$) showed a higher sensitivity to the presence of an obstacle in the avatar's line of sight over the occipital scalp region than the third-order component (IM: $2f_1 + f_2$). Whereas, the third-order component (IM: $2f_1 + f_2$) showed a higher sensitivity than the second-order component (IM: $f_1 + f_2$) over the centro-parietal scalp region. Thereby, lower-order IM components seem more linked to lower levels of the visual hierarchy and higher-order IM components seem more linked to higher level of the visual hierarchy and thereby, certainly with higher processing level (see also Alp et al., 2017).

Finally, the centro-parietal scalp activation was also observed in our previous study which investigated explicit VPT (same ROI used for analysis; Study 1 reported in Chapter 2) with an oddball paradigm. A visual scene alternated at 2.5 Hz and depicted an avatar facing an object except on every fifth stimulus (i.e., $2.5/5 = 0.5$ Hz). At every displayed stimulus, the avatar orientation and object position randomly changed. We observed FPVS response over the centro-parietal scalp region at the frequency corresponding to the avatar not facing the object (i.e., 0.5 Hz). Such response reflects the neural discrimination between the avatar facing or not an object. However, participants to perform the task had to pay attention to the avatar's visual experience (explicit VPT instruction) and it is interesting to note that the present paradigm showed even higher sensitivity to the avatar's visual experience albeit here, participant performed an implicit VPT task (stronger amplitude response of interest). The present paradigm seems particularly well adapted to explore the neural basis in human social cognition (Frith & Frith, 2007), the inter-individual difference and could be suitable as a diagnostic tool for detecting social cognition issues during infant development (e.g., high-functioning autism, schizophrenia).

5. Acknowledgements

We thank Andrea Conte for having created the application used to test the subjects. We also thank Talia Retter, Joan Liu-Shuang and Corentin Jacques for their comments on the manuscript.

6. Supplementary material

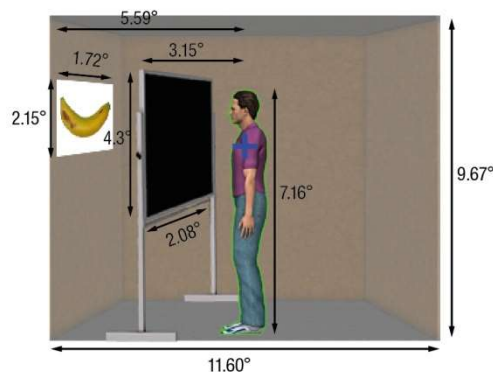


Figure S4.1. Original size of the images in visual angle at a viewing distance of 80 cm.

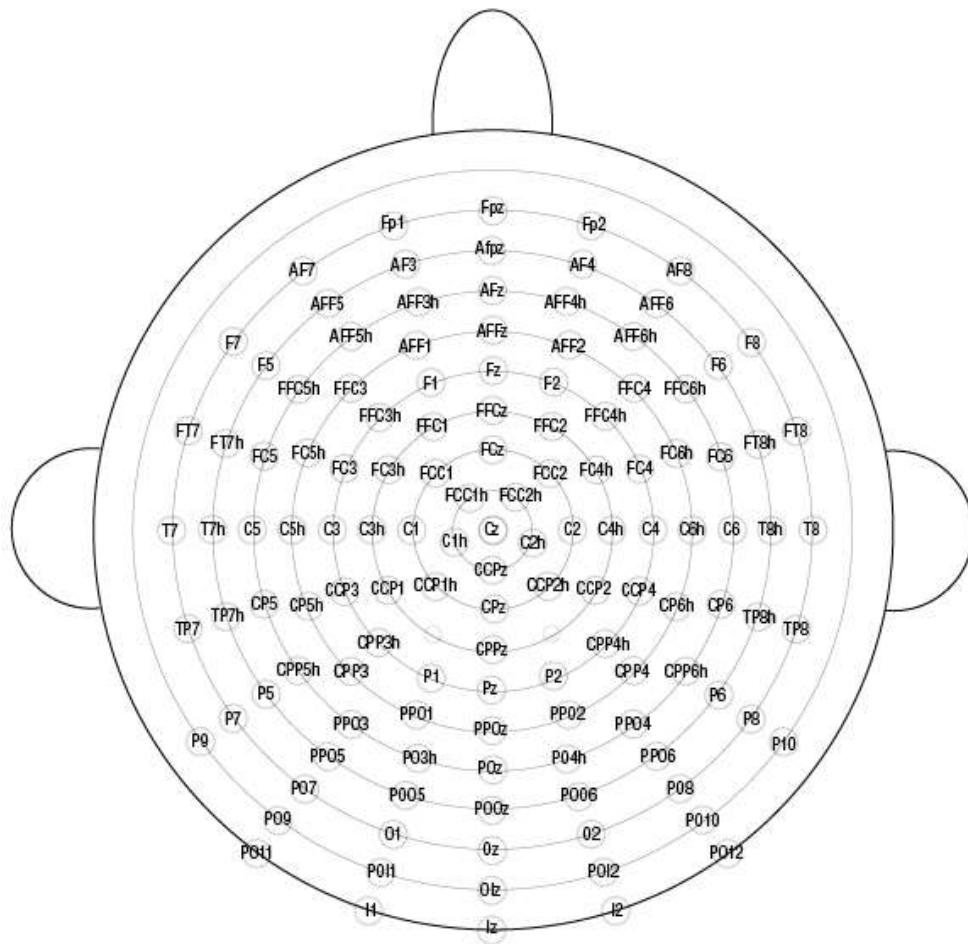


Figure S4.2. Electrodes names and position of the 128-channel Biosemi system.

Neural intermodulation responses when viewing an object in front of a person are affected by the presence of an obstacle in the person's line of sight

Table S4.1. Z-scores of the different harmonics of the avatar frequency. The Z-score was calculated on the 128 pooled electrodes from the grand average response of all conditions (one-tailed Z-score threshold set at 3.11, $p < 0.001$).

Harmonic	Z-score
$1f = 2.07$	173.04
$2f = 4.13$	259.86
$3f = 6.20$	158.45
$4f = 8.27$	75.92
$5f = 10.34$	41.49
$6f = 12.41$	87.92
$7f = 14.48$	27.94
$8f = 16.55$	31.79
$9f = 18.62$	14.65
$10f = 20.69$	5.45
$11f = 22.75$	1.79
$12f = 24.82$	8.08
$13f = 26.89$	0.97
$14f = 28.96$	0.02
$15f = 31.03$	-1.24

Table S4.2. Z-scores of the different harmonics of the object frequency. The Z-score was calculated on the 128 pooled electrodes from the grand average response of all conditions (one-tailed Z-score threshold set at 3.11, $p < 0.001$).

Harmonic	Z-score
$1f = 3$	98.74
$2f = 6$	48.77
$3f = 9$	45.03
$4f = 12$	18.13
$5f = 15$	6.26
$6f = 18$	5.27
$7f = 21$	0.50
$8f = 24$	2.06
$9f = 27$	1.88
$10f = 30$	0.96
$11f = 33$	-0.16
$12f = 36$	-0.06
$13f = 39$	1.19
$14f = 42$	1.37
$15f = 45$	0.53

Chapter 5

Discussion

1. Reminder of the goals of the thesis and the hypotheses

Humans are social agents: we represent other people's inner world and so we mentalize about others. What are the origins of such capacity? Inferring other people's visual experience is an important component to infer their mental states since what other people look at can be used as a gateway to access their inner world. For example, what someone is looking at can inform us about what that person might think, like or dislike etc. In the literature, several studies observed that when participants were merely asked to judge from their own point of view the number of targets that they could see in a scene, they were slower and more error prone when they could see a different number of targets than an avatar present in the scene. This delay of the participants' response was called altercentric intrusion and was first assumed to be caused by the spontaneous processing of the participants' and the avatar's perspective who both see a different number of targets (mentalizing hypothesis). However, an alternative explanation of the altercentric intrusion was soon proposed under the label of "submentalizing hypothesis". This alternative hypothesis assumes that participants simply process the spatial relationship between the avatar and the objects in his environment. In trials in which both the participant and the avatar can see the same number of targets, the avatar's orientation acts as a valid cue which facilitate participant's detection of the target, whereas when the participant and the avatar see a different number of targets, the avatar's orientation acts as a non-valid cue which conflicts with the target's actual location thereby delaying the detection of the target. Currently it is unknown whether when we merely look at someone, we simply process the spatial relationship between this person and the objects in his/her environment (submentalizing hypothesis), or if we go further and process the objects as seen or not seen by the person (mentalizing hypothesis). Indeed, the scientific literature provides evidence that support both hypotheses which makes it difficult to discriminate between a simple attentional effect and the fact of really processing the other's visual perspective.

The current thesis was undertaken to provide on the one hand, a new powerful methodological approach to investigate high-level socio-cognitive function and especially VPT. And on the other hand, thanks to our new methodological approach to provide new

evidence to feed into the mentalizing-submentalizing debate. The main findings will be summarized in the following section.

2. Summary of findings under the lenses of the mentalizing and the submentalizing hypotheses

Our different studies showed that the FPVS approach allows to record brain activity directly related to VPT processes. The empirical findings reported in this thesis brought new insights concerning the mentalizing-submentalizing debate. The present section discusses findings according to the two conflicting hypotheses and their divergent patterns of prediction.

Study 1 reported in Chapter 2 addressed a methodological issue. In Study 1, we were interested to see whether we could obtain a brain signature of explicit VPT with a simple EEG frequency-tagging approach (an oddball paradigm). In the oddball paradigm, participants watched scenes during a few minutes while performing an explicit VPT task (participants were explicitly asked to pay attention to the avatar's visual experience). We observed a strong EEG signal over the centro-parietal and frontal scalp regions each time the avatar depicted in the scene could not see an object that participants could see. However, it is possible that this response of interest (the discrimination response between an object seen/unseen by the avatar) simply reflected the processing of the spatial link between the avatar and the object. Thus, Study 1 of Chapter 2 does not allow us to discriminate between the submentalizing and the mentalizing hypotheses but provides the first important evidence that the FPVS approach can be used to investigate ToM.

In the literature, different studies had obtained the same altercentric intrusion whether they used an avatar or an arrow in the middle of the scene (Gardner et al., 2017; Conway et al., 2016; Santiesteban et al., 2014, 2017, Experiment 1). The submentalizing hypothesis as formulated in the literature assumes that the central cue, irrespectively of its nature (social or non-social), triggers an attentional orienting phenomenon. However, while a non-social cue (e.g., arrow) and a social cue (e.g., avatar) have similar effects on attention, this does not mean that the underlying mechanisms responsible of the effects are exactly the same. In the Study 2 reported in Chapter 3, we measured the specific brain responses

elicited while a frequency-tagged object was presented in the vicinity of an avatar versus an arrow. This manipulation was done in the absence of any explicit VPT instruction. We observed that even though the participants were not instructed to adopt the avatar's perspective (implicit VPT measure), the participants' FPVS responses corresponding to the object processing were stronger over the occipital scalp region when the object was presented in the vicinity of an avatar compared to an arrow. However, the responses did not differ whether the object was in front or behind the avatar. Interestingly, the EEG responses to the central stimulus remained similar whether it was an avatar or an arrow. Thus, while participants paid similar attention to the central cue, the way they processed the object differed. This pattern of results cannot be explained by a mere generic attentional process. Instead, Study 2 suggests a deeper processing of a visual scene that comprises social stimuli, an observation that goes against the submentalizing hypothesis.

In Study 3 of Chapter 4, for the first time in the literature we directly investigated the direct integration between an avatar and an object and wanted to quantify the invisible link which unites them in order to examine whether the link is modulated depending on whether the avatar could or could not see the object. In this study, the avatar always faced the object but without necessarily seeing it because of an obstacle which did or not obstruct the avatar's line of sight. On the one hand, according to the submentalizing hypothesis, the neural response should remain similar as long as the avatar always faced the object and that the distance between the avatar and the object remains constant. On the other hand, according to the mentalizing hypothesis, participant process more than the mere spatial relationship between the central cue and the object. In that case, and as observed in Study 3, the neural response was expected to be affected if the avatar's visual experience was modified by an obstacle in the avatar's line of sight. Therefore, again, our findings are not in line with the submentalizing hypothesis according to which the processing is a purely attentional orienting phenomenon.

All together our findings strongly reject the submentalizing hypothesis as purely attentional orienting phenomenon and suggest that participants processed more than just the spatial relation between the avatar and the object. However, as further discussed in the next section, our findings cannot be taken as a signature that participant attributed "full-

blown” mental states to the avatar since alternative interpretations of our results are also possible.

3. The alternative interpretations

In the literature, several studies have found pattern of results supporting the submentalizing hypothesis. Thereby, even if our findings reject the submentalizing hypothesis as purely attentional orienting phenomenon and support that implicit VPT involved perspective-taking mechanisms as proposed by the mentalizing hypothesis, it is important to consider whether there are any other explanations for our pattern of results.

Indeed, we acknowledge limitations with the interpretations of our results. For example, an oddball paradigm allows to discriminate between two types of stimuli (e.g., a face vs. a non-face or an object in front of an avatar vs. an object behind the avatar). However, as previously mentioned, our oddball paradigm as it is (Study 1 reported in Chapter 2 and in the Appendix 2) does not allow to discriminate between the mentalizing and the submentalizing hypothesis. Indeed, it is possible that the response of interest observed, i.e., the discrimination response between an object seen/unseen by the avatar simply reflects lower-level processing than VPT. Indeed, the mere processing of the spatial link between the avatar and the object could explain ours results without necessarily involving perspective-taking mechanisms. To explore further the nature of the processing involved, an interesting oddball follow-up study could be to present an avatar facing different targets but not necessarily seeing all of them because some would be hidden by an obstacle (as the stimuli used in Study 3). The oddball paradigm would be defined by a sequence of images (e.g., an avatar facing an object on the left or right wall but not seeing it because of an obstacle placed in the avatar’s line of sight). Inside the sequence, only at a fixed interval (and thus periodically and at a certain frequency), a different type of stimuli would be introduced (e.g., an avatar facing an object on the left or right wall and seeing it because the obstacle would not be placed in the avatar’s line of sight). Thus, the spatial relation between the avatar and the object remains constant and what defines the oddball is that only at periodic intervals the avatar can really see the object. An absence of oddball response in that paradigm would strongly support the submentalizing hypothesis, whereas observing an oddball response would strongly support the mentalizing hypothesis because

it would strongly suggest that participants processed more than just the spatial relation between the avatar and the object. This approach could provide an interesting replication of Study 3's findings. However, it is possible that the participant need more time to fully process the complexity of the stimuli (the position of the barrier) and so the visual stimulation frequency should be slowed down. Contrarily to the Study 3 of Chapter 4, the oddball paradigm offer to explore the speed of the VPT processing. For example, in Study 1 of Chapter 2 (and in Appendix 2), we presented the stimuli at 2.5 Hz which corresponds to a displayed of 400 ms per stimulus. Thus, participants were able to rapidly discriminate whether an avatar was facing or not an object but what if we increase the speed (see Appendix 3)? Thanks to this particularity, the oddball paradigm offers exiting future investigations (see below).

Our others paradigm and their results can lead to alternative interpretations which does not necessarily support the mentalizing hypothesis. For example, according to both the mentalizing and the submentalizing hypotheses, we expected a stronger response for an object in front than behind an avatar (or an arrow). However, we never found in our studies (Studies 2 and 3 reported in Chapter 3 and 4 respectively and in our different pre-tests, see Appendix 3) a significant different response whether the object was in front or behind the central stimulus (except in Appendix 1 in which the task instruction favored such difference between the object placed in front or in the back of the central stimulus). Thereby, this seems to highlight a limitation in the sensitivity of the FPVS response to measure implicit attentional shift. Furthermore, in Study 2 of Chapter 3, if the participant processed more than the mere spatial relationship between the central cue and the object and processed the object as "seen" or "not seen" by the avatar, we should have observed an interaction between the nature of the central stimuli (avatar or arrow) and the object being in front or behind the central stimuli (e.g., a stronger increase for an object in front of an avatar than an arrow). The absence of such interaction in Study 2 of Chapter 3 is a strong limitation to the claim that implicit VPT actually involved mentalizing mechanisms. Instead, it could be argued that the observed stronger response for the object presented in the vicinity of an avatar compared to an arrow would be merely due to the objects which become more engaging when put in a social context with a human agent because a person might do something with it (spectator to an action without involving perspective taking

processes), compared to the arrow that is just another inanimate object. Our paradigm which used an arrow as central cue instead of an avatar to investigate whether it was the directional rather than the agentic feature of the agent which was important could be addressed in another way. In fact, Study 3 reported in Chapter 4 show that IMs responses are particularly interesting neural responses to investigate the processes underlying implicit VPT. Thereby, the frequencies used in the Study 2 of Chapter 3 could be adapted in order to explore IMs responses (e.g., by using the same frequencies of the Study 3 of Chapter 4). Thus, as the IMs responses represent the common integration of the central cue and the object, we could expect that it is the IMs responses which should actually show sensitivity whether an object is in front or behind the avatar/arrow.

Finally, alternative interpretations can be proposed for the pattern of results observed in Study 3 of Chapter 4. Indeed, we observed that the IM responses were affected when the avatar's visual experience was modified by an obstacle in the avatar's line of sight. This strongly supports the mentalizing hypothesis. Nevertheless, it is possible that our pattern of results simply reflects a different type of perceptual integration. One alternative is that the visibility of the barrier changed depending on whether it was placed in the avatar's line of sight or not. Indeed, when the obstacle obstructed the avatar's line of sight, it was placed at the level of the avatar's head and the object, i.e., a location which corresponds also to the height of the participant's eye fixation point. Thereby, the obstacle is very salient. In contrast, when the obstacle is not placed at the level of the avatar's head and the object (i.e., when the obstacle does not obstruct the avatar's line of sight) it is placed below the participant's eye fixation point and could be less visible. One way to control for this would be to also frequency-tag the barrier (note that a fast frequency must be used otherwise the barrier might not always mask the object even if it is placed on the avatar's line of sight). In our study, we observed the same response for the fundamentals frequencies (response for the central cue and the object) whether the avatar sees or not the object. Thus, we could also expect that the response at the barrier frequency would remain unchanged when the barrier is placed below or at the level of the avatar's line of sight. This absence of difference at the barrier frequency between conditions would allow us to discard the barrier saliency interpretation. Even if this alternative interpretation was to be discarded, it could still be argued that the IM responses do not actually reflect perspective taking mechanisms.

Indeed, the IM responses could simply reflect a binding of objects without any mentalistic characteristic (for example a similar binding could be found between two neighboring objects that share semantic features). Comparing the IM responses when the central stimulus is an avatar versus an arrow would thus be yet a further step to understand the type of integration processes are at play.

