UNIVERSITAT DE BARCELONA

The self in the face of error: neural signatures of agency attribution

Alba Gómez Andrés

THE SELF IN THE FACE OF ERROR

Neural signatures of agency attribution

Alba Gómez Andrés

PhD Thesis

Cognition and Brain Plasticity Unit

IDIBELL
Institut d'Investigació Biomèdica de Bellvitge

UNIVERSITAT DE BARCELONA

THE SELF IN THE FACE OF ERROR

Neural signatures of agency attribution

Alba Gómez Andrés

Cognition and Brain Plasticity Unit

Department of Cognition, Development and Educational Psychology

University of Barcelona

Barcelona, May 2022

Doctoral program in Brain, Cognition and Behaviour

Supervisors:

Prof. Dr. Antoni Rodríguez Fornells

Prof. Dr. Toni Cunillera Llorente

AGRADECIMIENTOS

Muchísimas gracias a mis directores de tesis Antoni Rodríguez-Fornells y Toni Cunillera. Gracias por darme esta oportunidad única de poder dedicar unos años de mi vida a estudiar e investigar algo tan aparentemente poco pragmático como la consciencia y el sentido del yo. Gracias **Toni** por pensar en mi para este proyecto y darme la confianza para llevarlo a cabo. Admiro mucho tu inagotable curiosidad y motivación. Te agradezco que siempre hayas apoyado mis ideas y me hayas permitido llevarlas a cabo. Gràcies **Cuni** per tot el suport i ajuda durant aquests anys. Gràcies per trobar sempre un moment per parlar, per totes les teves múltiples revisions tant exhaustives i per tots els teus consells. En definitiva, he estat molt afortunada de poder compartir aquest camí amb tots dos, formeu un gran tàndem!

Sin duda, gran parte de mi experiencia durante este doctorado se ha desarrollado en **Brainvitge,** un gran grupo de investigación al que he tenido la suerte de pertenecer. Aun con todas las sorpresas que me ha traído la vida durante estos años, Brainvitge siempre ha sido un gran lugar al que volver con ilusión para seguir trabajando y aprendiendo. Gracias por haberme acogido y acompañado durante este tiempo. Gracias a todos los IPs, sois un gran ejemplo de constancia y trabajo. A la **Dra. Mireia Hernández,** gracias por tu ayuda en el proyecto de SMA y por tu dedicación a la investigación en neuropsicología. Gracias a la **Dra. Estela Cámara** por tus sabios consejos y tus visitas a la 1.7. Gracias a la **Dra. Ruth de Diego**, al **Dr. Lluís Fuentemilla** y al **Dr. Josep Marco** por vuestra amabilidad y simpatía. A **David Cucurell**, gràcies per tota la teva ajuda aquests anys. Per totes les converses sobre ciència, música, cinema, sèries i per les estones amb Radio 3 al zulo, gràcies QQ per donar-me suport quan ho necessitava. Gràcies **Joan** per tota la teva ajuda, carinyo, estones parlant d'òpera i pel teu riure contagiós capaç d'atravessar parets!

A todos mis compañeros y compañeras de la 1.7, presentes y pasados, gracias por todo! **Neus, Gemma, Jenny, Patri, Guillem, Emma, Clem, Lucía, Clara, Marc, Xim, Marta**, gracias por compartir esta experiencia conmigo, he aprendido muchísimo de todos vosotros. Gracias por vuestro apoyo, risas, abrazos y cafés a media tarde. A todos los Brainvitgeros con los que me he cruzado por el lab, hemos compartido momentos y preocupaciones: **Claudia, Tarrida, Berta, Marta Silva, Myriam, Brian, Alberto, Audrey, Jeison, Francesco, Helena T., Paula, Ludovico, Xiongbo, Gemma F., Marc D., Marc S., Lucia Vaquero, Helena A., Ane, Ernest,** gracias!

Durante estos años he tenido la suerte de poder colaborar con otros investigadores que me han aportado mucho personalmente, así como a la redacción de esta tesis. Gracias a todo el equipo de Mapping del Departamento de Neurocirugía y Neuroradiología del Hospital de Bellvitge, especialmente al **Dr. Andreu Gabarros** por dejarme colaborar en sus proyectos. Gràcies Andreu per la teva energia i disposició a fer coses noves i assolir nous reptes. Gracias al **Dr. Alejandro**

Fernández Coello por tu ayuda, profesionalidad y amabilidad. A la **Dra. Àngels Camins,** gràcies per la teva ajuda amb les resos i bona disposició. Gracias al **Dr. Pablo Naval** por tu ayuda, energía y ganas. A la **Dra. Imma Rico,** gràcies pel teu suport amb la neuropsico, informes i mappings. Gracias a la **Dra. Ruth Lau** por tu ayuda y colaboración en el proyecto de SMA. Gràcies també al **Dr. Pere Cardona** de la Unitat d'Ictus per la teva disposició a fer recerca i per la teva dedicació. Gràcies a la **Dra. Montserrat Juncadella** per donar-me l'oportunitat d'aprendre tant i formar-me com a Neuropsicòloga durant els dos anys de Màster.

A la **Dra. Joanna Sierpowska**, gracias por dejarme colaborar en el proyecto de 'S*witching*'. Gracias Asia por compartir tus guidelines de fMRI y mapping conmigo y por tu dedicación a la investigación con pacientes.

A la **Dra. Laura Ferreri**, gracias por dejarme formar parte del proyecto de *Dopamina* en el Hospital de Sant Pau ('*Pastillas de Sant Pau'*). Nunca olvidare los madrugones, efectos adversos de psicofármacos e interminables checklists. Gracias Laura por tu cariño durante estos años.

Gracias a la **Dra. Ana Tajadura** por darme la oportunidad de trabajar con sus *Magic shoes* en el Hospital de l'Esperança (AKA proyecto *Espardenya*). Gracias Ana por dejarme aportar mi granito de arena a tu proyecto, por tu profesionalidad y dedicación. Gràcies a la **Dra. Esther Duarte** per la teva col·laboració i amabilitat.

Manel, gràcies per compartir amb nosaltres els teus invents, no haguéssim pogut tirar endavant els últims experiments sense el teu coneixement.

Al **Dr. Joan Llobera** i a **David Pérez**, gràcies per la vostra ajuda i col·laboració en el projecte d'*Anosognosia i realitat virtual.*

Quiero dar las gracias a todos los participantes anónimos que aceptasteis pasar unas horas de vuestra vida dentro de una cámara de Faraday. Vuestra predisposición y colaboración es indispensable para seguir avanzando.

A mi familia. Gracias por apoyarme siempre, por dejarme hacer y deshacer, y por tener la certeza de que siempre puedo contar con vosotros.

A ti, vida, por estar siempre a mi lado y compartir tu vida conmigo.

"Y llegó el fin, aunque no acabemos."

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

1PP	First person perspective
ACC	Anterior cingulate cortex
AG	Angular gyrus
AHP	Anosognosia for hemiplegia
aIC	Anterior insular cortex
AlHS	Alien hand syndrome
ANHS	Anarchic hand syndrome
A-O	Action-outcome
aPE	Agency prediction error
BCI	Brain computer interface
Bil.	Bilateral
CC	Corpus callosum
DLPFC	Dorsolateral prefrontal cortex
EBA	Extraestriate body area
EE	External error
EEG	Electroencephalography
EMG	Electromyography
ERN	Error-related negativity
ERP	Event-related brain potentials
ErrPs	Error potentials
ESM	Electrical stimulation mapping
FCRP	Feedback correct-related negativity
fMRI	Functional magnetic resonance imaging
FoA	Feeling of agency
FRN	Feedback-related negativity
FRS	First-Rank Symptoms
GMFP	Global mean field power
IFG	Inferior frontal gyrus
IFS	Inferior frontal sulcus
IOG	Inferior occipital gyrus
IPL	Inferior parietal lobe
IPS	Intraparietal sulcus
JoA	Judgement of agency
L	Left
LC	Locus ceruleus

LRP	Lateralized readiness potential
MFC	Middle frontal cortex
MOG	Middle occipital gyrus.
mPFC	Medial prefrontal cortex
MRI	Magnetic resonance imaging
MTG	Middle temporal gyrus
NE	Norepinephrine
OFC	Orbital frontal cortex
PCC	Posterior cingulate cortex
PCG	Precentral gyrus
PES	Post-error slowing
PET	Positron emission tomography
PFC	Prefrontal cortex
PI	Posterior insula
PMC	Premotor cortex
PoCG	Postcentral gyrus
PPC	Posterior parietal cortex
Pre-SMA	Pre-supplementary motor area
R	Right
RHI	Rubber hand illusion
RP	Readiness potential
RT	Reaction time
SEM	Standard error of the mean
SE	Self-error
SMA	Supplementary motor area
SMG	Supramarginal gyrus
SoA	Sense of agency
SoO	Sense of ownership
SPL	Superior parietal lobe
STG	Superior temporal gyrus
STS	Superior temporal sulcus
SVM	Support vector machine
TMS	Transcranial magnetic stimulation
TPJ	Temporo-parietal junction
VR	Virtual reality

ABSTRACT

How do we recognize that "I" am the one acting upon the world? This experience of knowing oneself as the author of our actions and their consequences, the sense of agency, is a basic feature of our subjective experience. The evaluation of action outcomes is crucial for self-regulation, enabling us to adapt our behaviour based on the attribution of our actions and errors to internal or external causes. From the late 1990's to date, the interest on the self from the neuroscientific perspective has gained increasing attention due to its relevance to a variety of neurological conditions and psychiatric disorders.

In the present dissertation, I present a series of empirical work exploring agency attribution mechanisms. To this end, we implemented several modified versions of well-known experimental paradigms using electroencephalography, classification algorithms, functional magnetic resonance imaging and electrical stimulation brain mapping with the aim of unravelling the neural basis for internal and external agency attribution judgements. The first study describes neural signatures in terms of event-related brain potentials and time frequency decomposition. We report neurophysiological patterns associated with different agency prediction errors, introducing content incongruence and temporal delay, associated to an increase in external agency attributions and reduced internal agency judgments, respectively. In study 2, a support vector machine classification algorithm was implemented to decode, on a single trial basis, agency-error conditions based on the electroencephalographic time series data previously acquired in Experiment 1 from Study 1. We demonstrate the feasibility of using single-trial data to correctly decipher the causal agentic nature of correct and erroneous actions, providing new potential implementations, adaptations, and corrective mechanisms during brain-computer interactions. Lastly, the third study was aimed at examining the role of the anterior insular cortex in self-monitoring abilities in a patient undergoing awake brain surgery for tumour removal involving left insular regions. To do so, we employed electrical stimulation brain mapping at the anterior insula as well as functional magnetic resonance imaging while undergoing a modified Stroop task providing incongruent performance visual feedback. We show a dysfunction in monitoring self-produced outcomes when momentarily disrupting the anterior insular cortex, as well as specific patterns of activity involving this same region during the presentation of incongruent feedback inside the scanner.

Understanding how the brain processes the experience of agency is a very challenging and fascinating question. The current doctoral dissertation provides an important step towards comprehending how the subjective experience of being the agent comes about. Likewise, I believe that the results presented in this thesis will be useful to gain more insight on how the integration of predictive and retrospective information occurs for the elicitation of a coherent sense of self.

RESUMEN

¿Cómo reconocemos que "yo" soy el que actúa sobre el mundo? La experiencia de saberse a uno mismo como el autor de nuestras acciones y sus consecuencias, el sentido de agencia, es una característica fundamental de nuestra experiencia subjetiva. La evaluación de los resultados de las acciones es crucial para la autorregulación, ya que nos permite adaptar nuestro comportamiento en función de la atribución causal de nuestras acciones y errores a factores internos o externos. Desde finales de la década de los 90 hasta la fecha, el interés por el sentido del yo desde la perspectiva neurocientífica ha ido ganado atención debido a su relevancia en una variedad de condiciones neurológicas y trastornos psiquiátricos.

En la presente disertación, presento una serie de trabajos empíricos que exploran los mecanismos de atribución de la agencia. Con este fin, hemos implementado versiones modificadas de paradigmas experimentales bien conocidos utilizando electroencefalografía, algoritmos de clasificación, resonancia magnética funcional y mapeo cerebral mediante estimulación eléctrica con el objetivo de desentrañar las bases neurales de los juicios de atribución interna o externa de la agencia. En el primer estudio describo los correlatos neurales en términos de potenciales evocados relacionados con eventos y descomposición tiempo-frecuencia. Mostramos patrones neurofisiológicos asociados con diferentes tipos de errores de predicción de agencia, introduciendo incongruencias de contenido o retrasos temporales, directamente relacionados con un aumento en la atribución externa de agencia y una disminución de la agencia interna, respectivamente. En el estudio 2, implementamos un algoritmo de clasificación de Máquinas de vectores de soporte para descodificar, a nivel de cada ensayo, las diferentes condiciones de agencia del error en función de los datos de electroencefalografía en la dimensión temporal, previamente adquiridos en el Experimento 1 del Estudio 1. Con ello, conseguimos demostrar la viabilidad de usar datos de electroencefalografía a nivel de ensayo único para descifrar correctamente la naturaleza causal de las acciones correctas y erróneas, proporcionando nuevas potenciales adaptaciones y mecanismos de corrección en las interfaces cerebro-máquina. Por último, el tercer estudio tuvo como objetivo examinar el papel de la corteza insular anterior en las habilidades de monitorización en una paciente que se sometió a cirugía cerebral despierta para la extirpación de un tumour cerebral adyacente a la ínsula izquierda. Para ello, empleamos mapeo cerebral con estimulación eléctrica en la corteza insular anterior, así como resonancia magnética funcional durante la realización de una tarea Stroop modificada que proporcionaba 'feedback' visual incongruente. Demostramos una disfunción en la capacidad de monitorizar sus acciones al introducir una disrupción momentánea en la funcionalidad de la ínsula anterior, así como patrones específicos de actividad cerebral que involucran a esta misma región durante la presentación de 'feedback' incongruente.

Comprender cómo el cerebro procesa la experiencia de agencia es una pregunta compleja y a la vez fascinante. La presente tesis doctoral proporciona un paso importante hacia el entendimiento de cómo la experiencia subjetiva de ser el agente se origina, así como de la integración de información predictiva y retrospectiva para el desarrollo coherente del sentido del yo.

INTRODUCTION

1. INTRODUCTION

1.1. The self in action

How do I recognize my mental and bodily states as mine? Always a central topic in philosophy, human self-consciousness and its bodily foundations has recently become an increasingly outstanding topic in cognitive neuroscience. Wittgenstein (1958/2010) once said that no one could have an experience and wonder whose experience it was. An experience I feel has to be MY experience, and it is conceptually impossible to think differently. However, after a brain lesion or dysfunction, the conceptually impossible becomes possible for certain patients. Therefore, alongside its conceptual importance in the philosophical discourse, impairments in volition have prompted the scientific investigation of the psychological and neurobiological processes underlying the self.

1.1.1. Defining the self: What is the self and why does it matter?

Self-consciousness can be defined as the ability to become aware of one's own states, especially (but not only), mental states (i.e., perceptions, emotions, and attitudes), as one's own (Newen and Vogeley, 2003). Since the famous statement pronounced by René Descartes "*Cogito ergo sum*" (translated as "*I think, therefore, I am"*), many philosophers have attempted to elucidate what he described as a self. Descartes defined the self as a mental substance (i.e., *res cogitans*), with no extension in the world and not supported by causal laws, in continuous interaction with the body, a material substance (i.e., *res extensa*). In contrast, other philosophers such as David Hume (1739/1896) claimed that there is no self as a mental entity, but a set of complex perceptions of previously associated events that reflect the world in a unitary and causal format. In this view, the self is merely an illusion. At the other end, Immanuel Kant (1787/1929) argued that an adequate account of phenomenal consciousness required a richer structure of mental and intentional organization, highlighting that the conscious experience could not be a mere succession of associated ideas, but at a minimum, had to be the experience of a conscious self situated in an objective world structured with respect to space, time and causality. More contemporary philosophers such as Thomas Metzinger argues that there is no self as a mental entity (Metzinger, 2003), but only a dynamic, ongoing process creating very specific representational and functional properties.

Achieving an empirical understanding of the self also has a long history in psychology, dating back to William James' distinction between the self as the knower (*'I'*) and the self as the object that is known (*'me'*) (James, 1890). As the knower, the self is the subject who thinks, feels and acts. As

object, the self consists of the knowledge that we have about ourselves, encouraging us to reflect upon our actions and their consequences. This experience of the self as the object of attention is the cognitive state known nowadays as **self-awareness**, a key element for conscious self-monitoring or metacognition, by which we monitor and control our behaviour adjusting our beliefs and knowledge of the world.

Self-awareness is manifested to various degrees across the animal kingdom. For example, most animals demonstrate a tendency towards self-preservation (characteristic #1 of **Table 1**), while other species are also able to manipulate their environments to their own benefit (characteristic #4 of **Table 1**). A commonly employed test for self-awareness is the "mirror test" (Gallup, 1970; Gallup, Anderson & Shillito, 2002), during which a mirror is presented in front of the subject while observing its behaviour. Most animals initially display social behaviour (i.e., smiling, gesturing, threatening, bobbing, etc.), indicating that they perceive the image in the mirror as another animal. On the contrary, other animals such as chimps, dolphins, elephants and also humans >18 months, stop displaying social behaviour after several expositions displaying a more personal, self-directed behaviour, such as grooming, picking teeth, picking nose, etc., indicating that they identify the mirror image as themselves. Significantly, this continuum of self-awareness correlates well with animals' brain complexity, with more complex brains exhibiting greater degrees of self-awareness (Monat, 2017). Furthermore, it has been proposed that the evolution of self-awareness capabilities might be subjected to the **reafference principle**, understood as the notion that self-initiated actions evoke sensory effects that are correlated with these actions and, therefore, can be predicted and used by the system (Jékely, Godfrey-Smith & Keijzer, 2021). In this way, reafference provides a means by which organisms can evaluate their predictions and modify the activity involved. In animals with nervous systems, these predictive mechanisms, known as corollary discharge or efference copy (Sperry, 1950; von Holst & Mittelstaedt, 1950), might allow these species to compensate for predicted sensory changes by registering the particular action underway at a time. In animals without nervous systems, sponges for example, this reafference might involve less sophisticated mechanisms such as sensory cilia keeping track of the flow produced within the body signaling when this flow ceases (Ludeman et al., 2014). The evolution of these corollary discharge mechanisms in response to an expansion of behavioural repertoires or demands of control in animals, has been pointed out as a key factor for the evolution of active neuronal markings of the self vs. other distinction (Jékely, Godfrey-Smith & Keijzer, 2021). This 'proto-self' structure (Damasio, 1999) provides a platform for further stages in evolution, including the evolution of more complex neural systems and more complex and explicit forms of self.

Table 1: Operational characteristics of self-awareness. *Adapted from Monat (2017).*

	Characteristic	Manifested by which animal
1	The entity has a tendency toward **self-preservation**.	Most
2	The entity **recognizes that the world** around it can impact it.	Many
3	The entity recognizes that it has the **ability to observe and sense** the outside world through the senses.	Many
4	The entity recognizes that it has **the ability to impact the world** around it through conscious thought and action; and hence that it can modify the world around it to positively (or negatively) impact itself.	Some; all who use tools, build nests, trap food, and otherwise manipulate the environment
5	The entity identifies itself as **an entity**, complete, but separate and apart from the world around it.	Only those who pass the mirror test
6	The entity identifies the **physical boundaries** separating it from the world—that is, it knows where it stops and the world begins.	Only those who pass the mirror test
7	The entity identifies the **temporal boundaries** separating it from the world–that is, it knows that it was born and that it will die.	Humans only
8	The entity understands and identifies **its own constituents**, both visible and non-visible.	Humans only
9	A self-referential pattern exists such that the entity has and maintains **a model of itself within itself**.	Humans only

Self-awareness was until quite recently considered off-limits for empirical sciences. More specifically, neuroscientific research has avoided the "hard question of how and why consciousness and self-awareness arise from a physical basis", considering it inaccessible with current technology (Chalmers, 1996; Koch, 2012). The cognitive neuroscience approach to self-awareness views it as an operationalizable construct that can be broken down into paradigms amenable to neuroscience techniques (Christoff et al., 2011), focusing on different features of the self such as autobiographical memory, experiencing the self as a unit, labeling of stimuli as self-referential, feelings of agency or the experience of body ownership (see **Figure 1**).

1.1.2. Components of the self

Although our sense of self feels to us as being one solid entity, the phenomenology of the self involves many facets (see **Figure 1**). Recent work on the sensory-motor basis of the self distinguishes two key aspects: (i) the **sense of agency (SoA),** which enables us to recognize that we are the authors of our voluntary actions, and of their consequences, and (ii) the **sense of body ownership (SoO),** by which we recognize our body as our own and as the basis of our experience with the world (Gallagher, 2000).

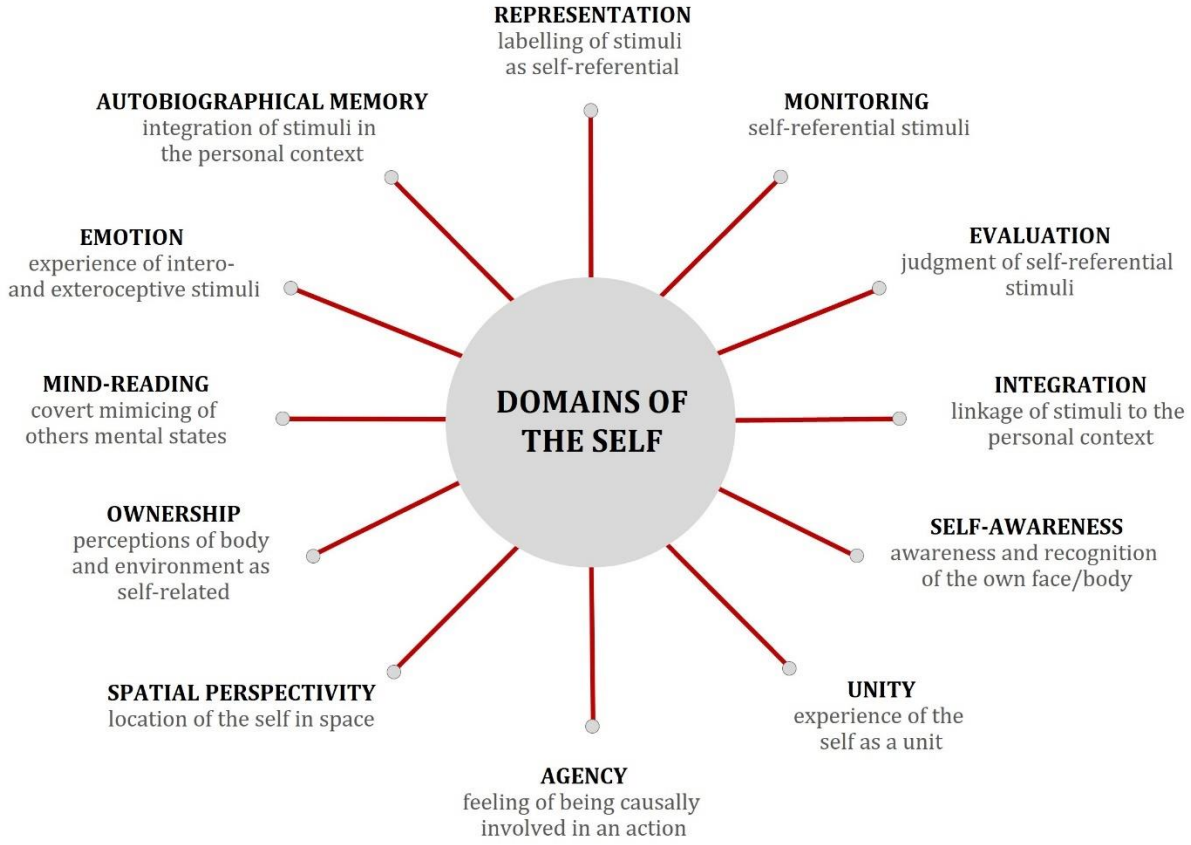


Figure 1. Domains of the self. *Adapted from Northoff and Bermpohl (2004).*

1.1.2.1. Sense of agency (SoA)

Whenever we perform a voluntary action, we tend not to feel as if it is simply happing to us but instead, we feel we are responsible for it. The SoA refers to this feeling/experience of being in the driving seat when it comes down to our actions, (e.g., "*I am the cause or author of the movement*").

As such, the SoA allows us to distinguish actions that are self-generated from those generated by others enabling self vs. other distinctions in the domain of action, contributing to the subjective phenomenon of self-consciousness (Gallagher, 2000; Georgieff & Jeannerod, 1998; Jeannerod & Pacherie, 2004, see ***Box 1***. *Developing a sense of agency* for a developmental perspective on the SoA). A major feature of the SoA is that is phenomenologically thin (Haggard, 2005), experienced online in a very fluid and unconscious manner which does not necessarily require the presence of a conscious reflection unless under metacognitive interrogation (Yeung & Summerfield, 2012).

Recently, an important distinction has been made between two different levels of the SoA (Synofzik, Vosgerau & Newen, 2008a) namely, the ***feeling*** **of agency (FoA)** and the ***judgment*** **of agency (JoA).** These two levels of agency processing, although related, have very distinct phenomenological characteristics and levels of processing. While the FoA is built upon the processing of internal cues, the JoA is more related to a higher-order, postdictive belief-based mechanisms influenced both by internal and external cues. (see subsection ***1.2.3. Multifactorial two-step account*** for more details on the FoA vs. JoA distinction).

1.1.2.2. Sense of Ownership (SoO)

As previously mentioned, the SoO refers to the experience or feeling of the body as one's own (e.g., "*This is my body*") (Tsakiris, 2010). While the SoA should only be elicited mostly during voluntary action, the SoO is typically continuous and omnipresent; we experience body-ownership not only during voluntary actions, but also during passive/involuntary movement (Van den Bos & Jeannerod, 2002). Over the last two decades, several experimental paradigms have been designed allowing the systematic manipulation of the SoO. The most well-known or predominant paradigm is the **rubber hand illusion (RHI***)*, developed by Botvinick and Cohen in 1998. In this experimental paradigm, an artificial rubber hand is placed at a visible and anatomically plausible position while the participants' real hand is hidden from his/her view. The experimenter then repeatedly strokes both the artificial hand and the real hand in a synchronous manner (both spatially and temporally). In most cases, this visuo-tactile synchronicity induces an illusory SoO over the artificial hand. The SoO over the artificial is typically measured both objectively, by means of the proprioceptive drift (i.e., the perceived location of one's own hand toward the rubber hand) and subjectively (i.e., the phenomenological experience of owning the rubber hand). Also, applying a potentially painful manipulation to the artificial hand (i.e., knife threat to the artificial hand by a fast and unpredictable move) can elicit a strong physiological fear response, which has been interpreted as implicit evidence for a successful embodiment of the artificial hand (Armel & Ramachandran, 2003; Braun et al., 2016; Ehrsson et al., 2007). Botvinicks' explanation for the RHI was that it reflected a strong multisensory integration between the efferent information provided by the stimulation (i.e., touch,

proprioception and vision) surpassing the pre-existing somatosensory representation of the subjects' own hand and vision of the fake hand, resulting in an induced unitary multisensory perception of the fake hand as being one's own (Armel & Ramachandran, 2003; Botvinick, 2004). Therefore, if the SoO was driven purely by bottom-up multisensory integration, we could expect that objects that do not resemble body-parts could induce SoO. Nevertheless, accumulating evidence suggests that the RHI is not induced when the artificial hand is replaced by a neutral noncorporeal object (Graziano, Cooke & Taylor, 2000; Haans, Ijsselsteijn & de Kort, 2008; Holmes, Snijders & Spence, 2006; Tsakiris, Costantini & Haggard, 2008; Tsakiris & Haggard, 2005). Hence, it has been proposed that although correlated multisensory stimulation and spatial proximity are necessary conditions for the SoO to emerge, they are not sufficient, adding to the traditional bottom-up explanation a top-down influence of pre-existing internal models of the body supporting the construction of body ownership (Tsakiris, 2007; Tsakiris & Haggard, 2005).

1.1.2.3. Integration of SoA and SoO

In the last two subsections, although SoO and SoA have been presented in isolation, it is important to highlight that we usually experience both at the same time in our everyday activities. While both the SoA and SoO are necessary for self-recognition (van den Bos & Jeannerod, 2002), several experimental studies have addressed the nature of this relationship. For example, several studies applying the RHI during active vs. passive movements have shown that voluntary action, or at least some component of it (e.g., efferent motor signals or SoA), have a promoting effect onto our SoO experience (Braun et al., 2014; Dummer et al., 2009; Kalckert & Ehrsson, 2014). Recent studies have also suggested that the SoA can override SoO (Tsakiris et al., 2007), by reporting that agency over a moving hand image itself generates SoO toward the image even when there is asynchrony between them (Imaizumi & Asai, 2015). Additionally, the SoA and SoO can be dissociated not only conceptually but also in terms of spatiotemporal factors (Gallagher, 2000; Tsakiris, 2010). Since the spatial component of visual feedback plays a more crucial role than the temporal component does in eliciting the SoA (Farrer et al., 2008), a certain amount of delay of visual feedback could be acceptable for eliciting SoA (Asai & Tanno, 2007; Bays, Wolpert & Flanagan, 2005; Miyazaki & Hiraki, 2006). On the contrary, SoO requires temporal congruence between inter-sensory stimuli (Botvinick & Cohen, 1998; Tsakiris & Haggard, 2005).

A working hypothesis regarding the influence of SoA over SoO states that voluntary action is an important source of information for self-recognition (Synofzik at al., 2008a; Van Den Bos & Jeannerod, 2002), and that by actually moving the body in an active way, the brain can test its predictions about which sensory events reflect the own body and which ones do not (Synofzik et al., 2008a), sharpening one's own bodily boundaries and consequently inducing more vivid SoO

experiences. In relation to the influence of SoO on SoA, our actions always originate in our body (Wong, 2010), henceforth it is reasonable that our brain attributes higher certainty levels of authorship to our immediate body actions than to less anticipatable effects on the world.

Moreover, neuroimaging evidence indicate that distinct brain networks may be responsible for processing the SoA and SoO. For example, the pre-supplementary motor area (pre-SMA) is involved in generating the conscious experience of motor intentions (Haggard, 2005; Lau et al., 2004, 2007), and the posterior parietal cortex (PPC) and cerebellum may be key regions for implementing the predicted vs. actual sensory feedback comparisons required for SoA computations (Blakemore & Sirigu, 2003; Farrer & Frith, 2002; Farrer et al., 2003, 2008; Tsakiris et al., 2010). In contrast, the SoO has been associated with activation in the premotor cortex (PMC) (Ehrsson, Spence & Passingham, 2004, 2005; Petkova et al., 2011), the intraparietal cortex (Ehrsson et al., 2004, 2005; Petkova et al., 2011), and the putamen (Petkova et al., 2011), which are cortical nodes for integrating multisensory information (i.e., visual, tactile, and proprioceptive) from the body and the space surrounding the body (Duhamel, Colby & Goldberg, 1998; Fogassi et al., 1996; Graziano & Gross, 1993).

Interim summary

We usually take it for granted that our bodies are spatially extended in the world and that we are the agents responsible for our actions. The SoA, this phenomenal experience of initiating and controlling our actions and the SoO, the feeling of *mineness* towards our body, are crucial for experiencing a self independent of the external world. The distinction between SoA and SoO has attracted considerable attention in various fields including psychology, philosophy, and cognitive neuroscience (Blakemore, Wolpert & Frith, 2002; Marcel, 2003; Tsakiris & Haggard, 2005). There is growing evidence that the SoA is based on that which precedes action and transforms intentions into actions, while the SoO can be explained in terms of ecological self-awareness built into movement and perception (Gallagher, 2000). A better understanding of how these experiences relate to each other is of great importance in basic research supporting and promoting new advances in different fields of neurorehabilitation.

Box 1. Developing a sense of agency

Newborns seem to have very little or no control at all over their actions and very little or no SoA, developing full intentional actions quite gradually (Hauf, 2007). Classical theories of developmental psychology suggest that voluntary control of actions emerge in the first months of life, guided by the reinforcement of parental affective communication (Ruvolo, Messinger & Movellan, 2015).

Previous studies on the development of SoA have shown that infants as young as 2 months of age increase the frequency of an action when this action is followed by an effect (Gergely & Watson, 1999; Watanabe & Taga, 2006, 2009; Watson, 1972). For example, Rochat (1998) reported observations of 3-month-old infants actively discovering the contingency between their own leg kicks and the sounds of a rattle, concluding that by 2–3 months, infants start to develop a sense of their body as an acting agent in the environment (Rochat, 1998). These findings have been interpreted as evidence for the ability to detect causal relations between actions and their effects (Kelso, 2016; Rochat & Striano, 1999; Rochat & Striano, 2000; Watanabe & Taga, 2009), an essential building block for instrumental learning and operant action. From around 9 months, infants start to perceive adult's movements as intentional (Jovanovic & Schwarzer, 2007), developing joint attention. At the same time, they start to engage in goal-directed behaviour (Babik et al., 2019) and continue to develop gradually throughout the entire second year, learning to solve complex tasks and transfer their knowledge across various contexts. Around this time, they begin to pass the classic mirror self-recognition test (Bard et al., 2006). Importantly, at this age, linguistic and symbolic competencies enable infants to mark contrasts between themselves and other people in their verbal production. (Rochat, 2003).

Around the fifth year, they begin to be capable of holding multiple representations and perspectives on objects and people allowing awareness in relation to the self (Williams & Happé, 2010).

1.2. Synthesis of SoA theories

The systematization of scientific evidence within a conceptual and theoretical framework is of great usefulness, allowing us to focus on specific aspects and mechanisms of the studied phenomena. During the last decades, two main theoretical perspectives have been proposed describing the underlying neurocognitive mechanisms responsible for the SoA. On one hand, the *Comparator model* (Wolpert, Ghahramani & Jordan, 1995) suggests that the SoA is formed mainly from processes related to motor control and that it is inferred by means of a predictive signal/forward model. On the contrary, the *Apparent Mental Causation theory* proposed by Wegner (Wegner & Wheatley, 1999) proposes that the SoA arises whenever external events are congruent with our intentions. More recently, other models such as the *Two Step model* (Synofzik et al., 2008a) or the *Cue Integration theory* (Moore, Wegner & Haggard, 2009; Moore & Fletcher, 2012) have been proposed according to which both internal and external cues are integrated in order to build a coherent sense of self.

1.2.1. Motor control-based theories

A key factor that distinguishes self/active vs. external/passive generated actions is the existence of an intended motor command in the case of active movement. One of the most prominent accounts regarding the SoA is the ***Comparator model*** (Blakemore, Wolpert & Frith, 1998; Blakemore et al., 2002; Frith, Blakemore & Wolpert, 2000; Miall & Wolpert, 1996; Wolpert & Flanagan, 2001; Wolpert et al., 1995; Wolpert & Kawato, 1998), which was originally formulated as a model for motor learning and motor control (see **Figure 2**). According to this theory, our central nervous system contains two types of internal models which mimic aspects of one's own body and the external world (Wolpert et al., 1995), namely, the *inverse* and *forward* models. The inverse models or controllers are the ones in charge of selecting the appropriate motor commands necessary to achieve a desired goal or outcome. These inverse models allow rapid and fluid evaluations of ongoing motor commands, enabling us to switch controllers as the context changes or when detecting deviations from the intended actions (i.e., prediction errors). Every time that we issue a motor command, an internal copy of this motoric signal, termed efference copy or corollary discharge, is issued in parallel allowing the forward models or predictors to build an estimation of the sensory consequences of the movement (for further details on the efference copy see subsection ***1.4.2. Prediction***). By means of a comparator mechanism, the predicted sensory outcomes are constantly compared to the actual sensory outcome. In case of a match/congruence between the predicted and the actual states of our system, the action is experienced as coming

from the self, whereas whenever an incongruence is detected, the action is attributed to an external agent (**Figure 2**).

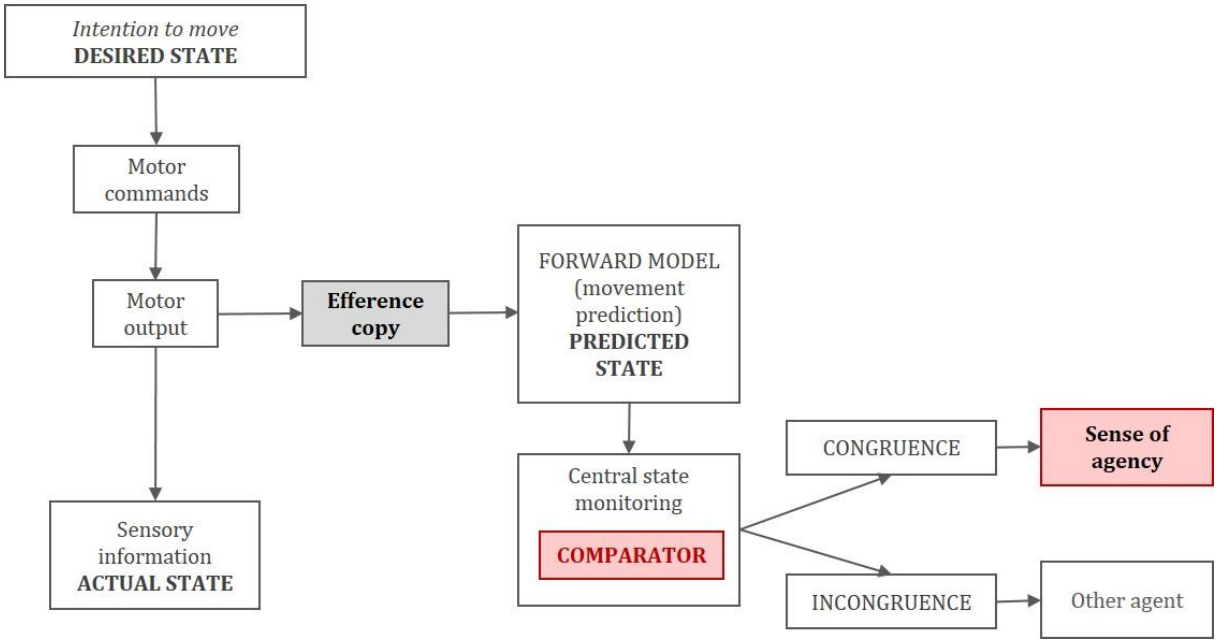


Figure 2. Comparator model. *Adapted from David, Newen and Vogeley (2008).*

The experimental paradigm most commonly employed to test the validity of the *Comparator account* consists in inserting either **spatial** (Blakemore et al., 1998; Ebert & Wegner, 2010; Farrer et al., 2003; Fourneret & Jeannerod, 1998; Hon, Poh & Soon, 2013; Kannape et al., 2010; Knoblich & Kircher, 2004; Nielsen, 1963; Tsakiris et al., 2005; Synofzik, Their & Linder, 2006), **temporal** (Ebert & Wegner, 2010; Farrer et al., 2008; Farrer, Valentin & Hupé, 2013; Hon et al., 2013; Kawabe, 2013; Kühn et al., 2011; MacDonald & Paus, 2003; Sato & Yasuda, 2005; Wen, Yamashita & Asama, 2015) or **content** (Gentsch, Ullsperger & Ullsperger, 2009; Mathalon et al., 2003; Padrao et al., 2016; Steinhauser & Kiesel., 2011) discrepancies between the action and the outcome (A-O) (see subsection ***1.5. Experimental manipulations of SoA*** for more details). For example, a commonly used spatial manipulation consists of providing the participant with distorted visual feedback of their hand moving a joystick or pressing a button (see **Table 4** for a description of experiments inducing A-O discrepancies). If the feedback provided by the virtual hand/computer screen does not match the participants' real movement, the participant tends to attribute the action to another external agent, showing a gradual reduction of the SoA as the temporal or spatial

discrepancies increase. Moreover, the *Comparator* framework has been extensively adopted to study several pathological conditions characterized by awareness abnormalities, such as schizophrenia, anosognosia for hemiplegia, alien/anarchic hand syndrome and utilization behaviour (Frith et al., 2000, see subsection ***1.3. Dysfunctions of the self*** for an explanation of these pathological conditions).

Although the *Comparator model* dominates the SoA research, several limitations have been recently pointed out (Synofzik, Vosgerau & Newen, 2008b; Synofzik, Vosgerau & Voss, 2013; Vosgerau & Synofzik, 2012). Mainly, it has been argued that the *Comparator* account might be useful to explain the low-level FoA, but that it cannot account for the conceptual attribution of agency/JoA, as it does not solely depend on sensorimotor cues but also on contextual cues, prior beliefs, and retrospective inferential mechanisms (Synofzik et al., 2008b). Moreover, the importance of the motor system in the generation of the SoA has also been questioned, for example, when Desmurget and colleagues (2009) induced illusory SoA when electrically stimulating the PPC in the absence of movement [see ***1.4.1.2. Anatomy of the intention to 'act': The supplementary motor area (SMA) and posterior parietal cortex (PPC)***].

1.2.2. High-level theories

In contrast to the computational low-level perspective, other theories approach the nature of the SoA from a higher-level cognitive perspective. Essentially, the question of agency attribution is a matter of inference and depends on several mental states such as prior beliefs, desires, and intentions. Therefore, while the *Comparator* account focuses mainly on the aspect of prediction and the importance of the motor system, these theories highlight the role of postdiction and inferential mechanisms. One of the most prominent high-level theories is the ***Apparent Mental Causation Theory or Inference Model*** (Wegner, 2003; Wegner & Wheatley, 1999). From this perspective, the SoA arises if three principles are met: (i) the thought or intention appears prior to the action (priority), (ii) is consistent with the action (consistency) and (iii) no other alternative causes can be present (exclusivity). Therefore, the congruence between the agents' intention and the outcome is sufficient for the emergence of a coherent SoA. It is the match between a prior conscious thought and the observed outcome of the action that makes the agent infer retrospectively that he/she was at the origin of the sensory event (see **Figure 3**). Early experimental support for the *Apparent Mental Causation Theory or Inference Model* comes from the "I spy" experiment by Wegner and Wheatley (1999). Participants and an experimental confederate jointly controlled a computer mouse that could be freely moved over a series of pictures shown on a computer screen. For each trial, the participants were asked to point with the

cursor at one of the images and then indicate his/her level of action contribution. When the participants had been primed before with the name of the chosen picture, they were more likely to attribute the action to themselves, even when the picture was chosen by the confederate and not by the participant, pointing out the key role of the inferential process in the adscription of agency.

As in the case of the *Comparator account,* this theory also has its limitations, mainly due to its focus in the postdictive inferential mechanisms (e.g., prior beliefs, expectations, etc.), which cannot give a satisfactory explanation for lower levels of FoA, as it does not depend on post-hoc inferential mechanisms (Synofzik et al., 2008a). Moreover, as Carruthers (2012) pointed out, the model of Wegner and colleagues cannot explain cases of young children, who displayed an intact SoA "despite not being able to infer that their mental states cause their action" (as cited by Carruthers, 2012, p. 342). This clashes with the model proposed by Wegner, which states that the ability to infer one's own mental states as the cause of one's actions is crucial for the SoA. Also, Desmurget et al. (2009) findings cannot be explained in terms of Wegner's model, as the basic postulation of this model (i.e., SoA is based on a causal inference about the relationship between thought and action) is not achieved in these patients.

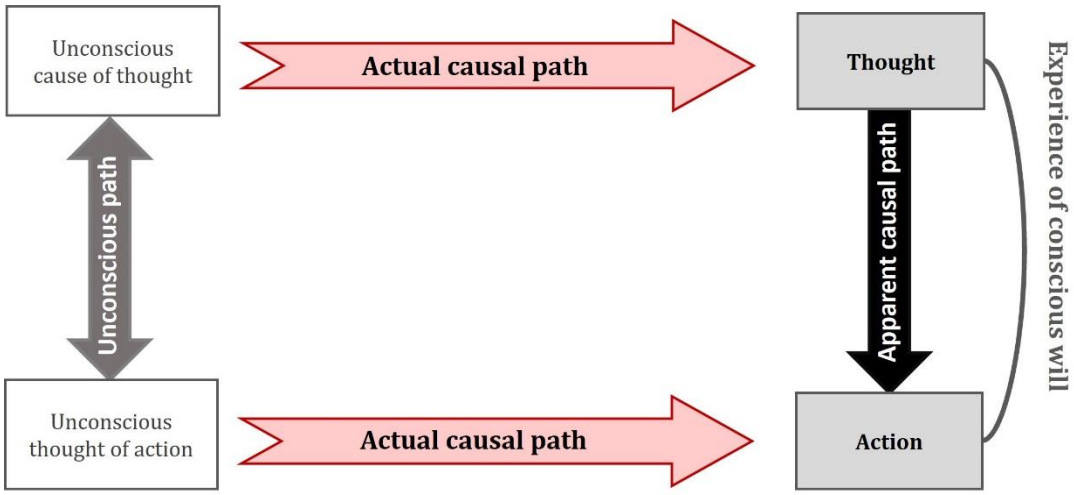


Figure 3. Apparent Mental Causation Theory or Inference Model. The experience of conscious will arises when the person infers the apparent causal path from the thought to action, based on unconscious mental events (thoughts). Therefore, the sense of being the agent is illusory, apparent, not causally involved in the generation of action. *Adapted from Wegner (2003).*

1.2.3. Multifactorial two-step account

Since both the predictive (i.e., internal motoric signals: *Comparator model*; Blakemore, Wolpert & Frith, 1999a; Wolpert & Ghahramani, 2000) and postdictive (i.e., external and situational cues: *High-level theories*, Wegner, 2003; Wegner & Wheatley, 1999) theories have major limitations and do not exhaustively depict the complexity of the SoA, Synofzik et al. (2008a) proposed a ***Multifactorial two-step account*** of the SoA involving both lower-level/internal motor processes and higher-level/external factors (see **Figure 4**). As previously mentioned, Synofzik and colleagues (2008a, 2008b) draw an important conceptual distinction in the SoA between the FoA, a "non-conceptual, low-level feeling of being the agent" of an action, that *what-it-is-like-for-me-ness* (Synofzik et al., 2008a, p.228) and the JoA, a "conceptual, interpretative judgment of being an agent" (Synofzik et al., 2008a, p.229), similarly to what Gallagher (2007) proposed as first-order experiences and second-order reflective attributions. While the FoA, which consists of the integration of multiple internal indicators -such as internal predictions, sensory feedback, and proprioception- provides us with a rather "diffuse sense of a coherent, harmonious ongoing flow of action processing" (Synofzik et al., 2008a, p.228), the JoA is foremost an explicit, reflective, conceptual, and inferential process. As a result, they can be dissociated from one another. For example, when an unexpected action outcome occurs, which would signal non-agency at the feeling level, it can nevertheless be attributed as coming from the self at the judgement level if our beliefs and contextual cues imply self-attribution (e.g., when an unexpected outcome occurs while being alone in a room).

Moreover, these authors proposed two levels of agency processing based on the weighting of different agency cues: (i) *predictive cues* (i.e., the forward model-based outcome prediction) that influence the FoA and throughout that the JoA, and (ii) *postdictive cues* (i.e., sensory feedback) which influence both our FoA and JoA and sometimes "retrospectively" change our beliefs about what our prior intentions for the action were (Haggard, 2008; Kühn & Brass, 2009). The question now is how all these different cues are processed and what is their respective contribution to the agentive experience.

In its initial proposal, the *Multifactorial two-step account* did not provide any criteria on how this cue integration process occurred to build and shape the SoA. To fill this gap, it has been recently proposed that this combination of the different cues depends on their relative reliability in a given situation as suggested in the *Bayesian Cue Integration Theory* (Moore et al., 2009; Moore & Fletcher, 2012; Synofzik et al., 2013). Following the *Multifactorial two-step account* explanatory gap on how agency cues are selected and integrated, the ***Bayesian Cue Integration Theory***

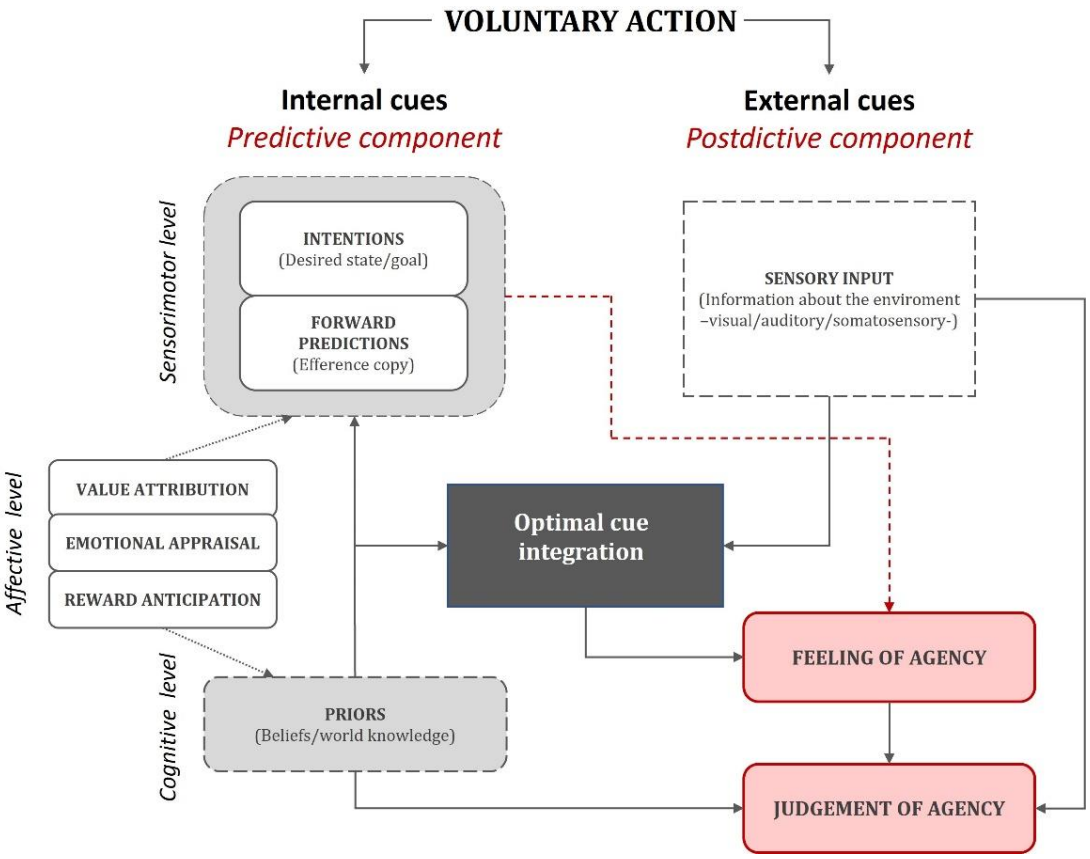


Figure 4. Multifactorial two-step account. The SoA arises from a complex interplay between predictive and postdictive components. At the sensorimotor level, the predictive component includes internal cues such as motor predictions, action selection, efference copy and motor output signals. These internal cues can directly give rise to the FoA depending on the context. On other occasions, internal cues/predictions are integrated with external cues (i.e., sensory input) resulting in a postdictive feeling of agency, based on an optimal integration. Moreover, this pre-reflective feeling of agency can lead to a more explicit, reflective judgement of agency, influenced by background information or beliefs. Furthermore, emotional appraisal, anticipation of reward or punishment or value attribution may influence the weighting of internal or external signals on both the sensorimotor and cognitive level. *Adapted from Synofzik et al. (2013).*

(Moore et al., 2009; Moore & Fletcher, 2012; Synofzik et al., 2013) proposed that the brain must integrate this information within a Bayesian framework (see ***Box 2.*** *The Bayesian brain*). We have access to many different agency cues (information channels), each one providing their own estimates about the agentic origin of an event. However, these agency estimates are highly noisy signals (i.e., have a high signal variance), so that for each agency cue there is an estimate uncertainty. Consequently, we must optimally integrate all relevant information coming from the different agency cues. These authors suggest that the brain applies a Maximum likelihood estimation to all its agency cues and thereby obtains an overall agency estimate, whose noise is

Box 2. The Bayesian brain

The *Predictive Processing Hypothesis* (Clark, 2013; Friston, 2010, 2012) states that we have internal statistical models that try to represent the causal structure of the world. Based on Bayesian inference, that is, the optimal and precision-weighted integration of information coming from different sources (e.g., expectation, vision, touch, etc.), we generate probabilistic internal models (i.e., forward models) that are then used to formulate predictions about the world. These predictions, which are then tested on sensory data, are then employed by our system to update our beliefs (i.e., empirical priors) about the causes of the sensations we experience (Friston, 2012). From this perspective, the cortex is hierarchically organized with lower levels representing low-level sensorimotor information (e.g., orientations, colors, etc.) and higher levels representing more and more abstract information (e.g., beliefs about abstract concepts, the world, myself, etc.). Importantly, the flow of information between the different levels is asymmetrical, with the top-down modulations carrying most information in the form of predictions while only *prediction errors* (thought to rely on dopamine) get sent as input to higher levels influencing our internal models (Rao & Ballard, 1999) (**Figure 5**).

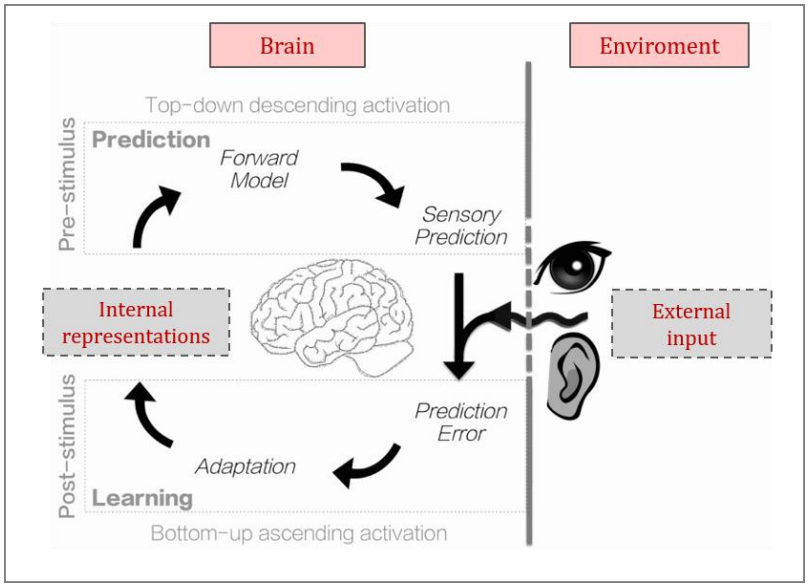


Figure 5. Graphic representation of the various stages involved in prediction according to the *Predictive Processing Hypothesis*. Predictions are sent top-down from higher to lower-level regions to 'explain away' the incoming sensory input (i.e., facilitate its processing). Then, the 'prediction error' – the portion of information that remains unexplained – is sent back and used to tune the internal model to improve future predictions. *Adapted from Molinaro, Monsalve and Lizarazu (2016)*.

much lower than the signal variances of any agency cue on its own. However, it should be pointed out the application of the Bayesian framework for explanations of the mind and brain is not without criticism, mainly because such models are hard to falsify (Bowers & Davis, 2012).

Interim summary

To summarize, during the last two decades different accounts have been proposed to explain the neurocognitive underpinnings of the agentic experience. While some accounts put a stronger emphasis on predictive processes as the central mechanism (e.g., the *Comparator* model by C. Frith and D. Wolpert), others emphasize more strongly the postdictive/retrospective inferences (e.g., *Inference Model* by D. Wegner). More recent theories propose an integration of both predictive and postdictive mechanisms through an optimal cue integration process, such as the *Multifactorial two-step account* of M. Synofzik and colleagues. The relationship and relative weight of all these aspects in the formation of a coherent SoA is still a matter of debate. In the following sections, I will provide further insight regarding some of these facets such as intentions, predictions, and action/state monitoring mechanisms.

1.3. Dysfunctions of the self

The understanding of pathological conditions characterized by a disrupted SoA such as schizophrenia, anosognosia for hemiplegia or alien/anarchic hand is not only of theoretical interest, but also has clinical implications of great importance. Self-awareness is key for monitoring our performance and adapting our behaviour to a changing environment. The presence of an aberrant or disrupted SoA represents a negative prognostic sign, compromising the course of recovery and rehabilitation in these patients (Gialanella et al., 2005; Pedersen et al., 1996). Likewise, the study of these pathological conditions can significantly contribute to our understanding of higher cognitive functions such as self-awareness and consciousness.

1.3.1. Schizophrenia

Schizophrenia is a psychiatric disorder characterized by prominent psychotic symptoms that include the false attribution of perceptual experience to an external source (hallucinations), distorted thinking (delusions), reduction in affect and behaviour (negative symptoms) and disorganization of thought and language (thought disorder) (Andreasen & Flaum, 1991; Meehl, 1962). In addition, patients with schizophrenia exhibit impairments in both basic sensory processing and higher cognitive functions, such as language, reasoning, and planning. Despite more than 100 years of research, the causes of schizophrenia are still unknown. However, it is becoming increasingly clear that the psychotic phenomena and cognitive dysfunctions that characterize this disorder are not due to circumscribed deficits in specific cortical regions but rather it might represent a distributed impairment involving many cortical areas and their connectivity.

A dysfunction in self-monitoring abilities has been used to explain First-Rank Symptoms (FRS) that are among the most distressing sensations encountered by patients suffering from schizophrenia. FRS are characterized by the feeling that actions as well as personal states are no longer under their own control (Schneider, 1959). The main FRS are auditory hallucinations, thought insertion, thought broadcasting, delusions of influence, and all the feelings that another is controlling the patient's thoughts, actions, or emotions (see **Table 2**). Such experiences are highly complex, but all of them seem to involve a disruption in the SoA processing. For example, in thought insertion, the individual believes that his/her thoughts don't belong to them but that they are actually those of another person. In delusions of motor control, an action is perceived to be controlled/initiated by another agent. There are other features also related to SoA disruption although in a less obvious manner, as for example auditory hallucinations, which clearly arise in the brain of the sufferer, perhaps as internal speech, but that are perceived as externally generated.

Table 2. First Rank Symptoms of Schizophrenia, initially described by K. Schneider (1959). *Extracted from Farrer and Frack (2007).*

Symptoms	Descriptions
Acoustic-verbal hallucinations	Voices heard arguing or commenting on patient's actions.
Audible thoughts	Patient's own thoughts heard by him/herself.
Thought broadcasting	Patient's thoughts are passively diffused to other people.
Thought insertion	Other people intrude their thoughts upon the patient.
Thought withdrawal	Other people actively take patient's thoughts in his/her mind.
Made affect and feelings	Experience of influences playing on patient's sensations.
Somatic passivity, delusions of influence, alien control	Experience of influences playing on patient's actions.
Delusional perception	Patient's experience of a peculiar, intense, convincing experience not shared by other people.

Several studies suggest that patients with **delusions of control** have problems with both motor and sensory aspects of the *forward/internal cues* (Blakemore et al., 2000; Franck et al., 2001). From the *Comparator perspective*, it has been stated that a failure to form a representation of the predicted consequences (i.e., efference copy) of an action would result in an impaired ability to distinguish between one's own and another's actions, resulting in SoA abnormalities and external agency attributions (Frith et al., 2000). Nevertheless, the absence of efference copy does not explain the full phenomenon of delusions of control since the anomalous experience may also feel alien (i.e., not belonging to oneself). One possible explanation for the alien feeling is that a disruption in the integration of somatosensory, visual and auditory signals with the efference copy may generate a sense of alien control at the level of first-order experience (De Vignemont & Fourneret, 2004; Gallagher, 2004; Pacherie, Green & Bayne, 2006), probably related to mechanisms involving self-other discrimination (Georgieff & Jeannerod, 1998) and associated to functional or structural abnormalities at the cortical level in the anterior insular cortex (aIC) or right PPC (Farrer & Frith, 2002) [see subsections ***1.4.1.2. Anatomy of the intention to 'act': The supplementary motor area (SMA) and posterior parietal cortex (PPC)*** and ***1.4.2.4. Reflecting on one's own performance: the anterior cingulate cortex (ACC), the premotor cortex (PMC) and the insula***].

In parallel, it has been proposed that an increased noise involving *internal/predictive cues*, perhaps arising from upregulated or chaotic dopamine firing, may lead to changes in the signal-to-noise ratio and a corresponding attenuation of the degree to which internal signals are able to contribute to the SoA (Fiorillo, Tobler & Schultz, 2003). If this is the case, it would also lead to an increased likelihood of mismatch between the predicted and the actual sensory consequence of a movement (i.e., prediction errors), which have been related to psychotic symptoms (Blakemore et al., 2002; Frith, 2005; Schultz & Dickinson, 2000; Shergill et al., 2005). Moreover, the idea that prediction errors are a critical updating signal has been used to explain abnormal updating of beliefs in schizophrenic patients (Corlett, Honey & Fletcher, 2007; Fletcher & Frith, 2009).

On the other hand, other models of schizophrenia have highlighted an excessive reliance on *external/environmental cues* to account for delusions of control and thought insertion. For example, Synofzik et al. (2010) reported that patients with schizophrenia tend to rely more on external cues such as visual feedback than in internal sensorimotor cues when doing agency attributions. This view, which has been linked to dopaminergic dysregulation, proposes that an alteration in the perceived salience of external events may alter the weightings applied to the *Cue integration* process, leading to an alteration in causal agentic attributions (Kapur, 2003; Palaniyappan & Liddle, 2012). For example, if a noise (e.g., car horn) is coincident with one's movement (e.g., hand movement), then the enhanced, inappropriate association between one's movement and the noise might lead to the experience that one had caused the noise or the sensation that the movement was caused by the noise, which could ultimately lead to the delusion of control.

Finally, *prior expectations or beliefs* have also been related to the appearance of illusions of control (Corlett, Frith & Fletcher, 2009; Fletcher & Frith, 2009). Patients suffering from delusions of control might have the tendency to use prior information in distinct, sub-optimal ways (Moritz & Woodward, 2006), forming aberrant or altered priors (i.e., beliefs), such as the belief that one is under external control. Once the belief that one is under external control has been formed, particularly in the setting of noisy internal cueing, this belief generally shapes the interpretation of one's experiences.

1.3.2. Anosognosia for hemiplegia (AHP)

The term anosognosia (a-noso-gnosia Greek for 'non illness knowledge') refers to a neurological symptom characterized by the unawareness or lack of concern over one's own deficits, such as paralysis or aphasia. One of its main common presentations is **anosognosia for hemiplegia (AHP),** a condition where patients fail to recognize a severe motor deficit despite direct

confrontation during neurological examination (Babinski, 1914; Langer & Levine, 2014). It is thought to be relatively common, encountered in at least 20-30% of hemiplegics after an acute stroke, typically involving the right hemisphere (Jehkonen et al., 2006). It can last from days to weeks at acute and subacute stages, and its clinical manifestation may vary in the following ways: while some patients admit their deficits but minimize their practical or emotional importance, others do not acknowledge any disability despite blatant evidence to the contrary. Other patients may acknowledge their deficits in a general manner while failing to detect specific motor inabilities and vice versa. AHP can also be accompanied by delusional beliefs (Barraquer Bordas, 1974), such as:

- *Misoplegia*: the patient seems to despise their own limb.
- *Hyperschematia*: the affected limb seems to be heavier and larger.
- *Asomatognosia*: patients reject the ownership of their limb completely.
- *Somatoparaphrenia*: attribute its ownership to someone else.
- *Personification*: talk about it in 3rd person.

Interestingly, some of these patients display an implicit awareness of their deficits in verbal (Fotopoulou et al., 2010) or behavioural tasks (Cocchini et al., 2010; Moro et al., 2011), and they may become more aware of their motor failures when taking a third-person perspective compared to a first-person perspective (1PP) (Fotopoulou et al., 2009; Marcel et al., 2004).

Initially, AHP was explained as being a secondary consequence of sensorimotor and cognitive impairments [i.e., primary sensorimotor deficits, generalized cognitive impairment, or neglect (Jehkonen et al., 2006; Orfei et al., 2007)]. Yet, dissociations have been reported (Bisiach et al., 1986; Marcel et al., 2004), suggesting that they are not a necessary condition for AHP to occur. More recently, the *Comparator* account has been proposed as an alternative explanation, suggesting that AHP patients, although being able to correctly form the appropriate motor intentions and predict the expected sensory consequences of their actions, they are unable to detect the discrepancies between the intended/predicted and the actual sensory feedback because of a damaged comparator mechanism (Berti et al., 2005; Frith et al., 2000). Empirical findings support the comparator explanation by showing that brain areas related to A-O monitoring are selectively damaged in AHP patients, such as the aIC, right PMC or the inferior frontal gyrus (IFG) (Berti et al., 2005; Fotopoulou et al., 2010; Kortte et al., 2015), corresponding to brain regions involved in motor initiation, preparation, and monitoring [see subsection ***1.4.2.4. Reflecting on one's own performance: the anterior cingulate cortex (ACC), the premotor cortex (PMC) and the insula***].

Nevertheless, although the *Comparator* explanation is effective in explaining the illusion of moving, it cannot account for other beliefs and attitudes accompanying this syndrome (i.e., their adherence to the delusional belief that they have functional limbs even when confronted with their disability). Here, the *Predictive Processing Hypothesis* (see ***Box 2**. The Bayesian brain*) seem to become useful, by proposing that AHP patients present aberrant perceptual inferences (suboptimal synaptic activity; Friston, 2010) that lead to a disrupted SoA due to their non-updated predictions about their motor abilities (Fotopoulou, 2012).

Moreover, recent lesion-mapping studies have highlighted other areas related to bodily salience and interoceptive awareness (Craig, 2003; Critchley et al., 2004), such as the insula, limbic structures, and subcortical white matter connections selectively damaged in AHP (Fotopoulou et al., 2010; Karnath, Baier & Nägele, 2005; Moro et al., 2011; Vocat et al., 2010). Therefore, it might be the case that weak/imprecise interoceptive signals about the current state of the body may provoke the adherence to past predictions of how the affected body parts should feel producing aberrant beliefs. Importantly, during the first days of acute illness, damage to the posterior insula is predictive for developing anosognosia and also patients with persistent awareness deficits usually present frontal and parietal damage (Vocat et al., 2010).

1.3.3. Alien/anarchic hand syndrome

Another interesting neurological syndrome is the **alien/anarchic hand syndrome**, where patients report the stressing experience of one of their arms or hands acting without the patient's conscious will [for a review, see Fisher (2000)]. The first description of alien/anarchic hand syndrome was made by Goldstein in 1908. It was the case was of a 57-year-old female who felt that her left hand had a will of its own and grabbed her throat, taking her a great effort to restrain her affected hand. Interestingly, the actions that the affected arm performs, although not intended, are though purposeful in themselves (e.g., open a door, take off clothes), and are typically completed successfully.

Noteworthy, there has been some taxonomical confusion about the terms 'alien' and 'anarchic'. Thus, Della Sala and colleagues (Marchetti & Della Sala, 1998) proposed restricting the term "anarchic" to denote those conditions where subjects perform involuntary movements with their hand but acknowledge the ownership and "alien hand" to conditions involving the lack of SoO over the hand.

- Anarchic hand (AnHS): Usually affects the dominant hand. Characterized by impulsive grasping, compulsive manipulation of objects and difficulty on releasing them. The patient is aware that the arm belongs to them but is unable to voluntarily suppress movements (they may try to restrain movements by holding it or sitting on it). It is usually associated to left medial frontal lesions (medial prefrontal cortex -mPFC-, supplementary motor area -SMA-) with or without corpus callosum -CC- damage. It has been proposed that these patients lack the ability to voluntarily suppress/inhibit the actions primed by the perceptual processing of objects due to a disruption in inhibitory mechanisms -SMA- over the cortical regions controlling the movement of the hand -primary motor regions-, causing utilization behaviour such as impulsive grasping and compulsive manipulation of objects (Biran et al., 2006; Frith, 2000; Giovannetti et al., 2005; Schaefer, Heinze & Galazky, 2010). Nevertheless, the areas that underlie the desired, predicted, and actual states of the body are considered intact in these patients, hence they know their arm executes actions that differ from their own conscious intentions.

- Alien hand (AlHS): Affects the non-dominant hand. Characterized by intermanual conflict (i.e., one hand is acting at cross purposes to the other) with minimal limb weakness and absence of frontal features (i.e., impulsive grasping, non-fluent speech, etc.). Other signs accompanying the AlHS are apraxia, tactile anomia, visual anomia, agraphia, neglect and/or alexia. It is mostly caused by isolated CC injury due to callosotomy, callosal hemorrhage or infarct or callosal demyelination (multiple sclerosis), which may lead to interhemispheric disconnection due to loss of transcallosal inhibition in the contralateral hemisphere resulting in the alien experience. In cases where the brain lesion also involves more parietal areas, such as the PPC, patients can also present with strong feelings of estrangement of the affected limb. It can also be accompanied by limb levitation, ataxia, non-conflicting movements, and parietal sensory deficits such as neglect, body schema dysfunction, hemiasomatognosia or spatial neglect. Some common etiologies are Corticobasal syndrome, Creutzfeldt-Jakob disease and stroke involving the posterior parietal artery vascular territory.

Although these two neurological signs share many common features, most notably, the tendency towards behaviour driven by external cues rather than internal goals, they differ greatly in the accompanying subjective experience these patients' report. Blakemore et al. (2002) have suggested that these two groups of patients differ in their ability to access the mental representations of their intended movements. While AnHS patients maintain this capacity (i.e.,

they recognize that external cues are in direct conflict with their own goals, to which they have access) AlHS do not, they fail to recognize that external cues are in direct conflict with their own goals as they are unable to access these goal representations. Therefore, AlHS patients accept these movements as their own, justifying them with post hoc explanations.

1.3.4. Phantom limb

Following limb amputation due to trauma or disease, more than 90% of amputees experience a **phantom limb**, a phenomenon where amputee patients report that they continue to feel the amputated limb (Ramachandran & Hirstein, 1998). Phantom limb often involves sensations such as position, touch, warmth, and coldness. In approximately 40–80% of cases, there are reports of experiencing chronic pain of the phantom limb (Hanley et al., 2009; Kooijman et al., 2000). The pain is frequently described as a burning, tingling, and cramping sensation (Sherman et al., 1989), and about 70% of them experience phantom limb pain even 26 years after the amputation (Sherman, Sherman & Parker, 1984).

Traditionally, it has been suggested that reorganization and plasticity involving the primary somatosensory and motor cortices play a crucial role in experiencing the phantom limb, due to a reorganization of the deafferented regions of cortex (Kew et al., 1997; Ramachandran & Hirstein, 1992) and/or to a preserved representation in the sensorimotor cortices corresponding to the amputated limb (Flor, Diers & Andoh, 2013; Makin et al., 2013). Several studies have shown that restoring the sensorimotor loop involving the phantom limb can lead to a significant improvement of the phantom limb pain, for example using the Mirror Box[1] therapy developed by Ramachandran and Rogers-Ramachandran (1996), implying a potential relationship between SoA and the phantom limb pain.

Advocates for the *Comparator model* have suggested that an adaptation in the forward models could explain how Ramachandran and Rogers-Ramachandran were able to reinstate voluntary movement of the phantom (Blakemore et al., 2002). The insertion of the false visual feedback supplied by the Mirror box allows the forward models to be updated, resulting in the elicitation of the efference copy of the motor commands generating changes in the predicted position of the missing limb corresponding to what the patient had seen in the mirror.

[1] Ramachandran and Rogers-Ramachandran (1996) developed the *Mirror box therapy* by which the patient is allowed to feel the imaginary movement of the phantom by providing false visual feedback of a moving limb corresponding to the phantom. This is achieved by placing a mirror in the mid-sagittal plane. With the head in the appropriate position, it is possible for the patient to see the intact limb at the same time as the mirror reflects his/her phantom limb. For most patients, moving their hand in this *Mirror box* rapidly leads to the perception that they are now able to move the phantom limb again.

On the other hand, it has also been proposed that reinstating the SoO over the phantom may also reduce the phantom limb pain, an effect known as "visual analgesia" (Longo et al., 2009). Vision of one's own body can decrease subjective pain and pain-related brain responses relative to the vision of another person's body or an object (Hansel et al., 2011; Martini, Perez-Marcos & Sanchez-Vives, 2014; Romano & Maravita, 2014), indicating that having a SoO over the phantom limb might also modulate phantom limb pain.