Altogether, our results strongly reject the submentalizing hypothesis as purely attentional orienting phenomenon. However, it is important to be cautious about interpreting this as evidence that participant actually attributed mental states to the avatar. The implications of our findings and how they can be conciliated with the findings of the literature are further discussed in the next sections.

4. The continuum scale model

The literature generally considers that the mentalizing and submentalizing hypotheses are mutually exclusive. Thereby studies interpret their results as if they reflect a binary system, we do mentalize or we do not. However, if our findings strongly support the idea that implicit VPT involves more than a purely attentional orienting phenomenon, the current empirical evidences cannot be taken as a signature of actually attributing “full-blown” mental states to the avatar. Instead, in the current section I will propose another theoretical framework, the continuum scale model (Figure 5.1; see also Christensen & Michael, 2016; Westra, 2017) which conciliates divergent findings observed in the literature.

We propose to re-think the “Two-system” model of ToM (Apperly & Butterfill 2009; Apperly 2011; Butterfill & Apperly, 2013). The “two-system” model proposes the existence of two distinct processes which exist alongside one another. One would rely on “low-level” and implicit mechanisms, and another one would rely on “high-level” and explicit mechanisms. According to this model, the implicit mechanisms would allow us to track others’ mental states in a relatively spontaneous manner. The underling processing would not be flexible albeit, fast and efficient. The second system on the other hand would be more cognitively effortful and linked to executive function but at the same time it would be more flexible. However, several studies showed results which cannot explain the “two-

system” model as proposed by Apperly & Butterfill (2009). For example, according to this model level-1 VPT rely on “low-level” mechanisms which are fast, efficient but not flexible and should be unaffected by external factors. However, Furlanetto et al. (2016)’s results suggest that level-1 VPT can be affected by external factors. Indeed, they observed that participants spontaneously took into account whether the avatar sees or not targets depending on whether he wore transparent or opaque goggles. To explain these results one must assume that participants used their knowledge about which goggle is transparent and which one is opaque. Similarly, as presented in the introduction, there are also different studies which suggest that participants who performed a gaze-cuing task might be affected by external factors (e.g., Teufel et al., 2010a, 2010b; Wiese et al., 2012). Thereby, level-1 VPT seems to be relatively flexible and more likely to rely on both “low-level” mechanisms and “high-level” mechanisms which operate in concert. We propose that VPT would rely on the dynamic combination between fast and efficient processes and mentalizing reasoning. In fact, instead of having a dichotomous process, one depending on generic attentional processes and another corresponding to the mentalizing processes, there could be a continuum of processing of increasing depth. More specifically, there would be an interactive relationship between attention processes and mentalizing processes from which would depend the amount of integrated information about the other’s perceptual experience. For example, our results showed evidence that participants can integrate a wide number of information from a visual scene. First of all, Study 2 of Chapter 3 showed that participants process differently an object stimulus in the presence or in the absence of a social cue (avatar vs. arrow). Furthermore, the Study 3 reported in Chapter 4 showed IM responses both when the avatar saw and did not see the object, however, the IM responses were stronger when the object was seen compared to when it was not seen by the social cue. Such observations strongly suggest that participants can indeed process information present in the visual scene at different level/depth/intensity. A second interesting observation comes from the fact that we observed different scalp regions involved in processing the social information (the occipital and centro-parietal scalp regions). This suggest that social information is processed (at least in part) by different population of neurons which taken together increase the amount of integrated information and reflect how deeply people might process the different information present in the visual scene.

The integrated information continuum scale model allows to conciliate the conflicting results presented in the introduction about the mentalizing/submentalizing debate, with some studies reflecting a mere attentional effect whereas other studies reflect a deeper processing which is akin to what is called mentalizing. In fact, according to the continuum scale model, it would be the amount of integrated information which would reflect more a submentalizing effect or a mentalizing effect. Crucially, the representation of someone else's inner world would not be only generated by bottom-up processes. Top-down factors would also influence the amount of integrated information. Indeed, the representation of someone else's mental state is a dynamic process whereby we constantly adjust our on-line representation based on bottom-up information but also top-down mechanism. Such top-down mechanisms include stored knowledge which will help to use past experiences and make associations with the current situation to construct more accurate representations of the other person's mental state. Some top-down processes (e.g., particular aspects of knowledge) would not be cognitively demanding and would permit to more accurately adjust the inferred mental states. Thereby, the amount of integrated information would be greatly modulated by top-down mechanisms. For example, at a certain time point, the motivation, the circumstances, the social mind set, the emotional state, the presence of a social agent will influence the observer's sensitivity to implicitly integrate a wide number of different information and to process up to a certain level the social information. The influence of these top-down mechanisms can explain several conflicting results found in the literature. More particularly, within a social context (e.g., when interacting or competing with another person, or when alternating between our own and the other's perspective), participants will more deeply associate different social relevant information together. This will be reflected in a pattern of data that fits with the mentalizing hypothesis (e.g., Furlanetto et al., 2016; Baker et al., 2015; Freundlieb et al., 2016; Bukowski et al., Bukowski et al., 2015, 2016; Ferguson et al., 2016). If the social context is low (e.g., participants only need to concentrate on their own perspective through the experiment), the treatment of the stimuli will strongly rely on domain general attentional processes. Participants will integrate basic information and not necessarily make associations between socially relevant information; hence the pattern of data will essentially reflect low-level processes (i.e., attentional effect) and suggest that there is no mentalizing per se but only submentalizing

mechanism that simulate the effects of mentalizing (e.g., Cole et al., 2016; Conway et al., 2016). In our studies, participants' social mind set was certainly increased by the fact that we always asked them to paid attention to the avatar. Thereby, we put ourselves in a favorable position to measure a certain level of processing that goes beyond mere generic attentional processing.

Investigations of level-2 VPT (i.e., how someone else would see something from a different position) offer a second illustration of the importance of the top-down processing. Indeed, studies which used a live human being instead of an avatar to investigate level-2 VPT found results which support that participants adopt more spontaneously the other perspective (Elekes, Varga, Király, 2016; Freundlieb et al., 2016; Surtees et al., 2016a vs. Surtees et al., 2016b; see also Westra 2016). The presence of another live human being is more likely to affect the participants' social mind set and increase their motivation to adopt the other's perspective which is in line with the continuum scale model. Furthermore, with sufficient level of motivation it will be possible to infer mental states to abstract shape. For example, participants who watch abstract film of simple shapes roaming around are quick to see a story unfold (Heider & Simmel, 1944) with the different moving shapes having a goal and possessing qualities of an intentional mind. Osaka et al. (2012) conducted an fMRI study in which participants watch animations modeled on that of Heider & Simmel (1944) and manipulated the degree of intention from low to high. They observed that different brain region also involved in the ToM network were activated while participants watched the high-intention animations (posterior STS, inferior temporal gyrus, inferior frontal gyrus, premotor, temporal pole, supramarginal gyrus, and superior parietal lobule), in contrast lower-intention animations did not activated these ToM network regions (occipital, lingual, and middle frontal gyri). The high-intentional animations are more likely to increase the participants' motivation to attribute qualities of an intentional mind to the moving shapes. According to the continuum scale model, top-down processing (e.g., motivation) would contribute to the stronger integration of the social information by the ToM network (see also Castelli, Happe, Frith, 2000; Grossman et al., 2005; Pavlova, Guerreschi, Lutzenberger, Krägeloh-Mann, 2010). I propose that the ToM network activation could actually reflect the association between the different social information together which would be necessary to attribute a "full-blown" mental states to the other. If the treatment of the social stimulus

strongly relies on generic attention, the ToM network would not be activated. Whereas, if the treatment of the social stimuli goes deeper, additional layers of processing would be involved to integrate the social information together which would correspond to the activation of the ToM network.

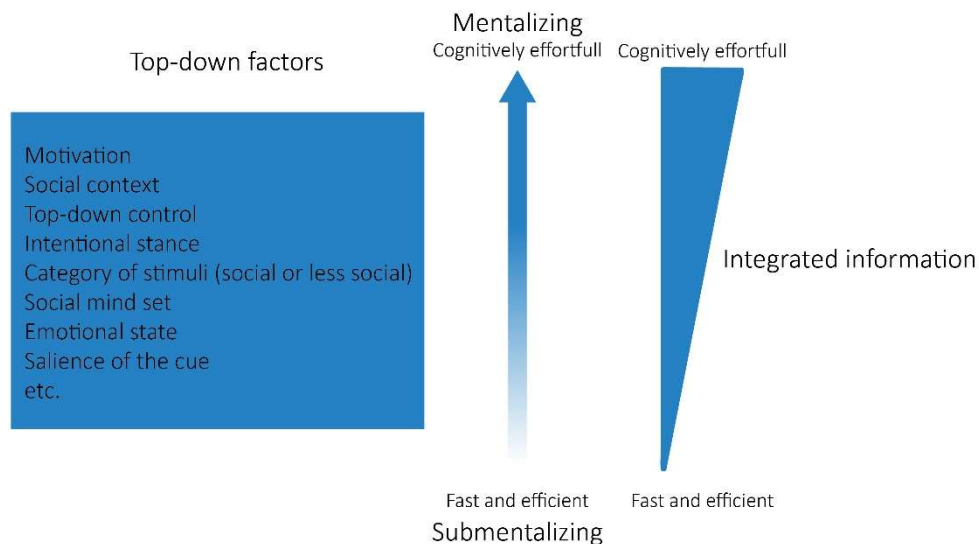


Figure 5.1. Representation of the integrated information continuum scale model. We propose that between the submentalizing and the mentalizing processes there could be a continuum of processing of increased depth. The bottom of the scale would correspond to the process of the mere spatial relationship between the central cue and the object, whereas the top of the scale would correspond to a conscious representation of the mental experience of the other person. In addition, top-down factors would strongly influence the amount of integrated information processed implicitly and explicitly.

From my point of view, there is maybe an issue in the terminology, mentalizing and submentalizing corresponding to the two extreme of the scale whereas when we implicitly process other's perspective we are more likely to be on the road between these two extreme positions. Thereby, it would be difficult to clearly draw a cut-off line between the "low-level" mechanisms and "high-level" mechanisms. Instead, we should maybe talk more about low or high social engagement and only use the term "mentalizing" in studies which explicitly ask participants for a conscious representation of the mental experience of the other person (e.g., explicit mentalizing task). The bottom of the continuum scale model would depend on generic attention mechanisms which regards how attention is directed toward relevant stimuli, for example, an arrow which is then spontaneously driven to the

direction induced by the directional cue (e.g., in front of the arrow). Deeper processing would involve social attention mechanisms which certainly play a fundamental role in the first steps of the mentalizing processes. Finally, the top of the scale would correspond to a conscious representation of the mental experience of the other person. For example, in the Study 1 of Chapter 2 and in Appendix 2, participants received explicit (Study 1) and implicit (Appendix 2) VPT instruction (high and low social engagement task respectively). For both conditions, we showed a brain response reflecting an objective signature that participants discriminated between an object placed in front or behind an avatar but the explicit VPT condition lead to stronger amplitude responses (at the frequency of interest and its harmonics) than the implicit VPT condition. This suggests that the different FPVS amplitude responses reflect to some degree different depth of processing of the avatar's perspective. In addition, the observed scalp activation in the implicit VPT condition was a subset of the scalp activation observed in the explicit VPT condition. This suggests that implicit VPT reflect processes somewhere between attentional and mentalizing mechanisms.

Finally, from my perspective, the integrated information continuum scale model is an ecological model. Indeed, we are often surrounded by several persons (e.g., in the street) and it is not necessarily relevant or cost-effective to infer mental states to everyone. Instead, we will treat other individual to a certain level, with different sensitivity, for example an eye contact between an observer and a gazing person will create a social connection and increase the observer's social mind set which will certainly engaged the observer to process additional information about the gazer and implicitly (or not) represent more deeply the other's mind set (note that it would be also true for the gazer watching the observer). Moreover, if the integrated information continuum scale model emerged principally from our studies which investigated visual perspective taking, it can be asked whether this model is valid for other type of mental states. I postulate it is the case and in the literature, studies which show results which support the mentalizing hypothesis (or not) were in fact dependent on the amount of social information integrated.

5. Strengths and limitations of the FPVS approach

The FPVS technique has many advantages which are interesting to remember and discuss further before elaborating on future research perspectives. One advantage of the

FPVS approach is that the response of interest can be objectively identified at the experimentally defined stimulation frequency and its harmonics. This is a valuable asset of the technique because as already emphasize, in the scientific literature, the subjectivity in the methods and the interpretations of the results led to difficulties in the replicability of the studies which slow down our understanding of the targeted mechanism and slow down the advancement of the scientific knowledge. In general, the use of a highly objective approach can avoid possible ambiguity in future studies and greatly help the understanding of the mechanism of interest.

In addition, another advantage of the FPVS approach is that it allows to measure the response of interest with a high signal to noise ratio (SNR). The signal of interest is concentrated on a particular frequency (and its harmonics) while the noise is distributed over the entire spectrum. Therefore, the localization of the response of interest may be clearly visible on the scalp of each subject as well as its identification in the EEG spectrum (in the frequency domain). Similarly, this resistance to the noise would help to obtain a good signal quality when testing infants or animals. One more particularity of the FPVS approach is that it allows to record a measure that is relatively behavioral-free and thus, it is possible to obtain an implicit measure of the investigated mechanism (Studies 2 and 3 of Chapter 3 and 4 respectively and Appendix 2). Indeed, the response of interest can be measured in the absence of an overt behavior as participants are only required to watch a train of images. Again, investigating a brain mechanism in the absence of an overt behavior would be particularly interesting when testing infants or animals. Finally, another property of the technique is that by frequency tagging multiple simultaneously presented stimuli it is also possible to investigate their separate and common integration which opens the way towards interesting follow-up studies (see below for examples).

However, a primary limitation of the use of the FPVS approach is the difficulty to preserve the timing information (Regan, 1989; Luck, 2005). Indeed, FPVS approach consists in presenting stimuli at fast rate. Therefore, the EEG measured response over time is not locked to one stimulus. There is no clear beginning or end of the brain response to one particular stimulus of the visual sequence. In addition, there are several different aspects of the FPVS response which would deserve to be investigated on the one hand to facilitate

paradigm development (e.g., which frequencies to use; see appendix 3) and on the other hand to envisage a more comprehensive interpretation of the FPVS responses. For example, as explained in the introduction, the fast fourier transform (FFT) used in our analysis, converts the EEG signal from the time domain to the frequency domain (and vice versa) and the different harmonics (in the frequency domain) of the FPVS response correspond to the shape of the response in the time domain. Thereby, in order to have the “full” response of interest, it seems justified to sum the harmonics of the response of interest (Retter & Rossion, 2016). Nevertheless, their respective responses on the scalp often change which suggests that they may correspond to different underlying processes. Thereby, the harmonics of a response would deserve to be also explored separately for a more comprehensive interpretation of the investigated mechanism.

Despite the limitations, it is interesting to note that the present thesis brings new insights regarding the FPVS response. In Study 3 of Chapter 4, the quantification of the IM responses (i.e., the brain signature of the common integration of the avatar and the object) showed stronger amplitude response over the occipital than over the centro-parietal scalp region whereas the opposite was found concerning the sensitivity to the avatar visibly (i.e., whether the avatar could or could not see the object), namely a stronger response in the centro-parietal than in the occipital scalp region. This suggests that a higher amplitude response over a scalp region does not necessarily reflect its higher sensitivity for the investigated function.

In the next section I will propose different axes of ToM investigation which could benefit from the FPVS approach.

6. Future studies

The FPVS technique used in all the studies of the present thesis opens new ways to explore social cognition and more generally to study high-level cognitive functions (see also Appendix 2 and 3). As example, I will propose extension studies for which the use of the FPVS can be particularly interesting.

Our results showed that besides a robust group effect, the responses were significant and quantifiable for most individual participants. Furthermore, a main advantage of the

FPVS approach is that the response can be recorded in the absence of overt behavior, allowing to test infants or clinical populations (e.g., in infants; de Heering & Rossion, 2015). As presented in the introduction, there is a debate whether infants have or not implicit mentalizing abilities. It could be interesting to conduct our studies with infants to investigate how innate VPT processing actually is. From what minimal age, can we actually find FPVS responses which show that infants represent other's visual perspective (Kampis et al., 2015)? It could prove insightful to examine the neural underpinnings of the implicit processing of others' visual perspective in infants (e.g., do infants and adults show the same response profiles, the same topographical maps?). Investigating these questions might help gaining insight into how VPT develops across the ages (e.g., how the FPVS response changes later in childhood around 4 years old, i.e. a landmark developmental stage at which children seem to gain new mentalizing abilities). Similarly, our studies might be used to investigate VPT in other animals such as chimpanzee and could inform about its phylogenetic emergence.

A second interesting set of follow-up studies could explore whether the use of our approach could be used as a diagnostic tool to detect early autism spectrum disorder (ASD) symptoms. Indeed, ASD individuals are known to have impairment in social cognition. Furthermore, the literature suggests that low-level and implicit ToM systems are specifically impaired in cases of ASD individuals (Callenmark, Kjellin, Rönqvist, & Bölte, 2014). Thereby, we can expect that ASD individuals will not present the same FPVS responses in our studies than control participants. Does an individual with ASD will show (as in Study 2 of Chapter 3) an object response modulated depending on whether the object is displayed in the vicinity of an avatar or an arrow? Furthermore, some studies have found in ASD individuals an intact level-1 VPT but an impaired level-2 VPT. However, with the only behavioral approach it is difficult to dissociate whether the pattern of results reflects a mere attentional shift and so if ASD individuals have an intact or an impaired level-1 VPT. I think that the FPVS technique could help to better understand the ASD symptoms and ensure that the intact level-1 VPT found in ASD individuals cannot merely reflect an attentional shift. For example, by testing ASD individuals with a similar paradigm as the one used in Study 3 of Chapter 4 we could investigate whether the IMs amplitudes are sensitive to whether the avatar can or cannot see the object. If the IM responses remain similar whether

the avatar can or not see the object would suggest that ASD individuals have in fact, an impaired level-1 VPT and that the intact level-1 VPT found in previous study in ASD individuals might reflect an attentional shift.

A third axis of research would investigate the continuum scale model. For example, according to this model several factors such as the motivation, the emotion, the social context will influence the amount of integrated information. Thereby, these factors should modulate the amplitude of the FPVS response. Future studies could clarify what influences perspective taking performances. Furthermore, we can hypothesize that the different factors which would influence perspective taking performances are not all equivalent and so some of them could influence more or less the FPVS response (do these factors have the same impact on the FPVS response across individuals or in individuals with ASD?).