Interim summary

In this section I have shown that SoA and self-awareness disturbances are quite common following brain damage, involving a range of different disorders, such as schizophrenia, AHP, AlHS, AnHS, or phantom limb. A clearer and deeper understanding of how sensorimotor, perceptual, and environmental cues interact or compete in altering the SoA may aid in deepening our understanding of SoA dysfunctions that characterise certain focal neurological disorders and mental illnesses. Future research focusing on the agentic processing of patients with brain disorders involving these alterations may prove useful in developing new therapeutical and clinical interventions.

1.4. Cognitive processes that drive the SoA and their neurophysiological and anatomical substrates

There are many different factors that affect the SoA. Among them, some of the most relevant ones are intentionality and volition (Haggard & Clark, 2003; Haggard, 2017), outcome predictability (Engbert & Wohlschläger, 2007; Moore & Haggard, 2008), performance/error monitoring (Wolpert et al., 1995) and action-effect coupling (Haggard, Clark & Kalogeras, 2002; Moore et al., 2009).

1.4.1. Intentionality and volition

Contrasting voluntary/intentional actions with reflexes provides several useful neuroscientific pointers. Voluntary actions involve the cerebral cortex, whereas some reflexes are purely spinal. Volition matures late in human's development, whereas reflexes can be present at or even before birth. Finally, voluntary actions involve two distinct subjective experiences that are generally absent from reflexes. These are the experience of **'intention'** and the experience of agency (Haggard, 2008).

Wittgenstein famously asked, '*What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?*' (Wittgenstein, 1958/2010). The conscious experience of intending to move is a crucial part of the answer. The term 'intention' covers several distinct processes within the chain of information processing that translates desires and goals into behaviour. Intentions can be defined as mental representations of an individual's upcoming volitional movement (Antusch, Aarts & Custers, 2019) which constitute an essential building block of human action preparation, action initiation and action awareness promoting the experience of self-agency (Frith et al., 2000), and has important ramifications concerning social responsibility and free will. An important distinction in the study of intentionality comes from Searle (1983), who made a distinction between prior intentions (i.e., when an action is planned to reach a goal or outcome) and intentions in action (i.e., when an action is prepared and initiated). While both types of intentions are important for actions to occur, intentions in action are crucial for the emergence of action awareness. Irregularities in the formation and attribution of intentions are associated with disturbances in action and self-awareness, reflected in both implicit and explicit SoA reports (Blakemore et al., 2002; Frith et al., 2000; Voss et al., 2010). More specifically, the SoA arises from intentional actions and is diminished for unintentional or reflexive actions (Damen et al., 2015; Miller & Ross, 1975).

Throughout history, there have been two main theoretical perspectives on intentionality, namely a *dualist/libertarian* perspective, corresponding to the folk psychology, and a more recent *materialistic/scientific* explanation. The dualist/libertarian perspective suggests that our brain is just the recipient of our conscious intentions that are formed elsewhere in a non-physical realm (mind), implying both dualism and mind–to-body causation. This idea is deeply embedded in human culture, and particularly in the European culture since Descartes.

1.4.1.1. Neural correlates of intentionality: Readiness potential (RP) and lateralized readiness potential (LRP)

The establishment of conscious motor intention as a valid object of scientific investigation can be traced back to the pioneering work of Benjamin Libet and colleagues which reported that preparatory motor activity, known as the **'readiness potential' (RP)**, preceded the emergence of the conscious intention to act (Libet, 1983; see **Figure 6**, top panel). In the classic study by Libet and colleagues (1983), a temporal judgement task was performed asking participants to flex a finger at a self-chosen moment while the EEG was recorded. Participants were asked to observe the screen of an oscilloscope on which a spot moved in a circle at a speed of approx. 2.5 s per rotation, simulating a fast-moving clock hand. The participants were asked to look at the clock and to remember the clock-position at which they noted their 'intention' or 'wanting' to move. This timing of the conscious awareness of 'wanting to move' was then reported after the self-initiated movement had been performed. The encountered results indicated that the average time of this 'wanting to move' was 206 ms before the onset of muscle activity (see **Figure 6**, middle panel). Moreover, the preparatory brain activity preceding voluntary action, the RP, preceded the 'wanting to move' judgement by several hundred milliseconds. These results suggested that the initiation of action involves an unconscious neural process, which eventually produces the conscious experience of intention, which led them to the conclusion that movement is initiated without conscious awareness (Libet, 1985), rising doubt about the putative view of free will. Nevertheless, despite many years of theorizing about Libet's findings, the precise role of the RP has not been empirically established. It is still unclear whether the RP is a neural correlate of planning a motor act, anticipating a motor act, preparing to perform a motor act, consciously willing a motor act, a combination of these, or even none of the above (Schurger et al., 2012).

In a replication study, Haggard and Eimer (1999) extended Libet's findings by using a modified version of Libet's paradigm that consisted in letting the participants choose whether to respond with the left or the right index finger. In addition to the classical RP, Haggard and Eimer (1999) also recorded the **Lateralized readiness potential (LRP),** which measures the additional negativity contralateral to the actual movement, over and above that in the ipsilateral cortex. The

LRP is a more specific marker of motor preparation thought to reflect the point in time at which the response side (left vs. right), is determined. These authors claimed to have found evidence supporting a causal relationship between LRP and conscious will, concluding that the 'wanting to move' reflects events pertaining to the implementation of a specific movement, rather than more abstract representations of action occurring at processing stages prior to selection of a specific movement (Haggard & Eimer, 1999).

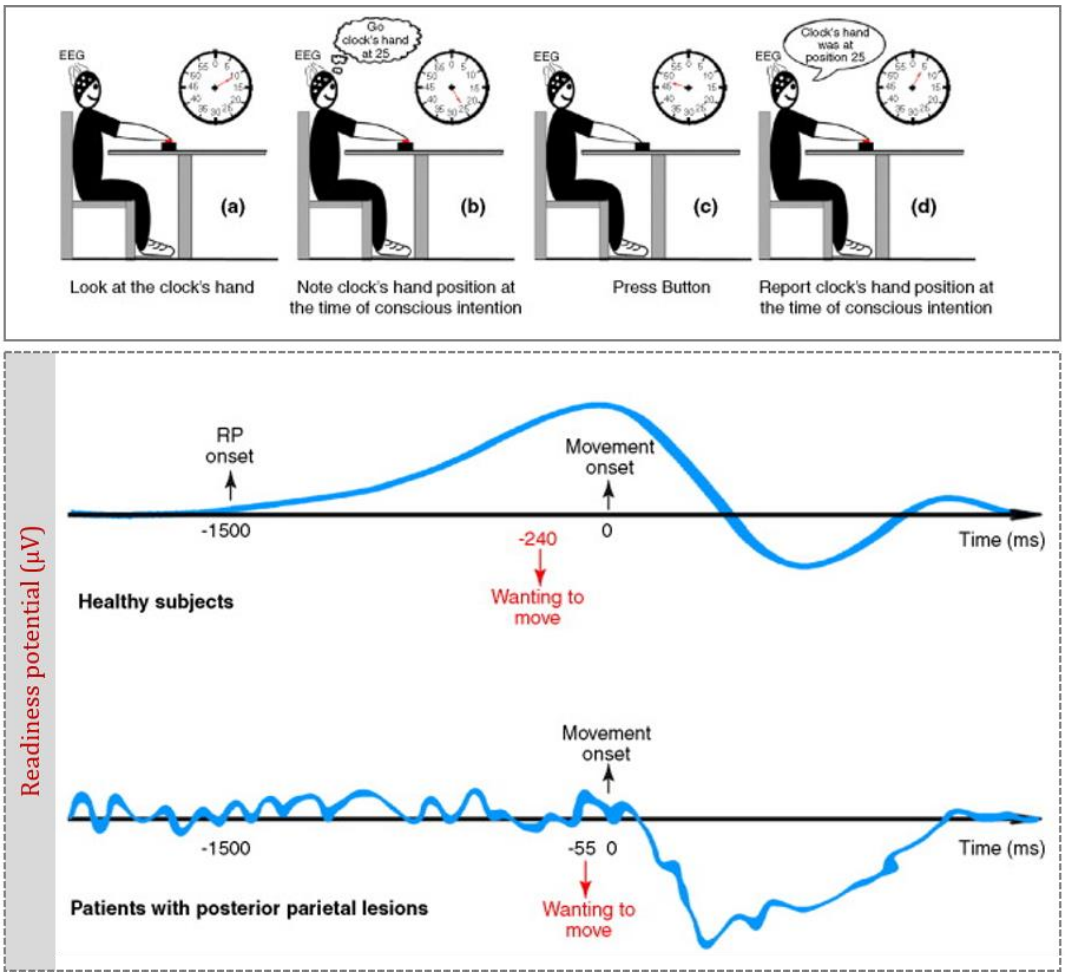


Figure 6. Readiness potentials (RP) recorded from Sirigu et al. (2004) experiment. Top panel: Illustration of Libet's paradigm. The task is to press a button with the right index finger. (a) Participants watch a clock-hand rotating on a screen. (b) They must identify the instant when they 'feel the intention to move'. (c) They press the button. (d) After a random delay, the clock stops and the subjects report the position of the clock-hand, identified in step (b). **Bottom panel**: In healthy subjects, the RP begins around 1.5 seconds before the conscious experience of wanting to move, which occurs itself around 240 milliseconds before movement onset. In patients with posterior parietal damage, the RP is absent and the conscious experience of wanting to move precedes movement onset by only a few tens of milliseconds. *Adapted from Desmurget and Sirigu (2009).*

Additionally, in a study using Libet's paradigm and RP measurements, Sirigu et al. (2004) reported that patients with parietal lesions involving the PPC (angular gyrus) can lose the early subjective experience of wanting/intending to move (see **Figure 6**, bottom panel). In this study, five healthy participants, five cerebellar patients and five patients with selective lesion in the PPC were asked to perform a task similar to Libet's (1983) original paradigm. Subjects were asked to perform a simple voluntary movement (i.e., button press), at a time of their own choosing, following a trial start cue. While performing the task, subjects were instructed, in separate blocks of trials, to focus their attention on either the actual onset of their finger movement ('movement' judgement) or their internal decision to execute it ('wanting to move' judgement). Their results showed that parietal patients had lost the ability to correctly estimate the instant in time when their intention to move was defined, showing that, the 'wanting to move' judgement onset only preceded the actual movement by 55 milliseconds, instead of ~250 milliseconds as in healthy controls or cerebellar patients (see **Figure 6**, bottom panel). The authors suggested this short delay may indicate that parietal patients did not know about their intention to move until the movement had been released once the disinhibition over the motor command by the SMA allowed them to access this subjective experience (Sirigu et al., 2004).

1.4.1.2. Anatomy of the intention to 'act': The supplementary motor area (SMA) and posterior parietal cortex (PPC)

In recent years, the neural bases of our self-conscious experiences have been extensively investigated [Appelros et al., 2007; Assal et al., 2007; Berti et al., 2005; Haggard & Magnos, 1999; Haggard et al., 2002; Lau et al., 2004; Singer, 2001; Sirigu et al., 1999, 2004, see **Figure 7** for meta-analysis results on external vs. self-agency (Sperduti et al., 2011), agency disruption (Zito, Wiest & Aybek, 2020) and motor intention and self-agency (Seghezzi et al., 2019)]. All these approaches have led to the identification of a complex interconnected network underlying conscious experience, organized around two major regions: the SMA and PPC.

The **pre-SMA** and **SMA** have been described as the most likely sources of the RP, strengthening its relationship to the intention to act (Ball et al., 1999; Eimer, 1998; Fried et al., 1991; Ikeda et al., 1992; Lau et al., 2004). Clinical studies also support this association, as it has been found that these cortical areas are commonly injured in patients suffering from AnHS (see subsection ***1.3.3. Alien/anarchic hand syndrome***) (Goldberg et al., 1981; Scepkowski & Cronin-Golomb, 2003).

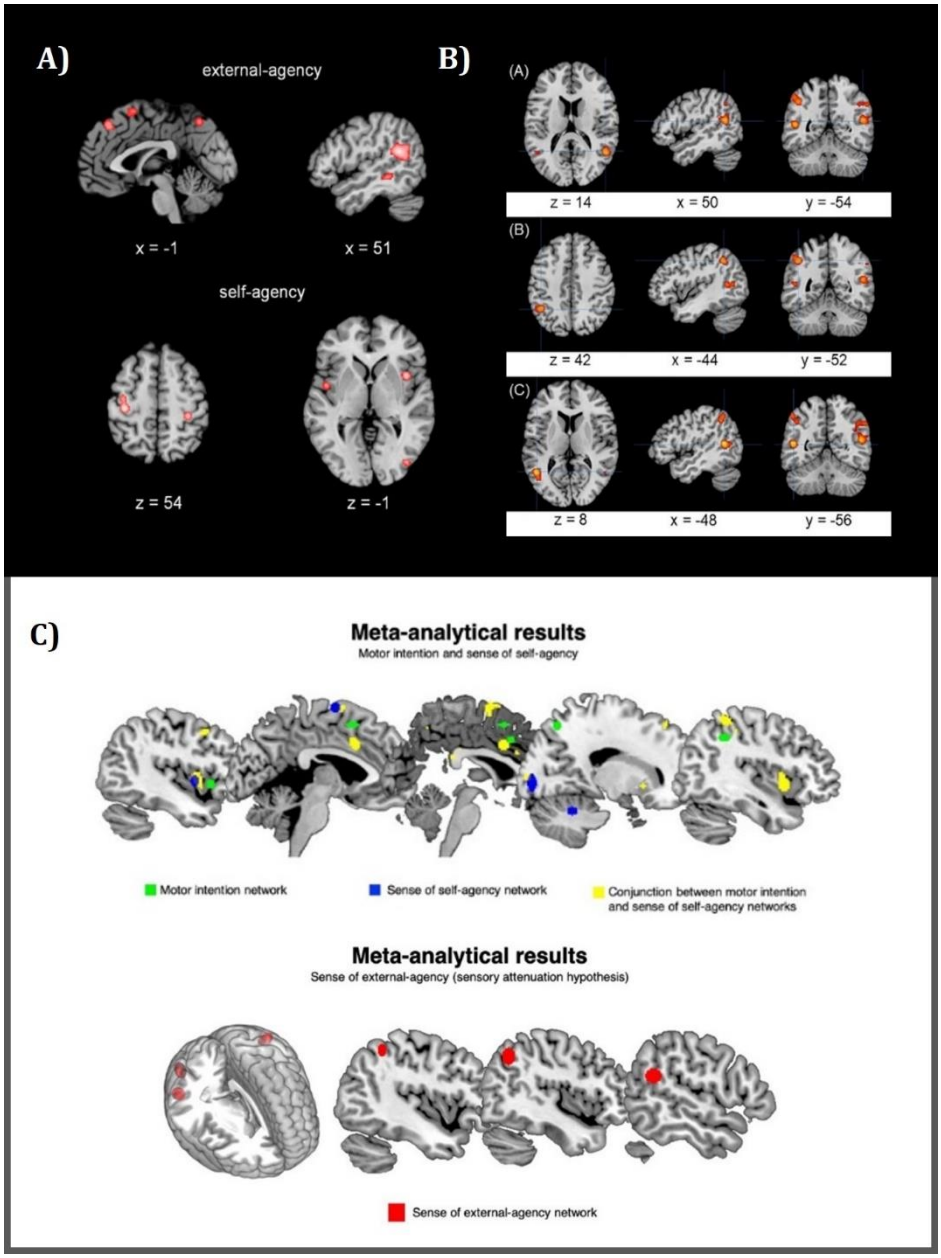


Figure 7. Neuroimaging meta-analyses. A) Results from the activation likelihood estimation for external and self-agency meta-analyses from Sperduti et al. (2011). External agency showed significant activation at the left PPC, precuneus, right superior and left middle temporal gyri -STG and MTG- and medial portions of the frontal cortex. Self-agency was related to activity in the posterior cingulate cortex -bilateral- and insula -bilateral-. All p-value < 0.05 FDR corrected. **B)** Results of the meta-analysis from Zito, Wiest & Aybek (2020), showing hyper-activation patterns in response to negative agency in the (a) right STG, (b) left PPC and (c) left MTG. All p-value < 0.05 FEW corrected. **C)** Meta-analytical results for motor intention at right middle cingulum, left pre-SMA, left aIC, superior and right superior and inferior parietal lobule (green clusters), sense of self-agency at left SMA, left posterior insula and right cerebellum (blue clusters) and conjunction of motor intention and sense of self-agency (yellow clusters). On the bottom part, meta-analysis results for external agency (red clusters), depicting the right PPC regions. *Adapted from Seghezzi et al. (2019).*

The SMA, (especially it's anterior portion –pre-SMA-) has been related to initiation and control of voluntary action (Chambon et al., 2013; David et al., 2007; Farrer et al., 2003, 2008; Leube et al., 2003; Matsuzawa et al., 2005; Miele et al., 2011; Nahab et al., 2011), especially in suppressing competing plans (Deiber et al. 1991; Hikosaka et al. 1996; Nachev et al. 2007). The pre-SMA has also been related to the phenomenon of intentional binding (Haggard et al., 2002; Haggard & Clark, 2003; for a review see Moore & Obhi, 2012, see ***1.4.2. Prediction***) and the conscious intention to act (Ball et al., 1999; Fried et al., 1991; Lau et al., 2004), highlighting the role of the pre-SMA in the predictive mechanisms underlying the SoA. For example, interruption of pre-SMA using transcranial magnetic stimulation (TMS) evoked a disruption in the SoA by showing a reduction of the temporal linkage between a voluntary key-press action and a subsequent electrocutaneous stimulus (i.e., intention binding effect) (Moore & Haggard, 2010, see ***1.4.2. Prediction***).

Direct evidence linking the SMA/pre-SMA regions to intentionality comes from *electrical stimulation mapping (ESM)* studies. ESM has a long history in medical and fundamental sciences (Penfield & Boldrey, 1937), becoming a standard technique during brain surgery for inferring the function of brain areas in humans (Duffau, 2005). In these resections, while the patient performs motor, language, or cognitive tasks under controlled conditions (e.g., opening-closing hand; naming objects), the cortical surface is electrically stimulated to provoke reproducible, transient changes in behaviour. The ability to map critical locations has tremendous clinical value, as the location of brain function varies between patients due to individual variability and to functional reorganization/plasticity in response to the individual's brain pathology (e.g., brain tumour). However, a common criticism of this method concerning the tendency of electrical currents to spread through white matter bundles has been posed. Yet, this signal propagation does not follow an anarchic path but spreads along physiologically meaningful pathways (Duffau, 2005), mimicking the normal function of the stimulated tissue.

In a study by Fried and colleagues (1991), ESM was applied within the SMA region in a group of epileptic patients, triggering an '**urge to move'** evoking movement intentions such as an '*urge to move the right leg inward*' or '*to lift the right elbow*' (**Figure 8A**). When the intensity of the stimulation was increased above the urge threshold, the actual movement occurred due to the suppression of the inhibition exerted on the primary motor cortex (M1) (Ball et al., 1999).

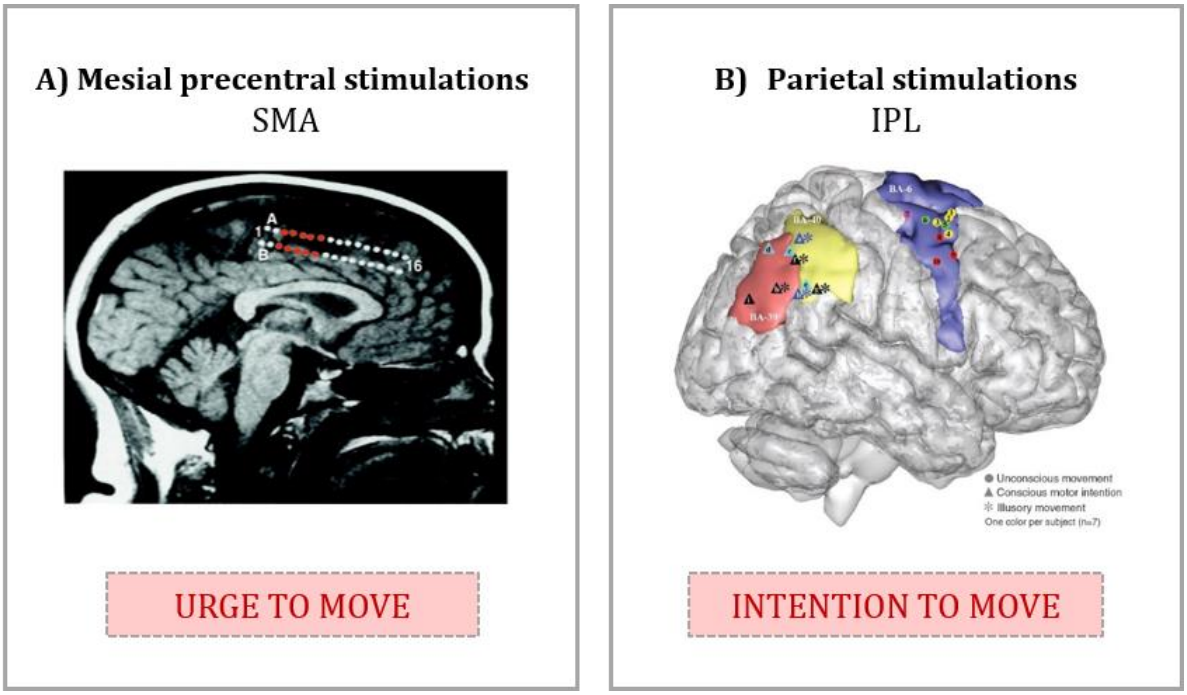


Figure 8. A) Mesial Precentral sites evoking feelings of 'urge to move' when electrically stimulated (red circles). Data are shown for a single subject on the individual Magnetic Resonance imaging (MRI) image of this subject. *Adapted from Fried et al. (1991).* **B)** Inferior parietal (IPL) sites evoking feelings of 'wanting to move' when electrically stimulated (triangles). Data are shown for 3 subjects after registration of the individual MRI to the MNI template. One colour per subject. *Adapted from Desmurget et al. (2009).*

Another significant brain region commonly found to be engaged during self-processing is the **PPC,** a brain region that has historically been considered an associative region, important for spatial attention and multisensory integration (Colby & Goldberg, 1999; Ungerleider & Mishkin, 1982), in charge of transforming sensory inputs into a representation that is useful for guiding actions in the external world. It has been suggested that the PPC, together with the cerebellum, might represent the neural correlates of the *Comparator* mechanism described earlier, due to its involvement in monitoring the concordance between self-produced movements and their sensory consequences (Agnew & Wise, 2008; Blakemore & Sirigu, 2003; Chaminade & Decety, 2002; David et al., 2007; Farrer & Frith, 2002; Farrer et al., 2003, 2008; Spence et al., 1997; Spengler, von Cramon & Brass, 2009a).

For example, there is evidence on the involvement of the PPC in online correction of reaching movements (Desmurget et al., 1999; Della-Maggiore et al., 2004; Pisella et al., 2000), during which both somatosensory and visual signals converge in the PPC providing feedback for making online

corrections. Efference copy signals (coming from cerebellar computations), which provide replicas of the issued motor commands, are also fed back to the PPC via feedback projections, leading to an integration of feedback and forward signals, aiding in the agency estimation (Blakemore & Sirigu, 2003; Wolpert et al., 1995).

Besides, it has been proposed that this region is also involved in action processing, including movement intention and decision making (Andersen, 1988; Andersen & Buneo, 2002; Burnod et al., 1999; Desmurget et al., 2009; Gold & Shadlen, 2007; Graziano & Gross, 1998; Kalaska et al., 1997; Rizzolatti, Fogassi & Gallese, 1997). In the fascinating ESM study by Desmurget et al. (2009), a group of seven patients underwent awake brain surgery for tumour removal involving premotor and parietal sites. These authors electrically stimulated different sites in the brain while simultaneously measuring electrical activity in 12 muscles in the face, hand, wrist, elbow, knee and foot using electromyography (EMG)[2]. The patients were asked to report when they felt an 'urge to move', and when they thought they had moved (see **Figure 8B**). They reported that when ESM was applied over the left and right PPC (Brodmann Areas – BA 40 and 39-, supramarginal and angular gyri, respectively), a strong desire to move the contralesionally hand, arm, or foot (right PPC) and lips (left PPC) was elicited, in the absence of actual movement. For example, a patient reported:

- Patient: *'I wanted to move my foot'.*
- Experimenter asked: *'Which foot'.*
- Patient (showing his left leg): *'This one'.*
- Experimenter: *'How did you want to move it?'*
- *Patient: 'I don't know, I just wanted to move it'.*

In addition, when the same area was re-stimulated at a higher intensity, patients reported they had already moved, even though they still had not. On the other hand, patients whose PMC was directly stimulated, actual movement was elicited, but without any awareness of having done so. Interestingly, in most patients these reports were much less specific than the precise intentions reported by Fried et al. (1991) when stimulating the SMA (see above). Desmurget et al. (2009) proposed that the PPC contains a 'map of intentions' so that when stimulated, an internal state

[2] EMG is a technique for recording electrical activity produced by skeletal muscles, which detects the electrical potential changes generated by the muscle cells.

that resembles what Searle called '*intention in action'* arises. Interestingly, when the stimulation intensity was increased, the patients reported that they had performed the movement they previously intended to do (in the absence of any muscle contraction), leading the authors to hypothesize that higher intensities of stimulation would not simply prime a motor representation to consciousness (giving rise to intention), but also recruit the executive network responsible for movement monitoring through forward modelling (Desmurget et al., 2009). These results suggest that an internal cue to agency, the intention to move, was sufficient to produce the experience of movement, challenging the *Comparator* explanation regarding the need of a motor/efferent signal (see **Table 3** for a complete overview of brain regions associated to SoA processing).

Moreover, clinical and neuropsychological data from lesion studies often associate damage in PPC regions to abnormalities in self-awareness such as hemi-spatial neglect (Mort et al., 2003), asomatognosia (Baier & Karnath, 2008), or AlHS/AnHS (Franck et al., 2001; Fourneret et al., 2002), as well as in AHP, related to the impossibility to access the information derived from the external sensory feedback following damage to components of the fronto-parietal network (Fotopoulou, 2014; Monai et al., 2020; Pacella et al., 2019). Experimental studies in neurological patients with lesions involving this region also show changes in awareness of voluntary action (Sirigu et al., 1999, 2004) (see subsection ***1.3. Dysfunctions of the self*** for more information on these neurological conditions), for example showing an hyperactivation of the right PPC in schizophrenic patients when experiencing delusions of control (Spence et al., 1997) during a positron emission tomography (PET) study.

1.4.2. Prediction

Prediction is a crucial aspect for adaptive behaviour. Contemporary models of brain function emphasize the predictive nature of the brain, constantly trying to make sense of the world by predicting what we will experience and updating these predictions as the context changes. Classical motor control theories such as the *Ideomotor theory* (Prinz, 1997) suggest that actions are represented in terms of their sensory consequences and that they are selected and initiated by anticipating these sensory consequences (for empirical evidence, see e.g., Elsner & Hommel, 2001; Kunde, 2001; Pfister, Kiesel & Hoffmann, 2011).

Table 3. Literature search of neuroimaging studies addressing the SoA.

Study	Task	Contrast	Brain region
Agnew and Wise (2008)	Finger tap	*Agency:* Active vs. passive finger tap	Cerebellum
		Agency disruption: Passive vs. active finger tap	Bil. IPL, L anterior parietal lobe
Bralslev et al. (2008)	Cursor movement	*Agency disruption:* Asynchronous vs. synchronous feedback	Bil. TPJ, L SPL, L IPL, R MTG
David et al. (2007)	Joystick movement	*Agency disruption:* Incongruent vs. congruent feedback	Bil. MFC, L precuneus, Bil. IPL, Bil SMA, Bil. DLPFC, L PCG, Bil. MTG
Decety et al. (2002) -PET	Manipulation of objects	*Agency:* Imitation of the self by the other vs. self-action	R SMG, Bil. MFC, R caudate Nucleus, R IFG, L pre-SMA
Farrer and Frith (2002)	Joystick movement	*Agency:* Self-attribution vs. other attribution	L aIC, L SMA, L PMC, Bil. primary somatosensory cortex, Cerebellum
		Agency disruption: Other attribution vs. self-attribution.	Bil. AG, Bil. Precuneus, L PMC
Farrer et al. (2003) -PET	Joystick movement	*Agency:* Parametric decrease self to other visual feedback	R IPI, R cerebellum
		Agency disruption: Parametric increase self to other visual feedback	Bil. IPL, Pre-SMA, R PMC, R ACC
Farrer et al. (2008)	Manual peg removal task	*Agency disruption:* Delay vs. no delay	Bil. IPL, L SMG, Bil. DLPFC, L IFG, L pre-SMA, L PCG, L Pulvinar
Farrer et al. (2008)	Finger movement	*Agency disruption:* Perturbed agency vs. preserved agency	Bil. IPL, R IPS, Bil. MFC, L IFG, R IFS, R DLPFC
Fink et al. (1999) -PET	Luria's bimanual coordination task	*Agency disruption:* Interaction effect of visual feedback modulation on out-of-phase bimanual movements	R DLPFC
Fukushima et al. (2013)	Key press	*Agency:* No agency judgement vs. agency judgement	R PCC, R IOG, L cerebellum, L precuneus, L STG, L PCG
		Agency disruption: No agency judgement vs. agency judgement	R PCG, L CC, L cerebellum
Kontaris, Wigget and Downing (2009)	Hand actions	*Agency:* Compatible vs. incompatible condition	R parieto-occipital sulcus
		Agency disruption: Incompatible vs. compatible visual feedback	Bil. STG, Bil. MTG, Bil. SFG, Bil. MFC, Bil. IFG, R Precuneus
Kühn, Brass and Haggard (2013)	Key press	*Agency:* Active vs. passive condition	SMA
Leube et al. (2003a)	Open and close the hand	*Agency:* Self vs. other perform-observe mismatch	L IOG, R ITL, Bil. MOG, R IFG, R SPL
		Agency disruption: Other vs. self perform-observe mismatch	R Lingual, R ACC, R precuneus
Leube et al. (2003b)	Open and close the hand	*Agency disruption:* Delay vs. baseline	Bil. Insula, R IFG, Bil. PCG, R STG, R SOG, L MOG, L SMA, L PoCG, Putamen
Matsuzawa et al. (2005)	Key press	*Agency:* Synchronous vs. delayed action-effect feedback	R cerebellum, R PPC, Bil. primary visual area, Bil. SMA, R PMC, L sensorimotor
		Agency disruption: Delayed vs. synchronous action-effect feedback	R cerebellum, R PPC, Bil. primary visual area, Bil. SMA, R PMC, L sensorimotor
Miele et al. (2011)	Cursor movement	*Agency disruption:* Conditions with turbulence vs. conditions without turbulence	Bil. TPJ, Bil. MTG, R SMA, L SPL
Nahab et al. (2011)	Virtual hand movement	*Agency disruption:* Incoherent vs. coherent hand movement	Bil. MFC, R STG, R MTG, R STG, R SMG, R AG, Bil. IPL, Cerebellum, L insula, R precuneus
Renes et al. (2015)	Key press	*Agency:* Agency over the square movement vs. no agency	L IPL, Bil. SFG, SMA
Schnell et al. (2007)	Bimanual joystick movement	*Agency:* Visuo-motor congruence vs. visuo-motor incongruence	Bil. PCG, L MFC, Bil. PoCG, Bil. STG, L PCC, Cerebellum
		Agency disruption: Visuo-motor incongruence vs. visuo-motor congruence	Bil. MFC, R SFG, L ACC, Bil. SMG, Bil. IPL, R precuneus, Bil. MTG, R ITG, L PCC, L Cerebellum
Spengler, von Cramon and Brass (2009a)	Finger movement	*Agency disruption:* Parametric increase with increasing discrepancy of action-effect prediction	Bil. TPJ, L IFS
Tsakiris et al. (2010)	Finger movement	*Agency:* Active movement with synchronous feedback vs. passive movement synchronous feedback	L Cerebellum, R PoCG, R Cerebellum, L PI, R Precuneus
		Agency disruption: Asynchronous vs. synchronous feedback	R AG, L Insula, L Cerebellum, R MFC, R MOG, R MTG, R IFG
Yomogida et al. (2010)	Joystick movement	*Agency disruption:* Agency error vs. sensory matching error	L cerebellum, Bil. SMA, R IPL, EBA

L: left; R: right; Bil.: bilateral; fMRI: functional magnetic resonance imaging; PET: positron emission tomography; IPL: inferior parietal lobe; SPL: superior parietal lobe; TPJ: temporo-parietal junction: MTG: middle temporal gyrus; MFC: middle frontal cortex; SMA: supplementary motor area; DLPFC: dorsolateral prefrontal cortex; PCG: precentral gyrus; MTG: middle temporal gyrus; IFG: inferior frontal gyrus; aIC: anterior insular cortex; PI: posterior insula; PMC: premotor cortex; AG: angular gyrus; PCG: precentral gyrus; ACC: anterior cingulate cortex; SMG: supramarginal gyrus; IPS: intraparietal sulcus; IFS: inferior frontal sulcus; PCC: posterior cingulate cortex; IOG: inferior occipital gyrus; STG: superior temporal gyrus; EBA: extraestriate body area; PoCG: postcentral gyrus; MOG: middle occipital gyrus.

Predictive mechanisms of this type have been reported across many different species and systems (Crapse & Sommer, 2008), under the terms of **corollary discharge** (Sperry, 1950) or **efference copy** (von Holst & Mittelstaedt, 1950), to capture the idea that, in addition to generating the signals that produce a movement, a copy of the motor commands is used to generate the prediction regarding the sensory consequences of that movement. The predictive capacity of our system increases the efficiency of attention and cognitive processing by preventing the central nervous system from wasting metabolic resources processing irrelevant sensory stimuli. These signals have been identified in different sensory domains (e.g., visual and auditory), and are generated at distinct levels of conscious processing, from nonconscious sensory predictions generated for movements, to conscious predictions generated during decision making (Friston & Kiebel, 2009). A key structure related to the corollary discharge pathway is the anterior cerebellar cortex (Ito, 2006; Ramnani, 2006; Streng, Popa & Ebner, 2018). Its correlation with primary and secondary somatosensory cortex activity (Blakemore et al., 1999a) and its role in processing anticipated vs. unanticipated somatosensory stimuli is well established (Gao et al., 1996). In addition, the cerebellum receives afferents from the inferior olivary complex which responds to unexpected proprioceptive or cutaneous stimuli (Gellman, Gibson & Houk, 1985).

Several SoA models, especially the *Comparator account* (Blakemore et al., 1998, 1999a, 2002, 2003; Frith et al., 2000; Miall & Wolpert, 1996; Wolpert & Flanagan, 2001; Wolpert et al., 1995; Wolpert & Kawato, 1998) suggest that the anticipation of action outcomes is a key aspect for the SoA (see previous description of the Comparator account on section ***1.2.1. Motor control-based theories***). This model proposes that whenever there's a **prediction error**, defined as a mismatch between prior expectations and reality (den Ouden, Kok & de Lange, 2012), the sensory information is attributed to an external event, and the SoA is then reduced or lost. Indeed, the SoA has been found to be reduced by movement-related prediction errors (i.e., observation of unpredicted movements that do not correspond to the ones executed by the participants) (Daprati et al., 1997; David et al., 2016; Farrer et al., 2008; van den Bos & Jeannerod, 2002), as well as by outcome-related prediction errors (i.e., by the occurrence of unexpected outcomes following participant's action) (Caspar et al., 2016; David et al., 2016; Kühn et al., 2011; Sato & Yasuda, 2005, see the following section ***1.4.3. Outcome monitoring***). However, the notion that prediction errors systematically reduce SoA has been challenged by Synofzik and colleagues (2008b), who stated that the *Comparator* explanation fails to account for situations where people experience SoA despite prediction errors, for example during A-O temporal delays or when individuals experience a vicarious SoA for observed actions in the absence of actual movements (Tieri et al., 2015; Wegner, Sparrow & Winerman, 2004).

Another interesting finding regarding the influence of predictive signals on self-related processing is the **intentional binding effect**[3] [Haggard et al., 2002; Haggard & Clark, 2003; for a review see Moore and Obhi (2012)]. Following the seminal work by Libet et al. (1983), Haggard et al. (2002) used the Libet clock method to study the perceived time of actions and their consequent effects. They reliably found that voluntary actions (i.e., key press) and external stimulus (i.e., tone) are shifted towards each other in temporal perception, causing the perceptual compression of the temporal interval between them. This phenomenon is often interpreted as related to the SoA (Haggard, 2005; Tsakiris & Haggard, 2005), since this temporal binding effect has consistently been reported to occur during intentional/self-made actions and to be absent or weakened for unintentional/external or induced passively (Borhani, Beck & Haggard, 2017; Haggard et al., 2002). Neurocognitive explanations for this effect highlight the role of forward predictions, arguing that preparatory motor processes lead to a pre-activation of the neural representation of the predicted sensory effect of one's action. When the sensory outcome occurs, it reaches the perceptual threshold faster due to the increased excitability of the appropriate sensory representation (Waszak, Cardoso-Leite & Hugues, 2012), resulting in a shortening of the A-O perceptual latency (Wolpe et al., 2013). The association between the SoA and intentional binding has been verified in several studies, showing that the greater the explicit sensation of having provoked a consequence through one's own action (SoA), the greater the subjective temporal compression of the interval that elapses between the two events (Imaizumi & Tanno, 2019; Pyasik et al., 2018).

Surprisingly, the magnitude of the intentional binding effect has been found to be significantly stronger in schizophrenia patients reporting SoA disruption compared to healthy controls (Haggard et al., 2003; Voss et al., 2010), contrary to what one would expect. These findings are also consistent with evidence showing that schizophrenia patients tend to hyper-associate their A-O contingencies and to over-attribute the consequences of their movements to themselves (Daprati et al., 1997; Franck et al., 2001). Moreover, an increase in this binding effect was also reported when infusing ketamine on healthy control subjects (Moore et al., 2011), which at sub-

[3] Intentional binding effect. There were four critical conditions in Haggard, Clark and Kalogeras (2002) original experiment. In baseline conditions, participants either made voluntary actions or listened to the occurrence of an auditory tone while they watched a rotating clock hand on a computer screen. They were asked to report the position of the clock hand when they moved or when the tone occurred. In operant conditions, participants made a voluntary key press on every trial, but this time it was followed by an auditory tone 250 ms later. In specific blocks, participants were asked to judge either the time of their action or the time of the tone. The key comparison was the perceived times of actions and tones in baseline conditions with the perceived time of actions and tones in operant conditions.

anesthetic levels it has been shown to produce a state that resembles schizophrenia in several key aspects (Corlett, Honey & Fletcher, 2007; Corlett et al., 2010). Taken together, these results highlight the complex relationship between SoA measures and higher-level phenomenological experiences of being the agent.

1.4.2.1. Sensorimotor attenuation and the N1 component.

Among other mechanisms, intact sensorimotor predictive mechanisms are a key aspect in being able to distinguish between self-generated and external generated actions. The difference between the perception of self vs. external sensations is portrayed by **sensorimotor attenuation**. Sensorimotor attenuation refers to the reduction in the perceived intensity of the consequences of one's own actions relative to externally caused sensations (Shergill et al., 2003), for example when we try to tickle ourselves (Blakemore et al., 1998). It has been suggested that it is directly linked to the efference copy, which is used by the internal/forward model to generate a prediction about the sensory intensity, which is then removed from the actual sensory feedback reducing the sensitivity to the actual sensory stimulus (Bays et al., 2006; Roussel, Hugues & Waszak, 2014). It has been suggested that these predictions about the sensory outcomes occur in the cerebellum (see subsection ***1.4.2.1. The cerebellum as the integration hub for the forward modelling***).

A coherent SoA relies on an intact predictive signal and its consequent sensorimotor attenuation, which may be employed to measure the integrity of SoA processing. For instance, reduced sensory attenuation has been reported in patients suffering from schizophrenia (Shergill et al., 2005) during a force matching task[4], showing that self-generated forces were attenuated less in the patient compared to the healthy control group. Moreover, a correlation between visual sensory attenuation and the severity of delusions of control has also been reported (Lindner et al., 2005), confirming that difficulties in sensorimotor predictive mechanisms are tightly linked to delusions of influence/control and abnormalities in the SoA.

One of the earliest indicators of SoA related to predictive processing is the attenuation of self-produced sensations or sensory outcomes (Blakemore et al. 1999a). Using the internal predictive signals based on the efference copy of motor commands, we as agents can anticipate and thereby attenuate the sensory response to predicted sensory outcomes (Blakemore et al, 1999a), allowing

[4] A target force is applied to the subject's left index finger by a torque motor. Subjects are then required to reproduce the force they just experienced, either directly by pressing with the index finger of their right hand or indirectly by using a joystick controlling the torque motor. Usually, the reproduced forces are larger than the forces that are applied by the torque motor. This degree of overcompensation has been used as a proxy for sensorimotor attenuation and the integrity of agency (Shergill et al., 2003).

us to distinguish self-generated vs. externally generated sensations (see section ***1.4.2. Prediction*** for more details on sensory attenuation mechanisms).

Recently, the N1 event-related potential (ERP) component has been shown to be a neural marker of sensory attenuation (e.g., Bäss, Jacobsen & Schröger, 2008; Kuhn et al., 2011; Gentsch & Schuetz-Bosbach, 2011), showing a reduced amplitude of this component during self-generated vs. externally generated effects (see **Figure 9**) [see ***Box 3****. Temporal processing dimension of the self: Event-related brain potentials (ERPs) and oscillations*]. However, because sensory attenuation depends on outcomes being highly predictable [for a review see Hughes, Desantis and Waszak (2013)], it may be less relevant when encountering situations with high uncertainty. As a result, the N1 component might not be a reliable indicator of agency when monitoring requires high-level processing to determine whether the sensorimotor predictions match the actual feedback.

Importantly, previous studies have shown a sensory and motor involvement of gamma band oscillations (30-60 Hz) during SoA processing, proposing that gamma oscillations represent the forward (ascending) connections carrying prediction errors (Arnal & Giraud, 2012; Bastos et al., 2012; Bauer et al., 2014). For example, Palmer, Davare and Kilmer (2016) recently described a negative correlation over the sensorimotor cortex between gamma-oscillatory activity and the magnitude of perceptual sensory attenuation (as perceptual sensory attenuation increased, that is, as matching became less veridical, the power of oscillatory activity within the gamma-frequency band decreased). Interestingly, these findings support theoretical accounts of perceptual sensory attenuation, which postulate that the difference in sensory attenuation comparing self vs. other is due to the ability to generate predictions about the sensory consequences of our own actions (Blakemore et al., 1999a).

Moreover, gamma band activity is often reported as related to movement onset, associated with the initialization of motor actions (Ball et al., 2008; Cheyne et al., 2008; Muthukumaraswamy, 2010; Pfurtscheller et al., 2003). It is thought to reflect the competition of distinct motor programs in tasks where alternative responses are required (Gaetz et al., 2013; Grent-'t-Jong et al., 2013; Heinrichs-Graham et al., 2018; Isabella et al., 2015), suggesting they might be critical for cognitive control in the motor domain.

Action monitoring has also been related to gamma-band activity, as shown by Ulloa (2021) using an intentional binding paradigm in which participant's actions (i.e., finger press/lift), which could be either congruent or incongruent with a displayed finger movement, subsequently triggered an outcome (i.e., a tone). Participants with greater intentional binding effect (i.e., greater time compression differences between congruent and incongruent actions) showed greater gamma

power decreases for congruent versus incongruent actions (**Figure 10**). These authors concluded that gamma power may be related to tracking the effects of A-O congruency affecting action and SoA processing.

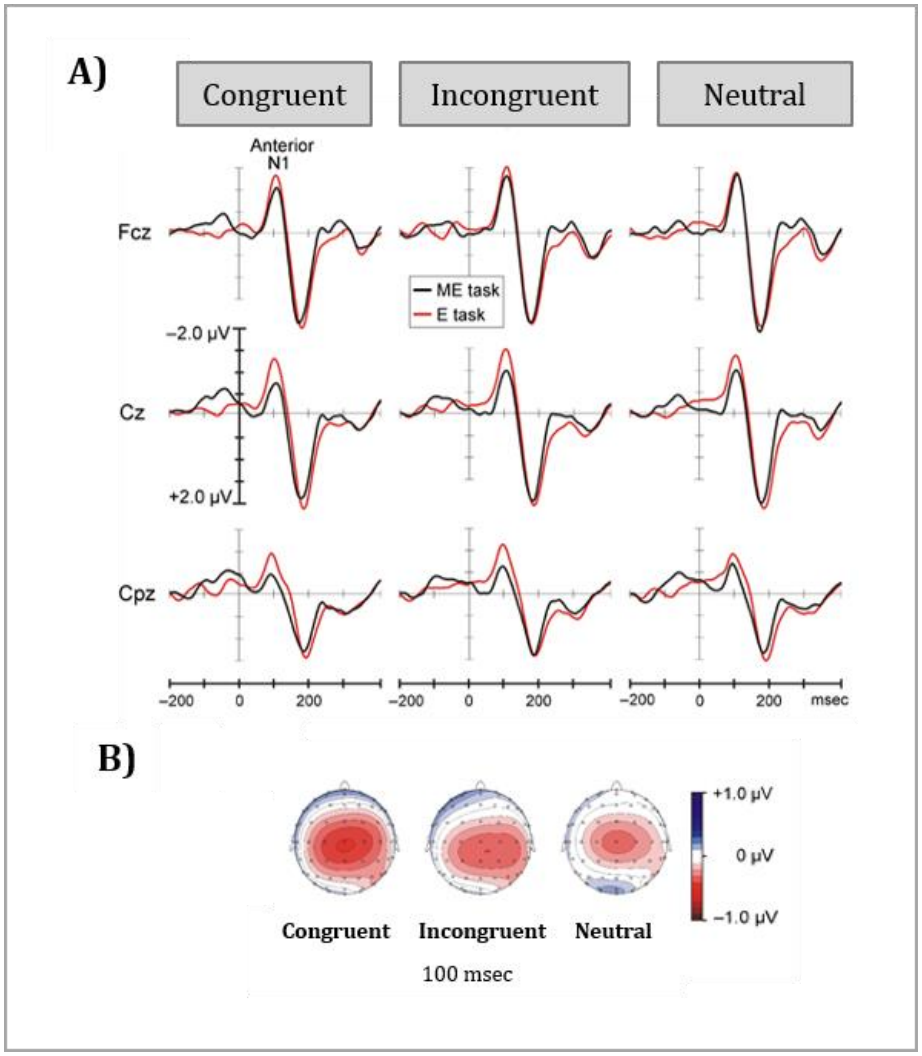


Figure 9. N1 component related to sensory suppression mechanisms. **A)** Stimulus-locked grand average ERP waveforms for the E (externally generated effect) and ME (self-generated effect) tasks at electrodes FCz, Cz, and CPz. **B)** Scalp topographies of the voltage difference between E and ME tasks, separately for the three priming conditions, showing an attenuation of the N1 to self-generated relative to externally generated effects. *Adapted* from *Gentsch and Schuetz-Bosbach (2011).*

Box 3. Temporal processing dimension of the self: Event-related brain potentials (ERPs) and oscillations.

Another technique broadly employed for studying the SoA are event-related brain potentials (ERPs). ERPs are voltage fluctuations in the electroencephalographic (EEG) signal recorded from the scalp that are time-locked to a specific event, such as the presentation of sensory feedback. Because these voltage fluctuations are very small, ERPs are usually observed after averaging multiple occurrences of the same event, forming an average ERP waveform (Donchin, Ritter, & McCallum, 1978). ERP waveforms reveal a series of reproducible positive- and negative-going deflections that delineate neural activity with millisecond precision, allowing to study the temporal dynamics of agency processing distinguishing between predictive/postdictive neural mechanisms. Crucially, these systematic bioelectrical responses, termed ERP components, reflect psychological processes related to the experimental task demands (Kappenman & Luck, 2011).

Besides ERPs, another measure of electrical brain activity that can be extracted from the EEG signal are neural oscillations. Oscillations are increases and decreases in power (i.e., wave amplitude) that reflect the synchronization and desynchronization of firing patterns in neural populations, respectively (Buzsáki & Watson, 2012). The more common approach to evaluate oscillatory activity is time-frequency (TF) decomposition, which allows to quantify power at multiple frequencies over time, including delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (13-30 Hz), and gamma (>30 Hz) bands. Interestingly, and unlike ERPs, oscillatory activity captures activity that is not phase-locked to the time-locking event (Tallon-Baudry et al., 1996), and thus can reveal additional and complementary information about the same underlying cognitive process.

1.4.2.2. Prediction errors and the error-related negativity (ERN)

As the Ideomotor theory suggests, predictions allow us to anticipate the sensory consequences of our actions, helping us to guide our behaviour (Hommel et al., 2001). For predictions to be useful, we must be able to detect if the intended action goal has been achieved or not and, subsequently, attribute this failure to a cause (Holroyd & Coles, 2002), supporting agency attribution. Like so, if errors are predicted by the forward model during action execution, these errors would most likely be due to internal causes (e.g., inappropriate motor commands) and attributed to the self. On the

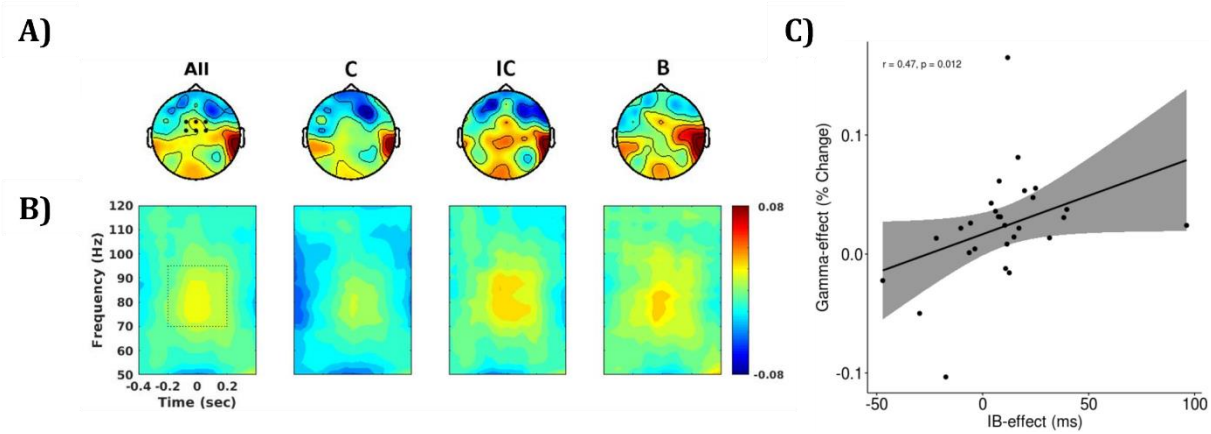


Figure 10. **A)** Topographic distributions of grand-average gamma-band (70-95 Hz) activity in the −0.2 and 0.2 s interval relative to response onset for all (All), congruent (C), incongruent (IC), and baseline (B) conditions, depicting decreased gamma power for the congruent relative to the incongruent and baseline conditions. **B)** Time-frequency representation of power estimates at electrodes depicted in A (all). **C)** Gamma power differences correlated positively with the intentional binding effects. Dispersion plot and tendency line for the gamma power effect versus the intentional binding effect (incongruent minus congruent). Each dot represents a participant. *Adapted from Ulloa (2021).*

other hand, if the intended goal is not achieved but no error was predicted by the forward modelling, this error should be attributed to external perturbations (e.g., the wind has changed the trajectory of the ball during a free kick in soccer). Based on the initial proposal by Del Cul et al. (2009) of a dual-route model for conscious and non-conscious decision making, it has been suggested that there are two distinct **error monitoring subsystems** running in parallel to the forward computations (Charles et al., 2013; Logan & Crump, 2010; Padrao et al., 2016; Wegner et al., 2004): i) an internal error-monitoring loop, in charge of detecting deviations between *intentions* and *predictions* and ii) an external error-monitoring loop involved in comparing *predicted* vs. *actual states* of our system (see **Figure 11**).

The **internal monitoring loop** oversees the implementation of fast and unconsciously triggered error-correction or compensatory mechanisms (Charles et al., 2013). Electrophysiological studies, which take advantage of its high temporal resolution, have identified the error-related negativity (ERN) component as a functional index of this loop (Falkenstein et al., 1991; Gehring et al., 1993; Holroyd et al., 2005; Rodriguez-Fornells, Kurzbuch & Münte, 2002). This early (<100 ms) ERP component recorded at fronto-central locations reflects a prediction of an event in the future (i.e., an upcoming error), emerging prior to the external feedback and therefore considered a neural

correlate of predictive error processing (Carter et al., 1998; Falkenstein et al., 1991; Gehring et al., 1993; Holroyd et al., 2005; Rodriguez-Fornells et al., 2002).

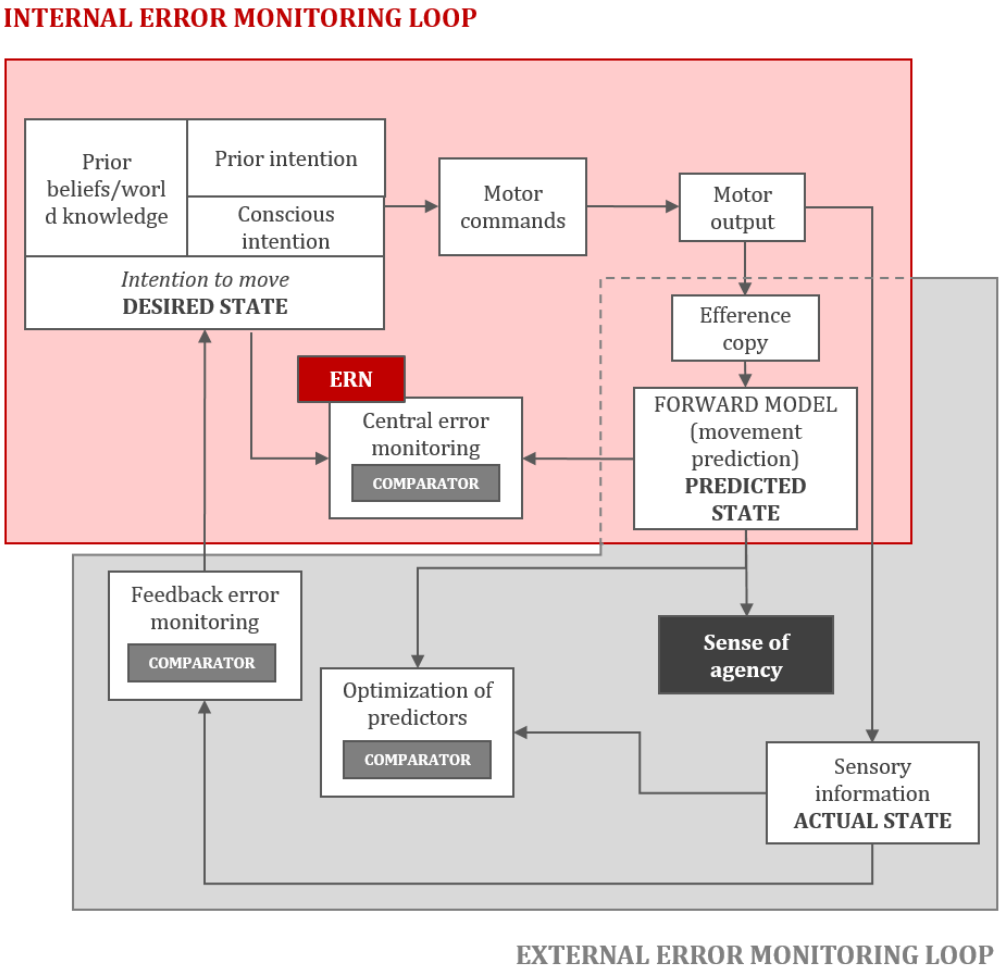


Figure 11. A summarized neurocognitive account for the error monitoring mechanisms underlying SoA processing. On one hand, internal cues such as prior beliefs, intentions to act and forward predictions are computed. The internal error monitoring loop (depicted in red) is continuously checking the congruence between intended-predicted states of the system by means of comparator mechanisms (depicted in grey). Whenever a mismatch is detected, an ERN component might be elicited signaling the need for fast compensatory error correction mechanisms. Depending on the environment and context, predictions are compared to the sensory inputs coming from the external world, resulting in the SoA. An external error monitoring loop (in blue) might be in charge of monitoring the congruency between predicted and actual states of the system, triggering the SoA for the performed action in case there's no mismatch detected. *Based on Synofzyk et al. (2008a) and Wolpert and Miall (1996).*

In the seminal experiment by Nieuwenhuis et al. (2001), the authors provided strong evidence on the unconscious nature of the ERN. Participants were asked to perform an antisaccade task during

which they had to move their eyes in the opposite direction of a visual target. Remarkably, in many cases the participants were not aware of their own errors (i.e., they had no idea that their eyes had transiently been to the wrong place), but even on such unconscious erroneous trials, the ERN continued to be elicited, providing a very clear example of the workings of the internal error monitoring loop.

The **ERN** is thought to reflect post-decisional conflict monitoring [for reviews, see Larson et al. (2014) and Yeung and Summerfield (2012)], originating from areas in the posterior MFC including ACC and pre-SMA (De Bruijn et al., 2009; Holroyd et al., 2004; Ridderinkhof et al., 2004; Ullsperger, Danielmeier & Jocham, 2014, see **Figure 12**). Studies on metacognition, that is, the cognitive processes by which individuals can reflect upon (monitor) their own internal mental states and apply their knowledge to evaluate and regulate (control) their own mental states (Nelson et al., 1999), have shown a modulation of this component with confidence judgements, showing more negative potentials for high error confidence (Boldt & Yeung, 2015; Scheffers & Coles, 2000). Because of its internal, early nature, the ERN might aid in the processing of the SoA by detecting deviations from the expected goal of our ongoing actions, implementing fast error-correction or compensatory mechanisms when needed (Gehring et al., 1993; Marco-Pallarés et al., 2008; Rodriguez-Fornells et al., 2002), before the outcome/sensory feedback appears. This readjustment might serve as a direct indication of agency, or it could influence post-hoc evaluations of performed actions (Knoblich & Natalie, 2005).

It has been proposed that some schizophrenic patients, particularly those who suffer from thought insertion and delusions of control, might have problems in accessing the representations elicited in the forward or internal error-monitoring system (Feinberg, 1978; Frith et al., 2000). Following the initial suggestion from Feinberg (1978), Frith et al. (2000) postulated that a problem might exist in the forward, action monitoring systems, showing abnormal internal monitoring brain potentials, reflected in an ERN amplitude reduction [Bates et al., 2002; Foti et al., 2012, see Manoach and Agam (2013) for a review].

Increases in **theta-band activity** (4-8 Hz) have been related to the ERN at medial-frontal locations (Cavanagh & Frank, 2014; Cavanagh et al., 2010; Cohen, 2011; Luu & Tucker, 2001), associated to error commission and negative prediction errors (**Figure 13**). These theta dynamics might act as temporal templates for organizing midfrontal neuronal workings, which are then enhanced following events indicating a need for cognitive control (Cavanagh & Frank, 2014). Hence, theta activity might reflect a common mechanism for implementing adaptive control in a variety of contexts involving uncertainty about actions and outcomes.

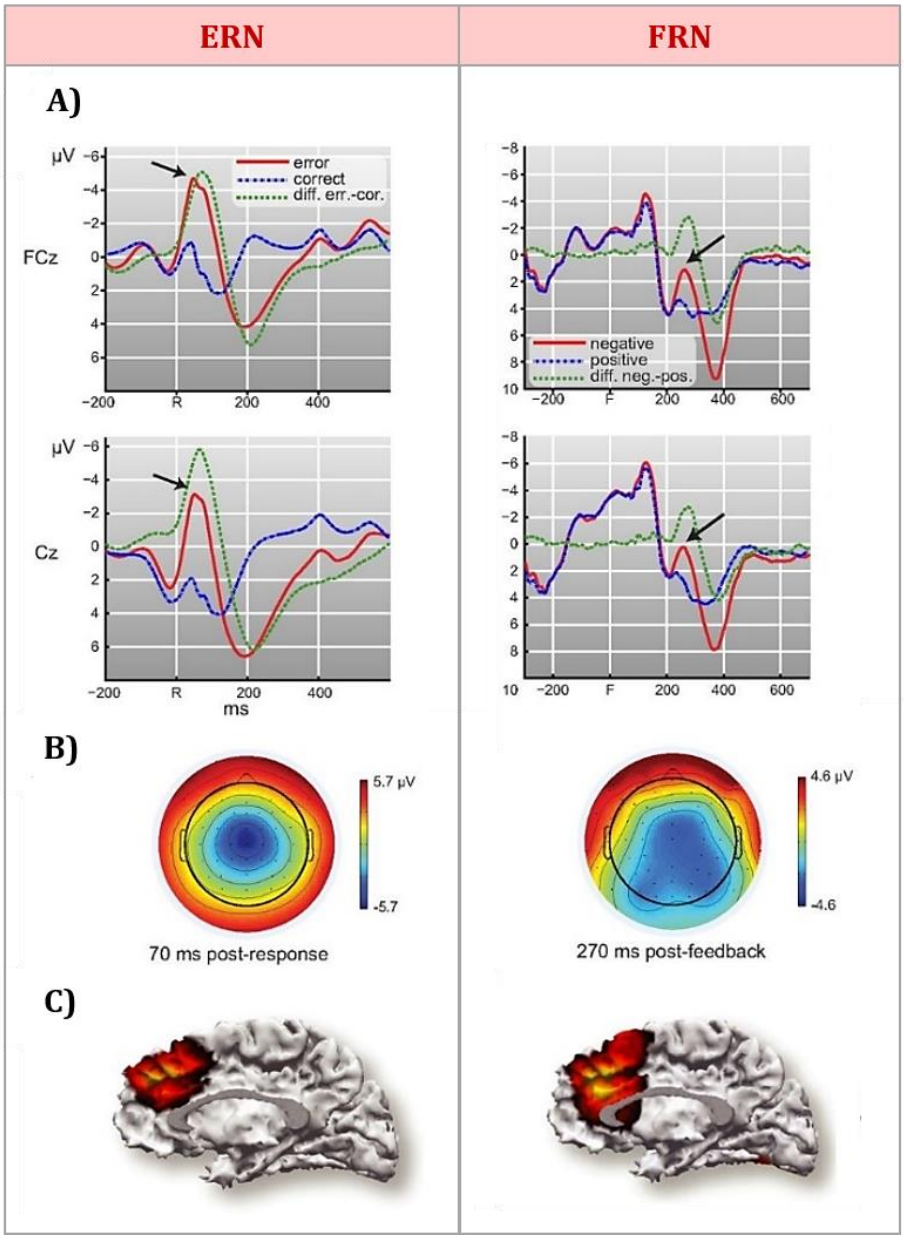


Figure 12. **ERN and feedback-related negativity -FRN- components. A)** ERPs associated to error commission (ERN, FRN) at midline electrodes Fz and Cz. associated with performance monitoring. **B)** Corresponding topographies of difference waves at the peak activity of the ERN (70 ms) and FRN (270 ms). **C)** Source localizations of the ERN and FRN depicting the MFC. *Adapted from Ullsperger et al. (2014).*

The need for cognitive control and executive processes appears to be conveyed by frontal midline theta activity (4-8 Hz). Theta band responses have been typically associated with control/monitoring functions including response inhibition, error monitoring and behavioural

adaptation mechanisms, such as the post-error slowing effect (PES)[5] (Cohen & Ridderinkhof, 2013; Gulbinaite et al., 2014; Nigbur et al., 2011; Rabbitt & Rogers, 1977).

Regarding SoA processing, in a recent EEG study using a spatial congruence paradigm to measure implicit SoA (i.e., intentional binding), Ulloa (2021) reported a strong theta-band response for the incongruent relative to the congruent condition, although it was not correlated with implicit measures of SoA. Nevertheless, these authors concluded that theta oscillations could have also contributed to implicit SoA by singling interference or conflict and affecting action fluency (Cohen, 2014; Ulloa, 2021).

1.4.2.1. The cerebellum as the integration hub for the forward modelling

A key brain structure commonly observed during fMRI studies addressing SoA processing is the **cerebellum** (Agnew & Wise, 2008; Farrer & Frith, 2002; Farrer et al., 2003; Fukushima et al., 2013; Kontaris et al., 2009; Matsuzawa et al., 2005; Nahab et al., 2010; Schnell et al., 2007; Tsakiris et al., 2010). It has been proposed that the cerebellum receives efferent signals from the primary motor cortex (von Holst & Mittelstaedt, 1950), and uses them to generate a motor-to-somatosensory prediction that mimics the sensory consequences of an action, that is, the efference copy, and prepares the musculoskeletal system to successfully execute a movement (Blakemore et al., 1998, 1999a; Blakemore & Sirigu, 2003; Imamizu & Kawato, 2008; Tseng et al., 2007; Wolpert & Kawato, 1998; see **Figure 14**).

Following the *Comparator* account, this corollary discharge/efference copy mechanism is then applied to compare the predicted vs. actual state of the system (Sperry, 1950). In case of a mismatch, the cerebellum receives feedback information from cortical and subcortical areas to reset its prediction and to apply an online correction of the movement [see Miall (1998) for a review].

[5] When our brain detects an error, this process changes how we react on ensuing trials. People show post-error adaptations, potentially to improve their performance in the subsequent incoming event. At least three types of behavioural post-error adjustments have been observed. These are post-error slowing (PES), post-error reduction of interference, and post-error improvement in accuracy. The PES describes the prolonged reaction time (RT) in trials after an error compared to RTs in trials following correct trials (Rabbitt, 1966). Several explanations for the PES have been proposed, such as i) that PES is a compensatory control mechanism serving to improve subsequent performance (Gehring & Fencsik, 2001), ii) PES reflects an orienting response -and so an extra processing- to an unexpected event (Notebaert et al., 2009) and iii) commission of an error leads to an increase in selective suppression in the immediately following trial (Ridderinkhof, 2002).

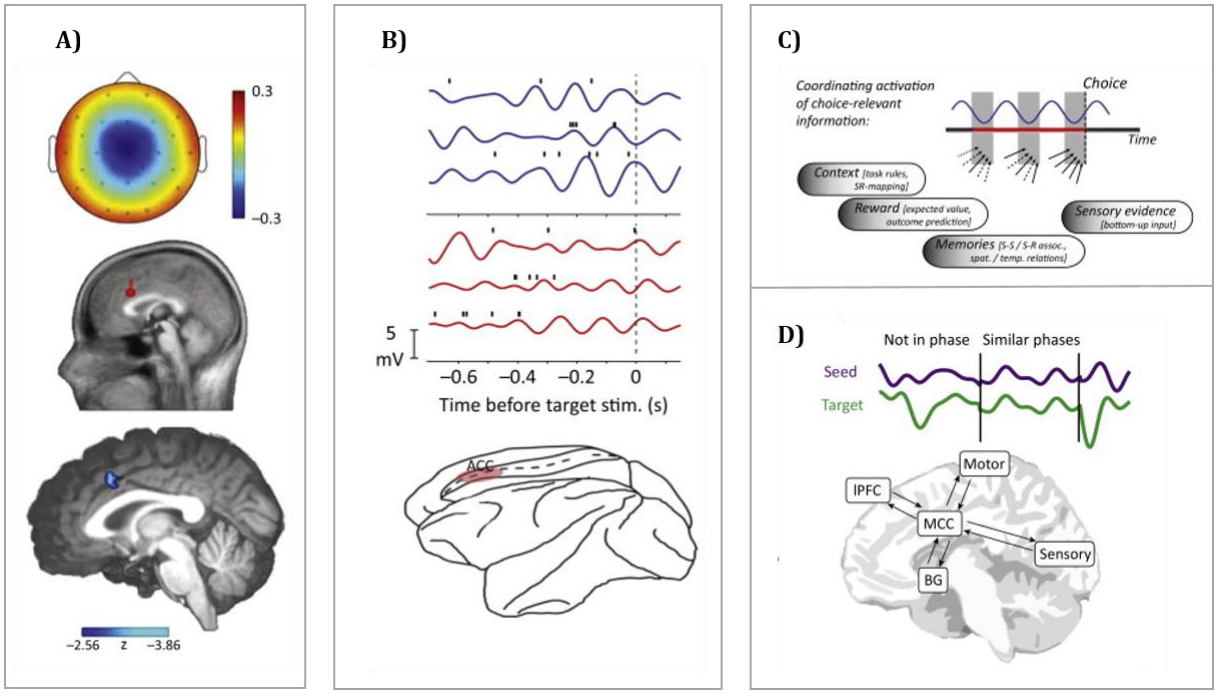


Figure 13. **Theta modulations and cognitive control. (A)** In humans, midfrontal theta evoked by errors (ERN) has been localized to midcingulate cortex (MCC) using dipole source modelling (red) and concurrent hemodynamic activity (blue). **(B)** Theta activity recorded from the rostral cingulate sulcus in rhesus macaques (red) during performance of an antisaccade task. Increased theta power on anti- versus pro-saccade trials (blue > red traces) was associated with stronger spike-field coupling within the theta rhythm. **(C)** Midfrontal theta is thought to reflect the synchronization of goal-relevant information around critical decision points. In this example, theta activities coordinate inputs across cortical areas (arrows), particularly at the trough of the oscillation (gray bars). **(D)** Theta band phase consistency is thought to reflect the instantiation of transient functional networks (purple and green traces). For instance, intersite theta band phase consistency following signals of the need for control have been observed between sources modelled in MCC, lateral PFC, motor areas, and sensory (i.e., extrastriate visual) cortex. *Adapted from Cavanagh and Frank (2014).*

Activity in the cerebellum has been shown to increase during motor errors (Diedrichsen et al., 2005; Schlerf, Ivry & Diedrichsen, 2012), as well as when manipulating the subjects' movements feedback. For instance, when introducing a variable and unexpected temporal delay between the subject's movement and its sensory consequences, activity in the cerebellum was found to be positively correlated with this delay (Blakemore et al., 1999a; van Kemenade et al., 2019). In another study, when subjects were instructed to make hand movements while receiving real-time visual feedback, which could be either synchronous or asynchronous with the subject's

movements, an increase in activity in the cerebellum was observed during asynchronous feedback, among other regions (Nahab et al., 2010).

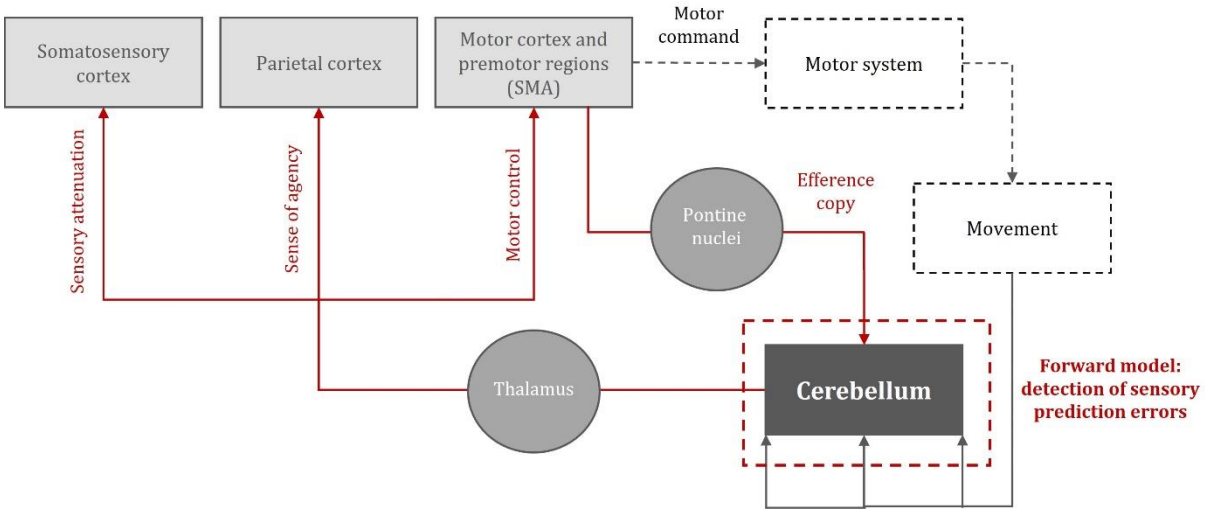


Figure 14. Hypothetical model for the cerebral-cerebellar loops involved in the forward model. The cerebellum is thought to integrate the efference copy (i.e., copy of the motor commands originated at motor and premotor regions that represent the intended action) and the actual sensory feedback generated by the movement. If a discrepancy occurs between the predicted and actual motor outcome (i.e., prediction error) the cerebellum would detect this mismatch and the error signal would then be sent to different cortical areas to serve different functions: motor control, sensory attenuation, and SoA. *Adapted from Welniarz, Worbe and Gallea (2021).*

Furthermore, it has been proposed that the cerebellum could play a particular role in sensory attenuation, by evaluating the degree of mismatch between the predicted and the actual sensorimotor feedback. Following this view, cerebellar activity is decreased in response to self-generated movements associated with tactile stimulation, while this activity is increased by external tactile stimulation (Blakemore et al., 1998). In a recent study using TMS to disrupt cerebellar activity, an interference with sensory attenuation of self-initiated sounds was reported (Cao et al., 2017). Moreover, electrophysiological recordings of the cerebellum in non-human primates have demonstrated that cerebellar neurons can cancel the reafferent sensory effects produced by self-generated movements (Brooks & Cullen, 2013).