A fourth axis of research would investigate inter-individual or gender ToM abilities difference. Indeed, more in-depth investigations of inter-individual or gender difference can be carried out thanks to the FPVS approach. For example, in the Study 1 of Chapter 2, we used an oddball paradigm with a visual stimulation at 2.5Hz, we could then investigate whether if increasing the speed of the visual stimulation (e.g., 3 Hz, 3.5 Hz, 4 Hz, etc.) we can still find responses at the periodic response of interest in some individuals (or gender), and if these individuals present better performances on some ToM task and more generally show better social abilities in their everyday life. A second study could investigate whether we can track on-line more than one visual perspective and if some participants would be able to track more perspectives than other participants. For example, we could replicate the Study 3 of Chapter 4 with more than one avatar (i.e., by frequency tagging the object and several avatars at different frequencies) and investigate if we can observe IM responses for each frequency-tagged avatar and object. We can then also examine whether the IM responses are sensitive whether the avatar can or cannot see the object. Other questions could also be investigated: can we identify distinct underlying processes involved during VPT (e.g., a particular harmonic of the FPVS response which would be linked with a specific aspect of the stimuli). Can the sensitivity of the different processes vary across different individuals/gender?

7. Conclusion

The current work used for the first time the FPVS approach to study VPT. Our findings led us to propose another theoretical framework, the continuum scale model which considers that submentalizing and implicit mentalizing are not dichotomous processes but reflect a continuum of processing of increased depth. The amount of integrated information would reflect submentalizing or mentalizing processes and would be influenced by top-down factors (as motivation, social context, social mind set, etc.). Such a new framework allows us to explain the conflicting results obtained so far in the literature in the context of the submentalizing/mentalizing debate. This novel theoretical framework also leads to testable predictions that could be examined with the FPVS approach. Overall, this thesis provides a promising new approach for exploring the neural basis in human social cognition (Frith & Frith, 2007) and their inter-individual difference.

References

- Abraham, A., Rakoczy, H., Werning, M., von Cramon, D. Y., & Schubotz, R. I. (2010). Matching mind to world and vice versa: Functional dissociations between belief and desire mental state processing. *Social Neuroscience*, 5(1), 1–18. <https://doi.org/10.1080/17470910903166853>
- Adrian, E. D., & Matthews, B. H. C. (1934). The interpretation of potential waves in the cortex. *The Journal of Physiology*, 81(4), 440–471. <https://doi.org/10.1113/jphysiol.1934.sp003147>
- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage*, 30(3), 1059–1068. <https://doi.org/10.1016/j.neuroimage.2005.10.026>
- Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., & Ladurner, G. (2008). Temporo-parietal Junction Activity in Theory-of-Mind Tasks : Falseness , Beliefs , or Attention. *Journal of Cognitive Neuroscience*, 21(6), 1179–1192.
- Akiyama, T., Kato, M., Muramatsu, T., Saito, F., Umeda, S., & Kashima, H. (2006). Gaze but not arrows: A dissociative impairment after right superior temporal gyrus damage. *Neuropsychologia*, 44(10), 1804–1810. <https://doi.org/10.1016/j.neuropsychologia.2006.03.007>
- Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, 4(7), 267–278. [https://doi.org/10.1016/S1364-6613\(00\)01501-1](https://doi.org/10.1016/S1364-6613(00)01501-1)
- Alonso-Prieto, E., Belle, G. Van, Liu-Shuang, J., Norcia, A. M., & Rossion, B. (2013). The 6Hz fundamental stimulation frequency rate for individual face discrimination in the right occipito-temporal cortex. *Neuropsychologia*, 51(13), 2863–2875. <https://doi.org/10.1016/j.neuropsychologia.2013.08.018>
- Alp, N., Kogo, N., Van Belle, G., Wagemans, J., & Rossion, B. (2016). Frequency tagging yields an objective neural signature of Gestalt formation. *Brain and Cognition*, 104, 15–24. <https://doi.org/10.1016/j.bandc.2016.01.008>
- Alp, N., Nikolaev, A. R., Wagemans, J., & Kogo, N. (2017). EEG frequency tagging dissociates between neural processing of motion synchrony and human quality of multiple point-light dancers. *Scientific Reports*, 7(January), 44012. <https://doi.org/10.1038/srep44012>
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, 7(4), 268–277. <https://doi.org/10.1038/nrn1884>
- Andersen, S. K., Hillyard, S. A., & Müller, M. M. (2008). Attention Facilitates Multiple Stimulus Features in Parallel in Human Visual Cortex. *Current Biology*, 18(13), 1006–1009. <https://doi.org/10.1016/j.cub.2008.06.030>
- Andersen, S. K., & Müller, M. M. (2010). Behavioral performance follows the time course of neural facilitation and suppression during cued shifts of feature-selective attention. *Proceedings of the National Academy of Sciences of the United States of America*, 107(31), 13878–82. <https://doi.org/10.1073/pnas.1002436107>
- Andersen, S. K., Muller, M. M., & Hillyard, S. A. (2009). Color-selective attention need not be mediated by spatial attention. *Journal of Vision*, 9(6), 2–2. <https://doi.org/10.1167/9.6.2>
- Andersen, S. K., Müller, M. M., & Martinovic, J. (2012). Bottom-up biases in feature-selective attention. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 32(47), 16953–8. <https://doi.org/10.1523/JNEUROSCI.1767-12.2012>

References

- Apperly, I. (2011). *Mindreading*. New York: Psychology Press.
- Apperly, I. (2012). What is “theory of mind”? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology*, 65(5), 825–839. <https://doi.org/10.1080/17470218.2012.676055>
- Apperly, I., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953–70. <https://doi.org/10.1037/a0016923>
- Arora, A., Schurz, M., & Perner, J. (2017). Systematic Comparison of Brain Imaging Meta-Analyses of ToM with vPT. *BioMed Research International*, 2017, 1–12. <https://doi.org/10.1155/2017/6875850>
- Astington, J. W., & Gopnik, A. (1991). Theoretical explanations of children’s understanding of the mind. *British Journal of Developmental Psychology*, 9(1), 7–31. <https://doi.org/10.1111/j.2044-835X.1991.tb00859.x>
- Avis, J., & Harris, P. (1991). Belief-desire reasoning among Baka children: Evidence for a universal conception of mind. *Child Development*, 62, 460–467.
- Bahnemann, M., Dziobek, I., Prehn, K., Wolf, I., & Heekeren, H. R. (2010). Sociotopy in the temporoparietal cortex: Common versus distinct processes. *Social Cognitive and Affective Neuroscience*, 5(1), 48–58. <https://doi.org/10.1093/scan/nsp045>
- Baitch, L. W., & Levi, D. M. (1988). Evidence for nonlinear binocular interactions in human visual cortex. *Vision Research*, 28(10), 1139–1143. [https://doi.org/10.1016/0042-6989\(88\)90140-X](https://doi.org/10.1016/0042-6989(88)90140-X)
- Baker, L. J., Levin, D. T., & Saylor, M. M. (2015). The extent of default visual perspective taking in complex layouts. *Journal of Experimental Psychology: Human Perception and Performance*, 42(4), 508–516. <https://doi.org/10.1037/xhp0000164>
- Bardi, L., Desmet, C., Nijhof, A., Wiersema, J. R., & Brass, M. (2016). Brain activation for spontaneous and explicit false belief tasks overlaps: new fMRI evidence on belief processing and violation of expectation. *Social Cognitive and Affective Neuroscience*, (October 2016), nsw143. <https://doi.org/10.1093/scan/nsw143>
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind.” *Cognition*, 21(21), 37–46.
- Bedny, M., Pascual-Leone, A., & Saxe, R. R. (2009). Growing up blind does not change the neural bases of Theory of Mind. *Proceedings of the National Academy of Sciences*, 106(27), 11312–11317. <https://doi.org/10.1073/pnas.1300178110>
- Bekhtereva, V., Craddock, M., & Müller, M. M. (2015). Attentional bias to affective faces and complex IAPS images in early visual cortex follows emotional cue extraction. *NeuroImage*, 112, 254–266. <https://doi.org/10.1016/j.neuroimage.2015.03.052>
- Belmonte, M. (1998). Shifts of visual spatial attention modulate a steady-state visual evoked potential. *Brain Research. Cognitive Brain Research*, 6(4), 295–307. <http://www.ncbi.nlm.nih.gov/pubmed/9593953>
- Biervoye, A., Dricot, L., Ivanoiu, A., & Samson, D. (2016). Impaired spontaneous belief inference following acquired damage to the left posterior temporoparietal junction. *Social Cognitive and Affective Neuroscience*, 11(10), 1513–1520. <https://doi.org/10.1093/scan/nsw076>
- Blakemore, S. J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, 9(4), 267–277. <https://doi.org/10.1038/nrn2353>
- Boremanse, A., Norcia, A. M., & Rossion, B. (2013). An objective signature for visual binding of face parts in the human brain. *Journal of Vision*, 13(11), 6–6. <https://doi.org/10.1167/13.11.6>

- Boremanse, A., Norcia, A. M., & Rossion, B. (2014). Dissociation of part-based and integrated neural responses to faces by means of electroencephalographic frequency tagging. *European Journal of Neuroscience*, 40(6), 2987–2997. <https://doi.org/10.1111/ejn.12663>
- Bräuer, J., Call, J., & Tomasello, M. (2004). Visual perspective taking in dogs (*Canis familiaris*) in the presence of barriers. *Applied Animal Behaviour Science*, 88(3–4), 299–317. <https://doi.org/10.1016/j.applanim.2004.03.004>
- Bräuer, J., Call, J., & Tomasello, M. (2005). All Great Ape Species Follow Gaze to Distant Locations and Around Barriers. *Journal of Comparative Psychology*, 119(2), 145–154. <https://doi.org/10.1037/0735-7036.119.2.145>
- Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition*, 10(4), 439–448. <https://doi.org/10.1007/s10071-007-0088-1>
- Bräuer, J., Kaminski, J., Riedel, J., Call, J., & Tomasello, M. (2006). Making inferences about the location of hidden food: Social dog, causal ape. *Journal of Comparative Psychology*, 120(1), 38–47. <https://doi.org/10.1037/0735-7036.120.1.38>
- Brooks, R., & Meltzoff, A. N. (2002). The Importance of Eyes: How Infants Interpret Adult Looking Behavior. *Motor Control*, 38(6), 959–966. <https://doi.org/10.1037/0012-1649.38.6.958>
- Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mind-evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience and Biobehavioral Reviews*, 30(4), 437–455. <https://doi.org/10.1016/j.neubiorev.2005.08.001>
- Bukowski, H., Hietanen, J. K., & Samson, D. (2015). From gaze cueing to perspective taking: Revisiting the claim that we automatically compute where or what other people are looking at. *Visual Cognition*, 23(8), 1020–1042. <https://doi.org/10.1080/13506285.2015.1132804>
- Bukowski, H., & Samson, D. (2016). Can emotions influence level-1 visual perspective taking? *Cognitive Neuroscience*, 7(1–4), 182–191. <https://doi.org/10.1080/17588928.2015.1043879>
- Bukowski, H., & Samson, D. (2017). New Insights into the Inter-Individual Variability in Perspective Taking. *Vision*, 1(1), 8. <https://doi.org/10.3390/vision1010008>
- Butler, S. C., Caron, A. J., & Brooks, R. (2000). Infant Understanding of the Referential Nature of Looking. *Journal of Cognition and Development*, 1(4), 359–377. https://doi.org/10.1207/S15327647JCD0104_01
- Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind and Language*, 28(5), 606–637. <https://doi.org/10.1111/mila.12036>
- Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., ... Eickhoff, S. B. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *NeuroImage*, 81, 381–392. <https://doi.org/10.1016/j.neuroimage.2013.05.046>
- Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12(5), 187–192. <https://doi.org/10.1016/j.tics.2008.02.010>
- Callenmark, B., Kjellin, L., Rönqvist, L., & Bölte, S. (2014). Explicit versus implicit social cognition testing in autism spectrum disorder. *Autism*, 18(6), 684–693. <https://doi.org/10.1177/1362361313492393>
- Canteloup, C., Piraux, E., Poulin, N., & Meunier, H. (2016). Do Tonkean macaques (*Macaca tonkeana*) perceive what conspecifics do and do not see? *PeerJ*, 4, e1693. <https://doi.org/10.7717/peerj.1693>

References

- Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive Function and Theory of Mind: Stability and Prediction From Ages 2 to 3. *Developmental Psychology*, 40(6), 1105–1122. <https://doi.org/10.1037/0012-1649.40.6.1105>
- Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, 87(4), 299–319. <https://doi.org/10.1016/j.jecp.2004.01.002>
- Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, 30(8), 2313–2335. <https://doi.org/10.1002/hbm.20671>
- Castelli, F., Happé, F., Frith, U., & Frith, C. D. (2000). Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns. *NeuroImage*, 12(3), 314–325. <https://doi.org/10.1006/nimg.2000.0612>
- Catmur, C., Santiesteban, I., Conway, J. R., Heyes, C., & Bird, G. (2016). Avatars and arrows in the brain. *NeuroImage*, 132, 8–10. <https://doi.org/10.1016/j.neuroimage.2016.02.021>
- Christensen, W., & Michael, J. (2016). From two systems to a multi-systems architecture for mindreading. *New Ideas in Psychology*, 40, 48–64. <https://doi.org/10.1016/j.newideapsych.2015.01.003>
- Clarke, J. L., & Daniell, H. (2011). Plastid biotechnology for crop production: present status and future perspectives. *Plant Molecular Biology*, 76(3–5), 211–220. <https://doi.org/10.1007/s11103-011-9767-z>
- Clayton, N. S., Dally, J. M., & Emery, N. J. (2007). Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 507–522. <https://doi.org/10.1098/rstb.2006.1992>
- Cole, G., Atkinson, M., D’Souza, A., & Smith, D. (2017). Spontaneous Perspective Taking in Humans? *Vision*, 1(2), 17. <https://doi.org/10.3390/vision1020017>
- Cole, G. G., Atkinson, M., Le, A. T. D., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica*, 164, 165–168. <https://doi.org/10.1016/j.actpsy.2016.01.007>
- Cole, G. G., Smith, D. T., & Atkinson, M. A. (2015). Mental state attribution and the gaze cueing effect. *Attention, Perception, & Psychophysics*, 77(4), 1105–1115. <https://doi.org/10.3758/s13414-014-0780-6>
- Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, 43(3), 454–465. <https://doi.org/10.1037/xhp0000319>
- Corkum, V., & Moore, C. (1998). The Origins of Joint Visual Attention in Infants. *Developmental Psychology*, 34(1), 28–38. <https://doi.org/10.1037/0012-1649.34.1.28>
- Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, 107(2), 705–717. <https://doi.org/10.1016/j.cognition.2007.08.001>
- David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., ... Vogeley, K. (2006). Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience*, 18(6), 898–910. <https://doi.org/10.1162/jocn.2006.18.6.898>

- Davis, M. H. (1980). A Multidimensional Approach to Individual Differences in Empathy. *Journal of Personality and Social Psychology*, (10), 85. <https://doi.org/10.1037/0022-3514.44.1.113>
- de Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *eLife*, 4(JUNE), 1–14. <https://doi.org/10.7554/eLife.06564>
- Ding, J., Sperling, G., & Srinivasan, R. (2006). Attentional modulation of SSVEP power depends on the network tagged by the flicker frequency. *Cerebral Cortex (New York, N.Y. : 1991)*, 16(7), 1016–29. <https://doi.org/10.1093/cercor/bhj044>
- Döhnell, K., Schuwerk, T., Meinhardt, J., Sodian, B., Hajak, G., & Sommer, M. (2012). Functional activity of the right temporo-parietal junction and of the medial prefrontal cortex associated with true and false belief reasoning. *NeuroImage*, 60(3), 1652–1661. <https://doi.org/10.1016/j.neuroimage.2012.01.073>
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, 6(5), 509–540. <https://doi.org/10.1080/135062899394920>
- Dzhelyova, M., Jacques, C., & Rossion, B. (2015). At a Single Glance: Fast Periodic Visual Stimulation Uncovers the Spatio-Temporal Dynamics of Brief Facial Expression Changes in the Human Brain. *Cerebral Cortex*, 47(1), 96–106. <https://doi.org/10.1093/cercor/bhw223>
- Dzhelyova, M., & Rossion, B. (2014). Supra-additive contribution of shape and surface information to individual face discrimination as revealed by fast periodic visual stimulation. *Journal of Vision*, 14(14), 15. <https://doi.org/10.1167/14.14.15>
- Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition*, 41, 93–103. <https://doi.org/10.1016/j.concog.2016.02.010>
- Emery, N. J., & Clayton, N. S. (2008). Comparative social cognition. *Annual Review of Psychology*, 60, 87–113. <https://doi.org/10.1146/annurev.psych.60.110707.163526>
- Ferguson, H. J., Apperly, I., & Cane, J. E. (2016). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *The Quarterly Journal of Experimental Psychology*, 70(8), 1646–1660. <https://doi.org/10.1080/17470218.2016.1199716>
- Flavell, J. H., Everett, B. a., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction. *Developmental Psychology*, 17(1), 99–103. <https://doi.org/10.1037/0012-1649.17.1.99>
- Flavell, J. H., Flavell, E. F., Green, F. L., & Wilcox, S. A. (1980). Young children's knowledge about visual perception: Effect of observer's distance from target on perceptual clarity of target. *Developmental Psychology*, 16(1), 10–12. <https://doi.org/10.1037/0012-1649.16.1.10>
- Fletcher, P. (1995). Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition*, 57(2), 109–128. [https://doi.org/10.1016/0010-0277\(95\)00692-R](https://doi.org/10.1016/0010-0277(95)00692-R)
- Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective? *Journal of Experimental Psychology: Human Perception and Performance*, 42(3), 401–412. <https://doi.org/10.1037/xhp0000153>
- Friesen, C. K., & Kingstone, A. (2003). Abrupt onsets and gaze direction cues trigger independent reflexive attentional effects. *Cognition*, 87(1), B1–B10. [https://doi.org/10.1016/S0010-0277\(02\)00181-6](https://doi.org/10.1016/S0010-0277(02)00181-6)
- Frith, C. D. (2007). The social brain of a teenager. *Philos Trans Roy Soc London B: Bio Sci.*, 20(10), 671–678. <https://doi.org/10.1098/rstb.2006.2003>