1.4.3. Outcome Monitoring

Although it has been largely demonstrated that the SoA requires the monitoring of internal predictive representations, it is important to notice that it can only be constructed after those signals/internal simulations (mental models) are compared with the external world information (i.e., reafferent visual, motor, or proprioceptive feedback). An **external monitoring loop** (**Figure 11**) might be in charge of comparing internal and external generated signals (i.e., predicted vs. actual outcomes), giving rise to the SoA in case congruency occurs (Blakemore, Wolpert & Frith, 1998; Blakemore et al., 2002; Frith, Blakemore & Wolpert, 2000; Miall & Wolpert, 1996; Wolpert & Flanagan, 2001; Wolpert et al., 1995; Wolpert & Kawato, 1998).

Some authors have recently argued that the experience of agency can only be retrospectively attributed (Moore & Haggard, 2008, Synofzik et al., 2008a; Wegner & Wheatley, 1999), suggesting that a reliable SoA may only arise when reafferent feedback signals are available that confirm or disconfirm its internal prediction or intention. Accordingly, inducing A-O discrepancies by creating mismatches between sensory, re-afferent feedback and anticipated A-O contingencies can induce a perturbed SoA. To do so, many studies have used temporal or spatial distorted sensory feedback or presented incongruent A-O contingencies in term of content consistency (see **Table 4** for an overview of experimental manipulations assessing the SoA and section ***1.5. Experimental manipulations of the SoA***).

1.4.3.1. Feedback monitoring: The feedback-related negativity (FRN) and feedback correct-related positivity (FCRP)

Outcome processing plays a key role in the SoA as it enables us to evaluate the consequences of our actions ascribing a congruent or incongruent agentic state based on our predictions and previous knowledge. A commonly reported ERP component signaling predicted vs. actual outcomes discrepancies is the feedback-related negativity **(FRN)** (Miltner et al., 1997). The FRN is a fronto-central ERP component peaking around 250–300 ms after outcome feedback (see **Figures 12** and **15**), generated at the ACC (Gehring & Willoughby, 2002; Luu et al., 2003; Miltner, Braun, & Coles, 1997; Ruchsow et al., 2002). The FRN can be observed after feedback about the action outcome is available, and it is generally larger following negative or unexpected feedback, reflecting a postdictive/retrospective correlate of error processing [for reviews, see San Martín (2012) and Ullsperger et al. (2014)].

From the *Reinforcement learning perspective* (Holroyd & Coles, 2002; Holroyd et al., 2008), the FRN is thought to reflect the degree of negative prediction error due to decreased mesencephalic

dopaminergic activity that is transmitted throughout the ACC to the mPFC from the basal ganglia (Holroyd & Coles, 2002; Nieuwenhuis et al., 2004). Thus, these signals conveyed in the mPFC might help the system to detect potential cognitive conflicts arising from previous expectations and unexpected outcomes, enhancing action monitoring and control processes signaling the need to update and modify internal models of A-O contingencies (Botvinick, Cohen & Carter, 2004; Holroyd & Coles, 2002; Ridderinkhof et al., 2004).

Recent studies addressing more retrospective aspects of agency attribution (i.e., outcome monitoring) have reported modulations of the FRN depending on SoA attribution. For example, Li et al. (2011) found that the amplitude of the FRN was enhanced when participants had a sense of responsibility over task outcomes. Experimental conditions such as omitting the occurrence of feedback (Gentsch, Ullsperger & Ullsperger, 2009), blocking the response buttons (Steinhauser & Kiesel., 2011) or manipulating participants' finger movements (Mathalon, Whitfield & Ford, 2003) have shown modulations of the FRN probably associated to agency prediction errors/A-O inconsistencies. Therefore, the FRN might be related to more posterior, postdictive, and retrospective aspects of SoA processing, signaling the contiguity and contingency mismatch between a self-made action and a sensory outcome, which may be used by the system to readjust the prediction of the sensory outcome, serving as a direct indication of SoA or influencing post-hoc JoA (Knoblich & Sebanz, 2005).

In addition to the FRN, Bednark et al. (2013) reported that the **feedback correct-related positivity (FCRP),** an ERP component related to feedback appearance indicating the achievement of a reward or task goal (Holroyd et al. 2008), is enhanced when a sensory outcome is elicited by a self-made action compared to when a sensory outcome's presentation is unrelated to the performance of a self-made action. More recently, these same authors reported that a smaller FCRP (or larger FRN) was associated with outcomes that were externally attributed in comparison to self-attributed outcomes (Bednark & Franz, 2014). Within the outcome/performance monitoring framework (Ullsperger et al., 2014; San Martín, 2012), the FCRP seems to be related to the retrospective component of SoA by tagging A-O contiguity or A-O coupling.

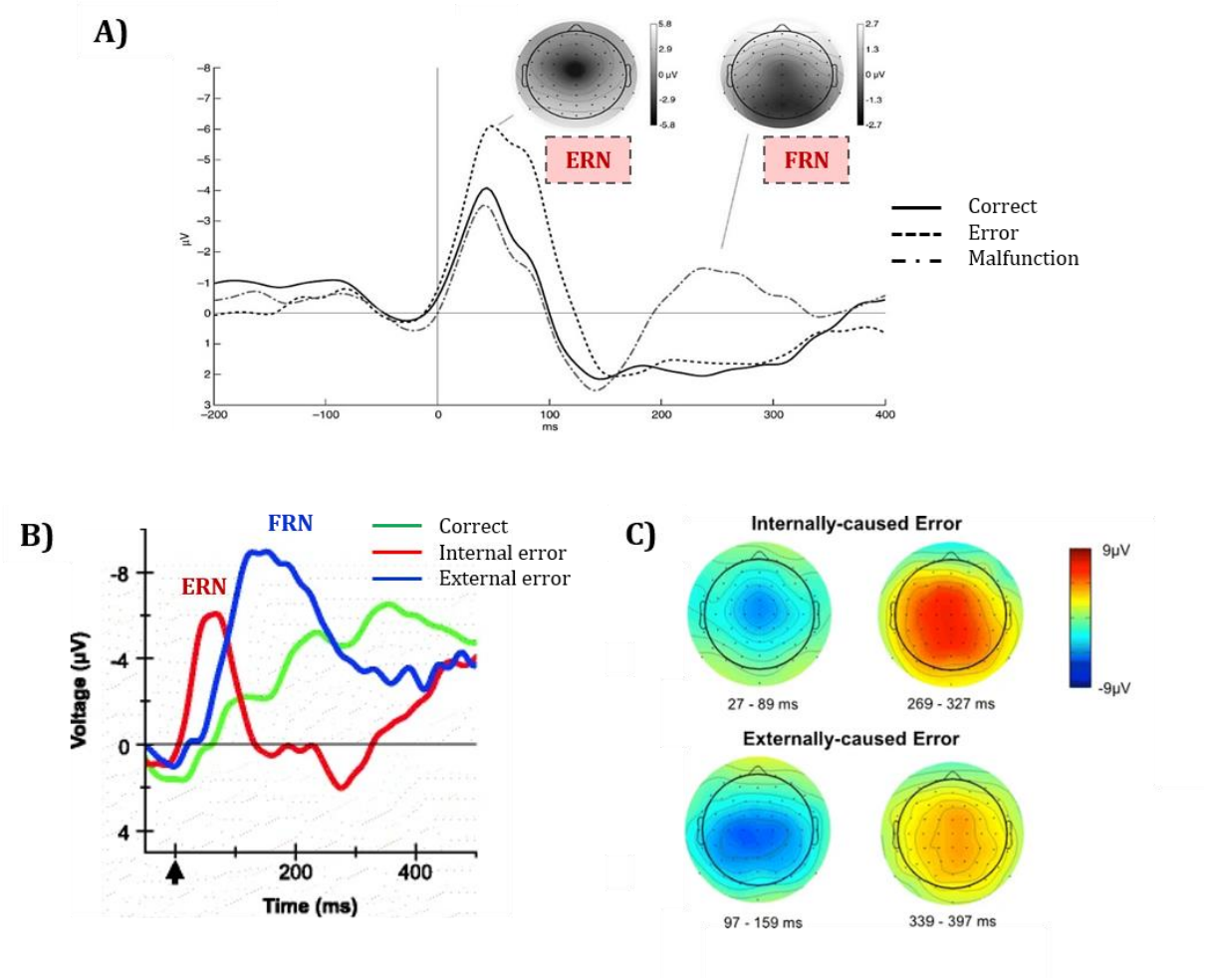


Figure 15. FRN component elicited after A-O discrepancies. A) Response-locked grand mean average and scalp topographies depicting the ERN and FRN components for self-made errors and technical malfunctions/external errors, respectively, at electrode FCz. *Adapted from Gentsch, Ullsperger and Ullsperger (2009).* **B)** Response-locked waveforms at channel Cz depicting the ERN and FRN components for internally caused errors and externally caused errors, respectively. **C)** Scalp topographies for difference waveforms representing internally caused errors minus correct trials and externally caused errors minus correct trials (FRN). *Adapted from Steinhauser and Kiesel (2011).*

1.4.3.2. Context updating: P300 component

In addition to the FRN/FCRP, another outcome monitoring electrophysiological deflection that has been related to SoA processing is the **P300/P3** component, a positive deflection of the ERPs peaking between 300 ms and 600 ms. The P3 has been related to attention-driven categorization of salient outcome-related information, as well as to the motivational salience of rewarding feedback [Donchin, 1981; Polish, 2007; Sutton et al., 1965, for a review see Glazer et al. (2018)]. The role of the P3 in the SoA processing is associated with A-O predictions over the sensory effects

of action, and it has its origin in the *Context-updating hypothesis* of the P3 (Donchin & Coles, 1988). According to this hypothesis, incoming stimuli are compared to an internal model of the current stimulus context, generating a P3 component when the incoming stimulus does not match the stimulus context, indicating that the internal model requires updating (Donchin & Coles, 1988; Polich, 2007). This *Context-updating hypothesis* of the P3 is very similar to the *Comparator account* for the SoA (Frith et al., 2000; Miall & Wolpert, 1996; Wolpert et al., 1995; Wolpert, & Frith, 2002; Wolpert & Flanagan, 2001), in the sense that both theories emphasize the comparison between an incoming stimulus and internal representations of the expected external sensory consequences as the key processing stage.

The P3 component can be divided into two separate subcomponents, namely the P3a and the P3b (Polich, 2007). The P3a has a more anterior distribution, and it is typically elicited by deviant or novel stimuli (e.g., the deviant stimulus in an oddball paradigm). Conversely, the P3b has a more posterior distribution, maximal over parietal regions, and is elicited by infrequent target stimuli (e.g., the target stimulus in an oddball paradigm).

Larger P3 amplitudes have been observed when comparing hearing one's own name to other words (Berlad & Pratt, 1995), or during autobiographical memories against other memories (Gray et al., 2004), as well as for self-referent vs. unrelated pronouns (Zhou et al., 2010), pointing out to its association with self-referential processing. Because the lack of contiguity between a self-made action and a sensory outcome may cause participants to consider the outcome as 'odd' or unexpected (Friedman, Cycowicz & Gaeta, 2001), the P3 has previously been shown to be associated to participants' ratings of agency for trials with the highest level of uncertainty. For example, Kuhn et al. (2011) found a stronger P3 component for tones that were judged to be generated externally compared to those judged to be generated by oneself (see **Figure 16**). Source localization analyses have located the P3 generators in the temporal or parietal lobes (Linden, 2005; Mulert et al., 2004; Volpe et al., 2007), which is in line with many fMRI studies associating SoA processing with the TPJ.

In the oscillatory domain, delta-band responses (1-4 Hz) have been shown to dominate the P3 response (Stampfer & Başar, 1985), being visible even without filtering in single-trial EEG-ERP analysis (McCarthy & Donchin, 1981). Delta-band oscillatory activity has been related to several cognitive functions, such as the motivational relevance of the task, the salience of the target stimulus (Knyazev, 2007, 2012), as well as with reward processing (Bernat et al., 2011; Bernat, Nelson & Baskin-Sommers, 2015; Cavanagh, 2015; Knyazev, 2007) and commission of motor errors (Yordanova et al., 2004). Moreover, parietal delta oscillations are considered to mediate signal

detection and decision-making (Başar et al., 1999; Schürmann et al., 2001) throughout the modulation of rhythmic gain of information accumulation (Lakatos et al., 2008; Schroeder & Lakatos, 2009). Furthermore, altered delta oscillations have also been associated with aberrant self-experience (Carhart-Harris et al., 2016; Kometer et al., 2015).

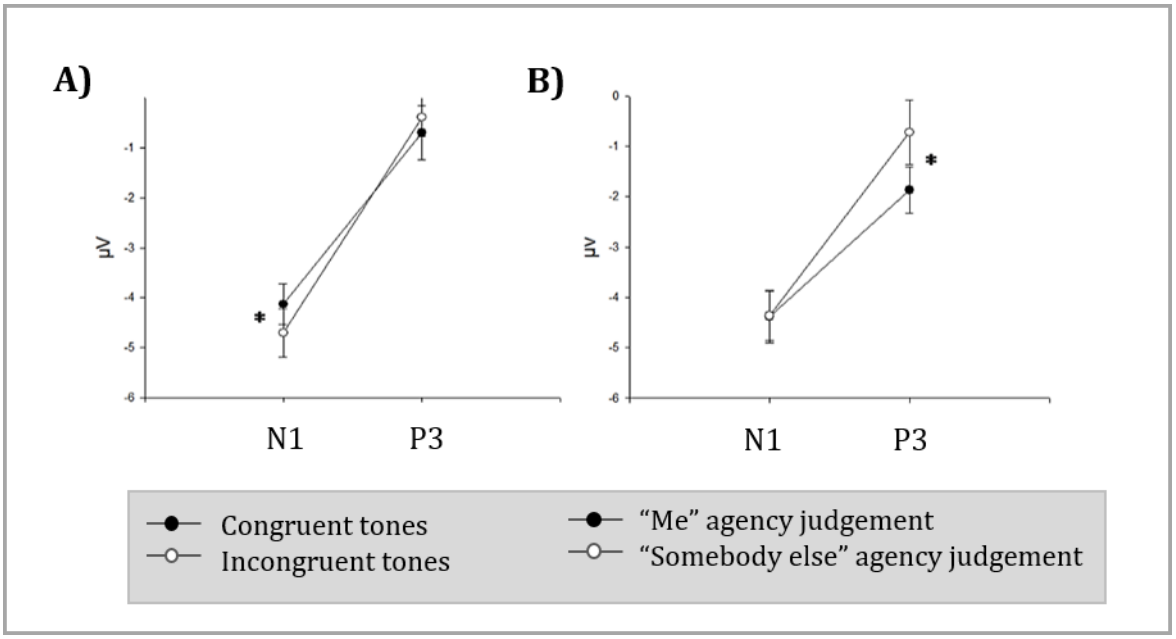


Figure 16. P3 component. A) Plot displaying mean signal averaged over electrodes F3, Fz, F4, FC1, FCz, FC2 showing more pronounced differences between congruent and incongruent tones for the N1 component. **B)** Plot displaying mean signal averaged over electrodes F3, Fz, F4, FC1, FCz, FC2 showing more pronounced mean amplitude differences between "me" and "somebody else" judgements for the P3a component, but not for the N1. *Adapted from Kuhn et al. (2011).*

The relationship of delta-band activity to SoA processing comes mainly from its association with the motivational significance of self-relevant information and its corresponding attentional involvement. Recently, in a dynamic reinforcement learning task, Cavanagh (2015) reported that delta activity at different time intervals reflected, separately, reward prediction errors and state prediction errors (Cavanagh, 2015). More specifically, early delta-band activity, related to reward positivity, may correspond to a surprising reward signal, while later delta activity, associated to the positive-going deflection of the P300 component, appeared to be related with behavioural adjustments. According to an influential hypothesis proposed by Nieuwenhuis, Aston-Jones and Cohen (2005), late large positive deflections, such as P3, might reflect the activity of the neuromodulatory *locus coeruleus–norepinephrine system*, which shows an enhancement in

response to motivationally significant events (Nieuwenhuis et al., 2005). From this perspective, late posterior delta-band activity could reflect the motivational significance of self-relevant information and its corresponding attentional involvement.

Importantly, it is to be noted that delta responses are only one of the determinants of the P3 component. P3 responses are also contributed to by theta (4-8 Hz) and gamma (30-60 Hz) oscillatory responses (Başar-Eroğlu & Başar, 1991; Demiralp et al., 1999, 2001; Karakaş et al., 2000; Kolev et al., 1997; Spencer & Polich, 1999; Yordanova et al., 2000).

1.4.3.3. The N400 component and semantic incongruence

Another ERP component that has been recently related to self vs. other distinction is the **N400** component, a negative deflection occurring approximately at 400 ms after stimulus onset. Although this component was first described as reflecting semantic anomalies/incongruences in the linguistic domain (Kutas & Hillyard, 1980), similar effects have been recently observed for non-linguistic material involving meaningful actions (e.g., Sitnikova, Kuperberg & Holcomb, 2003). For example, a modulation of the N400 has been previously found after the observation of erroneous actions (Bach et al., 2009; Balconi & Vitaloni, 2014), and other action-related information such as detection of mistakenly performed basketball actions in expert athletes (Proverbio et al., 2012). More recently, in a study using a combined ERP-Virtual Reality paradigm (Padrao et al., 2014, **Figure 17**) (full body immersion), the authors reported interesting findings regarding the N400 and the processing of the SoA.

While self-made errors showed a classical ERN component, the introduction of external errors made by the avatar (i.e., incongruent hand movements of the avatar relative to the real participant movement) diluted the SoA in participants and triggered a delayed negativity at parietal electrodes peaking at 400 ms after the occurrence of the incongruent action (see **Figure 17**). Based on these results, these authors proposed that the N400 might reflect a violation of the SoA in terms of sematic/conceptual knowledge about one's own action as a result of the mismatch between the internal predictions and the reafferent signals.

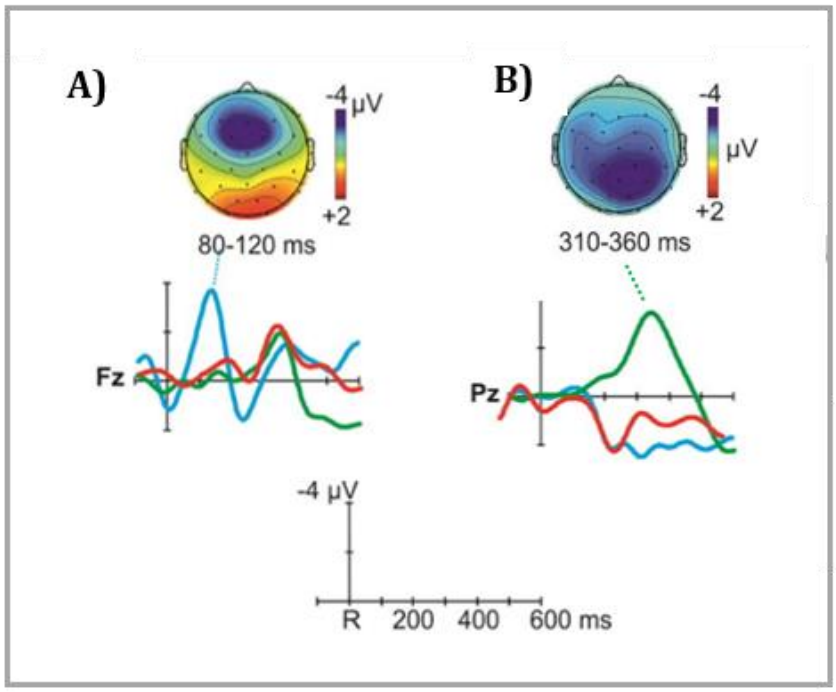


Figure 17. ERP results depicting the N400 component associated to false avatar errors. A) ERN component (red line) computed as the difference waveform for the contrast real errors minus correct responses. **B)** The N400 component (green line) computed as the difference in waveform for false (avatar) errors minus correct responses. *Adapted from Padrao et al. (2014).*

1.4.3.4. Reflecting on one's own performance: the anterior cingulate cortex (ACC), the premotor cortex (PMC) and the insula

Both the ACC and orbital frontal cortex (OFC) are two brain regions commonly involved in outcome evaluation, mainly implicating the **ACC** in action-outcome learned associations and the OFC in stimuli-outcome associations (Glascher et al., 2009; Izquierdo, Suda & Murray, 2004; Rudebeck et al., 2008). Specific groups of cells in the ACC have been shown to fire when an action is not followed by the expected reward, both in humans and monkeys (Shima & Tanji, 1998; Williams et al., 2004), as well as when they make errors (Yeung, Botvinick & Cohen, 2004). In accordance with theories of cognitive control [*Reinforcement learning perspective* (Holroyd & Coles, 2002; Holroyd et al., 2008)], whenever conflicting events, errors or negative feedback occur, the ACC may be responsible of alerting more top-down control regions, such as the DLPFC, to increase cognitive control and guide behavioural adjustments, probably by increasing attention to the task or informing other motor preparatory areas of the brain to make corrections or switch between action controllers (Botvinick, 2007; Holroyd & Coles, 2002; Kerns et al., 2004;

Ridderinkhoff et al., 2004). Furthermore, agency attribution mechanisms have been related with activity in the mPFC (Spengler, von Cramon & Brass, 2009), a brain region associated to higher-level, interpretative mechanisms, which might incorporate contextual knowledge and belief reasoning. Activity in the anterior PFC has also been associated with self-reflective processing (Arnodio & Frith, 2006), receiving input from subcortical structures related to the evaluation of self-relevant sensory stimuli. Also, the lateral PFC has been found to be relevant for conscious judgements about the self (Miele et al., 2011).

Furthermore, the ability to reflect on our own actions has been related to the workings of the **PMC.** For example, anatomic-clinical correlation studies of AHP indicate that a frontoparietal network in the right hemisphere, involving the PMC, plays a major role in the monitoring and awareness of left-sided motor deficits (Berti et al., 2005; Pia et al., 2004). In line with this hypothesis, it has been found that the PMC is involved in the computation of an expected sensory signal (Christensen et al., 2007), receiving many reafferent signals about the ongoing action (Hummelsheim et al., 1988).

The **insula** (Latin for 'island'), a brain region that lies folded deep within the lateral sulcus of each hemisphere hidden below parts of the frontal, parietal, and temporal lobes, has also been associated to self-monitoring/outcome processing. Farrer and Frith (2002) and Farrer et al. (2003) reported greater insula activation for self vs. other-agency attribution while other authors found insula-related activity when comparing active movement with synchronous feedback vs. passive movement synchronous feedback (Tsakiris et al., 2010). Interestingly, a unique feature of the insular cortex, especially the anterior insular cortex (aIC) of humans and a few other species (i.e., great apes, elephants, and some cetaceans), is the presence of clusters of large spindle-shaped neurons among the pyramidal neurons in layer 5, called von Economo neurons (Allman et al., 2005; Nimchinsky et al., 1999). While the precise function of these type of cells is not known, fMRI studies have shown that these types of neurons are selectively destroyed in disorders characterized by loss of emotional awareness and self-consciousness, such as frontotemporal dementia (Seeley et al., 2006; Sturm et al., 2006), schizophrenia (White et al., 2010), and autism (Minshew & Keller, 2010; Monk et al., 2009), which has led several researchers to postulate their involvement in empathy, self-awareness, and self-monitoring (Seeley et al., 2007). Moreover, dysfunction of the insular cortex and its interconnected regions are thought to be core features of several neurological disorders (Goodkind et al., 2015; Namkung, Kim & Sawa, 2017), such as patients with schizophrenia, AHP, major depression, and/or drug addiction, who show difficulties in self-monitoring as well as an abnormal SoA for their thoughts or actions (Eshel & Roiser, 2010; Karnath et al., 2005; Vocat et al., 2010; Ziauddeen & Murray, 2010).

1.4.4. Integrating agency, prediction and error monitoring

Agency attribution requires the monitoring of action-realted signals or cues (such as re-afferent sensory feedback) as they become available and comparing them with other relevant information for consistency (such as prior experiences or mental models) (Moore et al., 2009). As previously exposed on sections ***1.4.2.2. Prediction errors and the error-related negativity (ERN)*** and ***1.4.3. Outcome Monitoring,*** agency attribution processes might rely on comparisons between intended vs. predicted actions (internal error monitoring loop) and the predicted vs. actual states (external error monitoring loop) (see **Figure 11**). On one hand, the internal monitoring loop might be related to a rapid and rudimentary evaluation of evidence based on the subject's predictions and the processing of internal cues, probably associated to the subjective experience of fluently controlling the action (*FoA*) and giving rise to early modulations such as the ERN component. On the ther hand, the external monitoring loop might provide us with a more conscious agency attribution judgement (*JoA*), resulting from higher-order reflective inferences based on external cues and associated to later components such as the FRN/FCRP, P300 or the N400 (see **Figure 18**).

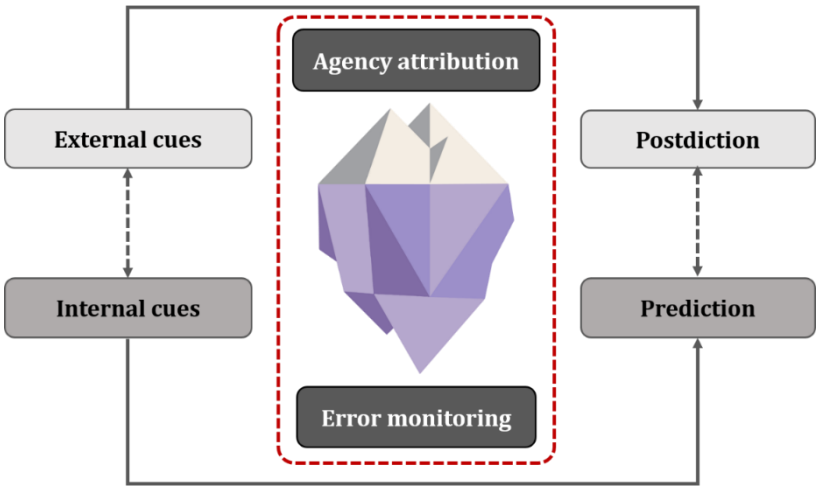


Figure 18. Relationship between agency attribution, prediction, and monitoring. Agency attribution relates to the error monitoring system by the constant comparisons between intended, predicted and actual states of the system we perform during voluntary movement. While the internal monitoring system ralies on one's own predicions and internal cues, the external monitoring loop may possibly depend on external cues and postdictive judgements.

Following this line of thought, the error monitoring system provides provide the appropriate link between the internal, prediction-based FoA and the external, postidictive-based JoA, contributing to the experience of agency and the perception of causality during voluntary action.

1.4.5. Anatomical dissociations for FoA vs. JoA

Following Synofzik et al. (2008a, 2008b) proposal on the distinction between different levels of agency, namely, FoA vs. JoA, we raise the question of whether it is possible to map this distinction in terms of their neural correlates. The **FoA** is conceptualized as an implicit, low-level, and pre-reflective experience, characterized by sensorimotor processes which usually runs outside awareness. In this way, the FoA might be related to brain regions associated to the comparator/monitoring mechanisms at a preconscious level, such as the cerebellum, PPC and TPJ (Blakemore & Sirigu, 2003; David et al., 2007) or visual association areas such as the EBA or posterior STS (David et al., 2007; Iacoboni et al., 2001; Leube et al., 2003b), which are not necessarily involved in the conscious detection of a sensorimotor mismatch (David et al., 2007). By contrast, the **JoA**, described by Synofzik et al. (2008a) as having a reflective, conscious, and conceptual nature, might be more related to brain regions such as the PFC, implicated in conflict monitoring and detection, between one's own intended action and the sensory outcome (Fink et al., 1999; Schnell et al., 2007; Slachevsky et al., 2001). Indeed, explicit JoA has been correlated with activity in DLPFC and mPFC (Spengler et al., 2009b), brain areas associated to higher-level processing, probably incorporating contextual knowledge and belief reasoning. Other neural correlates of the JoA have been found in the anterior PFC and the OFC (Miele et al., 2011), regions associated with self-reflective processing which receive inputs from subcortical structures, related to appraisal of self-relevant sensory stimuli (Arnodio & Frith, 2006), and from areas of the lateral PFC relevant for conscious judgements about the self (Miele et al., 2011).

1.4.6. A network perspective on SoA

Despite all the above evidence on the neural basis of the SoA, there is very little evidence on how these different neural correlates are functionally connected during the agentic processing. Increasing methodological and technological advances allow us to explore the neural underpinnings of the SoA by not only improving the knowledge of single brain regions associated to SoA processing but also to further understand the connectivity of these brain regions, especially for such complex cognitive processes as it is the case of the SoA.

Experiencing the SoA requires the interaction and integration of multiple sources of information generated at different brain regions and networks. But how do these brain sites communicate and interact with each other? A handful of studies have investigated the **functional connectivity** underlying the agentic experience. For example, David et al (2007) studied the SoA by asking participants to rate their sense of control over actions while providing congruent vs incongruent feedback about the action performed. Increased functional connectivity was observed between

the PPC, cerebellum and PMC regions when actions were correctly identified as being externally generated. On the other hand, increased connectivity between the insular cortex and the somatosensory cortex was observed when movements were correctly classified as coming from the self.

In another fMRI study by Nahab and colleagues (2011), the authors identified two main networks for SoA processing using a virtual-reality paradigm (i.e., *Cyberglove* recording voluntary sequential finger movements accompanied by real-time visual feedback). Firstly, a primary network consisted of the left anterior IPL, the right SMG, the right TPJ and the aIC. A second network entailed the CC, posterior IPL, and the PFC. While the former network would be engaged whenever a mismatch between the motor predictions and the sensory outcome was detected, the latter would be in charge of translating these A-O inconsistencies into a conscious experience of agency. These two networks likely represent a spatial and temporal flow of information, with the leading network serving the role of mismatch detection and the lagging network receiving this information and mediating its elevation to conscious awareness, giving rise to SoA (Nahab et al. 2011).

Ritterband-Rosenbaum, Nielsen and Christensen (2014) recently identified a PPC-pre-SMA network modulated by the experience of agency using dynamic causal modelling in an EEG experiment. When participants experienced SoA over their actions (i.e., congruent feedback with respect to cursor movement), a stronger gamma coupling from PPC to pre-SMA was encountered. This led the authors to propose that the PPC supplies the pre-SMA with information regarding the congruence between the sensorimotor/action and visual/outcome information once the action is performed, concluding that the attribution of agency (self vs other distinction) does not depend on the activity pattern of a single cortical area but on the coupled activity in a specific frequency band within this network in a retrospective manner. In the light of the *Comparator model*, the authors interpreted these findings as that the pre-SMA processes the intended outcome of the action, while the PPC would be responsible for the sensory integration of the visual and proprioceptive feedback.

Furthermore, Kang et al. (2015) recently used EEG power spectrum measures and phase coherence to estimate neuronal activity and functional connectivity during a virtual-reality paradigm (i.e., the *Cyberglove* experimental task). These authors reported that the alpha band was the most correlated frequency band with SoA modulation, particularly with anterior frontal regions, showing significantly larger desynchronization and lower functional connectivity as participants felt they could more effectively control the virtual hand movements. These results

suggest that the system might be more active when there is a disruption of the agentic experience. Moreover, these frontal regions were also functionally related to other cortical sites, supporting the notion that anterior frontal regions may be a sort of central hub in the generation of the SoA. Additionally, these authors also reported that the network communication of the alpha band in the anterior frontal area may be the main mechanism of SoA due to the parallel relationship between the phase synchronization and SoA modulation (Kang et al., 2015). Kang et al. (2015) also analysed the activity in the beta and gamma bands, however, their results suggest that these frequency bands might not be directly related to SoA as no significant correlations with SoA behavioural measurements (% control condition paradigm) were encountered. Therefore, they suggest that beta and gamma might not be directly related to the SoA but might have a role in general information processing (i.e., non-specific to SoA). Also, they proposed that these frequency bands might be associated with unconscious components of the SoA processing following recent distinctions between implicit/FoA vs conceptual/JoA (David, 2012; Synofzik et al., 2008a), probably associated with the non-conceptual, implicit FoA.

Finally, Buchholz et al. (2019) recently investigated the SoA from a network perspective, addressing changes in local oscillatory activity in magnetoencephalography (MEG) power spectra as well as connectivity changes as assessed by imaginary coherence between the motor cortex and the rest of the brain. A significant gamma power decrease in the right MTG and EBA related to SoA was reported, leading the authors to propose that this gamma-band activity might reflect more local feed-forward processing of bodily representations when participants aim to keep their behaviour synchronized to an external event. Moreover, a stronger functional connectivity to IPL and right MTG mediated by the beta band was encountered during the SoA condition, probably reflecting increased recurrent processing with motor cortices during SoA processing (Buchholz et al., 2019).

Interim summary

The SoA is very complex and ambiguous phenomenon. It has multiple contributories, some of which are consciously accessible and others which are pre-reflective and unconscious, such as outcome/feedback processing and predictive/forward processing, respectively. In this section I have explored the cognitive processes as well as the anatomical and functional neural bases underlying the experience of SoA. By means of fMRI and PET, several brain areas have been implicated in the SoA (see **Table 3**). These include brain regions known to be involved in the motor system such as the PMC, SMA and pre-SMA and the cerebellum. Moreover, other regions associated with self-processing and monitoring abilities such as the DLPFC, PPC or insula have

also been implicated in SoA processing. EEG has also proven very useful in the study of the temporal dynamics of SoA processing, allowing to distinguish between the different stages of pre-reflective/unconscious and reflective/conscious processing during the building-up of the agentic experience. For example, sensory attenuation signals, which allows us to distinguish self vs. external generated events, have been related to the elicitation of an N1 ERP component. Prediction errors, based on the comparison between intended and predicted states promoted by internal monitoring mechanisms have been shown to elicit the frontal ERN when mismatches at this level are detected. More posterior and conscious evaluation of outcomes have been associated to the FRN, fCRN, P3 and N4 components, associated to more reflective and retrospective judgements of agency attribution. Overall, the SoA seems to rely on a combination of bottom-up and top-down processes with dissociable neural and anatomical substrates, which require further theoretical and empirical investigation to achieve a more comprehensive mapping of its functional and structural functioning. Further exploration of the dynamic interactions between the different anatomical nodes and functional networks might aid in the construction of a comprehensive map underlying the experience of agency.

1.5. Experimental manipulations of the SoA

Many philosophers and cognitive scientists agree that our brain is not equipped to perceive causal relations directly from our environment but instead we infer them from the available cues and information (Hume, 1739/1888; Shanks, Holyoak & Medin, 1996). This commonality of properties is consistent with the view that the SoA is a product of the general determination of causality between an action and its effect (Buehner & Humphreys, 2009). Many experimental investigations have drawn on these principles by manipulating the consequences of subjects' actions, such as inducing temporal discrepancies, spatial deviations and content inconsistencies (see **Table 4** for an overview of experimental manipulations assessing the SoA). Based on this evidence, we designed our experimental paradigms presented in section **3. Empirical evidence.**

1.5.1. Temporal contiguity

One key factor for perceived causality is **temporal contiguity**. The longer the delay between two events, for example an A-O contingency, the less likely it is that the second event will be judged to be caused by the first one (Choi & Scholl, 2006; Greville & Buehner, 2010; Shanks, Pearson & Dickinson, 1987). Many behavioural and neuroimaging studies have reported a gradually decreasing SoA as delay increases (Ebert & Wegner, 2010; Farrer et al., 2008, 2013; Hon et al., 2013; Kawabe, 2013; Kühn et al., 2011; MacDonald & Paus, 2003; Sato & Yasuda, 2005; Tsakiris, Prabhu & Haggard, 2006; Wen, Yamashita & Asama, 2015).

In one of the first studies addressing SoA and temporal delay, Sato and Yasuda (2005) consistently showed a SoA reduction as A-O delay increased (see **Figure 19**). Throughout 3 consecutive experiments, these authors showed that the SoA, reported by the participants with the question "*I was the one who produced the tone*" [scale from 0-100 from "totally disagree" (score = 0) to "totally agree" (score = 100)], was modulated by the temporal delay between A-O. In Experiment 1, participants were instructed to make self-paced button presses, after which a certain tone (a 600 or 1000 Hz tone) was immediately presented for 200 ms through in-ear headphones. The tone presented could be either congruent (i.e., each button press evoked the same tone that had followed each button press in the learning session) or incongruent (i.e., a different tone from prediction followed each button press). Also, a delay condition was employed, during which the congruency was manipulated in terms of stimulus timing. The authors reported that the SoA was reduced when the presentation of the tone was unpredictable in terms of timing as well as its congruence (see **Figure 19A**). Following this experiment, in Experiment 2 participants were asked to press either the left or right button whenever a red or blue square appeared on the screen, respectively. Also, a specific tone was presented after the button press (a 600 or 1000 Hz

Table 4. Overview of experimental manipulations assessing the SoA.

Study	Feedback manipulation			Task
	Spatial	*Temporal*	*Content*	
Fourneret and Jeannerod (1998)	x			Computerized line drawing with spatially manipulated visual feedback.
Blakemore et al. (1998)	x	x		Self-controlled tickling with temporally and spatially delayed feedback.
Franck et al. (2001)	x	x		Virtual hand holding a joystick presented through a mirror superimposed on their real hand. Angular biases and temporal delays were introduced.
Haggard et al. (2002)	x			Judging onset of voluntary key presses inducing a finger twitch or tone.
Farrer et al. (2003)	x			Recognition of one's own limb from a spatially deviating alien hand.
MacDonald and Paus (2003)		x		Temporal delays in active/passive finger movements displayed as a virtual hand.
Knoblich and Kircher (2004)	x			Drawing task with varying visual feedback velocity.
Wegner et al. (2004)				Inducing a feeling of control over others' movements by prior instructions.
Tsakiris et al. (2005)			x	Self-/other-generated movements with manipulated visual feedback.
Sato and Yasuda (2005)	x			Button presses with associated auditory tones (congruent/incongruent).
Synofzik et al. (2006)	x			Pointing movements with spatially manipulated visual feedback.
Tsakiris et al. (2006)		x		Rubber hand illusion with active/passive finger movements, tactile stimulation & temporally delayed visual feedback.
Asai & Tanno (2007)	x	x		Control of a mouse device. Visual feedback biased either temporally or spatially.
Farrer et al. (2008)	x	x		Recognition of one's own movements from a spatially and temporally manipulated visual feedback.
Gentsch, Ullsperger and Ullsperger (2009)			x	External errors induced by lack of visual feedback during a Flanker paradigm.
Kannape et al. (2010)	x			Virtual spatial manipulation during locomotion.
Ebert and Wegner (2010)	x	x		Joystick pulling/pushing task to move an object visually displayed, congruent/incongruent feedback and variable temporal A-O delay.
Menzer et al. (2010)		x		Auditory paradigm inducing sensorimotor conflicts in free walking. Temporal delay of footstep sounds.
Steinhauser and Kiesel (2011)			x	External errors induced by blocking response button (i.e., lever).
Kannape & Blanke (2012)		x		Gait agency using life-size visual feedback of participants' ongoing locomotion manipulating temporal delay in virtual setting.
Hon et al. (2013)	x	x		Temporal delay and spatial incongruent feedback, up- or down-arrow key presses in response to visual stimuli (black dot).
Farrer et al. (2013)		x		Active/passive ball movement after button press with temporal delay.
Kawabe (2013)		x		Delay between a tactile vibration and a visual flash,
Wen, Yamashita and Asama (2015)		x		Key press in response to a moving dot inducing temporal delay. Self vs. external-control conditions
Imaizumi and Asai (2015)		x		Synchronous vs. asynchronous visual feedback during hand clasping with temporal delay.
Caspar et al. (2015)			x	Robotic hand performing congruent vs. incongruent (i.e., another finger) movement with respect to the participants' action.
Padrao et al. (2016)			x	Virtual reality scenario with avatar doing congruent vs. incongruent (i.e., hand to the left or right) movements.

tone). As in Experiment 1, the tone presented could be either congruent or incongruent with respect to the learning session, as well as temporally synchronized or delayed with respect to the button press. Additionally, a 'Other' condition was introduced, during which the tones were presented a certain ms after the presentation of target stimuli, irrespectively of the participant's response. Consistently with Experiment 1, participants misattributed self-generated tones to an external source when actual sensory feedbacks did not match predictions in terms of tone congruency as well as with increasing temporal delay. Moreover, during the 'Other' condition, high SoA scores were reported whenever the externally generated sensations happened to match the prediction made by forward model (**Figure 19B**).

Finally, during Experiment 3, participants were told to press either the left or right button whenever a "H" or a "N" appeared on the screen. Also, as it was the case for Experiments 1 and 2, tone congruency and temporal synchronicity were also manipulated. Additionally, after each trial, participants were asked whether they had committed an error or not in their preceding response. In this case, the SoA was reduced whenever there was a discrepancy between the predicted and actual sensory consequences (both during incongruent and delayed tone presentations), regardless of the presence or absence of a discrepancy between the intended and actual consequences of actions (correct or erroneous response) (**Figure 19C**). Taken together, the authors concluded that the SoA might mainly depend on a comparison between the predicted and actual consequences of actions, rather than on the comparison between intended and actual consequences (Sato & Yasuda, 2005). This is in line with the *Comparator account*, which postulates that for each action that is executed, a prediction of its sensory consequences is generated. When these predictions are compared with the actual consequences of an action, the larger the discrepancies, the less likely the action will be attributed as coming from the self (Frith et al., 2000). Interestingly, the SoA reports were affected in a similar way independently of whether the action was freely chosen (self-paced task –Exp. 1-) or performed in reaction to an external signal (reactive task –Exp. 2 and 3-). This finding has been interpreted as how the extent to which a person freely chooses an action does not modulate his/her experience of agency (Knoblich & Sebanz, 2005). However, it has been pointed out that the effects of the temporal delay were more pronounced when actions were performed in response to a stimulus than when actions were freely chosen (compare **Figures 19A** and **19B**, left). This finding may indicate that during stimulus-driven tasks, the agent might completely loose the SoA if there is a long delay between the action and its effect, whereas during 'active'/self-paced tasks, agents might be willing to wait longer for an action effect when they deliberately chose to produce it (Knoblich & Sebanz, 2005).

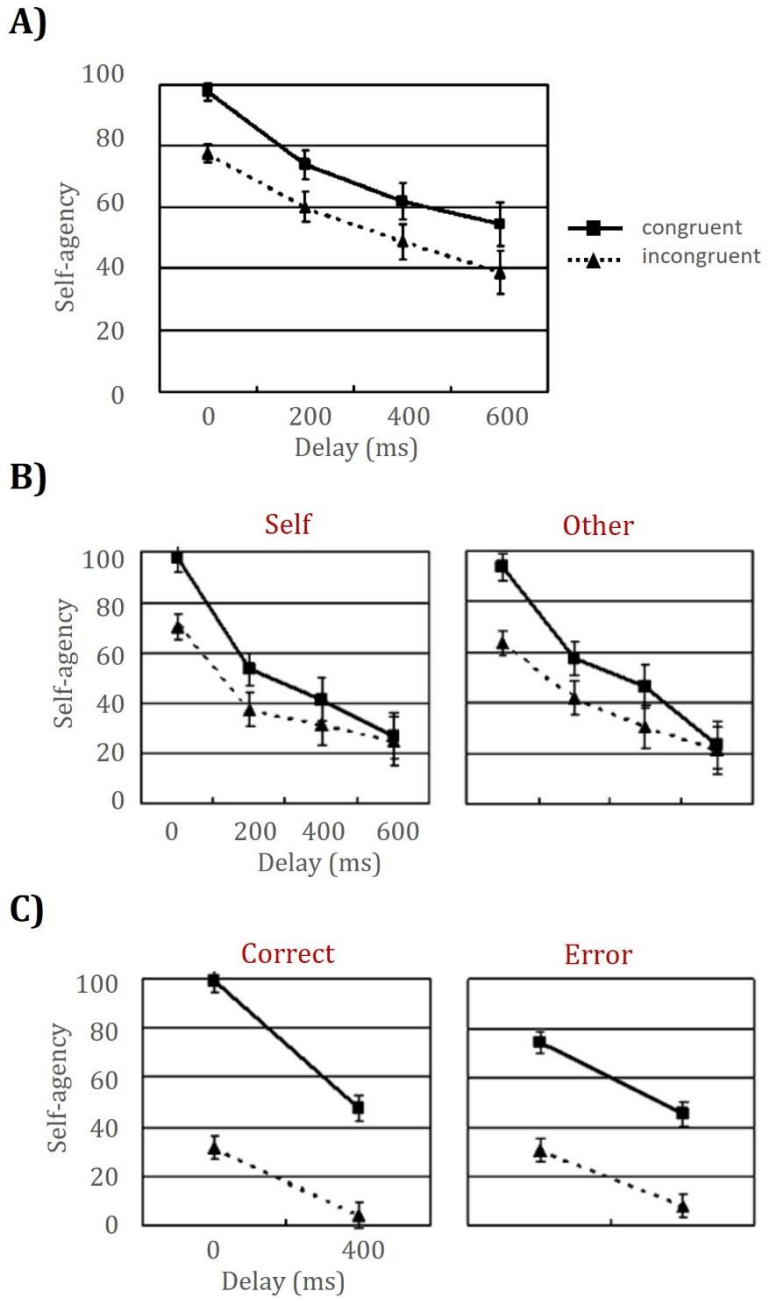


Figure 19. SoA reports from all 3 experiments by Sato and Yasuda (2005). A) Mean (± SEM) rating scores of self-agency during Experiment 1 showing a SoA reduction as a function of congruency and temporal delay during self-paced actions. **B)** Mean (± SEM) rating scores of self-agency during Experiment 2 confirming SoA reductions as a function of temporal delay during intended actions in response to stimuli. Additionally, participants seem to experience SoA even during externally generated sensations ('Other'), if the predictions made by forward model are congruent with the actual outcomes. **C)** In Experiment 3, the SoA was reduced whenever there was a discrepancy between the predicted-actual sensory consequences, independently of the intended-actual comparison (correct or erroneous response). Button-press and tone coupling could be either congruent or incongruent in relation to what they had previously learned on the training session. *Adapted from Sato and Yasuda (2005).*

More recently, Farrer et al. (2013) performed two experiments where a movement (i.e., button press) elicited an effect (i.e., ball appearing on a screen) after a varying delay, while also manipulating both internal cues (active vs. passive conditions -premotor information-) and external cues (contextual and prior information). Participants were asked to estimate, on each trial, what they thought happened, using three-choice responses: (1) Self/full control: "*my button press directly triggered the ball*"; (2) Delay/partial control: "*my button press triggered the ball but it appeared with a time lag*"; (3) Other/no-control: "*my button press did not trigger the ball, it's the computer that triggered it.*" In Experiment 1, participants were asked to press a button which triggered the appearance of a ball on the screen, at different temporal delays. Two tasks, an active and a passive task, were performed. These authors reported a significant decrease in SoA rating during the active task as delay increased, describing that participants gave a majority of 'full-control' responses when the ball occurred within a delay ranging between 0 and 334 ± 27 ms (mean and SEM), a majority of 'partial control' responses for intermediate delays (ranging between 334 and 708 ± 42 ms) and a majority of 'no-control' responses for longer delays (superior to 708 ± 42 ms) (**Figure 20**). When comparing active vs. passive conditions, the delay at which agency experiences shifted was increased during the active compared to the passive conditions, suggesting that premotor signals might have induced an extension of the time window, within which the effect can be combined with the action, in line with previous findings showing a reinforced SoA with premotor signals (Sato & Yasuda, 2005; Sato, 2009). This reinforcing effect of premotor signals could be explained by the higher reliability of internal cues compared to other agency cues (Sato, 2009), resulting in a larger weight of these signals in the SoA estimation (Synofzik, Vosgerau & Lindner, 2009; Synofzik et al., 2013), which could lead to an extension of the time windows for integration. It could be the case that because these high-reliable cues are maintained available longer for the integration of the action with the effect, participants could experience SoA (either full or partial) over a longer time period (Farrer et al., 2013). Moreover, when introducing further contextual information (i.e., a brief beep at the button-press onset), the temporal delay at which the SoA varied was shortened for both the 'full control' and 'no-control' responses.

Finally, during Experiment 2, participants were required to perform the same active and passive tasks but varying the degree of contingency between the participant's movement and its subsequent effect [condition 1: action and effect are present, with the ball appearing at variable delays; condition 2: the effect was absent, that is, the ball did not appear; condition 3: no action but with effect (i.e., the ball appeared without action present); condition 4: no action and no effect]. In this case, the authors replicated the results of Experiment 1 showing an effect of temporal delay on the SoA ratings, varying from full control to partial control and to no-control as

the delay increased, revealing longer delays at which the responses varied in the active task compared to the passive task. Nevertheless, manipulating the degree of action-effect contingency had no significant impact on the delays at which the responses varied.

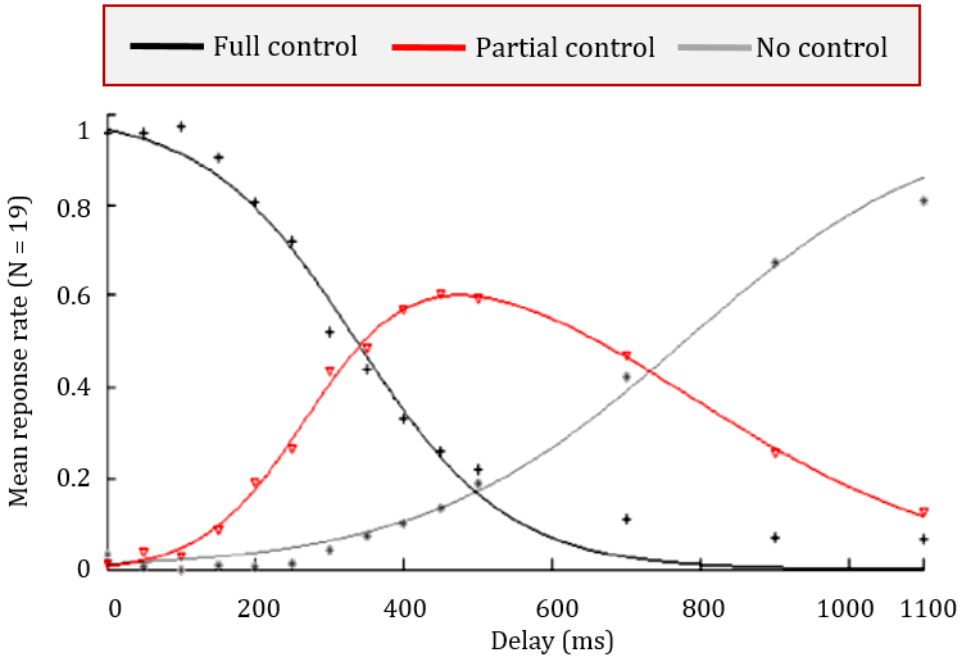


Figure 20. SoA ratings for Experiment 1 of Farrer, Valentin and Hupé (2013). Mean response rates (N = 19) for full-control (black), partial control (red), and no-control (grey) responses as a function of the delay (ms) in the active task for the no-sound session. Each data point represents a single measure of response rate (aggregated data of all subjects). For data analysis, curves were fitted independently for each subject and condition. *Adapted from Farrer et al. (2013).*

Furthermore, temporal delay has also been employed using auditory cues. For example, Menzer et al. (2010) used an auditory paradigm to induce online sensorimotor conflicts in free walking conditions, showing that confirmatory SoA judgments about gait (the percentage of "Yes" responses as to whether the walking that they heard over the headphones corresponded to the walking they had just performed), rapidly decreased for delays >120 ms reaching its minimum at 400–500 ms.

Nevertheless, several studies have reported conflicting results regarding the influence of temporal contiguity in SoA. For example, in a previous study by Farrer et al. (2008), the authors employed a device that allowed the introduction of A-O discrepancies between the movements performed

(i.e., joystick movement) and the sensory feedback (i.e., visual feedback displayed on a computer screen) in terms of space and time. Participants were asked to rate whether they were viewing 1) their own movements (Self), 2) their own movements modified (spatially or temporally displaced) (Bias), or 3) those of another agent (i.e., the experimenter) (Other). Interestingly, Farrer et al. (2008) found that even when the delay in visual feedback was at the maximum condition (1100 ms), people choose 'Bias' rather than 'Other' for the delayed visual feedback, providing strong evidence that people clearly know that they are the ones who caused the feedback even when there is a temporal discrepancy (up to 1100 ms) (Farrer et al., 2008, see **Figure 21**).

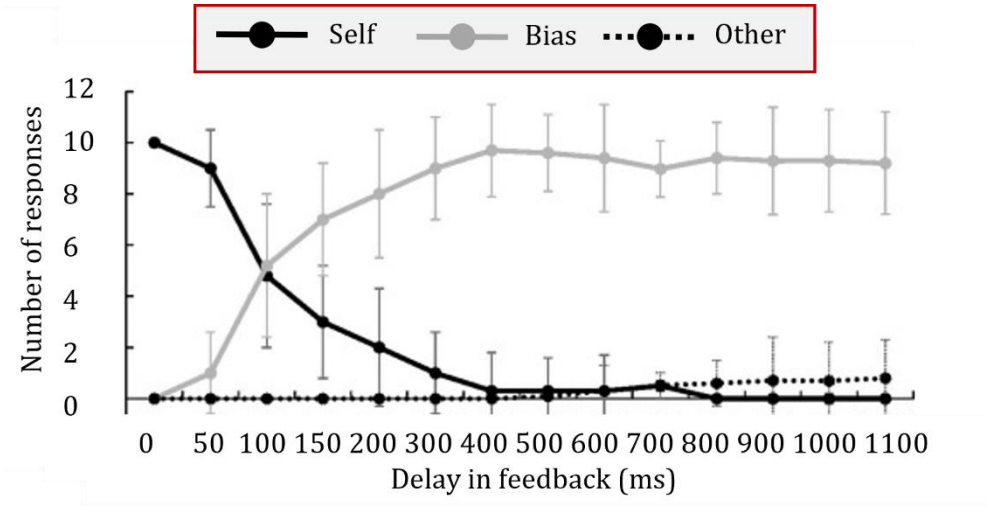


Figure 21. **Means and standard deviations of the number of 'Self', 'Bias', and 'Other' responses as a function of the temporal bias in Farrer et al. (2008).** Participants barely gave 'other' judgment even in the longest delay conditions (1100 ms). *Adapted from Wen (2019).*

Consequently, it has been proposed that temporal contiguity between A-O might be non-reliable due to the effect of other factors (Wen, 2019), such as task performance, strict time windows of internal models (e.g., we might be willing to wait several seconds after pressing an elevator call button for the elevator to arrive, but not after pressing a button to turn on a light), influence of other high-cognitive factors such as motor control, or even the way we ask participants to classify several conditions into limited categories (i.e., we tend to give different responses for different conditions, and a lower agency rating is easily associated with a longer delay).

1.5.2. Spatial consistency

In the spatial dimension, when the visual outcome of actions is modified by generating **spatial distortions** such as angular deviations, the SoA seems to diminish significantly (Blakemore et al., 1998, 1999b; Ebert & Wegner, 2010; Farrer et al., 2003; Fourneret & Jeannerod, 1998; Haggard et al., 2002; Hon et al., 2013; Kannape et al., 2010; Knoblich & Kircher, 2004; Nielsen, 1963; Tsakiris et al., 2005; Synofzik et al., 2006).

In the pioneering work by T. Nielsen (1963), the author developed an experimental setting that allowed to present the participants with real-time feedback about the action performed and distort it so that it could be either congruent or incongruent with the subjects' actual movement. In his classical experiment, subjects were asked to draw a line on a piece of paper. They could either see their own hand or, unknowingly to the subject, an ''alien hand" (i.e., the experimenter's hand). The alien hand's movements spatially deviated from the subject's own movement. Interestingly, subjects adjusted their own actual movement to the false visual feedback without being aware of the adjustment. Relatedly, in the study of Farrer et al. (2003), a task consisting of executing a series of simple movements with a joystick was employed using a modified version of the Nielsen paradigm. Spatial congruency was manipulated by altering the visual feedback displayed on a computer screen introducing several angular biases (5°, 10°, 15°, 20°, 30°, 40°, and 50°) (see **Figure 22A**). After each trial, participants were asked if the movements they visualised on the mirror were concordant with the ones they had executed. Their results showed that the threshold value below which the subjects gave more than 50% of "No" responses (i.e., '*The movement displayed is not my movement'*) was found to be located between biases of 10° and 15° (**Figure 22A**).

Furthermore, these results were replicated a few years later by Farrer and colleagues (2008), who confirmed that by modifying the apparent direction of the movement performed introducing spatial incongruences resulted in the attribution of one's own actions to another person ("Other" responses) (see **Figure 22B**), highlighting the influence of the visual sensory feedback processing in the SoA attribution.

Interestingly, several versions of the Nielsen paradigm have been employed to examine self-recognition abilities in patients suffering of schizophrenia, with and without delusions of control (Daprati et la., 1997; Franck et al., 2001; Posada et al., 2007). For example, in the study by Franck et al. (2001), participants had to execute a joystick movement in the direction indicated by the position of a green spot while introducing visual feedback which could either be congruent, temporally delayed or spatially distorted (angular bias). In the case of angular biases,

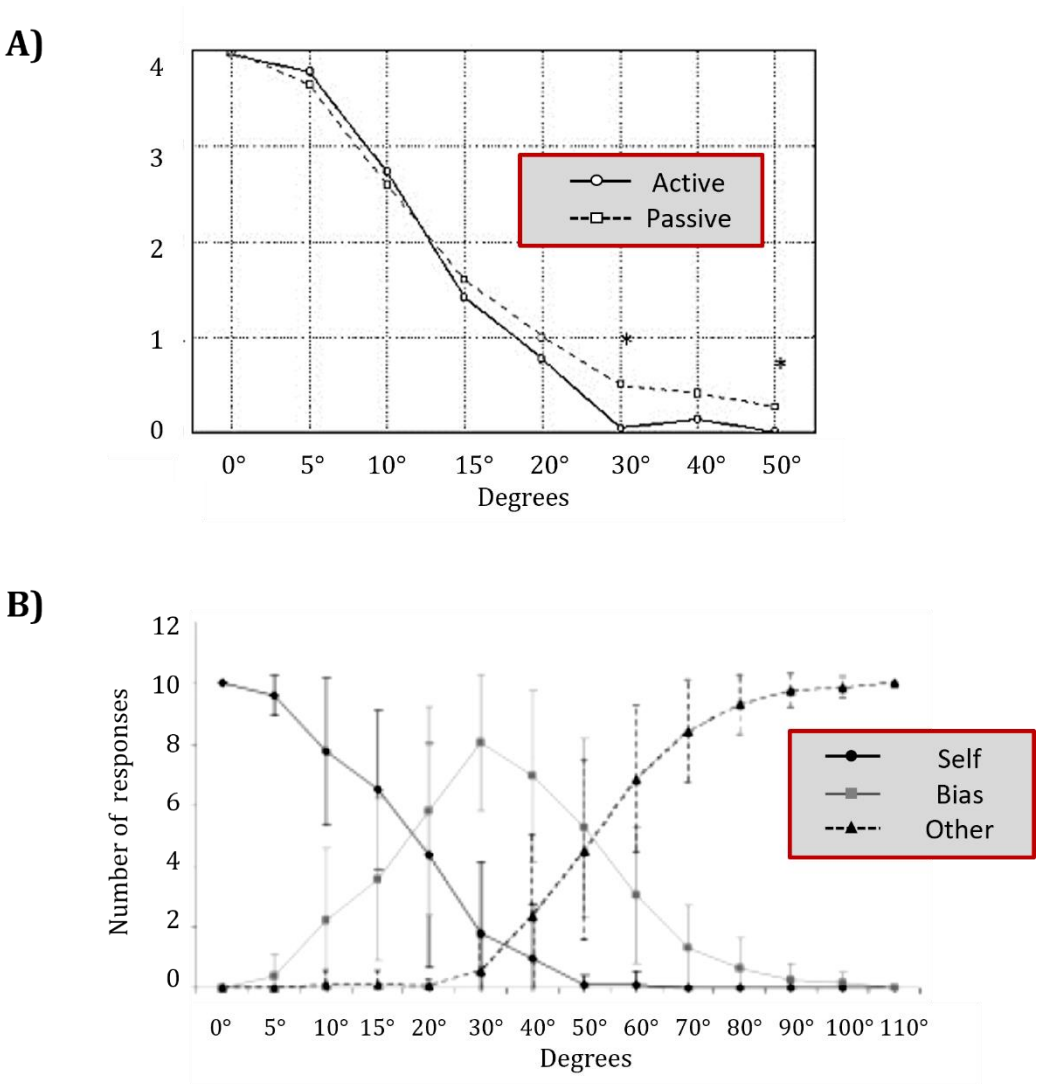


Figure 22. Modulation of SoA with angular bias. A) Number of "Yes" responses ('The movement displayed is my movement') as a function of angular bias in the active and in the passive conditions. *Adapted from Farrer et al. (2003).* **B)** Means and standard deviations of the number of "Self" responses, "Bias" responses and "Other" responses as a function of the angular bias. *Adapted from Farrer et al. (2008).*

schizophrenic patients suffering from delusions of influence gave globally more "Yes" responses than schizophrenic patients without delusions of influence or healthy subjects (see **Figure 23**), showing clear difficulties in detecting that the deviated movements displayed were distinct from their own movements.

These findings have been interpreted as resulting from difficulties with the *Comparator* mechanisms in the formation of the efference copy/predictive signals, causing a lack of awareness

of certain aspects of motor control derived from these internal models (Blakemore et al., 2002; Frith, 1992).

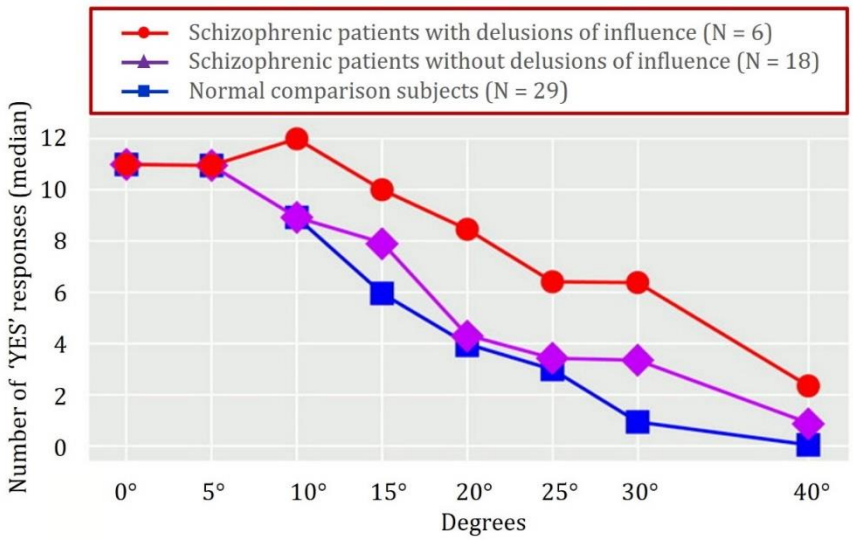


Figure 23. **Self-recognition in schizophrenic patients.** Number of "Yes" responses by schizophrenic patients with and without delusions of influence and normal comparison subjects when asked whether the hand movements displayed on a computer screen matched their own hand movements. *Adapted from Franck et al. (2008).*

More recently, a virtual reality (VR) study using full-body tracking was performed to investigate how agents consciously monitor actions of the entire body in space during locomotion (Kannape et al., 2010). While fully immersed in a virtual scenario, participants were instructed to move their virtual body to a virtual target location. In some trials, the walking trajectory of the virtual body was systematically deviated either towards the left or the right (by 5°, 10°, 15°, or 30°). In line with previous evidence (Farrer et al., 2003; Franck et al., 2001; Jeannerod & Pacherie, 2004; Posada et al., 2007; Slachevsky et al., 2001; Synofzik et al., 2006), participants only detected angular biases when surpassing a 10°–15° threshold, at which they become conscious and walking parameters may be adapted through conscious control

Moreover, spatial manipulation of auditory cues (i.e., tapping sounds) were recently employed by Tajadura-Jiménez et al. (2015). By altering the sound produced by tapping on a surface with one's hand, as though they were occurring from double or quadruple the distance of the origin of the actual sound, resulted in a diminished feeling of agency (i.e., the sounds are coming from my

tapping) [see Stanton and Spence (2020) for a review on the influence of auditory cues on movement perception].

1.5.3. Content congruency

Another interesting manipulation addressing A-O coupling is the insertion of **content inconsistencies**. For example, Tsakiris et al. (2005) addressed the contribution of efferent information in self vs. other distinction designing a task where participants experienced a passive extension of the right index finger, either as an effect of moving their left hand via a lever ('self-generated action') or imposed externally by the experimenter ('externally-generated action'). The visual feedback was also manipulated so that subjects saw either their own right hand ('view own hand' condition) or someone else's right hand ('view other's hand' condition). Their results evidenced that participants were more accurate in correctly recognizing their own hand when their index finger movement was self-generated compared to when it was externally generated. When the action was externally generated, subjects incorrectly attributed the experimenter's hand to themselves in 55% of the trials (for the self-generated action, incorrect attribution to the self occurred in 38% of the trials). This difference might indicate that efferent information makes a specific contribution to self-recognition, probably facilitating the generation of sensorimotor predictions in the forward modelling improving sensory processing and facilitating spatial precision and temporal discrimination (Wolpert, 1997). Interestingly, subjects tended to misattribute the experimenter's hand to themselves, as previously reported in several studies (Daprati et al., 1997; Sirigu et al., 1999; Van den Bos & Jeannerod, 2002), a fact which might indicate that self-attribution could be a default mode of attribution, when no clear cues for self-recognition are available (Van den Bos & Jeannerod, 2002).

Experimental conditions inducing external error by omitting the occurrence of feedback (Gentsch, Ullsperger & Ullsperger, 2009) or blocking the response buttons (Steinhauser & Kiesel, 2011) have resulted in modulations of the causal attribution of actions. In a recent VR study, participants were embodied into an avatar body in 1PP while performing an error-prone reaction time task (Padrao et al., 2016). Interestingly, externally generated errors (i.e., incongruent hand movements performed by the avatar) resulted in SoA disruption reporting significantly lower scores on several SoA items (e.g., "*The movements of the virtual hand seemed to be my movements*").

Interim summary

As agents, we are continuously learning the contingent relationships between our actions and their sensory effects on the world surrounding us throughout repeated experience, allowing us to feel the SoA for the actions performed. Experimental manipulation of these A-O contingencies has

been found to reduce the experienced SoA, allowing to investigate the interaction between internal and external cues. Both temporal and spatial modulations have significant effects in the SoA, as well as the introduction of incongruent feedback. Importantly, both spatial consistency and the insertion of incongruent outcomes seem to be more disruptive of the agentive experience, while the temporal aspect does not seem to produce comparable effects. While a very large temporal delay is needed to ascribe an action as not coming from the self [i.e., 1100 ms in Farrer et al. (2008)], a small spatial deviation [(i.e., 10-15° angular bias in Farrer et al. (2008)] or the presence of a clear content incongruence in the feedback (i.e., external error in Padrao et al. (2016)] is enough to induce an external agency attribution. It has been proposed that this difference might be related to the perception of the goal in the movement, preserved during A-O temporal delays but altered for A-O spatial inconsistencies or content incongruences, leading one to attribute the movement as coming from another person, as the perception of the final goal would no longer correspond to the goal of the intended action (Farrer et al., 2008). Generally, empirical evidence points out that detection of discrepancies based on these signals is quite different between aspects, probably relying on different comparisons and computations depending on the nature of the distortion. For example, sensitivity to alterations in the temporal aspect might rely on comparison of onset timing relations between visual and sensorimotor signals, related to parietal and cerebellar regions (Blakemore & Sirigu, 2003; Farrer et al., 2008; Leube et al., 2003b; MacDonald & Paus, 2003; Salomon, Malach, & Lamy, 2009), while detecting discrepancies in the spatial aspects may require taking into account continuous action plans and somatotopic representations associated with the activity of frontal and parietal regions (David et al., 2007; Farrer et al., 2003).

RESEARCH AIMS

2. RESEARCH AIMS

The research questions and the empirical studies of this thesis are based on the theoretical framework and experimental evidence outlined in the previous chapter. In recent years, there has been a significant amount of research and discussion about the concept of agency from several perspectives, including psychology, psychopathology, and neuroscience. Most neuroscientific evidence has been derived from the predominant account on explaining the SoA, that is, the *Comparator model* (Blakemore et al., 1998, 2002; Frith, 1992; Kawato, 1999; von Holst & Mittelstaedt, 1950; Wolpert et al., 1995), which suggests that the SoA strongly depends on the degree of congruence between the predicted and the actual sensory outcome. Based on this background, the goal of the present dissertation was to **investigate the electrophysiological signatures underlying agency attribution processing by manipulating action-outcome (A-O) congruencies inducing agency prediction errors (aPE),** defined as a mismatch between the predicted and the actual sensory consequence of a movement. To this end, we developed and tested a series of modified well-known experimental paradigms to unveil the neural activity involved in the processing of agency attribution mechanisms. From this starting point, the specific research aims of each study were the following:

The **first aim** was to **(1) investigate the different neural mechanisms underlying the internal vs. external attribution of agency by modulating visuo-motor congruencies [Study 1, Gomez-Andres et al. (2022a)].** In this study, we addressed the influence of aPE for the self in action via a novel set-up allowing the illusion of ownership and SoA over a pair of digital hands. Error processing is an important aspect of learning, as it allows us to compensate and adapt our behaviour. To do so, it is necessary to distinguish between self-performed and externally induced erroneous actions. During self-generated actions, intentions and efferent information not only predict the consequent multisensory signals produced by our movements (Wolpert & Miall, 1996), but also modulate their perception and the agentic experience (Tsakiris & Haggard, 2005). While efferent signals have been usually implicated in the unconscious function of internal models of the motor system, responsible for motor learning, motor prediction and motor correction, afferent information provides us with the specific content of our bodily self-awareness (Wolpert & Miall, 1996). In the present study, we performed three ERP experiments inducing aPE in terms of content consistency (i.e., we induced externally generated errors in the visual domain providing incongruent outcome feedback -Experiments 1 and 3-) and temporal synchronicity (i.e., introducing A-O delays -Experiments 2 and 3-). Relying on the precise timing of EEG, we aimed to provide new evidence on the ERPs and oscillatory activity underlying the computation and subsequent attribution of agency, distinguishing between self-internally attributed and other-

externally attributed agency for the actions performed. Based on the premise that outcome-related prediction errors influence agentic inferences, we hypothesized that when inducing content or temporal aPE, a modulation of the attribution of agency would be experienced by the participants, leading to an increase in external agency attribution reports associated to specific neural markers.