References

- Frith, C. D., & Frith, U. (1999). Interacting Minds — A Biological Basis. *Science*, 286(November), 1692–1695. <https://doi.org/10.1126/science.286.5445.1692>
- Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, 50(4), 531–534. <https://doi.org/10.1016/j.neuron.2006.05.001>
- Frith, C. D., & Frith, U. (2007). Social Cognition in Humans. *Current Biology*, 17(16), 724–732. <https://doi.org/10.1016/j.cub.2007.05.068>
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358(1431), 459–73. <https://doi.org/10.1098/rstb.2002.1218>
- Fuchs, S., Andersen, S. K., Gruber, T., & Müller, M. M. (2008). Attentional bias of competitive interactions in neuronal networks of early visual processing in the human brain. *NeuroImage*, 41(3), 1086–1101. <https://doi.org/10.1016/j.neuroimage.2008.02.040>
- Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance*, 42(2), 158–163. <https://doi.org/10.1037/xhp0000138>
- Gagne, D. L., & Coppola, M. (2017). Visible Social Interactions Do Not Support the Development of False Belief Understanding in the Absence of Linguistic Input: Evidence from Deaf Adult Homesigners. *Frontiers in Psychology*, 8(June), 1–21. <https://doi.org/10.3389/fpsyg.2017.00837>
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of “theory of mind.” *Trends in Cognitive Sciences*, 7(2), 77–83. [https://doi.org/10.1016/S1364-6613\(02\)00025-6](https://doi.org/10.1016/S1364-6613(02)00025-6)
- Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of “theory of mind” in verbal and nonverbal tasks. *Neuropsychologia*, 38(1), 11–21. [https://doi.org/10.1016/S0028-3932\(99\)00053-6](https://doi.org/10.1016/S0028-3932(99)00053-6)
- Gardner, M. R., Hull, Z., Taylor, D., & Edmonds, C. J. (2017). “Spontaneous” Visual Perspective-Taking Mediated by Attention Orienting that is Voluntary and not Reflexive. *The Quarterly Journal of Experimental Psychology*, 0(0), 1–33. <https://doi.org/10.1080/17470218.2017.1307868>
- Ge, S., Matsuoka, A., Ueno, S., & Iramina, K. (2007). Transcranial magnetic stimulation study of temporal aspect of posterior parietal cortex in visual search. *International Congress Series*, 1300, 287–290. <https://doi.org/10.1016/j.ics.2006.12.069>
- Gentile, F., & Rossion, B. (2012). Temporal frequency tuning of the cortical face-sensitive network for individual face perception. *Journal of Vision*, 12(9), 1171–1171. <https://doi.org/10.1167/12.9.1171>
- George, N., Driver, J., & Dolan, R. J. (2001). Seen Gaze-Direction Modulates Fusiform Activity and Its Coupling with Other Brain Areas during Face Processing. *NeuroImage*, 13(6), 1102–1112. <https://doi.org/10.1006/nimg.2001.0769>
- German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in “theory of mind”: Evidence from compromised belief-desire reasoning in old age. *Cognition*, 101(1), 129–152. <https://doi.org/10.1016/j.cognition.2005.05.007>
- Giani, A. S., Ortiz, E., Belardinelli, P., Kleiner, M., Preissl, H., & Noppeney, U. (2012). Steady-state responses in MEG demonstrate information integration within but not across the auditory and visual senses. *NeuroImage*, 60(2), 1478–89. <https://doi.org/10.1016/j.neuroimage.2012.01.114>

- Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two Takes on the Social Brain: A Comparison of Theory of Mind Tasks. *Journal of Cognitive Neuroscience*, 19(11), 1803–1814. <https://doi.org/10.1162/jocn.2007.19.11.1803>
- Gopnik, A., Slaughter, V., & Meltzoff, A. (1994). Changing your views: how understanding visual perception can lead to a new theory of the mind. *Children's Early Understanding of Mind: Origins and Development*, (April 2014), 157–181.
- Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2016). Implicit and explicit false belief development in preschool children. *Developmental Science*, 1–15. <https://doi.org/10.1111/desc.12445>
- Grosse Wiesmann, C., Schreiber, J., Singer, T., Steinbeis, N., & Friederici, A. D. (2017). White matter maturation is associated with the emergence of Theory of Mind in early childhood. *Nature Communications*, 8, 14692. <https://doi.org/10.1038/ncomms14692>
- Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research*, 45(22), 2847–2853. <https://doi.org/10.1016/j.visres.2005.05.027>
- Gundlach, C., & Müller, M. M. (2013). Perception of illusory contours forms intermodulation responses of steady state visual evoked potentials as a neural signature of spatial integration. *Biological Psychology*, 94(1), 55–60. <https://doi.org/10.1016/j.biopsycho.2013.04.014>
- Gzesh, S. M., & Surber, C. F. (1985). Visual perspective-taking skills in children. *Child Development*, 56(5), 1204–1213. <https://doi.org/10.2307/1130235>
- Hamilton, A., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition*, 113(1), 37–44. <https://doi.org/10.1016/j.cognition.2009.07.007>
- Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59(4), 771–785. <https://doi.org/10.1006/anbe.1999.1377>
- Hartwright, C. E., Apperly, I., & Hansen, P. C. (2015). The special case of self-perspective inhibition in mental, but not non-mental, representation. *Neuropsychologia*, 67, 183–192. <https://doi.org/10.1016/j.neuropsychologia.2014.12.015>
- Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *American Journal of Psychology*, 57(2), 243–259. https://doi.org/10.4992/jjpsy.20.2_67
- Herrmann, C. S. (2001). Human EEG responses to 1-100 Hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, 137(3–4), 346–353. <https://doi.org/10.1007/s002210100682>
- Heyes, C. (2014). Submentalizing. *Perspectives on Psychological Science*, 9(2), 131–143. <https://doi.org/10.1177/1745691613518076>
- Heyes, S. B., Sebastian, C., Cohen Kadosh, K., & Blakemore, S. J. (2010). The social brain in adolescence: Evidence from functional magnetic resonance imaging and behavioural studies. *Neuroscience and Biobehavioral Reviews*, 35(8), 1654–1664. <https://doi.org/10.1016/j.neubiorev.2010.10.011>
- Hietanen, J. K. (1999). Does your gaze direction and head orientation shift my visual attention? *NeuroReport*, 10(16), 3443–3447. <https://doi.org/10.1097/00001756-199911080-00033>
- Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-aho, K., & Ruuhiala, H. J. (2008). Seeing direct and averted gaze activates the approach-avoidance motivational brain systems. *Neuropsychologia*, 46(9), 2423–2430. <https://doi.org/10.1016/j.neuropsychologia.2008.02.029>

References

- Harris PL. 1989. *Children and Emotion: The Development of Psychological Understanding*. Cambridge, MA: Basil Blackwell. 243 pp.
- Hodges, S. D. (2015). Perspective taking. In *The International Encyclopedia of Social and Behavioral Sciences*, 2nd Edition. Elsevier.
- Hood, B. M., Willen, J., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science*, 9(2), 131–134. <https://doi.org/10.1111/1467-9280.00024>
- Hutchins, T. L., & Prelock, P. A. (2016). The theory of mind atlas, 1–9.
- Kalla, R., Muggleton, N. G., Juan, C.-H., Cowey, A., & Walsh, V. (2008). The timing of the involvement of the frontal eye fields and posterior parietal cortex in visual search. *Neuroreport*, 19(10), 1067–71. <https://doi.org/10.1097/WNR.0b013e328304d9c4>
- Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proceedings of the Royal Society B: Biological Sciences*, 282(1819), 20151683. <https://doi.org/10.1098/rspb.2015.1683>
- Keil, A., Moratti, S., Sabatinelli, D., Bradley, M. M., & Lang, P. J. (2005). Additive effects of emotional content and spatial selective attention on electrocortical facilitation. *Cerebral Cortex (New York, N.Y. : 1991)*, 15(8), 1187–97. <https://doi.org/10.1093/cercor/bhi001>
- Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology*, 1(December), 1–12. <https://doi.org/10.3389/fpsyg.2010.00213>
- Kim, Y. J., Grabowecy, M., Paller, K. a, Muthu, K., & Suzuki, S. (2007). Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nature Neuroscience*, 10(1), 117–25. <https://doi.org/10.1038/nn1821>
- Kovács, Á. M., Kühn, S., Gergely, G., Csibra, G., & Brass, M. (2014). Are All Beliefs Equal? Implicit Belief Attributions Recruiting Core Brain Regions of Theory of Mind. *PLoS ONE*, 9(9), e106558. <https://doi.org/10.1371/journal.pone.0106558>
- Kovacs, A. M., Teglás, E., & Endress, A. D. (2010). The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, 330(6012), 1830–1834. <https://doi.org/10.1126/science.1190792>
- Laube, I., Kamphuis, S., Dicke, P. W., & Thier, P. (2011). Cortical processing of head- and eye-gaze cues guiding joint social attention. *NeuroImage*, 54(2), 1643–1653. <https://doi.org/10.1016/j.neuroimage.2010.08.074>
- Law, A. S., Langton, S. R. H., & Logie, R. H. (2010). Assessing the impact of verbal and visuospatial working memory load on eye-gaze cueing. *Visual Cognition*, 18(10), 1420–1438. <https://doi.org/10.1080/13506285.2010.496579>
- Le Bouc, R., Lenfant, P., Delbeuck, X., Ravasi, L., Lebert, F., Semah, F., & Pasquier, F. (2012). My belief or yours? Differential theory of mind deficits in frontotemporal dementia and Alzheimer's disease. *Brain*, 135(10), 3026–3038. <https://doi.org/10.1093/brain/aws237>
- Legrand, D., & Ruby, P. (2009). What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychological Review*, 116(1), 252–282. <https://doi.org/10.1037/a0014172>
- Lempers, J. D., Flavell, E. R., & Flavell, J. H. (1977). The development in very young children of tacit knowledge concerning visual perception. *Genetic Psychology Monographs*, 95(1), 3-53.
- Leslie, A. M. (1984). Infant perception of a manual pick-up event. *British Journal of Developmental Psychology*, 2, 19–32. <https://doi.org/10.1111/j.2044-835X.1984.tb00531.x>

- Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task : Two conjectures. *Developmental Science*, 1(2), 247–253. <https://doi.org/10.1111/1467-7687.00038>
- Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, 40(3), 225–251.
- Libenson, M.H. *Practical Approach to Electroencephalography*. Saunders Elsevier.
- Liu-Shuang, J., Norcia, A. M., & Rossion, B. (2014). An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia*, 52, 57–72. <https://doi.org/10.1016/j.neuropsychologia.2013.10.022>
- Liu-Shuang, J., Torfs, K., & Rossion, B. (2015). An objective electrophysiological marker of face individualisation impairment in acquired prosopagnosia with fast periodic visual stimulation. *Neuropsychologia*, 1–14. <https://doi.org/10.1016/j.neuropsychologia.2015.08.023>
- Liu, G., Qin, Q., Chan, K. W. Y., Li, Y., Bulte, J. W. M., McMahon, M. T., ... Gilad, A. A. (2014). Non-invasive temperature mapping using temperature-responsive water saturation shift referencing (T-WASSR) MRI. *NMR in Biomedicine*, 27(3), 320–331. <https://doi.org/10.1002/nbm.3066>
- Lochy, A., Belle, G. Van, & Rossion, B. (2015). A robust index of lexical representation in the left occipito-temporal cortex as evidenced by EEG responses to fast periodic visual stimulation. *Neuropsychologia*, 66, 18–31. <https://doi.org/10.1016/j.neuropsychologia.2014.11.007>
- Lochy, A., Van Reybroeck, M., & Rossion, B. (2016). Left cortical specialization for visual letter strings predicts rudimentary knowledge of letter-sound association in preschoolers. *Proceedings of the National Academy of Sciences*, 201520366. <https://doi.org/10.1073/pnas.1520366113>
- Lopes da Silva FH (2002) Electrical potentials. In: Ramachandran VS (ed) *Encyclopedia of the human brain*. Elsevier, New York, pp 147–67
- Luck SJ (2005). An introduction to the event-related potential technique.
- Mar, R. A. (2011). The Neural Bases of Social Cognition and Story Comprehension. *Annual Review of Psychology*, 62(1), 103–134. <https://doi.org/10.1146/annurev-psych-120709-145406>
- Mattan, B., Rotshtein, P., & Quinn, K. a. (2016). Empathy and visual perspective-taking performance. *Cognitive Neuroscience*, 8928(January), 1–12. <https://doi.org/10.1080/17588928.2015.1085372>
- McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 62(3), 203–208. [https://doi.org/10.1016/0168-5597\(85\)90015-2](https://doi.org/10.1016/0168-5597(85)90015-2)
- McCleery, J. P., Surtees, A., Graham, K. a, Richards, J. E., & Apperly, I. (2011). The neural and cognitive time course of theory of mind. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 31(36), 12849–54. <https://doi.org/10.1523/JNEUROSCI.1392-11.2011>
- Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31(5), 838–850. <https://doi.org/10.1037/0012-1649.31.5.838>
- Michael, J., Wolf, T., Letesson, C., Butterfill, S. A., & Skewes, J. (2017). Seeing is both: Using a double-cuing task to investigate the role of spatial cueing in level-1 visual perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance*.
- Michel, C., Dricot, L., Lhommel, R., Grandin, C., Ivanoiu, A., Pillon, A., & Samson, D. (2013). Extensive Left Temporal Pole Damage Does Not Impact on Theory of Mind Abilities. *Journal of Cognitive Neuroscience*, 25(12), 2025–2046. https://doi.org/10.1162/jocn_a_00488

References

- Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics*, 68(2), 327–337. <https://doi.org/10.3758/BF03193680>
- Moll, H., & Tomasello, M. (2004). 12- and 18-Month-Old Infants Follow Gaze To Spaces Behind Barriers. *Developmental Science*, 7(1), 1–9. <https://doi.org/10.1111/j.1467-7687.2004.00315.x>
- Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, 24(3), 603–613. <https://doi.org/10.1348/026151005X55370>
- Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS ONE*, 7(5), 34–42. <https://doi.org/10.1371/journal.pone.0037527>
- Morgan, S. T., Hansen, J. C., & Hillyard, S. a. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10), 4770–4.
- Müller, M. M., Andersen, S. K., & Keil, A. (2008). Time course of competition for visual processing resources between emotional pictures and foreground task. *Cerebral Cortex*, 18(8), 1892–1899. <https://doi.org/10.1093/cercor/bhm215>
- Müller, M. M., & Hillyard, S. A. (2000). Concurrent recording of steady-state and transient event-related potentials as indices of visual-spatial selective attention. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, 111(9), 1544–52. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10964063>
- Müller, M. M., & Hubner, R. (2002). Can the Spotlight of Attention Be Shaped Like a Doughnut? Evidence From Steady-State Visual Evoked Potentials. *Psychological Science*, 13(2), 119–124. <https://doi.org/10.1111/1467-9280.00422>
- Müller, M. M., Picton, T. W., Valdes-Sosa, P., Riera, J., Teder-Sälejärvi, W. a, & Hillyard, S. A. (1998). Effects of spatial selective attention on the steady-state visual evoked potential in the 20-28 Hz range. *Brain Research. Cognitive Brain Research*, 6(4), 249–61. <http://www.ncbi.nlm.nih.gov/pubmed/9593922>
- Nielsen, M. K., Slade, L., Levy, J. P., & Holmes, A. (2015). Inclined to see it your way: Do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process? *The Quarterly Journal of Experimental Psychology*, (April 2015), 1–21. <https://doi.org/10.1080/17470218.2015.1023206>
- Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottureau, B. R., & Rossion, B. (2015a). The steady-state visual evoked potential in vision research : A review. *Journal of Vision*, 15(6), 1–46. <https://doi.org/10.1167/15.6.4>
- Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottureau, B. R., & Rossion, B. (2015b). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, 15(6), 4. <https://doi.org/10.1167/15.6.4>
- Northoff, G., & Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences*, 8(3), 102–107. <https://doi.org/10.1016/j.tics.2004.01.004>
- Nuku, P., & Bekkering, H. (2008). Joint attention: Inferring what others perceive (and don't perceive). *Consciousness and Cognition*, 17(1), 339–349. <https://doi.org/10.1016/j.concog.2007.06.014>
- Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, 13(3), 135–143. <https://doi.org/10.1016/j.tics.2008.12.006>

- Nunez, P. L., Srinivasan R (2006). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press. 611pp
- Onishi, K. H. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308(5719), 255–258. <https://doi.org/10.1126/science.1107621>
- Osaka, N., Ikeda, T., & Osaka, M. (2012). Effect of Intentional Bias on Agency Attribution of Animated Motion: An Event-Related fMRI Study. *PLoS ONE*, 7(11), e49053. <https://doi.org/10.1371/journal.pone.0049053>
- Painter, D. R., Dux, P. E., Travis, S. L., & Mattingley, J. B. (2014). Neural responses to target features outside a search array are enhanced during conjunction but not unique-feature search. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(9), 3390–401. <https://doi.org/10.1523/JNEUROSCI.3630-13.2014>
- Pastor, M. A., Artieda, J., Arbizu, J., Valencia, M., & Masdeu, J. C. (2003). Human cerebral activation during steady-state visual-evoked responses. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(37), 11621–7. <http://www.ncbi.nlm.nih.gov/pubmed/14684864>
- Pavlova, M., Guerreschi, M., Lutzenberger, W., & Krägeloh-Mann, I. (2010). Social Interaction Revealed by Motion: Dynamics of Neuromagnetic Gamma Activity. *Cerebral Cortex*, 20(10), 2361–2367. <https://doi.org/10.1093/cercor/bhp304>
- Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
- Perner, J., & Ruffman, T. (1995). Episodic memory and autonoetic consciousness: developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology*. <https://doi.org/10.1006/jecp.1995.1024>
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3–25. <https://doi.org/10.1080/00335558008248231>
- Poulin-Dubois, D., Sodian, B., Metz, U., Tilden, J., & Schoeppner, B. (2007). Out of sight is not out of mind: Developmental changes in infants' understanding of visual perception during the second year. *Journal of Cognition and Development*, 8(4), 401–425. <https://doi.org/10.1080/15248370701612951>
- Premack, D., & Woodruff, G. (1978). Premack and Woodruff : Chimpanzee theory of mind. *Behavioral and Brain Sciences*, 4(1978), 515–526.
- Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, 117(2), 230–236. <https://doi.org/10.1016/j.cognition.2010.08.003>
- Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalography and Clinical Neurophysiology*, 20(3), 238–248.
- Regan, D. (1977). Steady-state evoked potentials. *Optical Society of America*, 11(October 1975), 1475–1489.
- Regan, D. (1989). *Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine* (672 pp.). New York: Elsevier
- Regan, D., & Heron, J. R. (1969). Clinical investigation of lesions of the visual pathway: a new objective technique. *Journal of Neurology, Neurosurgery, and Psychiatry*, 32(5), 479–83.
- Retter, T. L., & Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia*, 91, 9–28. <https://doi.org/10.1016/j.neuropsychologia.2016.07.028>

References

- Rossion, B., & Boremanse, A. (2011). Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. *Journal of Vision*, 11(2), 16–16. <https://doi.org/10.1167/11.2.16>
- Rossion, B., Prieto, E. A., Boremanse, A., Kuefner, D., & Van Belle, G. (2012). A steady-state visual evoked potential approach to individual face perception: Effect of inversion, contrast-reversal and temporal dynamics. *NeuroImage*, 63(3), 1585–1600. <https://doi.org/10.1016/j.neuroimage.2012.08.033>
- Samson, D. (2013). Published in the Oxford Handbook of Cognitive Psychology (2013). D. Reisberg (Ed.).
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1255–1266. <https://doi.org/10.1037/a0018729>
- Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, 7(5), 499–500. <https://doi.org/10.1038/nn1223>
- Samson, D., Apperly, I. A., & Humphreys, G. W. (2007). Error analyses reveal contrasting deficits in “theory of mind”: Neuropsychological evidence from a 3-option false belief task. *Neuropsychologia*, 45(11), 2561–2569. <https://doi.org/10.1016/j.neuropsychologia.2007.03.013>
- Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain*, 128(5), 1102–1111. <https://doi.org/10.1093/brain/awh464>
- Samson, D., Houthuys, S., & Humphreys, G. W. (2015). Self-perspective inhibition deficits cannot be explained by general executive control difficulties. *Cortex*, 70, 189–201. <https://doi.org/10.1016/j.cortex.2014.12.021>
- Santesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 929–937. <https://doi.org/10.1037/a0035175>
- Santesteban, I., Kaur, S., Bird, G., & Catmur, C. (2017). Attentional processes, not implicit mentalizing, mediate performance in a perspective-taking task: Evidence from stimulation of the temporoparietal junction. *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2017.04.055>
- Santesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic Bulletin & Review*, 22(3), 844–849. <https://doi.org/10.3758/s13423-014-0716-0>
- Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, 16(2), 235–239. <https://doi.org/10.1016/j.conb.2006.03.001>
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporoparietal junction in “theory of mind.” *NeuroImage*, 19(4), 1835–1842. [https://doi.org/10.1016/S1053-8119\(03\)00230-1](https://doi.org/10.1016/S1053-8119(03)00230-1)
- Saxe, R., & Powell, L. J. (2006). It's the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind. *PSYCHOLOGICAL SCIENCE*, 17(8), 692–699. <https://doi.org/10.1017/S0140525X00053607>
- Scaife, M., & Bruner, J. S. (1975). © 1975 Nature Publishing Group. *Nature*, 253, 265–266.

- Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, 19(2), 65–72.
<https://doi.org/10.1016/j.tics.2014.11.007>
- Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, 129(2), 410–417.
<https://doi.org/10.1016/j.cognition.2013.08.004>
- Schneider, D., Slaughter, V. P., Becker, S. I., & Dux, P. E. (2014). Implicit false-belief processing in the human brain. *NeuroImage*, 101, 268–275. <https://doi.org/10.1016/j.neuroimage.2014.07.014>
- Scholz, J., Klein, M. C., Behrens, T. E. J., & Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nature Neuroscience*, 12(11), 1370–1371. <https://doi.org/10.1038/nn.2412>
- Schuller, A.-M., & Rossion, B. (2004). Perception of static eye gaze direction facilitates subsequent early visual processing. *Clinical Neurophysiology*, 115(5), 1161–1168.
<https://doi.org/10.1016/j.clinph.2003.12.022>
- Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in Human Neuroscience*, 7(November), 1–14. <https://doi.org/10.3389/fnhum.2013.00712>
- Schurz, M., Kronbichler, M., Weissengruber, S., Surtees, A., Samson, D., & Perner, J. (2015). Clarifying the role of theory of mind areas during visual perspective taking: Issues of spontaneity and domain-specificity. *NeuroImage*, 117, 386–396. <https://doi.org/10.1016/j.neuroimage.2015.04.031>
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, 42, 9–34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>
- Schuwerk, T., Döhnell, K., Sodian, B., Keck, I. R., Rupprecht, R., & Sommer, M. (2014a). Functional activity and effective connectivity of the posterior medial prefrontal cortex during processing of incongruent mental states. *Human Brain Mapping*, 35(7), 2950–2965.
<https://doi.org/10.1002/hbm.22377>
- Schuwerk, T., Schecklmann, M., Langguth, B., Döhnell, K., Sodian, B., & Sommer, M. (2014b). Inhibiting the posterior medial prefrontal cortex by rTMS decreases the discrepancy between self and other in Theory of Mind reasoning. *Behavioural Brain Research*, 274, 312–318.
<https://doi.org/10.1016/j.bbr.2014.08.031>
- Scott, R. M., & Baillargeon, R. (2014). How fresh a look? A reply to Heyes. *Developmental Science*, 17(5), 660–664. <https://doi.org/10.1111/desc.12173>
- Sebastian, C. L., McCrory, E. J. P., Cecil, C. A. M., Lockwood, P. L., De Brito, S. A., Fontaine, N. M. G., & Viding, E. (2012). Neural Responses to Affective and Cognitive Theory of Mind in Children With Conduct Problems and Varying Levels of Callous-Unemotional Traits. *Archives of General Psychiatry*, 69(8), 814. <https://doi.org/10.1001/archgenpsychiatry.2011.2070>
- Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, 18(9), 668–671. <https://doi.org/10.1016/j.cub.2008.03.059>
- Shibata, H., & Inui, T. (2011). Brain activity associated with recognition of appropriate action selection based on allocentric perspectives. *Neuroscience Letters*, 491(3), 187–191.
<https://doi.org/10.1016/j.neulet.2011.01.033>
- Shillito, D. J., Shumaker, R. W., Gallup, G. G., & Beck, B. B. (2005). Understanding visual barriers: Evidence for level 1 perspective taking in an orang-utan, *Pongo pygmaeus*. *Animal Behaviour*, 69(3), 679–687. <https://doi.org/10.1016/j.anbehav.2004.04.022>

References

- Silberstein, R. B., Shier, M. A., Pipingas, A., Ciociari, J., Wood, S. R., & Simpson, D. G. (1990). Steady-State Visually Evoked Potential Topography Associated with a Visual Vigilance Task. *Brain Topography*, 3(2), 337–347.
- Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews*, 30(6), 855–863. <https://doi.org/10.1016/j.neubiorev.2006.06.011>
- Singla, R., & Haseena, B. A. (2014). Comparison of SSVEP Signal Classification Techniques Using SVM and ANN Models for BCI Applications. *International Journal of Information and Electronics Engineering*, 4(1), 6–10. <https://doi.org/10.7763/IJIEE.2014.V4.398>
- Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science*, 10(2), 199–204. <https://doi.org/10.1111/j.1467-7687.2007.00580.x>
- Sommer, M., Döhl, K., Sodan, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *NeuroImage*, 35(3), 1378–1384. <https://doi.org/10.1016/j.neuroimage.2007.01.042>
- Stormer, V. S., Winther, G. N., Li, S.-C., & Andersen, S. K. (2013). Sustained Multifocal Attentional Enhancement of Stimulus Processing in Early Visual Areas Predicts Tracking Performance. *Journal of Neuroscience*, 33(12), 5346–5351. <https://doi.org/10.1523/JNEUROSCI.4015-12.2013>
- Surtees, A., & Apperly, I. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development*, 83(2), 452–60. <https://doi.org/10.1111/j.1467-8624.2011.01730.x>
- Surtees, A., Apperly, I., & Samson, D. (2013a). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, 129(2), 426–438. <https://doi.org/10.1016/j.cognition.2013.06.008>
- Surtees, A., Apperly, I., & Samson, D. (2013b). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in Human Neuroscience*, 7(698), 698. <https://doi.org/10.3389/fnhum.2013.00698>
- Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, 150, 43–52. <https://doi.org/10.1016/j.cognition.2016.01.014>
- Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, 148, 97–105. <https://doi.org/10.1016/j.cognition.2015.12.010>
- Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010a). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception & Psychophysics*, 72(3), 695–705. <https://doi.org/10.3758/APP.72.3.695>
- Teufel, C., Alexis, D. M., Todd, H., Lawrence-Owen, A. J., Clayton, N. S., & Davis, G. (2009). Social Cognition Modulates the Sensory Coding of Observed Gaze Direction. *Current Biology*, 19(15), 1274–1277. <https://doi.org/10.1016/j.cub.2009.05.069>
- Teufel, C., Fletcher, P. C., & Davis, G. (2010b). Seeing other minds: Attributed mental states influence perception. *Trends in Cognitive Sciences*, 14(8), 376–382. <https://doi.org/10.1016/j.tics.2010.05.005>
- Tiffany L. Hutchins & Patricia A. Prelock. (2016). The theory of Mind Atlas

- Todd, A. R., Cameron, C. D., & Simpson, A. J. (2017). Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition*, 159, 97–101.
<https://doi.org/10.1016/j.cognition.2016.11.010>
- Todd, A. R., Hanco, K., Galinsky, A. D., & Mussweiler, T. (2011). When focusing on differences leads to similar perspectives. *Psychological Science*, 22(1), 134–41.
<https://doi.org/10.1177/0956797610392929>
- Todd, A. R., & Simpson, A. J. (2016). Anxiety impairs spontaneous perspective calculation: Evidence from a level-1 visual perspective-taking task. *Cognition*, 156, 88–94.
<https://doi.org/10.1016/j.cognition.2016.08.004>
- Tomasello, M., Call, J. (1997). Primate cognition. Oxford University Press.
- Tomasello, M. 1999: The Cultural Origins of Human Cognition. Cambridge, MA: Harvard University Press.
- Tomasello, M., Call, J., & Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Animal Behaviour*, 55(4), 1063–1069. <https://doi.org/10.1006/anbe.1997.0636>
- Tremblay-Leveau, H. (1999). Avant les croyances. *Enfance*, 52(3), 313–321.
<https://doi.org/10.3406/enfan.1999.3156>
- Van der Meer, L., Groenewold, N. A., Nolen, W. A., Pijnenborg, M., & Aleman, A. (2011). Inhibit yourself and understand the other: Neural basis of distinct processes underlying Theory of Mind. *NeuroImage*, 56(4), 2364–2374. <https://doi.org/10.1016/j.neuroimage.2011.03.053>
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30(3), 829–858. <https://doi.org/10.1002/hbm.20547>
- Vlaar, M. P., van der Helm, F. C. T., & Schouten, A. C. (2015). Frequency Domain Characterization of the Somatosensory Steady State Response in Electroencephalography. *IFAC-PapersOnLine*, 48(28), 1391–1396. <https://doi.org/10.1016/j.ifacol.2015.12.327>
- Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., ... Zilles, K. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *NeuroImage*, 14(1 Pt 1), 170–181. <https://doi.org/10.1006/nimg.2001.0789>
- Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., & Fink, G. R. (2004). Neural Correlates of First-Person Perspective as One Constituent of Human Self-Consciousness. *Journal of Cognitive Neuroscience*, 16(5), 817–827. <https://doi.org/10.1162/089892904970799>
- Walter, S., Quigley, C., Andersen, S. K., & Müller, M. M. (2012). Effects of overt and covert attention on the steady-state visual evoked potential. *Neuroscience Letters*, 519(1), 37–41.
<https://doi.org/10.1016/j.neulet.2012.05.011>
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, 72(3), 655–684.
<https://doi.org/10.1111/1467-8624.00304>
- Westra, E. (2017). Spontaneous mindreading: a problem for the two-systems account. *Synthese*, 194(11), 4559–4581. <https://doi.org/10.1007/s11229-016-1159-0>
- Whiten, A. 1994: Grades of mindreading. In C. Lewis and P. Mitchell (eds.), *Children's Early Understanding of Mind*. Mahwah, NJ: Erlbaum, 47–70.
- Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I See What You Mean: How Attentional Selection Is Shaped by Ascribing Intentions to Others. *PLoS ONE*, 7(9), 1–7.
<https://doi.org/10.1371/journal.pone.0045391>

References

- Wieser, M. J., & Keil, A. (2014). Fearful faces heighten the cortical representation of contextual threat. *NeuroImage*, 86, 317–325. <https://doi.org/10.1016/j.neuroimage.2013.10.008>
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128.
- Young, L., & Saxe, R. (2009). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, 21(7), 1396–405. <https://doi.org/10.1162/jocn.2009.21137>
- Zacks, J. M., Gilliam, F., & Ojemann, J. G. (2003). Selective disturbance of mental rotation by cortical stimulation. *Neuropsychologia*, 41(12), 1659–1667. [https://doi.org/10.1016/S0028-3932\(03\)00099-X](https://doi.org/10.1016/S0028-3932(03)00099-X)
- Zemon, V., & Ratliff, F. (1982). Visual evoked potentials: evidence for lateral interactions. *Proceedings of the National Academy of Sciences*, 79(18), 5723–5726. <https://doi.org/10.1073/pnas.79.18.5723>
- Zemon, V., & Ratliff, F. (1984). Biological Cybernetics Intermodulation Components of the Visual Evoked Potential : *Biological Cybernetics*, 408, 401–408.
- Zhu, D., Bieger, J., Garcia Molina, G., & Aarts, R. M. (2010). A survey of stimulation methods used in SSVEP-based BCIs. *Computational Intelligence and Neuroscience*, 2010, 702357. <https://doi.org/10.1155/2010/702357>

Appendices

Appendix 1: Opportunities and limitations of fast periodic visual stimulation for studying covert social attention

Beck, Alexy-Assaf, Rossion, Bruno & Samson, Dana

Abstract

Presenting visual stimuli at a fast periodic (i.e. fixed) rate leads to high signal-to-noise ratio (SNR) responses in the electroencephalogram (EEG) known as “steady-state visual evoked potentials” (Regan, 1966). This fast periodic visual stimulation (FPVS) approach has been successfully used to explore low-level vision, spatial and selective attention, and more recently face perception. The purpose of this study was to assess the potential of FPVS to explore the processes underpinning visual perspective taking. Particularly, we wanted to examine if FPVS can be used to track whether participants attend at the same object as the one another person fixates. We combined EEG recording with two frequencies-tagged checkerboards. Each checkerboard was located in one visual hemifield and each checkerboard’s contrast was reversed at different temporal frequencies (4 Hz and 4.5 Hz). A person was presented in the scene facing one checkerboard and turning its back to the other checkerboard. In the first experiment, participants were explicitly told to attend to the checkerboard seen by the person while fixating a central fixation cross. Frequency domain analysis revealed a significant higher SNR response for the frequency matching the stimulus seen by the person compared to the frequency matching the stimulus not seen by the person. In the second experiment, participants were told to attend to the checkerboard that was either in front or behind the person (spatial rather than visual perspective taking instructions). We hypothesized that if the SNR response for the frequency matching the stimulus congruent with the task instruction was higher in the “front” compared to the “behind” instruction, this could be a signature of implicit perspective taking (i.e., participants would still look at the stimulus in front of the person despite the instructions to attend to the stimulus behind the person). The results showed a significant higher SNR response for the frequency matching the stimulus congruent with the task instructions but no difference according to the type of instructions. We discuss the potential of the technique to investigate the cognitive mechanisms underlying visual perspective taking.

I. Study 1: Explicit perspective taking instructions

In the first study, we investigated if FPVS is a technic sensitive enough to detect whether participants attend to the location the person is looking at when they are explicitly instructed to do so (explicit VPT instructions). The paradigm that was chosen was inspired by Morgan et al., (1996) in which participants were cued to attend to a sequence of letter/number in one visual hemifield and to ignore the similar sequence in the opposite hemifield. The sequences of stimuli were superimposed on squares which were flickering at 8.6 Hz in one hemifield and 12 Hz in the opposite hemifield. EEG response showed an increase of the FPVS response over the occipito-temporal scalp area elicited by the square positioned within the attended visual hemifield. Therefore, the authors demonstrated that visual spatial attention plays an important role in modulating the FPVS response magnitude.

Based on previous results in the literature, we thus expected a significant larger response for the stimulus attended to (i.e. the stimulus seen by the person) than the stimulus unattended to (i.e., the stimulus not seen by the person).

1. Methods

Participants

Twenty-two healthy volunteers (female: 16, mean age: 23) with normal or corrected-to-normal vision took part in the experiment in exchange of a small honorarium. All subjects gave written informed consent prior to the experiment. Five participants were subsequently excluded from the sample because of their poor task performances (< 80 % correct responses). This experiment received the approval from the ethics committee of the Psychological Sciences Research Institute.

Stimuli

We used Adobe Photoshop CS6 (Adobe Systems Inc.) to create the stimuli. We used 18 photographs of student volunteers (9 females) which were taken in profile view. The picture of the person was placed in the middle of a room with the left, back, and right walls visible. The vanishing point of the room was localized at the level of the person's head. Half

of the stimuli depicted the person looking at the left wall and the other half of the stimuli were horizontally flipped so that the person was looking at the right wall.

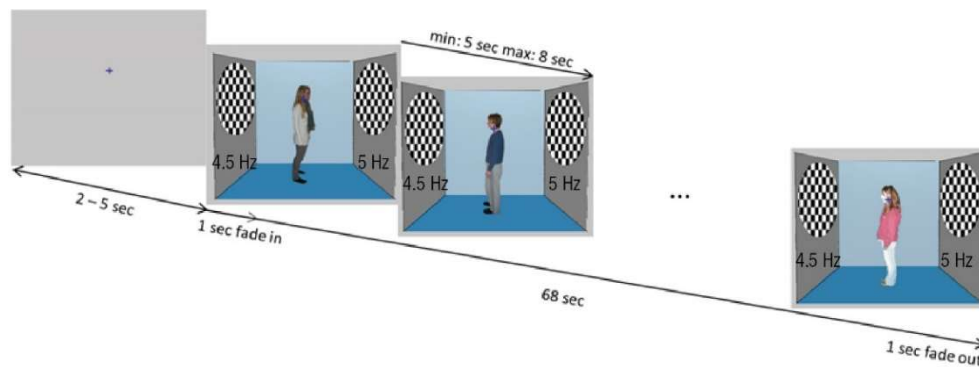


Figure 1. Illustration of the experimental procedure (example of one trial). Numbers indicating flickering frequency were not displayed during the experiment. Experiment 1: participants were instructed to focus their attention on the checkerboard which was seen by the person (explicit perspective taking instruction). Experiment 2: participants were instructed to focus their attention on the checkerboard in front of the person in one experiment part, and in the other part, participants had to focus on the checkerboard behind the person while doing the fixation cross task (explicit spatial perspective taking instruction).

Two identical oval flickering checkerboard discs were displayed ($3.76^\circ \times 2.05^\circ$ of visual angle), one on the right wall and the other on the left wall. One checkerboard was flickering at 4 Hz and the other one at 4.5 Hz, with the side (left or right) flickering at 4 or 4.5 Hz being counterbalanced across trials. The two flickering checkerboard were squarely contrast-modulated in a pattern reversal stimulation (black parts switched to white and white parts switched to black). Pattern reversal presentation keeps the mean luminance of the pattern constant and elicits a FPVS response at twice the rate of stimulation, i.e., in our study 8 Hz and 9 Hz, respectively. We used a FPVS response at 8 Hz and 9 Hz because these frequencies lead to the greatest amplitude response (Fuchs et al., 2008). The luminance of the right and left walls corresponded to the average luminance of the checkerboard (127 cd/m^2). All stimuli were displayed on a light grey background (198 cd/m^2), and dimension images were 480×480 pixels and were shown on a screen with a 120 Hz refresh rate. The position of the person in the room meant that s/he always saw only one of the checkerboards.

Procedure

Figure 1 shows a schematic illustration of the experimental design. The experiment comprised 12 trials each lasting 68 s. Within a trial, the flickering rate of the left and right checkerboard remained unchanged but the person in the middle of the room changed identity and/or orientation (looking either to the left or the right checkerboard).

There were 4 different conditions: (1) the person looking to the left to the 4 Hz flickering checkerboard (the 4.5 Hz flickering checkerboard on the right wall was not seen by the person); (2) the person looking to the left to the 4.5 Hz flickering checkerboard (the 4 Hz flickering checkerboard on the right wall was not seen by the person); (3) the person looking to the right to the 4 Hz flickering checkerboard (the 4.5 Hz flickering checkerboard on the left wall was not seen by the person) and; (4) the person looking to the right to the 4.5 Hz flickering checkerboard (the 4.5 Hz flickering checkerboard on the left wall was not seen by the person).

Stimuli were displayed using an in-house application (SinStim) written in Matlab (The MathWorks). Each trial started with the onset of a blue fixation cross. After a randomly time period of 2 to 5 s, the scene gradually appeared and reached full contrast after 1 s (contrast increase; fade-in 1 s). During each trial the person's identity and/or orientation (to the left/right wall) changed randomly every 5 s (minimum) to 8 s (maximum). Each trial always contained 5 persons oriented towards the right wall and 5 persons oriented towards the left wall, with an equal duration of presentation of each orientation within a trial. At the end of each trial, the scene gradually disappeared (contrast decrease; fade-out 1 s; see Figure 1).

During the EEG recording, participants were seated in a light- and sound-attenuated room, at a viewing distance of 70 cm from the CRT 17-in (43-cm) computer screen. The participants were instructed to fixate the fixation cross placed slightly below the person's head and to respond as quickly and accurately as possible when they detected a brief shape change (cross to square, 10 to 16 random changes per trial, duration of a change: 120 ms) by pressing a response key. The goal of this task was to avoid participants doing ocular saccades and to ensure that they maintained their attention throughout the trials. A

response was considered as correct if it was given within 3 standard deviations of the subject's mean reaction time after the fixation cross change. In addition to the fixation cross task, participants were instructed to focus their attention on the checkerboard which was seen by the person (explicit perspective taking instruction) without moving their eyes from the fixation cross so that they could detect simultaneously the change of shape of the fixation point.

EEG acquisition

EEG data was recorded using a 128 scalps channel at a sampling rate of 512 Hz with an ActiveTwo Biosemi electrode system (Biosemi, Amsterdam, Netherlands). Two additional electrodes, CMS (Common Mode Sense) and DRL (Driven Right Leg), were used as reference and ground. Each individual electrode's direct current (DC) offset were adjusted below ± 50 μ V prior to recording by injecting the electrode with a conductive gel (Signa). Eye-movements were recorded by four additional electro-oculograms around the subjects' eyes. During the experiment, triggers were sent via the parallel port from the stimulation computer to the EEG recording computer at the start/end of each trial and at each person's change.

EEG pre-processing

Off-line analysis was performed with Letswave 5 (<http://nocions.webnode.com/letswave>; Mouraux & Iannetti, 2008) and Matlab 2012 (Math-Works, Natick, MA). Individual EEG data were first filtered with a bandpass between 0.1 and 120 Hz (fourth order). Next, noisy channels containing other artefact than blinks (which were identified based on the electro-oculograms placed around the subjects' eyes) were replaced by the nearest neighboring electrodes (no more than 5% of channels were interpolated per participant) and in addition 2 noisy trials were excluded. Then, the signal was segmented for each trial, from 2 sec prior to the start of the trial to the end of the trial 70 sec later (68 sec + 2 sec). Afterwards, a common average reference computation was applied to all channels excluding ocular channels for each subject independently.

Frequency-domain analyses

Frequency domain analysis was performed to quantify the magnitude of the response to the 2 checkerboards. For that, each epoch was segmented a second time, from each person change to 2560 bins, corresponding to exactly 5 sec (minimum display time before the person in the room changed identity or orientation). In order to exclude the evoked response potential due to the onset of the stimulus, the first segmentation of every trial was excluded from the analysis. Then, each resulting segments were averaged separately for each participant and condition. A Fast Fourier Transformation (FFT) was applied on the averaged segments to extract the amplitude of the overall response for each electrode. Finally, to correct for noise level, for each participant, we subtracted the average voltage amplitude of the 3 surrounding bins (3 on each side, excluding the 2 immediately adjacent bins) from the amplitude of each frequency bin (Dzhelyova & Rossion, 2014) and obtained baseline-corrected amplitude.

ROIs selection

First, to determine the electrodes to take into account we averaged the subtracted baseline amplitude of all participants, then we calculated the Z-score in order to estimate the significance of the response for each electrode. Thus, we defined a region of interest (ROI) centered on Oz and including 15 channels, corresponding to the best ROI for the different harmonics (PO4/3h, POO5/6, O1/2, POI1/2, I1/2, POz, POOz, Oz, Olz, Iz). Secondly, in order to explore the scalp distribution of the EEG response we defined 2 more ROIs of 15 channels, centered in the parieto-temporal cortex (contralateral to the side the checkerboard was located: CPP3/4, CPP3/4h, CP5/6h, CCP3/4, CP5/6, CPP5/6h, PPO3/4, PPO5/6, P5/6, TP7/8h, TP7/8, P7/8, PO7/8, P9/10, PO9/10). Indeed, these regions are known to be involved in several cognitive domains and more specifically in a social context (Bzdok et al., 2013; Aichhorn et al., 2008; David et al., 2006).

Harmonics selection

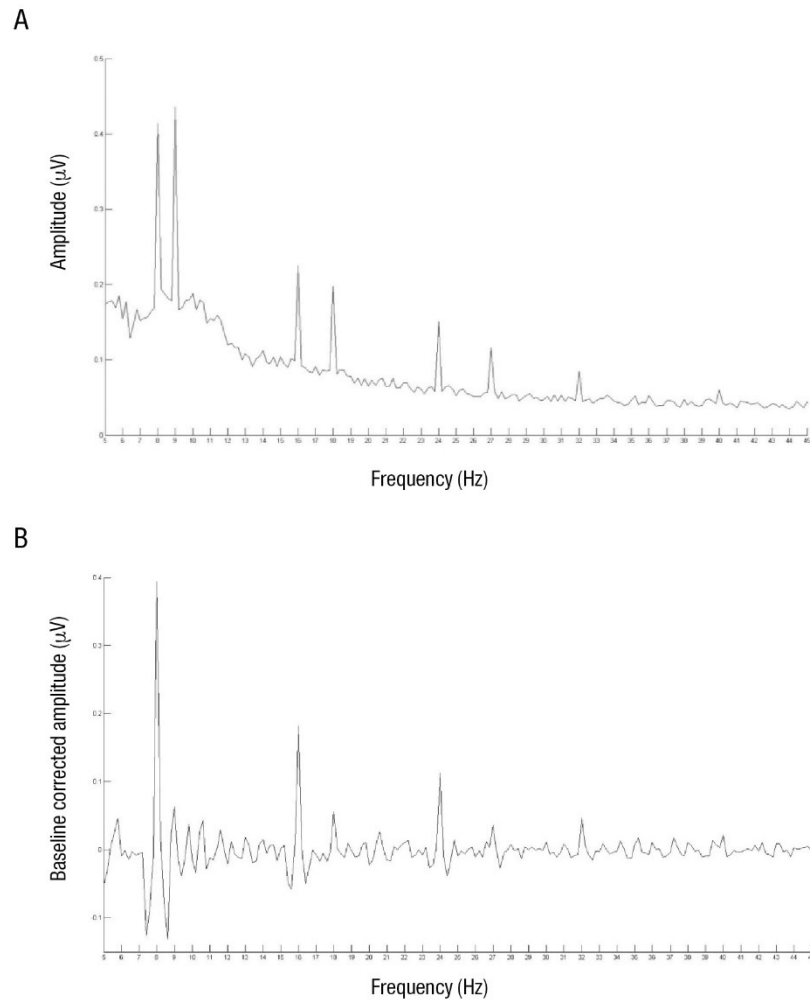


Figure 2. (A) Average across participants and across all conditions of the amplitude on Oz. (B) Average across participants of the baseline corrected amplitude on Oz, in the condition in which the person looked left to the 4 Hz flickering checkerboard.

To determine the presence of an FPVS response, irrespectively of the channel location, we averaged the amplitude responses across all conditions and all participants, then we pooled across all channels and finally we calculated the Z-score. Z-scores were estimated in order to assess the significance of the response and to determine the number of harmonics to take into account for computing the sum of the harmonics. Z-score threshold was 2.58

($p < 0.01$). The 2f, 4f, 6f, 8f had a Z-score > 2.58 for checkerboard flickering at 4 Hz and 4.5 Hz. (Figure 2.A show the amplitude of the response of different harmonics when all conditions are averaged together, 2.B show an example of a response of different harmonics in a specific condition). Thereby, in order to do statistical comparisons across the 4 conditions (person looking left/right to the 4/4.5 Hz flickering checkerboard), we summed the baseline-corrected amplitude of the responses of the 2f, 4f, 6f, 8f from the channels included in the ROIs for each participant.

2. Results

Figure 3 depicts the mean baseline-corrected amplitude values of the sum of the harmonics (2f + 4f + 6f + 8f) for the checkerboards flickering at 4 Hz and 4.5 Hz depending on whether the flickering stimulus was on the right or left and whether it was seen or not seen by the person in the room. The topographical maps showed a lateralization of the response, with a larger response contralateral to the side of the flickering checkerboard. This contralateral activation showed that participants did fixate the fixation cross.

We conducted a three-way repeated measures ANOVA with as independent variables *Visibility* (whether the person saw or not the checkerboard), *Side* (whether the checkerboard is located on the left wall or the right wall), and *Frequency* (whether the checkerboard flickers at a frequency of 4 Hz or 4.5 Hz). Of particular interest was any effect of *Visibility* as this would show that the technique is sensitive enough to detect whether participants attended to where the person is looking at.

Analysis on central occipital area

The 2 (Visibility: checkerboard seen versus the checkerboard not seen by the person) x 2 (Side: left wall versus right wall) x 2 (Frequency of stimulation: 4 Hz versus 4.5 Hz) repeated measures ANOVA on the baseline-corrected amplitude revealed a trend for a main effect of *Visibility*: $F(2, 16) = 3.940$, $p = 0.065$, $\eta p^2 = 0.198$, with a stronger response for the frequency matching the checkerboard seen by the other person. There was no main effect of side $F(2, 16) = 0.447$, $p = 0.513$, $\eta p^2 = 0.027$ but a main effect of *Frequency* $F(2, 16) = 10.424$, $p = 0.005$, $\eta p^2 = 0.394$ with a stronger response for checkerboard flickering at 4 Hz than 4.5 Hz. None of the interactions were significant ($F < 0.729$, $p > 0.406$).

Analysis centered in parieto-temporal area

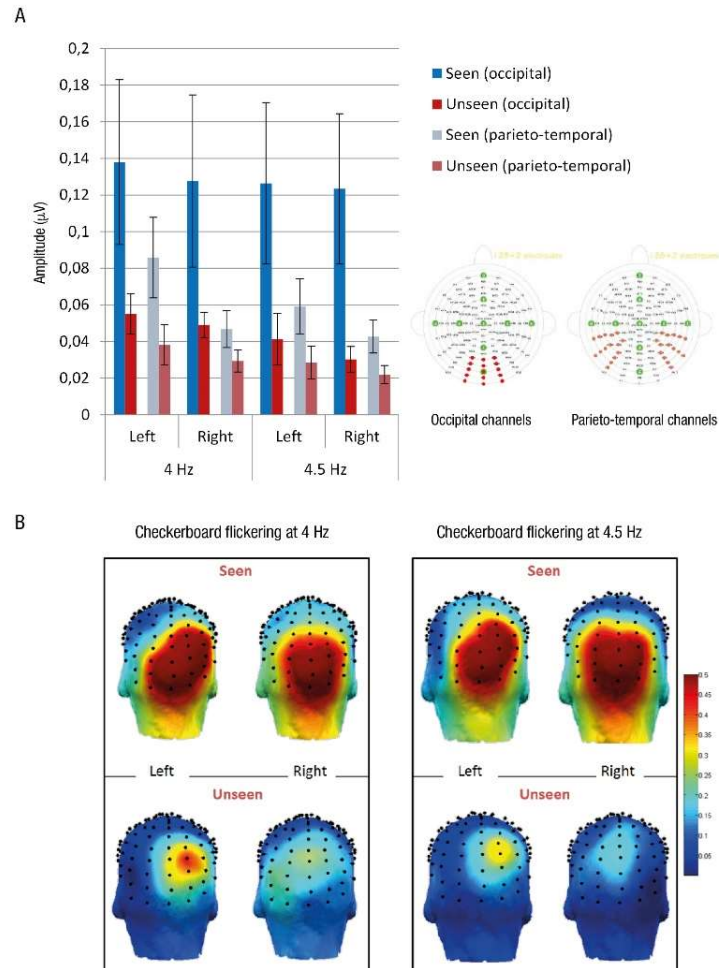


Figure 3. Baseline-corrected amplitude values of the sum of the harmonics ($2f + 4f + 6f + 8f$) for the checkerboards flickering at 4 Hz and 4.5 Hz separately display in function of the side (left wall, right wall) and depending if the checkerboard is seen or not by the person, “Seen” means the person does see the stimulus flickering while “Unseen” means the person does not see the stimulus. **(A)** Bar graph of the baseline-corrected amplitude values taken from the central occipital electrodes and parieto-temporal electrodes. Bars represent standard errors. **(B)** Topographical maps with the colour scale.

The same $2 \times 2 \times 2$ ANOVA performed on the baseline corrected amplitude, revealed a main effect of Visibility: $F(2, 16) = 8.603$, $p = 0.010$, $\eta p^2 = 0.350$, with a stronger corrected amplitude for the frequency matching the checkerboard seen by the person. There was a

trend for an effect of Side $F(2, 16) = 4.209, p = 0.057, \eta p^2 = 0.208$ with a stronger amplitude response for the checkerboard on left than right wall, and a main effect of Frequency $F(2, 16) = 5.627, p = 0.031, \eta p^2 = 0.260$ with a stronger response for checkerboard flickering at 4 Hz than 4.5 Hz. None of the interactions were significant ($F < 1.964, p > 0.180$).

3. Discussion of Experiment 1

Topographical maps (Figure 3) show a larger response to the checkerboard seen by the person compared with the checkerboard not seen by the person. If we consider the parieto-temporal region, the protocol of FPVS we used was sensitive enough to detect participants' computation of what another person is looking at when participants were explicitly instructed to attend to what the other person is looking at. Interestingly we only found a trend in central occipital region to support this effect.

Moreover, statistical analyses highlighted also a different sensibility to the frequencies we used, both on occipital and parietal electrodes. However, this effect was strongest in occipital region. These results could support the fact that occipital region is more involved in processing low level visual feature whereas parieto-temporal region is more involved in higher level information processing.

To deepen the effect of the person's visibility we conducted a second experiment to examine whether FPVS is sensitive enough to capture not only on-command perspective taking (as shown by Experiment 1) but also spontaneous perspective taking. In this second experiment, participants were asked to attend to the person without being asked to attend to the person's visual experience.

II. Study 2: Explicit spatial perspective taking instruction

The rationale of the Experiment 2 was to make the person important in the task but without explicitly asking participants to compute what the person was looking at. This was achieved by asking participants to either focus on the checkerboard located in front of the person or to attend to the checkerboard located behind the person (spatial perspective taking instructions). Again, participants were instructed not to move their eyes away from the central fixation cross.

We expected a larger neuronal response for the checkerboard attended by the participant than the checkerboard not attended by the participant and this would signal that participants were following the instructions. However, when participants were instructed to attend to the checkerboard in front of the person, the checkerboard they were asked to attend to was also seen by the person. However, when participants were asked to attend to the checkerboard behind the person, the checkerboard they were asked to attend to was this time not seen by the person. Thus, if participants spontaneously compute what the person is looking at, we should expect a larger response for the checkerboard to attend by the participant in the Front-instruction condition (checkerboard seen by the person) than in the Back-instruction (checkerboard not seen by the person). Furthermore, we should expect the reverse results for the checkerboard to ignore, i.e., a larger response for the Back-instruction condition (checkerboard seen by the person) than Front-instruction (checkerboard not seen by the person).

1. Methods

Participants

Forty-two healthy volunteers (female: 29, mean age: 22.1) with normal or corrected-to-normal vision took part in Experiment 2 in exchange of a small honorary. All subjects were different from Experiment 1 and gave written informed consent prior to the experiment. Four participants were subsequently excluded from the sample because of their poor task performances (< 80 % correct responses) and 2 other because of their abnormal noisy EEG recording, leading to 20 subjects who attended in the first part of the experiment the checkerboard located in front of the person and in the second part of the experiment the checkerboard located behind the person while 16 subjects performed the two parts of the experiment in the reverse order.

Stimuli

The stimuli used were exactly the same as those used in Experiment 1.

Procedure

The procedure was the same as for Experiment 1, only the instructions were different.

The experiment was divided in two parts of 12 trials each. Written instructions were given to the subject before starting each part of the experiment. In one part (Front-instruction condition), the subject had to focus on the checkerboard in front of the person while doing the fixation cross task. In the other part (Back-instruction condition), the subject had to focus on the checkerboard behind the person while doing the fixation cross task.

EEG

The equipment was the same as in Experiment 1. Nineteen noisy trials were excluded from the analysis due to abnormal artefacts. All steps of the analysis were the same as in experiment 1.