The **second aim** of this thesis was to **(2) evaluate the feasibility of employing a Support Vector Machine (SVM) classifier to accurately disentangle self vs. externally generated actions from the EEG signal at a single-trial level [Study 2, Gomez-Andres et al. (2022b)]**. To this end, we examined our time-series data from Experiment 1 (Study 1) and applied a computerized linear classifier to investigate whether we could decode the different defined conditions (Correct vs. self-errors vs. external errors) at the single-trial level. Previous literature has highlighted that whenever a person makes or perceives an error, a series of error-related potentials (ErrPs) can be detected in EEG signal, such as the ERN component (Falkenstein et al., 2000). Recently, ErrPs have gained a lot of interest for the use in BCI applications, which give the user the ability to communicate by means of decoding his/her brain activity. Several studies have reported that when a BCI delivers erroneous feedback, ErrPs can be detected at the single-trial level (Chavarriaga, Sobolewski & Millán, 2014; Iturrate et al., 2015; Kim et al., 2017; Usama et al., 2021; Zander et al., 2016), allowing the online correction of errors (Schmidt et al., 2012; Spüler et al., 2012) and/or to improve BCI adaptations (Llera et al., 2011; Spüler et al., 2012). In the present study, the implementation of a classification algorithm allowed us to test the viability of correctly disentangling self/internal vs. other/external agency-error attributions at different stages of brain processing based on the latency, amplitude and spatial topographical distribution of key ErrP features, namely, the ERN and P600.

Finally, the **third aim** of this thesis consisted in **(3) investigating the role of the insular cortex in the monitoring of self-generated actions in patients undergoing electrical stimulation mapping (ESM) during awake brain surgery [Study 3, Gomez-Andres et al. (in press)]**. More specifically, we wanted to examine the involvement of the aIC region in the detection and awareness of A-O incongruences during the realization of a modified version of the Stroop task (Stroop, 1935). Cortical ESM is an intrasurgical technique developed to detect functional areas within the exposed, living human brain cortex (Ojemann et al. 2008). Its main goal was to provide an anatomical guideline by delineating any eloquent cortex and its relationship with the lesioned tissue. During cortical ESM, a mild electrical current is applied to small patches of the patients' cortical tissue simultaneously to the patients' sensory, motor and cognitive functions. In the present study, a patient with a fronto-insular astrocytoma underwent awake brain surgery with

ESM. Amongst other regions, the aIC was electrically stimulated while the patient performed a modified Stroop paradigm assessing her ability to correctly detect A-O discrepancies in terms of congruent/incongruent feedback. To do so, the ESM was synchronously applied to the feedback presentation, followed by a test phase where the patient was asked to report if the feedback presented was either congruent or incongruent with respect to her performance. Based on previous evidence, we expected to elicit a disruption in performance monitoring mechanisms when applying ESM over the aIC region by interfering with the ability to correctly detect incongruences between the action performed by the patient and the subsequent feedback appearing on the screen. Moreover, we hypothesized that the fMRI analysis would show an activation pattern involving the aIC region when contrasting Incongruent vs. Congruent feedback during the realization of an adapted version of the experimental Stroop task inside the scanner.

EMPIRICAL STUDIES

STUDY 1

The complex nature of agency attribution: Neurophysiological signatures associated to monitoring self vs. external erroneous actions

3. EMPIRICAL STUDIES

3.1. STUDY 1

The complex nature of agency attribution: Neurophysiological signatures associated to monitoring self vs. external erroneous actions.

3.1.1. INTRODUCTION

In our daily-life experience, we usually do not have trouble distinguishing our actions from those of others. Recent work on the sensory-motor basis of the Self distinguishes two key aspects of self-awareness: (i) the sense of ownership (SoO), by which we recognize our body as our own and as the basis of our experience with the world (Gallagher, 2000) and (ii) the sense of agency (SoA), the feeling of being the author of one's own actions and their effects (Gallagher, 2000). This experience of being the agent is a very fluid and non-conscious ongoing flow of anticipations and sensory feedback (Pacherie, 2001) about which we practically never have any doubt. Indeed, this positive agency attribution might index a default state of being in control of our actions (Haggard & Chambon, 2007), only made accessible to consciousness under specific meta-cognitive interrogation (Yeung & Summerfield, 2012) or after the elicitation of agency prediction errors (aPE) (i.e., mismatch between the predicted and the actual sensory consequence of a movement).

When explaining the elicitation of aPE, classic accounts rely on the comparator model (Blakemore, Frith & Wolpert, 1999; Wolpert, Ghahramani & Jordan, 1995), which suggests that agency attribution basically arises from internal cueing determined by the comparison between the sensory predicted outcome (i.e., efference copy) and the actual external consequences. From this perspective, a default internal agency attribution is usually present, only disturbed whenever a mismatch between the intended/predicted vs. actual outcome occurs, leading to an aPE and to an external attribution of agency for the action. As the forward mechanism allows us to make very precise predictions on the expected action outcomes, any evidence of content inconsistency or temporal asynchrony might elicit an aPE. But this account has encountered difficulties explaining important evidence: (a) internal agency attribution seems resistant even when long delays (e.g., 1100 ms) are interposed between our actions and their outcomes (A-O) (Farrer et al., 2008, 2013), (b) we tend to overestimate our internal attribution even when we are not the actual agents responsible for the actions observed (Wegner and Wheatley, 1999) and (c) it cannot easily accommodate clinical cases reporting agency disturbances in presence of preserved efference signals (Fourneret et al., 2002; Farrer et al., 2008; Farrer, Valentin & Hupe, 2013). To account for these objections and the complex nature of agency, alternative theories emphasize the

involvement of higher-level cognitive processes that evaluate the effect of outcome-related information without the need to access internal predictive signals (Wegner & Wheatley, 1999). Synofzik and colleagues (2008) proposed an integrative two-step model, including a lower-level sensorimotor predictive process, giving rise to the feeling of agency (FoA), built upon the processing of internal cues, and a higher-order, postdictive belief-based mechanism supporting judgements of agency, influenced by internal and external cues. From this integrative perspective, the attribution of agency can be interpreted as posterior according to the Bayesian cue integration framework (Moore & Fletcher, 2012), determined by the weighting and reliability of both internal and external cues.

One of the main problems on arbitrating between these different accounts is that most of the behavioral measures employed are unable to differentiate between predictive (internal) and postdictive (external) components that contribute to the emergence of agency. Furthermore, fMRI studies are unable to disentangle the specific functional role of certain regions commonly observed during agency attribution [e.g., angular gyrus vs. pre-motor/supplementary motor area (SMA)/dorsolateral prefrontal cortex (DLPFC)], mostly due to a lack of temporal resolution (Chambon et al., 2013; Chambon, Moore & Haggard, 2015; Desmurget & Sirigu, 2009; Sperduti et al., 2011). In the present study, relying on the precise timing of electroencephalography (EEG), and the fact that predictive and postdictive signals become available to the cognitive system at different time points, we aimed to provide new evidence on the neurophysiological basis implicated in agency attribution by inducing aPE using a modified version of a Flanker task (Padrao et al., 2016). We measured ERPs and oscillatory activity across three experiments (**Table 1**) inducing aPE in terms of A-O content consistency (Experiment 1 and 3) and A-O temporal asynchrony (Experiments 2 and 3) while evaluating agency attribution as well as the SoO. Along the experiments, we inquired participants on different aspects of agency attribution, namely, self-attribution (hereafter referred to as internal attribution) and other attribution (i.e., external attribution) regarding their actions and outcomes.

3.1.2. MATERIALS AND METHODS

Participants

Right-handed graduate students were paid to participate in the study. Some participants were excluded due to artefact removal and/or a low rate of errors (Experiment 1: 2 exclusions; Experiment 2: 3 exclusions; Experiment 3: 2 exclusions). The final samples for each experiment were as follows (mean ± SD): i) Experiment 1: n = 25 (15 females, M = 24.3 years ± 4.31, ii)

Experiment 2: n = 24 (16 females, M = 23.5 years ± 4.21) and iii) Experiment 3: n = 18 (12 females, M = 23.9 years ± 4.12).

In all cases, participants were naïve with respect to the aims of the experiment. The procedures of the experiment were approved by the Biomedical Research Institute of Bellvitge (IDIBELL) and Hospital Universitari de Bellvitge ethics committee (CEIC, Ref. PR254/15). Informed consent in accordance with the Declaration of Helsinki was obtained from all participants prior to the commencement of the study.

Table 1. Summary of all three experiments and their resulting neurophysiological correlates as a function of outcome congruency and synchronicity with the action performed. In light grey we highlight the experimental conditions where an external action attribution was induced. Predicted ratings for Experiment 3 regarding subjective reports are shown based on our previous results on Experiment 1 which depicted the same electrophysiological results.

	Experimental condition		aPE	ERP signatures	Oscillatory activity	
					Delta (1-3 Hz)	Theta (3-8 Hz)
Experiment 1 **CONTENT aPE**	SE		∅	ERN		↑
	EE		Content	N2/P600	↑	↑
Experiment 2 **TEMPORAL aPE**	No-Delay A-O		∅	ERN		↑
	150-Delay A-O		Temporal	ERN		↑
	400-Delay A-O	Correct	Temporal	FCRP	↑	
		SE		ERN/FRN		↑
Experiment 3 **DELAYED CONTENT aPE**	Synchronous A-O	SE	∅	ERN		↑
		EE	Content	N2/P600	↑	↑
	Asynchronous A-O	SE	∅	ERN		↑
		EE	Content + Temporal	N2/P600	↑	↑

aPE: agency Prediction Error; ERP: Event-related brain potentials; SE: Self errors; EE: External errors; A-O: Action-outcome; ERN: Error-related negativity; FCRP: Feedback correct-related positivity; FRN: Feedback related negativity. ∅: No aPE

A**pparatus**

An apparatus inspired by Nielsen's (1963) paradigm and the Rubber Hand Illusion (Kalckert & Ehrsson, 2012, 2014; Krämer et al., 2007) was built, a representation of which is shown in **Figure 1**. Participants were seated comfortably inside a Faraday chamber. A full HD 24.5-inch monitor displaying the experimental task at 200 Hz refresh rate was mounted on a wooden stand and adjusted to the subjects' body. The monitor was oriented with an inclination of 30º on the

horizontal plane. The participants were asked to put on a pair of white latex gloves and to place their hands on top of the wooden stand surface (hidden from their view due to the monitor overlap). Two fixed joysticks with a button at the top were attached to the wooden stand so that he/she could grab the joysticks and push the button with his/her thumbs (see **Figure 1A**). After general instructions were given, the EEG cap was set up and the state of each electrode was checked. Finally, the room lights were turned off for the realization of the experiment.

Experimental design

A modified version of the Eriksen Flanker task (Rodriguez-Fornells, Kurzbuch & Münte, 2002) was presented (**Figure 1B**). Stimulus presentation was controlled with EPrime (Psychology Software Tools Inc., Pittsburgh, PA) and MATLAB software version R2017a (The MathWorks, Inc. Natick, MA). Participants were required to focus on the central target arrow from a visual array of three arrows allocated vertically and to respond with the right or left hand (thumb press on button placed on the high-end of the fixed joystick) depending on the directionality of the target arrow. The flanker arrows located above and below the target arrow were presented to either facilitate the target response (compatible trials: all arrows in the same direction) or to prime the other response pattern (incompatible trials: flanker arrows pointing in the opposite direction than the target arrow). To optimize the number of errors, a proportion of 40-60% of compatible and incompatible trials, respectively, were presented in a random order. During the whole duration of a trial, a pair of life-sized hands mimicked the participants' hands actions (closed hands holding the response device and performing the thumb button press movement) at a coherent position with respect to the participants' hands and body posture. To remove morphological cues that could affect self-identification, the 'virtual' hands were real hand (adult size) photographs wearing the same white latex gloves that participants were asked to wear (see **Figure 1**).

Participants were instructed to respond both accurately and as fast as possible to the direction of the target arrow. No performance feedback (correct/error) was provided throughout the task. The duration of the target presentation for all trials was fixed to 150 ms followed by a response threshold of 1000 ms (Reaction Time -RT-). At the same time participants were responding, the virtual hand response movement -Observed Response- appeared for 100 ms (see **Figure 1B**). A variable fixation slide (depending on the RT) appeared at the end of the trial. Participants were given instructions about how to reduce muscle artifacts by minimizing movement and to wait for a visual signal, an array of five asterisks appearing every 10 trials, to free blink for 5 seconds.

Every 80 trials a block pause of 10 seconds was presented. A training phase (20 trials) was always performed before the experiment began to ensure an adequate speed-accuracy rate.

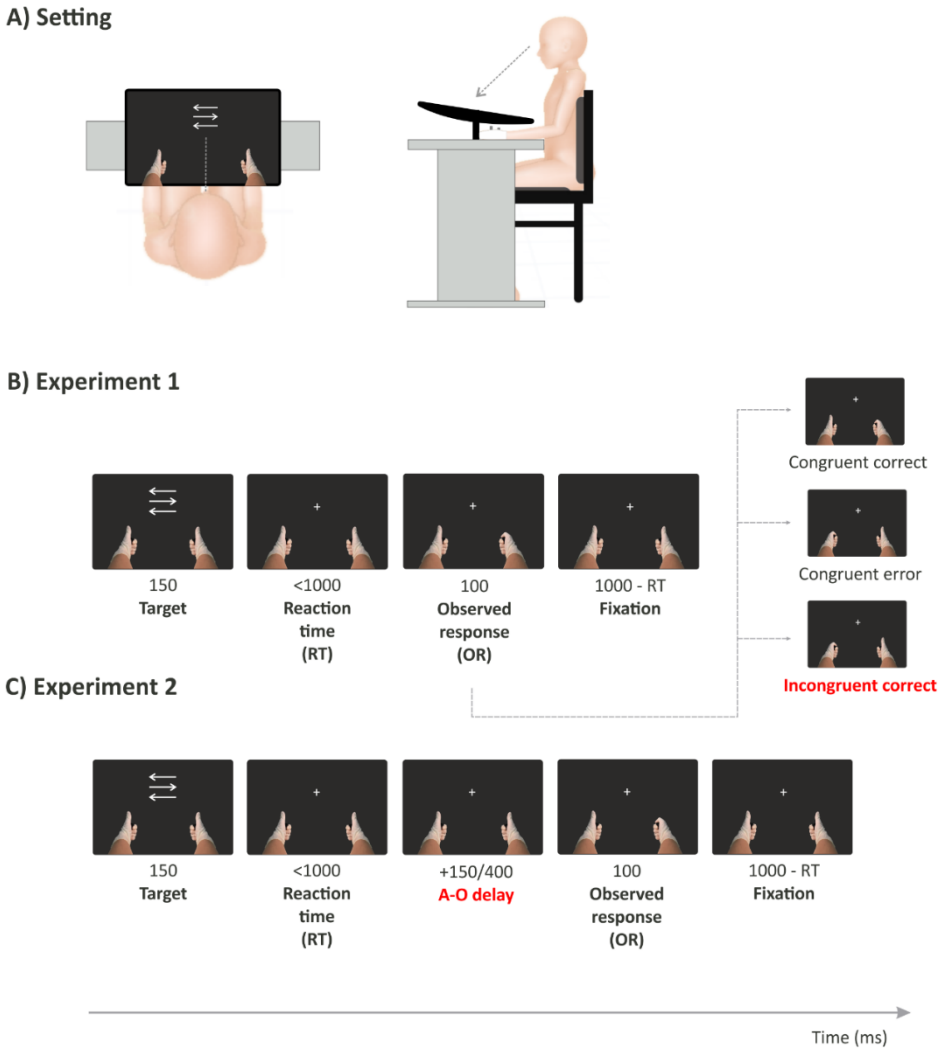


Figure 1. Apparatus and experimental paradigm. A) Schematic representation of the experimental apparatus. Bird's eye view (left) of the experimental setting showing the modified version of the Eriksen Flanker task (Padrao et al., 2016; Rodriguez-Fornells et al., 2002) and the visual feedback display (hand images). Lateral perspective (right) of the participants' position with respect to the wooden-mounted monitor. The participant's hands are hidden underneath the monitor, in a coherent position with respect to the 'digital' hand displayed, holding the response joysticks. **B)** Experimental paradigm depicting a modified version of the Eriksen Flanker task (Padrao et al., 2016; Rodriguez-Fornells et al., 2002). All trials started with the target presentation (150 ms) followed by a response threshold (<1000) in which participants had to respond to the target arrow as fast as possible (RT: Reaction Times). The visual feedback (Observed Response -OR-) corresponding to the virtual hand response movement was displayed for 100 ms, showing either congruent correct, congruent error or incongruent correct visual displays. **C)** Experimental manipulation corresponding to Experiment 2 including an action-outcome delay (A-O delay) between the actual response and the observed response of 150 and 400 ms.

Procedure and experimental aims

Experiment 1. Content aPE. Experiment 1 was designed to examine the electrophysiological and oscillatory activity associated to content aPE and their effects on the different levels of agency attribution processing (internal vs. external agency attribution) (see **Table 1**). Following Padrao et al. (2016), the experiment always started with one Standard block (160 trials) during which a complete spatial and temporal congruence between A-O was presented allowing the illusion of ownership and a coherent internal attribution for the perceived movements without the need of being fully immersed in a virtual scenario (Padrao et al., 2016). Following this first Standard block, two Error induction blocks of 640 trials each were performed. During the Error induction blocks, and only in some compatible trials (5% of the total number of trials, i.e., 64 trials in total for all Error induction blocks), an A-O content incongruence was inserted during which the 'virtual' hands produced an unexpected response movement inducing an external error (EE) and therefore, a content aPE (see **Figure 1B).** Thus, in those EE trials, when the participants pressed the button with one hand, it was the opposite 'virtual' hand which performed the response movement, causing an aPE. We avoided introducing EE in incompatible trials to avoid the pre-activation of incorrect motor channels responsible for the increase of error rates in the incompatible trials (compatibility effect). Immediately after the Standard and Error induction blocks, the participant was asked to rate his/her subjective experience of internal vs. external attribution of agency for the action performed (see subsection below **Subjective experience questionnaire**).

Experiment 2. Temporal aPE. With the aim of examining the temporal sensitivity for internal agency attribution in our experimental paradigm, we designed the following experiment introducing temporal aPE using different A-O temporal delays (see **Table 1**). The experiment was divided into 8 blocks of 160 trials each (maintaining the same proportion of compatible -40%- and incompatible -60%- trials as in Experiment 1). Two out of these 8 blocks were presented with a delay of 150 ms for the virtual hands' response movements with respect to the participants' action (A-O 150-Delay), and 2 other blocks with a delay set at 400 ms (A-O 400-Delay), following previous studies showing a gradually decreasing internal agency attribution as delay increases (Farrer et al., 2008; Farrer, Valentin & Hupe, 2013; Hara et al., 2015; Krugwasser, Harel & Salomon, 2019; Sato & Yasuda, 2005) (see **Figure 1C**). In the remaining 4 blocks, no delay was introduced (No-Delay). The experiment always started with a No-Delay block, followed by either A-O 150-Delay or A-O 400-Delay block, in an interleaved manner and counterbalanced among participants. No trials were presented in which an EE was imposed (as in Experiment 1), and therefore, content aPE were not induced in this experiment. At the end of each block, participants were requested to

rate their subjective experience of agency by answering a specific questionnaire (see subsection below **Subjective experience questionnaire**).

Experiment 3. Delayed content aPE. Following the evidence obtained in Experiments 1 and 2, this new experiment aimed to further investigate the workings of agency attribution by delaying the occurrence of content aPE (**Table 1**). For such a purpose, an A-O 400-Delay was chosen based on our results on Experiment 2 showing a significant decrease in internal agency judgements during the A-O 400-Delay condition, which are in line with previous reports showing a decreased internal attribution after 300-400 ms delay (Farrer et al., 2008; Farrer, Valentin & Hupe, 2013; Hara et al., 2015; Krugwasser, Harel & Salomon, 2019; Sato & Yasuda, 2005). Two blocks of 600 trials (compatible - 40% - and incompatible - 60% -, 1200 trials in total) were performed consecutively following the same experimental setting as Experiment 1. In this case, 50% of the trials were A-O 400-Delay trials (600 trials), and the other 50% No-Delay trials (600 trials). As in Experiment 1, we introduced EE randomly in 5% of the trials. Because all types of trials (No-Delay and A-O 400-Delay) and error conditions (SE and EE) were randomized along the whole experiment, the subjective experience could not be evaluated after the completion of each block. Instead, we opted for inserting an Error awareness question in 50% of the expected erroneous responses and in the same number of Correct responses for each A-O delay condition by asking participants: *"¿Have you committed an error?".* This response allowed us to study the relationship between temporal delay and error attribution for SE, without priming participants' responses by presenting the question for both erroneous and correct answers (we did not introduce it during EE as we already know from Experiment 1 that EE are perceived and processed as not coming from the self).

Subjective experience questionnaire

After each experimental condition (Experiments 1 and 2), participants were asked to rate their subjective experience (see Table 2). In the present study, we focused on examining the subjective experience of agency. This ability has been described by Georgieff and Jeannerod (1998) as the "Who" system, which allows us to track the origin of an action to its proper agent and therefore to distinguish the 'self' vs. 'other'. As exposed on the introductory section, this system might play an important role in self-awareness and consciousness of action, and it will partially depend on the sensorimotor input resulting from the action performed (Haggard, 2017). With the purpose of examining the modulation of agency judgements, we distinguished two levels of processing: (i) 'self' attribution of actions (referred to as internal attribution) and (ii) 'other' attribution of actions (referred to as external attribution).

A 7-item questionnaire (in Spanish) addressing internal (Q1 and Q2, i.e., *'Most of the time, the movements of the digital hands seemed to be my own movements'*), and external agency attribution

(Q3 and Q4, i.e., *'It sometimes seemed as if the errors were not caused by me'*) was created (Banakou, Groten & Slater, 2013; Botvinick & Cohen, 1998; Padrao et al., 2016) (Table 2). Two control questions (Q5 and Q6, i.e., *'It seemed as if I had more than two hands'*) about which we had no prior assumptions were also included. Moreover, an item addressing body ownership (Q7: *'I felt as if the digital hands were my hands'*) was also included. Participants were asked to rate their level of agreement with these 7 statements using a 7-level Likert-type response, ranging from "strongly disagree" (1) to "strongly agree" (7). For the present study, we will assess both aspects of agency attribution to investigate the possible dissociations between internal and external agency attribution for actions depending on the nature of the aPE (Experiments 1 and 3: content aPE; Experiments 2 and 3: temporal aPE).

Table 2. Item description of the subjective experience questionnaire. Description of internal attribution (Q1 and Q2), external attribution (Q3 and Q4) and Control (Q5, Q6 and Q7) items employed during Experiments 1 and 2.

Subjective feeling	*Questionnaire item*	*Description*
Internal attribution	*My movements*	**Q1**. Most of the time, the movements of the digital hands seemed to be my own movements.
	Feeling of control	**Q2**. I felt I could control the movements of the digital hands most part of the time.
External attribution	*Not my movements*	**Q3.** Sometimes, the digital hands seemed to be moving by themselves.
	external errors	**Q4**. It sometimes seemed as if the errors were not caused by me.
Control	*Influence*	**Q5.** Sometimes I felt as if the movements of the digital hands were influencing my own movements.
	More than 2 hands	**Q6.** It seemed as if I had more than two hands.
	My hands (SoO)	**Q7**. I felt as if the digital hands were my hands.

SoO: Sense of ownership.

EEG recording

The electroencephalographic (EEG) signal was recorded from the scalp using sintered Ag-AgCl ring electrodes mounted in an elastic cap (Easycap, International 10–20 System locations) and located at 27 standard positions (Fp1/2, Fz, F3/4, F7/8, Fc1/2, Fc5/6, Cz, C3/4, Cp1/2, Cp5/6, T7/8, Pz, P3/4, P7/8, O1/2). Biosignals were referenced on-line to the right mastoid electrode and posteriorly re-referenced off-line to half of the signal acquired on the left mastoid electrode.

Electrode impedances were kept below 5 kΩ. For all experiments, vertical eye movements were monitored with an electrode at the infraorbital ridge of the right eye. The electrophysiological signals were filtered online with a notch-filter (50 Hz) and a high-pass filter (0.016 Hz) and digitized at a rate of 250 Hz.

Data and statistical analysis

Behavioural analyses

To inspect the behavioral effects, we computed the RT and accuracy rates of the different kinds of trials (Correct, SE and EE) and conditions (No-Delay, A-O 150-Delay, A-O 400-Delay). We explored the compatibility effect (Danielmeier & Ullsperger, 2011; Eriksen & Schultz, 1979), in terms of RT and accuracy rates. To inspect compensatory mechanisms for erroneous responses, we computed the post-error slowing -PES- effect (Rabbitt & Rogers, 1977), by which errors are generally followed by more accurate and slower responses, reflecting cognitive control and monitoring mechanisms involved in behavioral adaptation (Botvinick et al., 2001; Cavanagh, Cohen & Allen, 2009; Marco-Pallares et al., 2008; Rabbit, 2002). Compatibility effects were also considered for the PES computations (Derrfuss et al. (2021), evidencing no significant differences in the distribution of compatible and incompatible trials for each condition (Chi-square test, all *p*-values > .10). Paired samples *t*-tests and *d*-Cohen for significant *t*-test results were employed.

ERPs

EEG analyses were conducted using routines taken from the ERPLAB toolbox V6.1.4 (Lopez-Calderon & Luck, 2014) and custom routines from MATLAB (The MathWorks, Inc. Natick, MA). A high-pass filter of 0.1 Hz (second order Butterworth filter, 12-40 dB) was applied to the raw EEG data. To perform artifact rejection, we excluded epochs with step-like artifacts when the amplitude jumps in the electro-oculograms exceeded 25 μV (moving window = 400 ms, moving step = 10 ms) or in which activity was ±100 μV in any channel. No additional filtering was applied for the statistical analyses. Only for illustrative purposes a low pass filter of 15 Hz (second order Butterworth filter, 12-40 dB) was applied on the grand average waveforms.

ERPs were time-locked to the onset of the response and baseline-corrected to its preceding 100 ms. Time windows for statistical analyses of ERP voltages were chosen based on the peak-based method (Luck, 2014) to determine the time window of interest by localizing the peak on the grand average waveform and then defining a symmetric time window centered on the peak (see *Supplementary Tables* at section **7. ANNEX**).

Time frequency

To study the brain oscillatory activity, we first generated large epochs including 2000 ms before and after the time window of interest. The same filtering as in the ERP analyses was applied. Data from each single trial were convoluted with a variable cycle complex Morlet wavelet between 4 and 10 cycles (logarithmic steps). Changes in time varying power (square of the convolution between the signal and the wavelet) in the frequency range from 1 to 40 Hz (linear increase) with respect to the baseline were computed for each trial and averaged for each participant before performing the grand average for the whole group.

In all experiments, to detect reliable differences and to adequately control for Type I errors when multiple comparisons were involved, the conditions of interest were submitted to a one-sample/repeated measures *t*-statistic two-tailed permutation test based on the cluster mass statistic (Bullmore et al., 1999; Maris & Oostenveld, 2007) using a family-wise alpha level of 0.05 by means of the Mass Univariate ERP Toolbox (Groppe, Urbach & Kutas, 2011).

In addition, we applied repeated measures ANOVA's (rmANOVA) to perform confirmatory analysis. In all cases, the Greenhouse-Geisser correction was applied whenever the sphericity assumption was not met. Furthermore, uncorrected degrees of freedom but corrected *p*-values are reported as p_{GG}. In case a theoretically relevant interaction was found to be significant, we disentangled the effect by means of post-hoc *t*-tests (p_{BC}). Effect sizes are reported as Partial eta-squared (η_p^2) for significant rmANOVA results, and as *d*-Cohen for significant *t*-test results.

Source localization

Source localization analyses of the major generators of the ERP components of interest was performed using Brain Electrical Source Analysis software (BESA Research 7.0). Cortical source estimation sources using the minimum-norm estimation (MNE) technique (Hämäläinen & Ilmoniemi, 1994) was performed. This approach involves the computation of the source activities of a several regional sources evenly distributed over 1420 standard locations of the smoothed surface of a standard brain. A unique current distribution explaining the surface measurements is obtained by applying the minimum norm constraint, which selects the solution with minimum overall intensity using a regularization procedure that sets the balance between fitting the measured data (minimizing the residual) and minimizing the contributions of noise [in our case we employed the Tikhonov–Philips approach for spatial regularization ($\lambda = 0.01$)]. Finally, distributed source model analysis using CLARA ("Classic LORETA Recursively Applied") (Hoechstetter, Berg & Scherg, 2010) was applied, performing a weighted LORETA with a reduced source space at each iteration. Compared to LORETA, this iterative approach reduces the blurring of the estimated sources while keeping the advantage of a predefined distributed source model,

making it easier to determine the location of the source with maximal activity (Pascual-Marqui, Michel & Lehmann, 1994).

Subjective feelings questionnaire

In order to evaluate the participant's subjective agentic experience over their performed actions and the influence of A-O incongruences, we directly compared the participant's ratings given immediately after finishing each experimental condition (Experiment 1: Standard vs. Error Induction block; Experiment 2: No-Delay vs. 150-Delay vs. 400-Delay blocks). Wilcoxon test (pairwise comparisons) were employed for testing the possible differences regarding the participants' scores on the questionnaire, with the significance alpha level adjusted to multiple comparisons (Bonferroni correction, $p < 0.003$ for Experiment 1 and $p < 0.002$ for Experiment 2).

3.1.3. RESULTS

EXPERIMENT 1. CONTENT AGENCY PREDICTION ERRORS

Behavioural performance

The participants' performance was as expected for this paradigm, with a mean percentage of self-generated errors (SE) approx. of 10.5% ± 6.1 (mean ± SEM). As previously reported, a compatibility effect (Danielmeier & Ullsperger, 2011; Eriksen & Schultz, 1979) was found, with participants responding more accurately for compatible than for incompatible trials [mean percentage of SE, 1.7% ± 1.7 vs. 8.8% ± 4, respectively; $t(24) = -11.6$, $p < .001$, $d = 2.3$] as well as faster [mean RT: 284 ± 4 ms vs. 301 ± 5 ms; $t(24) = -10.5$, $p < .001$, $d = 3.8$]. As expected, participants also showed significantly faster reaction time (RT) for SE compared to correct responses [mean RT correct: 293 ± 5 ms vs. mean RT SE: 237 ± 5 ms; $t(24) = 22.7$, $p < .001$, $d = 2.3$]. Altogether, these results indicated a correct implementation of the Flanker task. As expected, when computing the PES effect (Rabbitt & Rogers, 1977), SE were followed by significantly slower correct responses than those correct reactions following correct responses [PES SE = 15 ± 3 ms; $t(24) = -4.3$, $p < .001$, $d = 2.5$]. Interestingly, we replicated Padrao et al. (2016) findings, showing a significantly slower RT for correct responses following EE compared to correct [PES EE = 17 ± 3 ms; $t(24) = -4.8$, $p < .001$, $d = 3.1$], suggesting the engagement of control/monitoring systems despite having responded correctly.

Subjective Agency questionnaire

Crucially for the present set-up, the appearance of EE in Error induction blocks lead to an increase in external agency attribution (see **Figure 2A**). Results on the Wilcoxon test revealed significant differences between the Standard and Error induction blocks for both external attribution

questions: Q3 *("Sometimes, the digital hands seemed to be moving by themselves"*) (Z = -3.96, p < .001) and Q4 ("*It sometimes seemed as if the errors were not caused by myself*") (Z = -3.88, p < .001), but without affecting the internal attribution judgements (Q1: Z = -1.85, p = .065; Q2: Z = -2.46, p = .014). High levels of body ownership were observed for both conditions, and no other significant differences were found for any of the other control measures (**Figure 2A**).

ERP signatures for EE and SE

As can be observed in **Figures 3A** and **3B**, EE triggered a slight frontal N2-like component with a peak latency around 350 ms, although it was not confirmed by cluster permutation analysis (**Figure 3B**). Following the N2, a large P600 deflection was observed, peaking at centro-parietal sites 580 ms after the response onset during the EE trials (**Figure 3A** and **3B**). Mass permutation analysis confirmed a very extensive positive cluster at centro-parietal sites starting at 400 ms (**Figure 3B** -right-, voltage scalp topographic map), illustrating the P600 component (see **Table S1** for a full outline of significant rmANOVA results at section **7. ANNEX)**.

In contrast, participant's SE elicited the well-known fronto-central ERN component peaking at about 95 ms after error commission (**Figures 4A** and **4B**) (see **Table S1** for a complete description of significant rmANOVA results at section **7. ANNEX**). Time-by-time point cluster-based permutation analysis showed a significant negative frontal cluster (p-values < .05), spreading up to 200 ms and involving up to 15 electrodes around Fz at its peak (see **Figure 4B**, voltage scalp topographic map depicting significant frontal cluster), exhibiting the same timing and topographical distribution as the ERN component (Falkenstein, 1990; Gehring et al., 1993).

Association between the P600 component and external agency attribution

Importantly, we found that the amplitude of the P600 at central locations elicited when an EE was introduced (difference EE minus Correct) was associated to increased external agency attribution reports, both for Q3 and Q4 items separately [computed as the difference Q3/Q4 after Error induction blocks minus Q3/Q4 after Standard block; Q3: $r(23) = 0.43$, p = .02; Q4: $r(23) = 0.47$, p = .02], as well as for the compound score Q3/Q4 [median of both differences; $r(23) = 0.54$, p = .003] (see **Figure 4A**). No significant association was encountered for the P600 amplitude during EE and the internal attribution rating, neither for Q1 and Q2 items separately [computed as the difference Q1/Q2 after Error induction blocks minus Q1/Q2 after Standard block; Q1: $r(23) = -0.09$, p = .33; Q2: $r(23) = -0.07$, p = .73] nor for the compound score for the internal agency attribution category [median of both differences; $r(23) = 0.22$, p = .29].

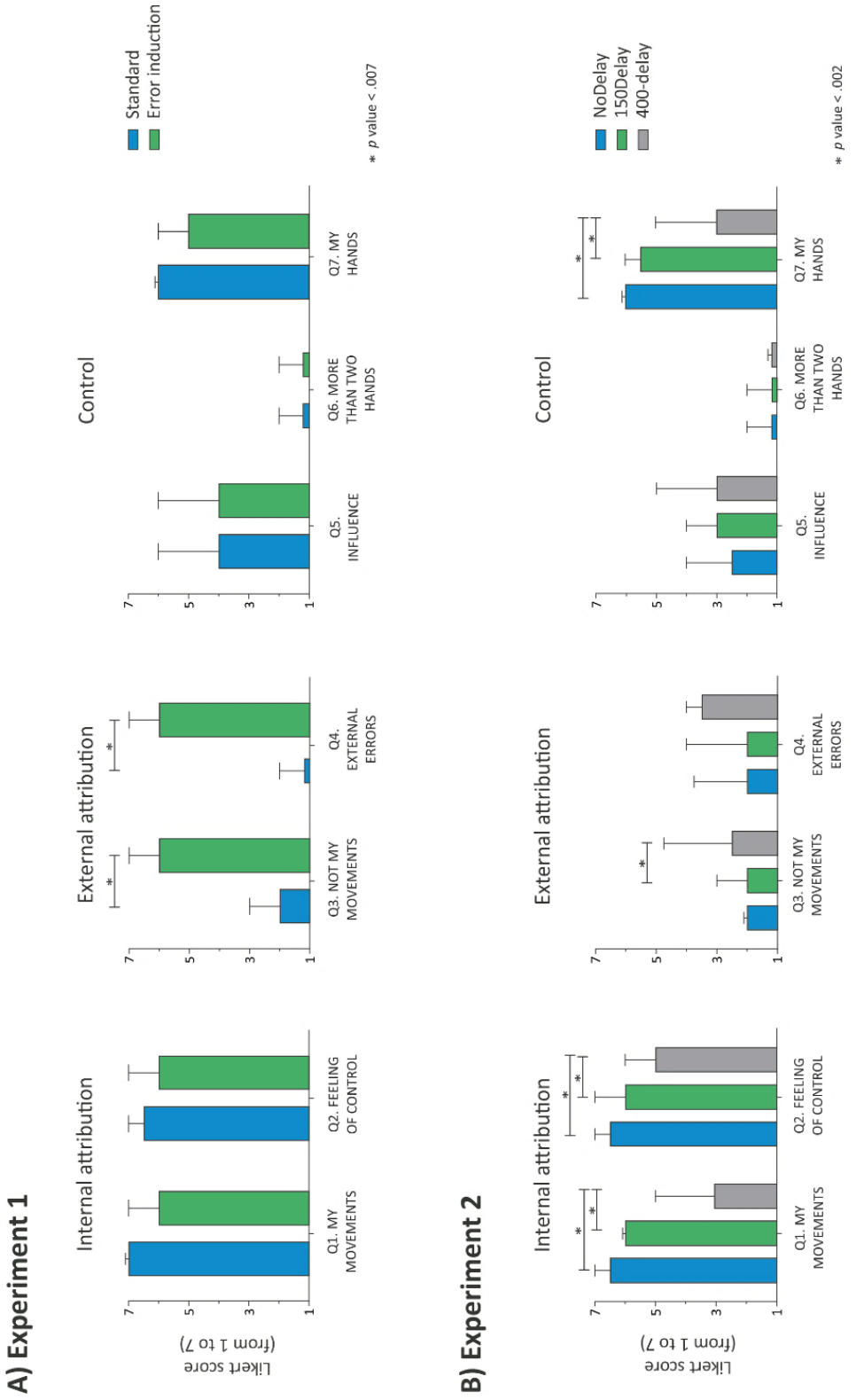


Figure 2. Results at the Subjective feelings questionnaire regarding internal attribution, external attribution and Control items for Experiments 1 and 2. Seven-point Likert scale scores, from "strongly disagree" (1) to "strongly agree" (7) are represented with Median scores (with interquartile range) for all items on the questionnaire.