Based on Z-score calculated from the pool of channels, we determined the number of harmonic frequencies to take into account. We analyzed the 2f, 4f, 6f, 8f and 10f, the common harmonics with a Z-score > 2.58 for checkerboard flickering at 4 Hz and 4.5 Hz (Figure 4 shows the amplitude of the response of different harmonics when all conditions are averaged together in the front instruction condition). We used the same occipital ROI of 15 channels as in Experiment 1, centered on Oz and the 2 additional parieto-temporal ROIs.

Harmonics selection

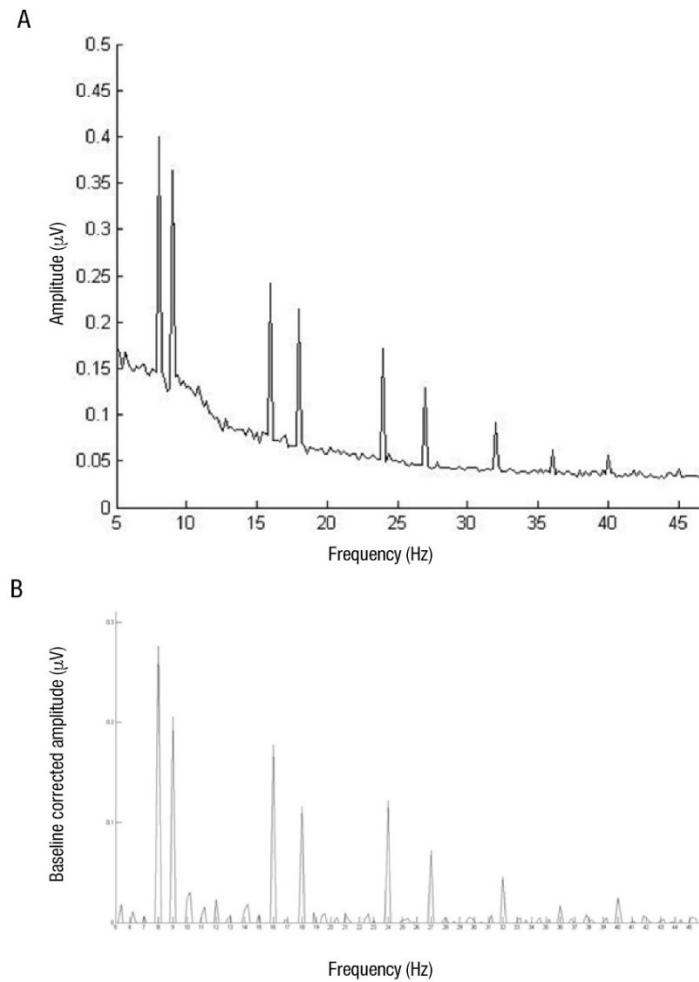


Figure 4. (A) Average across participants and across all conditions of the amplitude on Oz in the instruction front instruction condition. (B) Average across participants of the baseline corrected amplitude on Oz, in the condition in which the person looks right to the 4.5 Hz flickering checkerboard in the Back-instruction condition.

2. Results

Figure 5 represents baseline-corrected amplitude of the sum of the harmonics ($2f + 4f + 6f + 8f + 10f$) values taken from the central occipital and from the parieto-temporal

electrodes for each experimental part, depending on whether the flickering stimulus was attended to or not by the participants.

We conducted a four-way repeated measures ANOVA with the factors *Instruction* (attended front, attended back of the person), *Congruency* (checkerboard congruent with the instruction versus incongruent with the instruction), *Side* (left wall, right wall), and *Frequency* (4 Hz, 4.5 Hz) and *Order* (Front-instruction first versus Back-instruction first) as between subjects factors. As a reminder, evidence for spontaneous perspective taking should result in an *Instruction* x *Congruency* interaction effect.

Analysis on central occipital area

The 2 (*Instruction*: checkerboard attended in front of the person versus checkerboard attended in the back of the person) x 2 (*Congruency*: checkerboard congruent with the instruction versus incongruent with the instruction) x 2 (*Side*: left wall versus right wall) x 2 (*Frequency*: 4 Hz versus 4.5 Hz) repeated measures ANOVA with *Order* (Front-experiment first versus Back-experiment first) as between subjects factors showed a main effect of *Congruency* with a larger response for the checkerboard attended by the participant compared with the checkerboard ignored by the participant $F(3, 34) = 5.317, p = 0.027, \eta p^2 = 0.135$. There was no main effect of *Side* $F(3, 34) = 0.622, p = 0.436, \eta p^2 = 0.018$. There was a main effect of *Frequency* $F(3, 34) = 23.649, p < 0.000, \eta p^2 = 0.410$, with a stronger response for the checkerboard flickering at 4 Hz than the one flickering at 4.5 Hz. There was no main effect of *Order* $F(3, 34) = 0.434, p = 0.514, \eta p^2 = 0.013$, and no significant *Instruction* x *Congruency* interaction $F(3, 34) = 0.434, p = 0.514, \eta p^2 = 0.013$.

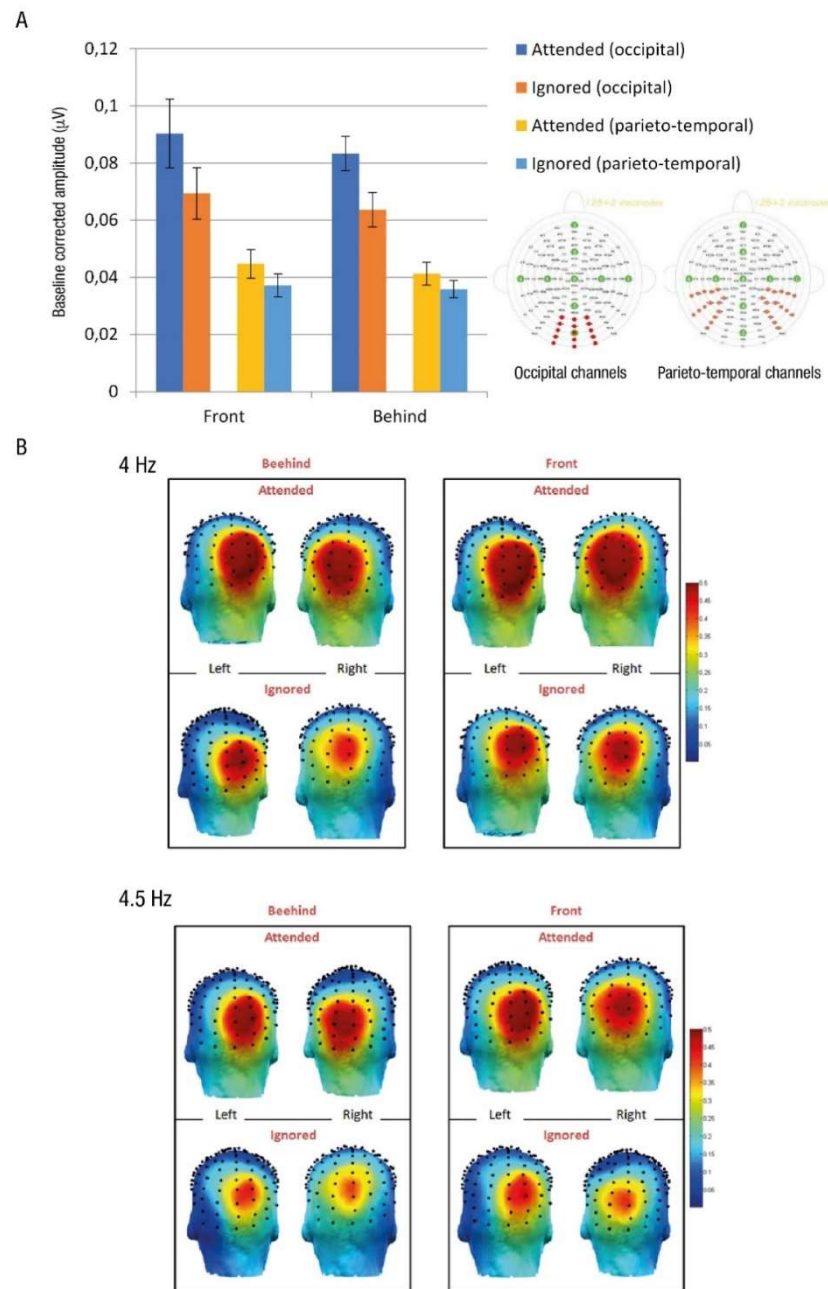


Figure 5. (A) Bar graph of the baseline-corrected amplitude of the sum of the harmonics ($2f + 4f + 6f + 8f + 10f$) separately display for experiment in Front and experiment Behind and depending if the checkerboard was attended or ignored by the participants. Bars represent standard errors. **(B)** Topographical maps with the colour scale of the sum of the harmonics for the 4 Hz and 4.5 Hz flickering checkerboard.

Analysis centered in parieto-temporal area

The same $2 \times 2 \times 2 \times 2$ ANOVA performed on the baseline corrected amplitude, revealed a main effect of Congruency $F(3, 34) = 5.357, p = 0.027, \eta p^2 = 0.136$, with a stronger response for checkerboard attended than ignored by the participant. The Frequency effect was significant $F(3, 34) = 10.545, p = 0.003, \eta p^2 = 0.237$, with a larger response for the checkerboard flickering at 4 Hz than the one flickering at 4.5 Hz. There was no significant Instruction x Congruency interaction $F(3, 34) = 0.483, p = 0.492, \eta p^2 = 0.014$.

3. Discussion of Experiment 2

The main effect of Congruency was present and equivalent in both, occipital and parieto-temporal region. This results support the fact that when participants received the instruction to attend to a specific checkerboard, the FPVS paradigm we used was sensitive enough to detect a larger baseline corrected amplitude response for the checkerboard attended by the participant, compared with the checkerboard ignored by the participant.

However, we did not found the expected Instruction x Congruency interaction that could show that participants pay attention to the stimulus seen by the person even when instructed to look at the stimulus behind him. One explanation of this lack of results is that our FPVS protocol is not well adapted to study spontaneous visual perspective taking. One could argue that the fact that participants were under attentional load (they must attend a specific stimulus and in addition doing the fixation cross task) could have interfered with the calculation of the person perspective. This seems, however, unlikely to us since Qureshi et al. (2010) showed that participants continue to calculate what the person sees when they are under cognitive load. There are several other factors, however, that make our protocol perhaps unsuitable. First, flickering stimuli are salient (and numerous participants reported movement illusion from the flickering checkerboards). It is known that salient stimuli attract attention (Corbetta et al., 2002) and that attentional resources capacity is limited (Treisman et al., 1980). This could lead to a competition of the checkerboards and the person to catch automatic attentional resources. If this is the case, the flickering checkerboards could prevent the involuntary computation of the person's perspective and explain our absence of amplitude difference between the checkerboards seen and not seen by the person.

Secondly, attention capture and orienting are transient effects (Wilschut et al., 2011), and this could lead the person's perspective to have a very small impact over time on the checkerboard participants attend to in the absence of explicit instruction, making FPVS not sensitive enough to measure this effect.

Thirdly, a bias reported in the literature (Andersen et al., 2013; Störmer & Alvarez, 2014) is that when participants attend to a feature (e.g., colored dots) in one visual hemifield flickering at a specific frequency and the exact same feature is also presented in the opposite visual hemifield flickering at a different frequency, the FPVS amplitude response to these stimuli is increased, both for the attended and the unattended stimuli (but the increase of the response is more pronounced for the attended stimulus). In our experiment, we used 2 identical checkerboards and we can assume that the same effect is happening, an increase of the FPVS response for the checkerboard seen but also for the checkerboard unseen by the person. The attentional bias to the incongruent (not attended) stimulus could make the attentional effect of the visual perspective taking harder to measure with our protocol.

Fourthly, others properties of the FPVS could affect the amplitude of the response to the stimuli: the stimulation frequency used (Ding et al., 2006; Toffanin et al., 2009), the stimulation type (sinusoidal or squared wave modulation), stimuli in colors or in greyscale. The influence of all these factors on the FPVS response is not yet fully understood (Russo et al., 2002; Zhu et al., 2010). They could independently improve or decrease the sensitivity to capture visual perspective taking. Finally, if using long sequence trial as we do increases the frequency resolution, we do not know how FPVS amplitude response evolve over time and the impact it has on the participant's mental fatigue.

Despite these methodological limitations, it is, nevertheless interesting to note that we found a reliable signature of visual and spatial perspective taking in the explicit perspective-taking instruction condition: similarly to previous studies investigating spatial attention (Morgan et al., 1996; Andersen et al., 2013) we found an enhanced baseline corrected amplitude for the attended stimulus. To our knowledge, we are the first to show this effect even when, within a trial, participants had to successively attend different target and switch rapidly between visual hemifield.

Appendix 1

Given the unsuitability of the paradigm to capture spontaneous perspective taking, we turned to other paradigms in subsequent studies.

Appendix 2: Replication of the Oddball paradigm (Study 1 of Chapter 2) without VPT instruction

Beck, Alexy-Assaf¹, Rossion, Bruno¹ & Samson, Dana¹

¹ Psychological Sciences Research Institute (IPSY), Institute of Neuroscience (IoNS), University of Louvain, Belgium.

Abstract

We develop a fast-periodic visual stimulation (FPVS) paradigm to explore implicit visual perspective taking. As in Study 1 of Chapter 2, we presented images of a person in a room with an object on one of the walls at a rapid rate (2.5 images/s) while recording electrophysiological brain activity. Stimuli consisted of 4 consecutive images in which the person could see the object followed by a 5th image in which the person could not see the object. We identified specific brain responses at 0.5 Hz (2.5 Hz/5) which represent a spontaneous discrimination response between an object seen/unseen by the other person even though participants were not asked to pay attention to the person's visual experience. In comparison with the results of Study 1 of Chapter 2, the findings here suggest that less deeper processes are involved during Implicit- than Explicit- visual perspective taking.

1. Materials and methods

Participants

Sixteen healthy volunteers (female: 9, mean age: 21 ± 2.21 , all right-handed) with normal or corrected-to-normal vision took part in the experiment in exchange of a small honorarium. All subjects were different from those who participated in Study 1 of Chapter 2 and gave written informed consent prior to the experiment. One participant was subsequently excluded from the sample because of his large number of blinks within each trial (mean: 21.6 blink/trial).

Stimuli

The stimuli used were the same as those used in Study 1 of Chapter 2 with the exception that the red objects were not present anymore.

Procedure

Within one trial, there were always 13 different person's identity (8 females), and 12 different objects. Participants performed 24 stimulation sequences (12 per condition). Periodic and non-periodic conditions alternated between sequences, with half of the participant starting with a periodic one and the other half starting with a non-periodic.

Participants were asked to always fixate the fixation cross placed in the middle of the screen and to respond by pressing the space bar as quickly and accuracy as possible when they detected a male person on the screen (15 targets per trial). They were not asked to pay attention to any object on the wall. The goal of this task was to ensure that participants maintained their attention throughout the trials and paid attention to the person on the screen. A response was considered as correct if it was given within three standard deviations of the participant's mean reaction time.

EEG

The equipment and the data analysis were the same as in Study 1 of Chapter 2.

2. Results

2.1 Behavioral Data

All participants successfully performed the task (mean = 93.44 ± 4.82 ; false alarm = $2.79\% \pm 1.46$; respond time = $0.59 \text{ s} \pm 0.043$). There was no significant performance difference between the periodic and the non-periodic conditions ($t(14) = 1.090$, $p = 0.294$).

2.2 EEG analysis

Base frequency analysis

As in Study 1 of Chapter 2, in order to know the number of harmonics to take into account for further base rate analysis, we pooled the 128 electrodes for each participant, then we merged the 2 conditions (periodic and non-periodic conditions) and computed the grand average, and then computed the Z-score. There were significant responses up to the 14th harmonic (35 Hz; threshold of significance placed at a Z-score of 3.30). We cropped the individual FFT spectrum, centered at the periodic frequency of the base rate up to the 14th harmonic with the 10 surrounded neighboring bins on each side. We summed the spectrum of the harmonics and their neighboring bins and computed the SNR and the baseline corrected amplitude then we computed the grand average of the 2 conditions independently and the 2 conditions together. The topographical maps were very similar to Study 1 of Chapter 2, therefore we selected the same ROI (POOz, POz, POO5, POO6, Oz). Then, we performed a pairwise t-test between the periodic and non-periodic conditions. There was no significant difference in the baseline corrected amplitude across the two conditions ($t(14) = 0.775$; $p = 0.451$; Figure 1). Thus, as expected the response at the base frequency rate did not differ between the periodic and non-periodic conditions.

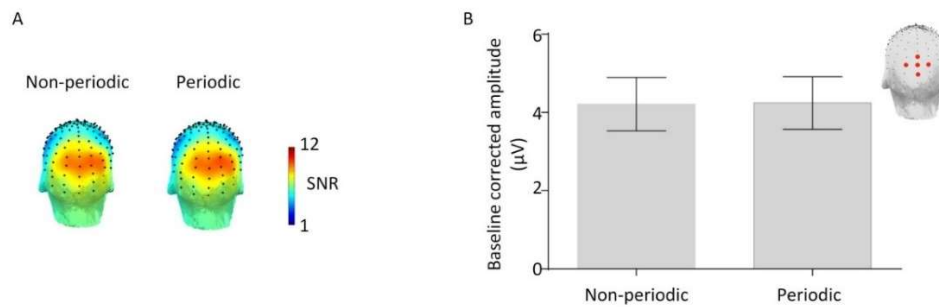


Figure 1. Base frequency response in the non-periodic and periodic conditions. **(A)** Scalp topography of the sum of the 14th first harmonics of the base frequency for both conditions (SNR). **(B)** Bar graph of the sum of the baseline corrected amplitude of the 14 first harmonics of the base rate over the occipital ROI. Error bars represent mean with 95% CI.

Periodic response of interest analysis

We computed the Z-score for the 6 first harmonics over the 2 same ROIs as in Study 1 of Chapter 2. There were no significant harmonics in the non-periodic condition. In contrast, for the periodic condition, the responses were significant for the 2nd harmonic over the Prefrontal ROI and the 1st and the 3rd harmonic were significant over the Centro-parietal ROI (Table 1).

We summed the same number of harmonics as in Study 1 of Chapter 2, i.e., the first fifth harmonics of the periodic frequency of interest (0.5 Hz, 1 Hz, 1.5 Hz, 2 Hz, 3 Hz) and we computed the Z-score on the 2 same ROI as in Study 1 of Chapter 2 (one-tailed Z-score threshold set at 1.96, $p < 0.025$). The Z-scores were not significant in the non-periodic condition over the Prefrontal ROI ($Z = 1.89$) and the Centro-parietal ROI ($Z = -0.31$). For the periodic condition, Z-score was not significant over the Prefrontal ROI ($Z = 0.87$) but was significant over the Centro-parietal ROI ($Z = 1.96$).

Then, we computed the SNR and merged the 2 conditions. Based on the resulting topographical map, we selected one ROI of 8 channels (FCC1, FCC1h, FCC2, FCC2h, C1, C1h, C3, C3h; Figure 2). Then, we computed the SNR over the selected ROI for the 2 conditions separately (Figure 3).

Table 1. Z-score response over the 2 ROIs defined in Study 1 of Chapter 2 for the fifth first harmonic of the periodic frequency of interest for the 2 conditions separately (one-tailed Z-score threshold set at 1.96, $p < 0.025$). Numbers in bold indicate significant responses.

Frequency (Hz)	Non-periodic Z-score		Periodic Z-score	
	Prefrontal ROI	Centro-parietal ROI	Prefrontal ROI	Centro-parietal ROI
1F = 0.5	1,12	1.00	0,66	2,13
2F = 1	0,38	-2.07	9.17	0,82
3F = 1.5	0,39	-1,38	0,37	3.03
4F = 2	0,17	1,10	0.75	0,26
6F = 3	-1.92	1,01	1,25	-1,69

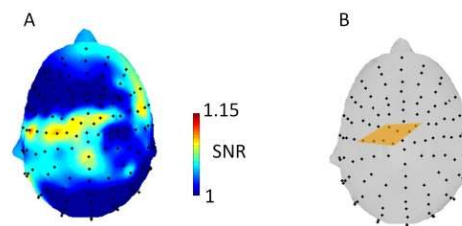


Figure 2. (A) Scalp topography of the grand average of the 2 conditions of the sum of the 5th first harmonic (excluded 2.5 Hz; SNR). (B) Selected ROI for further analysis.

From the ROI, we computed the SNR (Figure 3) and the Z-scores for each harmonic at the periodic frequency of interest for the two conditions. We took into account the same number of harmonics as in Study 1 of Chapter 2 (i.e., until the sixth harmonic (3 Hz)).

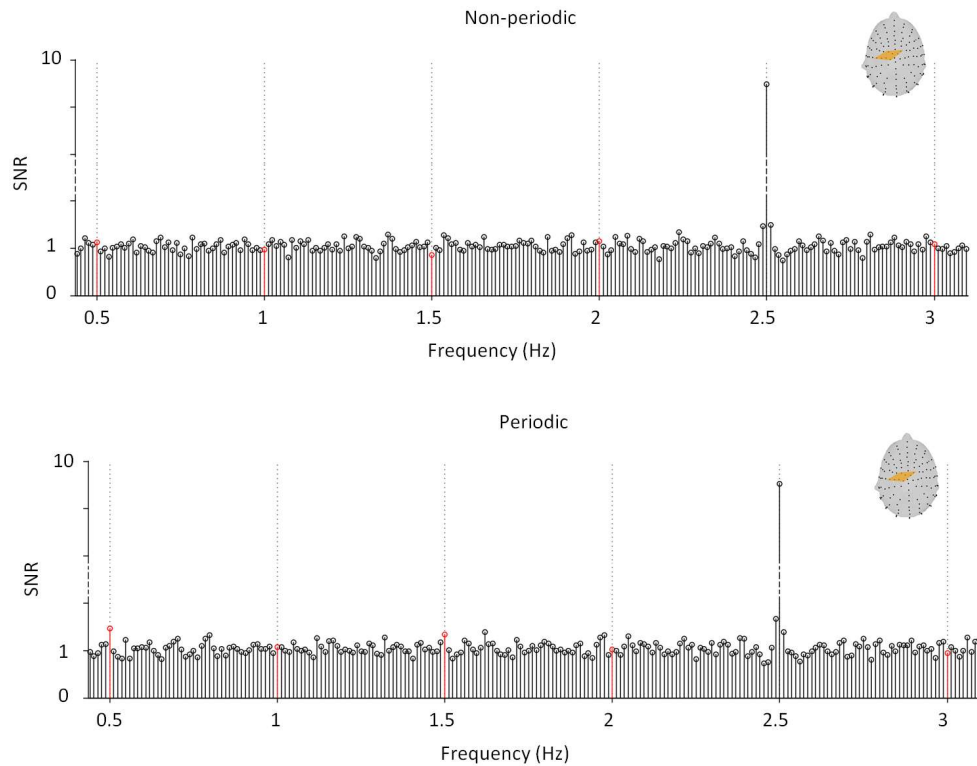


Figure 3. Grand average of the snr spectrum over the ROI and for each condition (non-periodic/periodic) independently.

There were no significant harmonics in the non-periodic condition. In contrast, for the periodic condition, the responses were significant for the first, and the third harmonics over the ROI (Table 2). We cropped the grand average FFT spectrum, centered at the periodic frequency of interest over the ROI (0.5 Hz, 1 Hz, 1.5 Hz, 2 Hz, 3 Hz) with the 10 surrounded neighboring bins on each side. We excluded the fifth harmonic (2.5 Hz) because it corresponded also to the base frequency rate. We summed the spectrum of the harmonics and their neighboring bins and computed the Z-score and the SNR over the ROI (Figure 4). The resulting Z-score were significant in the periodic condition only; $Z = 3.13$ versus $Z = -0.21$ for the periodic and non-periodic conditions, respectively.

These results showed a differential response of the brain for the consistent and the inconsistent perspective stimuli even in absence of explicit VPT instruction.

Table 2. Periodic frequency of interest. Grand average Z-score responses for the 2 conditions for different harmonics of the periodic frequency of interest. p-Values were calculated on the basis of Z-scores, numbers in bold indicate significant responses (one-tailed Z-score threshold set at 3.30, $p < 0.0005$).

Frequency (Hz)	ROI Z-score	
	Non-periodic	Periodic
1F = 0.5	0.73	3.39
2F = 1	-1.28	0.93
3F = 1.5	-2.46	3.61
4F = 2	1.56	-0.43
6F = 3	0.18	-1.39

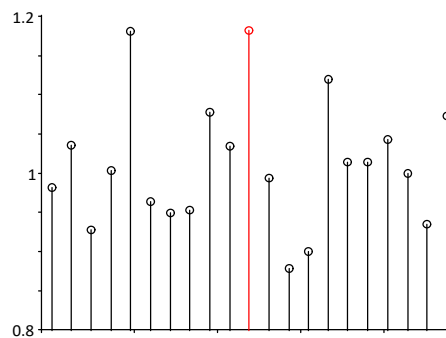


Figure 4. Response of the sum of the harmonics of the periodic frequency of interest of the periodic condition (SNR). The spectrum is centered at the response of the sum of the 5 harmonics, surrounded by the summed of the neighboring bins, indicating the noise levels.

Individual analysis

As in Study 1 of Chapter 2, for each participant, we cropped the FFT spectrum, centered at the periodic frequency of interest (0.5 Hz, 1 Hz, 1.5 Hz, 2 Hz, 3 Hz) with the 10 surrounded neighboring bins on each side. We summed the spectrum of the harmonics and their neighboring bins and computed the Z-scores and the SNR. The significant threshold was placed at one-tailed Z-score of 1.65 ($p < 0.05$) because we assumed that the signal response would then be above the noise level (Table 3). To illustrate the individual results, we show the SNR of a single electrode belonging to the ROI (Figure 5).

Appendix 2

Table 3. Individual participants' Z-score response computed on the sum of the harmonics of the periodic response of interest over the ROI. *p*-Values were calculated on the basis of Z-scores (one-tailed Z-score threshold placed at 1.65, $p < 0.05$).

Participant	ROI Z-score
P1	4.20
P2	3.46
P3	0.61
P4	-0,17
P5	-5.09
P6	1.70
P7	0,19
P8	1.14
P9	6.09
P10	0.54
P11	0,23
P12	2.60
P13	0.66
P14	1.04
P15	2.34

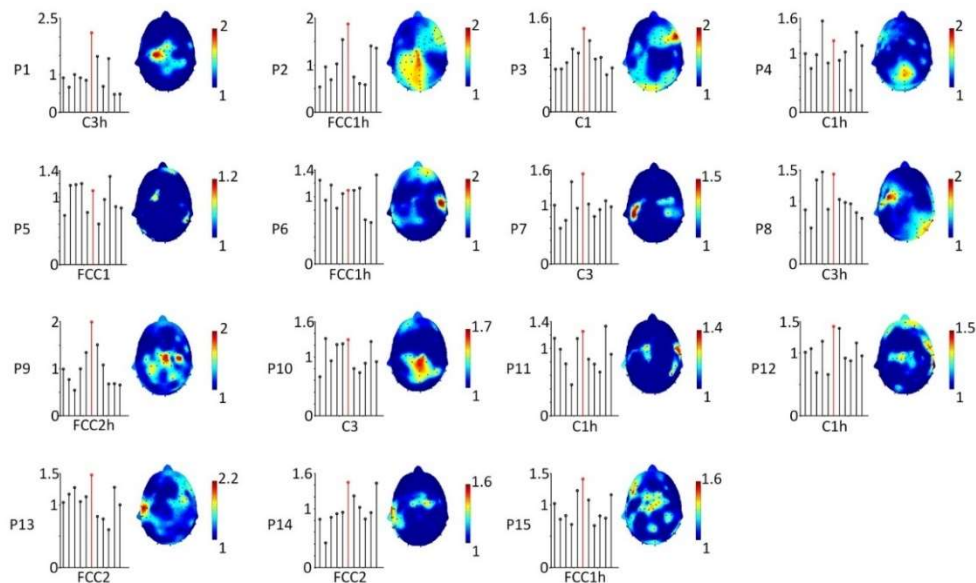


Figure 5. Illustration of individual participants' SNR response to changes of perspective consistency over one electrode that belongs to the ROI. The spectrum is centered at the response of the summed of the 5 harmonics, surrounded by the summed of the neighboring bins which indicate noise levels (for visibility only the 5 neighboring bins on each side are shown).

3. Discussion

The aim of this study was to compare the brain activity associated with the processing of stimuli while participants received explicit perspective-taking instructions (as examined in Study 1 of Chapter 2) versus without such instruction (as examined in the current Implicit-condition).

We found an objective brain signature at the exact frequency at which participants could witness the perspective of another person changing from being consistent with their perspective (the other person saw the object on the wall that participants could also see) to becoming inconsistent (the other person could not see the object that participants could see). Such brain signature shows that the brain discriminated between consistent and inconsistent perspectives. This result agrees with previous behavioral studies that showed that we compute other people's visual perspective regardless of the task instructions (Samson et al., 2010). Furthermore, the current response (i.e., in the Implicit-condition) at

the frequency of interest suggests the parallel activation of the computation of our own visual perspective and the perspective of someone else.

The centro-parietal scalp region responded to the change of perspective consistency in both the Explicit- and Implicit-conditions. However, the response was larger in the Explicit-condition and more harmonics were involved (significant harmonic up to the 6th versus up to the 3rd harmonic in Implicit-condition) suggesting prolonged and deeper processing of information in the Explicit-condition. This pattern of neural response is compatible with the hypothesis that a set of brain regions calculate another person's perspective with a differential level of depth of processing depending on the explicit or implicit nature of the task. The localization of the response on the scalp is compatible with the involvement of temporo-parietal brain areas in the computation of another person's perspective (McCleery et al., 2011). However, a strong word of caution is required here as EEG alone is not suitable to make strong brain localization claims.

At the individual level, the perspective consistency discrimination signature was clearly visible for participants who performed the Explicit-condition, whereas in Implicit-condition the discrimination signature was not visible in all participants. One explanation could be that there are larger inter-individual differences in the ability or likelihood to spontaneously compute other people's perspective. Another explanation could be that the discrimination response between perspectives is less easily detectable by the FPVS technique in the implicit condition due to the overall weaker brain activity.

Overall our results support the hypothesis that implicit VPT may engage lower levels of processing (highlighted by the overall lower response and the fact that less harmonics are involved).

Appendix 3: Pre-test studies

Our different paradigms (attentional orienting, oddball, distraction, intermodulation paradigms) necessitated considerable runs of pre-tests to find suitable parameters. Here I present a list of these pre-tests.

1. Attentional paradigm

Different frequencies (8 Hz and 9 Hz or 22 Hz and 27 Hz) were tested as well as different types of stimuli displayed on the wall (e.g., small circles, big circles - red or black - , the wall itself flickered) to finally choose the checkerboards and the 4 and 4,5 Hz frequencies.

2. Oddball paradigm

We ran several oddball paradigms in which we varied either the task instruction (e.g., fixation cross task or detecting each time the person saw a food image or each time the person was a male), the frequencies for the base or the oddball (e.g., base images were presented at 6 Hz and an oddball image was presented every 7 base images which corresponds to a frequency of 0.86 Hz or base images were presented at 2.35 Hz and an oddball image was presented every 3 base images which corresponds to a frequency of 0.78 Hz). Note that it was not possible to test participants with explicit VPT instruction and fast base frequencies (i.e., 6 Hz) because it made it impossible to respond to the task. Furthermore, we also varied whether the base images depicted a person seeing or not an object and vice versa for the oddball images.

We finally chose to present images of a person in a room seeing an object on the left or right wall at 2.5 Hz. Every 5th image the person could not see the object i.e., at 0.5 Hz (Study 1 of Chapter 2 & Appendix 2).

Table 1. Summary of the parameters of the oddball paradigm tested. *strong SNR response in the EEG spectrum

	Explicit VPT instruction				
	Task: detect avatar facing target image				
Base = Avatar facing the object	2.35 Hz*	2 Hz*	2 Hz*	3 Hz*	
Oddball = Avatar not facing the object	0.78 Hz*	1 Hz*	0.66 Hz*	0.6 Hz*	
Base = Avatar not facing the object	2.5 Hz*	2.5 Hz*			
Oddball = Avatar facing the object	0.833 Hz*	0.5 Hz*			

	Implicit VPT instruction		
	Task: detect fixation cross change		
Base = Avatar facing the object	6 Hz*	2 Hz*	2.5 Hz*
Oddball = Avatar not facing the object	1.2 Hz	0.66 Hz	0.5 Hz
Base = Avatar not facing the object	6 Hz*	6 Hz*	
Oddball = Avatar facing the object	0.86 Hz	1.2 Hz	

	Task: detect male avatar				
Base = Avatar facing the object	2 Hz*	1 Hz*	2.35 Hz*	2 Hz*	2 Hz*
Oddball = Avatar not facing the object	0.5 Hz	0.5 Hz	0.78 Hz*	1 Hz*	0.66 Hz

3. Distraction paradigm

The distraction paradigm consisted in placing randomly moving dots flickering on the stimuli's region of interest. For example, in our pre-test, superimposed on the left and right wall were randomly moving dots which flickered at 17 Hz on one side and 20 Hz on the other (Figure 1). Within a trial the person always faced the same side. Participant had to detect fixation cross change.

We did not observe difference whether the person was facing or not the object.

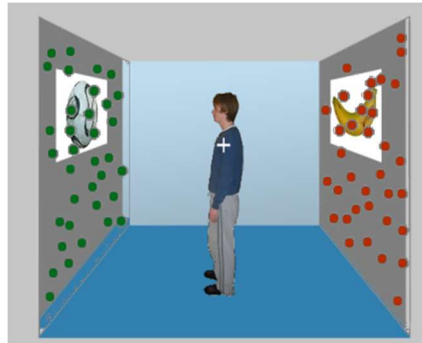


Figure 1. Example of stimuli used in the distraction paradigm.

4. Intermodulation paradigm

The stimuli of the intermodulation paradigm as used here depicted the person with 2 objects (one in front and one in the person's back). We frequency tagged at 3 different frequencies the person and the two objects displayed. First, the person was frequency tagged at 2 Hz, the objects were frequency tagged at 2.4 Hz and 2.6 Hz. We observed clear IMs on the EEG spectrum for both the implicit and explicit VPT instruction tested (e.g., fixation cross task or detecting each time the person saw a food image or each time the person was a male). Whereas we did not observe IM responses in our second pre-test condition in which the person was frequency tagged at 2.35 Hz, and the objects at 8 and 9 Hz (implicit VPT instruction).

In another pre-test, we create a totally new set of stimuli which were similar to the ones used in Study 3 of Chapter 4 except that the room was larger (and so, the distance between the person and the object was increased). The person was frequency-tagged at 2 Hz and the object at 2.4 Hz. Participants' task was to detect each person's target contour and each object target (same instruction as in Studies 2 and 3 of Chapter 3 and 4 respectively). We observed small IM responses and we decided to reduce the size of the room.

Table 2. Summary of the parameters of the intermodulation paradigm tested. Note that we used similar stimuli as the one used in the distraction paradigm without the dots.

Avatar Object IM	Explicit VPT instruction			
	Task: detect avatar facing target image			
	2 Hz 2.4 Hz & 2.6 Hz yes			
Avatar Object IM	Implicit VPT instruction			
	Task: detect fixation cross change		Task: detect avatar's target contour and target object	
	2.35 Hz 8 Hz & 9 Hz no	2.35 Hz 8 Hz & 9 Hz no	2 Hz 2.4 Hz & 2.6 Hz yes	2 Hz 2.4 Hz yes

5. Discussion

To sum up we identified three parameters which seem to influence the presence of an oddball or IM responses in the EEG spectrum. A first parameter which seems to clearly influence the presence of an oddball or IM responses is the type of task instructions used. The task had an effect on the amplitude of the response of interest. For example, the explicit VPT instruction led to the strongest oddball (and IM) responses. Furthermore, when participant performed an implicit VPT instruction, we observed an oddball response only when participant focus on the person (i.e., detection of male person; the detection of the change of the fixation cross color did not generate an oddball response). Such results are in line with the studies in the literature (Bukowski et al., 2016; Gardner et al., 2017, Experiment 1) which showed that the mere presence of a person is not sufficient to elicit VPT processes. Our pre-tests emphasize that VPT (i.e., the calculation of the perspectives) might not be always spontaneous and would be dependent on the attention deployment in response to the task instruction.

A second factor which seems to influence the amplitude of the response of interest was the frequency used. For example, faster base frequencies (e.g., 6 Hz) did not seemed to lead to oddball responses. The oddball paradigms generally showed oddball responses when the base frequency used was around 2 Hz. Certainly, lower frequencies give more

time to fully process the stimulus and so, allow a discrimination response between stimuli categories.

Finally, a third parameter was identified by the decrease of the amplitude of the IM response when the different frequency-tagged elements (i.e., the person and the object) were more distant from each other. Therefore, the distance between the frequency-tagged elements can play also a crucial role in the amplitude of the response of interest. Certainly, closer spatial relationships between the frequency-tagged elements would facilitate their common integration.