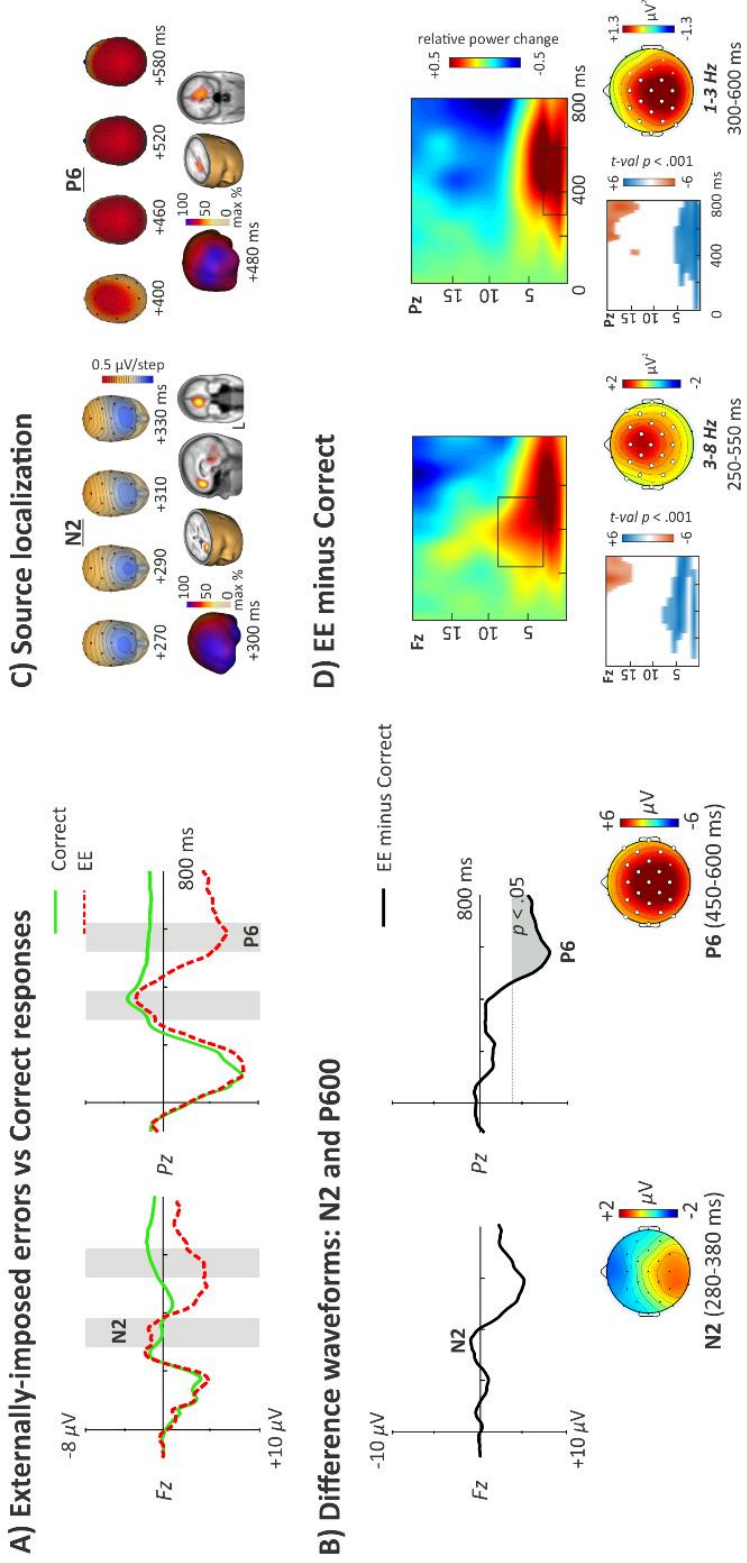


Figure 3. ERP and time frequency results for externally generated errors (EE) and correct responses in Experiment 1. A) Response-locked grand average waveforms for Correct (green line) and external errors (EE) (red dotted line) at Fz and Pz electrodes. The grey vertical segments indicate the time intervals that were separately subjected to statistical analysis -rmANOVAs-. **B)** Difference waveforms for

EE minus Correct at Fz and Pz electrodes depicting the N2 and P600 components. Below scalp topography maps for the significant clusters depicting the P600 component. The grey horizontal segment corresponds to the statistically significant cluster (p < .05) at the mass cluster permutation testing. Significant electrodes conforming the clusters are shown in white. **C)** Source localization analysis showing surface voltage maps, minimum norm estimates and CLARA distributed source model for the N2 and P600 components. **D)** Changes in power spectrum with respect to baseline (100 ms period prior to the response onset) for the difference EE minus Correct at Fz and Pz showing a power increase in the theta band (3-8 Hz) frontally distributed followed by a more posterior increase in delta power (1-3 Hz), similar to the topographical and latency distribution of the N2/P600 complex waveform described above. Black dotted contours show the areas where ANOVAs were computed. Below are illustrated the t-value maps (p < .001) for the difference EE minus Correct and the topographical scalp maps. Significant electrodes conforming the clusters are shown in white. EE: external errors. See also **Table S1** and **Figure S1** at section **7. ANNEX** for more details on rmANOVA results and time frequency data for Experiment 1.

Moreover, the P600 component was also positively correlated with the PES effect after EE (computed as the difference PES after EE minus PES after Correct), showing larger P600 amplitudes associated with longer PES [$r(23) = 0.37$, $p = .04$; **Figure 4B**], corroborating the relationship between external agency attribution/content aPE and the engagement of control/error monitoring adjustments.

Source localization analysis

Following the ERP analysis, we performed source localization analyses to locate the source of the N2/P600 (difference EE minus Correct) and the ERN (difference SE minus Correct) components. Firstly, we computed the surface voltage maps corresponding to the N2 component (270-330 ms) (**Figure 3C**). Surface voltage maps revealed an increase in negative voltage frontally distributed. The minimum-norm estimate at the 300 ms latency depicted a frontal source distribution, and a single main source located at the ACC is reported (Talairach: x = -10.5, y = 46.1, z = 9.7).

For the P600 component (450-600 ms) (see **Figure 3C**), the cortical surface voltage maps illustrated a centro-parietal distribution, as well as a parietal distributed source when using the minimum-norm estimate (480 ms). A main source located at the middle cingulate cortex (MCC) (Talairach: x = -2.5, y = 13.2, z = 21) was found when computing CLARA.

On the other hand, we also examined the ERN at the time window 60-135 ms after the response onset. As can be seen in **Figure 4C,** the ERN surface voltage distribution was characterized by an increased negative voltage at fronto-central locations when estimating the cortical sources using the minimum-norm estimation (MNE) technique. In this case, the minimum norm image at the peak latency 90 ms after an error slip revealed a frontal midline location for the ERN (see **Figure**

4C, bottom-left for the CLARA distributed model, Classic LORETA Recursively Applied) with a main source located at the ACC (Talairach: x = 5, y = 22.3, z = 34) at the time window 60-135 ms (**Figure 4C**, bottom-right).

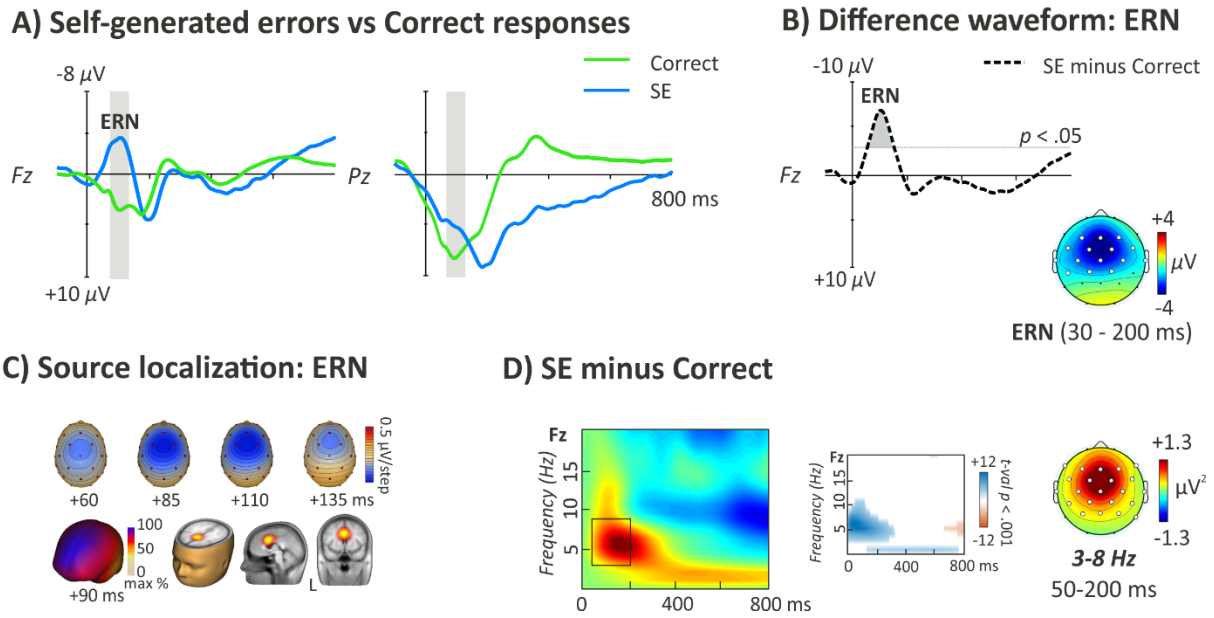


Figure 4. ERP and time frequency results for self-generated errors (SE) and correct responses in Experiment 1. A) Response-locked grand average waveforms for Correct (green line) and SE (blue line) at Fz and Pz electrodes. The grey vertical segments indicate the time intervals that were separately subjected to statistical analysis -rmANOVAs-. **B)** Difference waveform for SE minus Correct at Fz electrode depicting the ERN component. Scalp topography map for the significant cluster depicting the ERN component. The grey horizontal segment corresponds to the statistically significant cluster (p < .05) at the mass cluster permutation testing. Significant electrodes conforming the clusters are shown in white. **C)** Source localization analysis showing surface voltage maps, minimum norm estimates and CLARA distributed source model for the ERN component. **D)** Changes in power spectrum with respect to baseline (100 ms period prior to the response onset) for the difference SE minus Correct at Fz showing a power increase in the theta band (3-8 Hz) at the ERN time window. Black dotted contours show the areas where rmANOVAs were computed. On the right the t-value map (p < .001) for the difference SE minus Correct and the topographical scalp map are shown. Significant electrodes conforming the clusters are shown in white. SE: Self errors; ERN: Error-related negativity. See also **Table S1** and **Figure S1** at section **7. ANNEX** for more details on rmANOVA results and time frequency data for Experiment 1.

Time frequency analysis

*N2 and P600 (*time windows *200-400 and 400-600 ms): EE oscillatory activity*

When introducing an EE, an increase in relative power for the delta band (1-3 Hz) as well as for the theta (3-8 Hz) frequency band were registered, compared to Correct and SE trials (see **Figure S1** for EE condition at section **7. ANNEX**). Importantly, when computing the contrast EE minus Correct, a frontal-theta and parietal-delta dissociation was observed (see **Figure 3D**), correspondingly to the N2/P600 ERP pattern previously reported (see **Table S1** for rmANOVA results at section **7. ANNEX**).

Cluster-based permutation analysis for the difference EE vs. Correct corroborated the existence of a significant frontal positive cluster at the theta frequency band (3-8 Hz) (see **Figure 3D**), in accordance with the latency and topographical distribution of the N2 component. When inspecting the delta frequency band (1-3 Hz), a significant cluster with a centro-parietal distribution was confirmed between 300 ms onwards (**Figure 3D**), resembling the latency and topographical characteristics of the P600 component (see **Figure 3B**).

*ERN (*time window *50-150 ms): SE oscillatory activity*

As previously reported after error commission (Cohen, 2011; Luu, Tucker & Makeig, 2004), a significant increase in the theta band (3-8 Hz) was encountered for SE during the 50-200 ms time window (see **Figure S1** at section **7. ANNEX**), displaying a frontal topographical distribution (see **Table S1** for a complete outline of rmANOVA results at section **7. ANNEX**). Repeated-measures cluster permutation tests for the contrast SE vs. Correct confirmed a significant positive cluster ($p < .05$) at the theta frequency band (3-8 Hz) at the ERN time window (0-500 ms) (**Figure 4D**).

Association between delta oscillatory, P600 component and external agency attribution

Importantly, a positive correlation was found between centro-parietal delta oscillatory activity (1-3 Hz) recorded after EE (difference between EE minus Correct trials) and the compound score for external agency attribution [$r(23) = 0.41$, $p = .027$] (see **Figure 5C**), confirming a significant increase in delta activity when eliciting an aPE, following our ERP results for the P600 component. Interestingly, this centro-parietal delta power increase was also positively correlated with the PES effect for EE [$r(23) = 0.38$, $p = .03$] (see **Figure 5C**), which is clearly in line with our findings regarding the P600 component.

Noteworthy, when examining the relationship between the P600 component and delta power (1-3 Hz) at centro-posterior electrode locations, a significant positive correlation was observed [$r(23) = 0.80$, $p < .001$], supporting the P600-delta association (see **Figure 5E**).

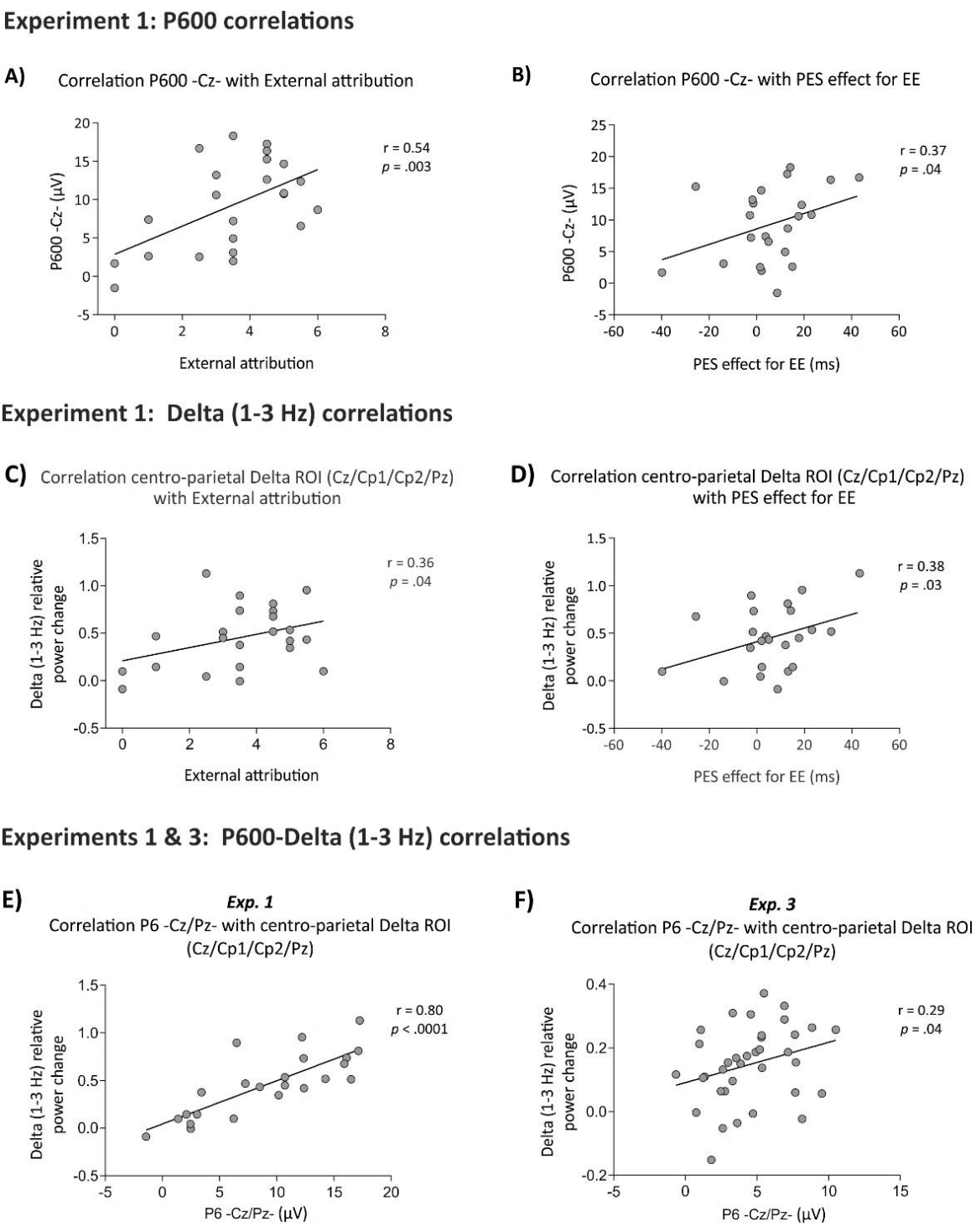


Figure 5. Correlational analyses. A) Correlation between the P600 amplitude -Cz- (difference EE minus Correct) and the compound score for external attribution [computed as the median of the two differences (Error induction minus Standard) for Q3 and Q4 items]. **B)** Correlation between the P600 amplitude -Cz- (difference EE minus Correct) and the post-error slowing effect (PES) for EE (computed as the PES after EE minus PES after Correct). **C)** Correlation between delta power (1-3 Hz) (centro-parietal ROI including Cz/Cp1/Cp2/Pz electrodes) and external attribution compound score. **D)** Correlation between delta power (1-3 Hz) (centro-parietal ROI including Cz/Cp1/Cp2/Pz electrodes) and PES effect for EE. EE: external errors. **E)**

Correlation between the P600 amplitude -mean Cz/Pz electrodes- (difference EE minus Correct) and delta power (1-3 Hz) (centro-parietal ROI including Cz/Cp1/Cp2/Pz electrodes) for Experiment 1. **F)** Correlation between the P600 amplitude -mean Cz/Pz electrodes- (difference EE minus Correct) and delta power (1-3 Hz) (centro-parietal ROI including Cz/Cp1/Cp2/Pz electrodes) for Experiment 3.

Interim conclusions Experiment 1

Altogether, the current results show the elicitation of a specific neural pattern consisting of a delta/P600 component significantly associated to the external attribution of agency during the insertion of content aPE. Besides, the latency of these components highlights the key role of retrospective inferencing on the processing of agentic inferences, appearing later in the processing stages compared to internally attributed actions (i.e., ERN component after SE).

EXPERIMENT 2. TEMPORAL AGENCY PREDICTION ERRORS

Behavioural performance

Participants' accuracy rates showed a mean percentage of SE of ~10% (No-Delay: 11% ± 1.3%; 150-Delay: 12% ± 1.4%; 400-Delay: 10% ± 1.3%) (mean ± SEM) did not differ between conditions and were comparable with that reported in Experiment 1. A larger percentage of SE were made for incompatible vs. compatible trials [No-Delay: Compatible 2% ± 0.4%, Incompatible 9% ± 0.9%, $t(24) = -11.5$, $p < .001$, $d = 10$; 150-Delay: Compatible 2% ± 0.5%, Incompatible 9% ± 0.9%, $t(24) = -6.5$, $p < .001$, $d = 9$; 400-delay: Compatible 2% ± 0.4%, Incompatible 8% ± 0.9%, $t(24) = -14.22$, $p < .001$, $d = 8$].

Also, faster RT for SE compared to correct responses were observed for all conditions ($p < .001$ in all cases). Likewise, a compatibility effect was found for all conditions when performing 3 x 2 rmANOVA [factors: A-O condition (No-Delay vs. 150-Delay vs. 400-Delay) and Compatibility (Compatible vs. Incompatible)] in correct responses, showing a significant main effect of Compatibility [$F(1,23) = 90$, $p < .001$, $\eta_p^2 = 0.80$] as well as a main effect of A-O condition [$F(2,46) = 15$, $p_{GG} < .001$, $\eta_p^2 = 0.40$]. A non-significant A-O condition x Compatibility interaction was observed [$F(2,46) = 16$, $p = .541$].

Subjective feelings questionnaire – internal agency attribution

As shown in **Figure 2B,** a significant decrease was encountered regarding internal attribution of agency during the 400-Delay condition compared to both the No-Delay and 150-Delay conditions, highlighting the influence of A-O asynchrony on internal attribution mechanisms (Blakemore, Frith & Wolpert, 1999; Farrer et al., 2008; Farrer, Valentin & Hupe, 2013; Hara et al., 2015;

Krugwasser, Harel & Salomon, 2019; Sato & Yasuda, 2005 Bonferroni-corrected Wilcoxon pairwise comparisons revealed significant differences ($p < .002$) for all internal attribution items when comparing No-Delay vs. 400-Delay conditions, as well as for 150-Delay vs. 400-Delay conditions, but not for the external agency category (Q3 and Q4) (see **Figure 2B**). Significant differences were encountered for item Q3 between 150-Delay vs. 400-Delay ("*Sometimes, the digital hands seemed to be moving by themselves*" –external agency attribution-), although depicting median values <4 in both cases equivalent to the statement "Neutral/Nor agree not disagree".

Importantly, and crucial to the present study, no external agency attribution was reported in any case (see **Figure 2B**), as opposed to the pattern observed during Experiment 1 (EE). When comparing the subjective experience of internal vs. external attribution of agency across Experiments 1 and 2, a clear dissociation was encountered. When introducing aPE in terms of content consistency (Experiment 1), a strong external attribution experience was reported compared to temporal aPE (Experiment 2) (Q3: Z = -3.09, $p = .002$; Q4: Z = -3.87, $p < .001$). On the contrary, temporal aPE (Experiment 2) resulted in a significant reduction of internal-related reports during the 400-Delay condition in contrast to the reports obtained during content aPE which did not show significant reductions (Q1: Z = -2.97, $p = .003$; Q2: Z = -3.26, $p = .001$).

Moreover, a significant decrease in the reported SoO was observed for the 400-Delay condition compared to the No-Delay and 150-Delay trials (see **Table S2** at section **7. ANNEX** for a complete description of all comparisons).

ERPs related to the temporal synchronicity of the action performed and the sensory feedback

ERP signatures for Correct and SE for each A-O delay condition (No-Delay, 150-Delay, 400-Delay) are shown in **Figure 6**. Significant differences were encountered for Correct responses regarding the A-O condition, evidencing an increased frontal positivity for the 400-Delay condition appearing between 600-800 ms after the response onset (see **Figures 6A** and **6C** -top- depicting the difference waveform Correct 400-Delay minus Correct No-Delay) and **Table S3** at section **7. ANNEX** for a complete outline of significant rmANOVA results). This frontally distributed positivity appeared around 200 ms after the feedback onset during the 400-Delaty condition, resembling the feedback correct-related positivity (FCRP)/reward positivity (RewP). Additionally, this result was additionally supported by a cluster-based permutation analysis that revealed a significant ($p < .05$) positive fronto-central cluster at the 600-700 ms time window (see **Figure 6C**).

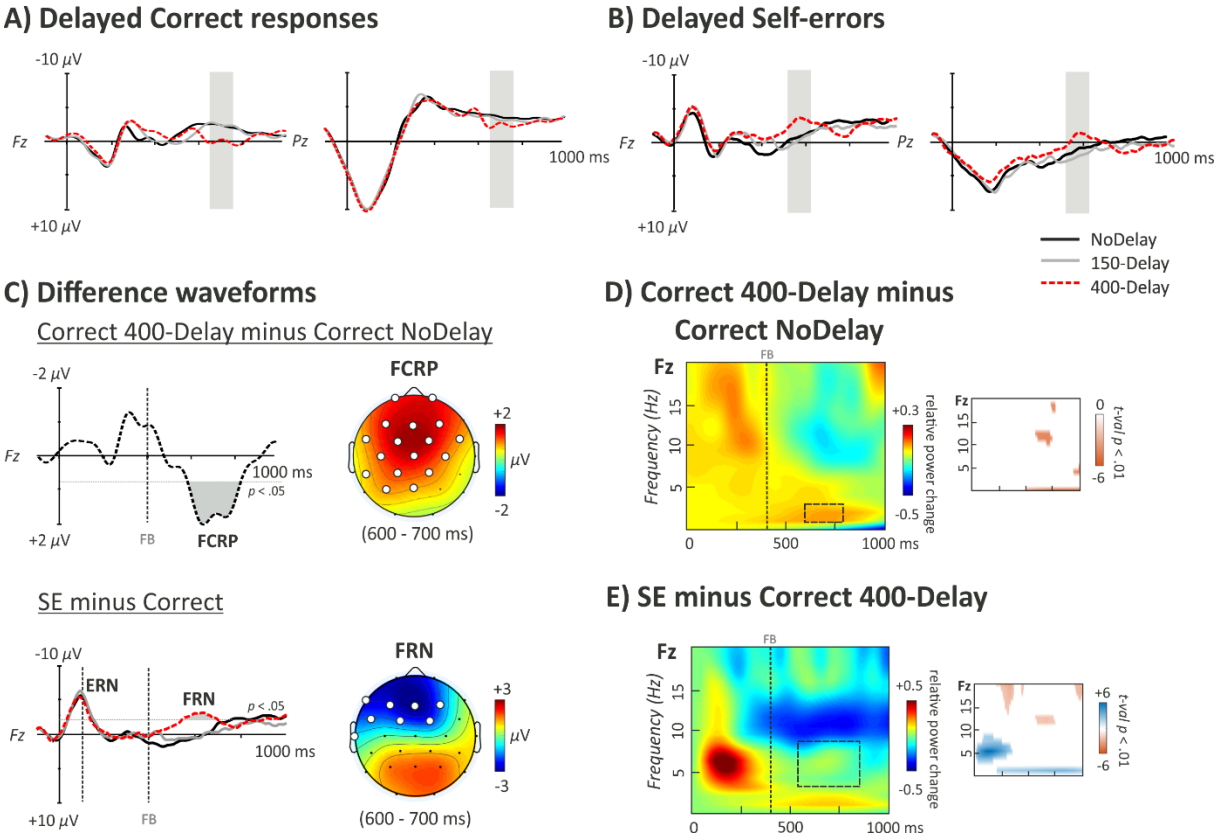


Figure 6. ERP and time frequency results for Experiment 2. **A)** Response-locked grand average waveforms at frontal (Fz) and parietal (Pz) locations for Correct responses (green line) and Errors (blue line) for all A-O delay conditions. **B)** Difference waveforms for the contrast SE minus Correct time-locked to the response onset for all conditions. Difference in scalp distribution map for the contrast SE minus Correct at the A-O 400-Delay showing a significant frontal negative cluster illustrating the FRN component. **C)** Source localization analyses for the FRN component showing surface voltage maps (time window 550-700 ms), minimum norm estimates, and CLARA distributed source model showing a clear frontal voltage increase with a main source located at the anterior cingulate cortex (ACC). **D)** Changes in power spectrum with respect to baseline (100 ms period prior to the response onset) for the differences SE minus Correct at Fz for all A-O delay conditions, showing an increase in theta (3-8 Hz) during the 400-Delay condition at the FRN (T.W: 550-800 ms). Black dotted contours show the areas where ANOVA's were computed. The grey vertical segments indicate the time intervals that were separately subjected to statistical analysis -rmANOVAs- (see **Table S3**). The grey horizontal segments correspond to the statistically significant cluster (p < .05) at the mass cluster permutation testing. Significant electrodes conforming the clusters are shown in white. SE: Self errors; FRN: Feedback-related negativity. See also **Table S3** and **Figure S2** at section **7. ANNEX** for more details on rmANOVA results and time frequency data for Experiment 2.

Moreover, participant's errors displayed the well-established ERN component at frontal locations in all conditions peaking around 80 ms after the response onset (see **Figure 6B**). Remarkably, only during the 400-Delay feedback interval condition, a later frontal negative component for SE peaking at 630 ms appeared (**Figure 6C** -bottom- for the difference waveform SE minus Correct

during the A-O 400-Delay), bearing a clear resemblance to a Feedback Related Negativity (FRN) component latency-dependent on the visual feedback appearance (virtual hand response movement at 400 ms) (see **Table S3** at section **7. ANNEX** for a complete outline of significant rmANOVA results). Moreover, the contrast SE vs. Correct in the 400-Delay A-O condition was submitted to cluster-based permutation testing, confirming an FRN-like negative frontal cluster ($p < .05$) involving several frontal electrodes at the time window 550-700 ms (see **Figure 6C**).

Time frequency analysis: delayed feedback of errors (FRN)

Results for the time-frequency analyses for Correct and SE at all A-O delay conditions are shown on **Figure S2** at section **7. ANNEX**. To examine the effect of the A-O delay on the oscillatory activity we studied the changes in power at the delta frequency (1-3 Hz) band during correct responses across all A-O conditions, which resulted in a significant increase in delta power at frontal electrode locations during the 400-Delay A-O condition compared to the No-Delay condition at the 600-800 ms time window [$t(24) = 2.02$, $p = .05$, $d = 4.3$], in line with our ERP results regarding the FCRP component [see **Figure 6D** for the contrast Correct 400-Delay vs Correct No-Delay and **Table S3** at section **7. ANNEX** for a complete outline of significant rmANOVA results at the delta frequency (1-3 Hz) band]. Nevertheless, no significant clusters were encountered at the delta frequency band (1-3 Hz) for the contrast Correct 400-Delay vs Correct No-Delay at the cluster permutation testing.

Additionally, we investigated the oscillatory activity related to the effect of temporal asynchrony on error processing, evidencing larger frontal theta power (3-8 Hz) increases for SE during the 400-Delay condition compared to the No-Delay [$t(24) = -4.53$, $p < .001$, $d = 1.03$] and 150-Delay [$t(24) = -3.14$, $p = .005$, $d = 0.70$], conditions between 550-850 ms [see **Figure S2A** and **Table S3** at section **7. ANNEX** for a complete outline of significant rmANOVA results at the theta frequency (3-8 Hz) band]. Moreover, we computed the contrasts SE minus Correct for all A-O conditions, time-locked to the response onset (see **Figure 6B** and **S2B** at section **7. ANNEX**), evidencing a significant larger theta power increase during the 400-Delay condition. This result is in line with the topographical and latency characteristics of the ERP findings depicting an FRN-like deflection during the A-O 400-Delay SE trials (see the topographic scalp map in **Figure 6B**), confirming previous reports describing FRN-related medial-frontal theta power increases (Cavanagh et al., 2010; Cavanagh & Frank, 2014; Cohen, 2011; Luu, Tucker & Makeig, 2004). Nevertheless, the results for the cluster permutation analysis at the theta band (3-8 Hz) for the difference SE 400-Delay minus Correct 400-Delay were not significant [see **Table S3** at section **7. ANNEX** for a complete outline of significant rmANOVA results at the theta frequency (3-8 Hz) band].

Interim conclusions Experiment 2

Overall, during the realization of this experiment we observed a decrease in internal agency attribution and SoO as the A-O delay increased (temporal aPE), however, no increase was observed in external agency attribution as previously observed in Experiment 1 (content aPE). During the execution of correct responses, a frontally distributed FCRP component was observed during the 400-Delay A-O condition. Also, an FRN modulation during SE committed at the longest delay was registered. Importantly, no parietal delta/P600 component was observed in this experiment, confirming that no content aPE/external agency judgement was elicited.

EXPERIMENT 3. DELAYED AGENCY CONTENT PREDICTION ERRORS

Behavioural performance

As in Experiment 1, both accuracy and RT were found to be influenced by the compatibility effect, showing larger error rates for Incompatible trials [No-Delay: Compatible: M = 1.2% ± 0.3%; Incompatible: M = 10.6% ± 3%; $t(17) = -3.3$, $p = .004$, $d = 4.4$; 400-Delay: Compatible: M = 1.1% ± 0.3%; Incompatible: M = 10% ± 3%; $t(17) = -3.2$, $p = .006$, $d = 4.1$] as well as faster RT for Compatible vs. Incompatible trials -correct responses- [Compatibility effect No-Delay = 19 ms; $t(17) = -5.94$, $p < .001$, $d = 2.36$; Compatibility effect 400-Delay = 20 ms; $t(17) = -5.8$, $p < .001$, $d = 2.2$]. No differences between A-O conditions were observed regarding neither accuracy nor RT for correct, SE or EE trials (all $p > .05$).

Assessment of error attribution and temporal synchronicity

Signal detection theory analysis was applied to calculate hit (SE detected as SE), miss (SE not detected), false alarm (FA) (Correct responses detected as SE), and correct rejection rates (CR: Correct responses detected as Correct) for each participant regarding the error awareness question ("¿Have you committed an error?") during SE and Correct trials. The percentage of aware and unaware errors did not differ between conditions [No-Delay: 88.1% ± 0.16; 400-Delay: 86.6% ± 0.18; $t(16) = 1.28$, $p > 0.2$]. No significant differences were found between A-O conditions for neither Hit, Miss, FA, nor CR ratios (all $p > .05$). We then computed the d-prime sensitivity measure for both A-O conditions, which revealed no significant differences between A-O delay conditions [No-Delay: M = 2.66 ± 0.12; 400-Delay: M = 2.38 ± 0.18; $t(16) = 1.82$, $p = 0.09$], indicating that A-O temporal synchronicity did not influence error perception and detection mechanisms.

ERP signatures for EE with and without temporal synchronicity

Averaged ERP voltages for Correct and EE during the No-Delay and 400-Delay A-O conditions are shown on **Figure 7A**. Importantly, we replicated the results encountered in Experiment 1 regarding the N2/P600 complex (**Figures 3A** and **3B**), illustrated by a frontal negativity followed by a large positive parietal deflection when an EE was introduced during the No-Delay trials. Interestingly, the same N2/P600 complex was evoked during the 400-Delay trials, although latency-dependent on the delayed incongruent feedback onset (see difference waveform for EE minus Correct during 400-Delay condition at **Figure 7B**). No significant main effect of A-O Condition was encountered when performing the rmANOVA (see **Table S4** at section **7. ANNEX**).

When applying cluster permutation analysis for the contrast EE minus Correct during both A-O conditions, a significant negative cluster at the N2 time window was only found for the No-Delay condition (**Figure 7B**). Notably, a large positive cluster depicting a posterior distribution was found for both A-O conditions at latter time windows (see **Figure 6B**), confirming the elicitation of the P600 deflection for both conditions although temporally dependent on the feedback appearance/outcome delay.

Time frequency analysis

EE and temporal synchronicity

As it can be seen at **Figure 7C**, the contrast EE minus Correct elicited the frontal-theta/parietal-delta complex previously described in Experiment 1. When examining theta oscillatory activity power increases for the contrast EE minus correct, the analysis revealed a significant positive cluster at frontal locations during the 400-Delay trials between 300-600 ms although it did not reach significance for the No-Delay condition (**Figure 7C**). Results of the rmANOVA showed no significant main effect regarding A-O condition, which is in line with our previously reported ERP results.

SE and temporal synchronicity.

An increase in the theta band (3-8 Hz) was observed for SE for both A-O delay conditions at frontal electrode locations (see Figure S3), illustrating the ERN component. No significant differences regarding A-O condition were found at Fz electrode at the time window 50-200 ms ($p > .05$).

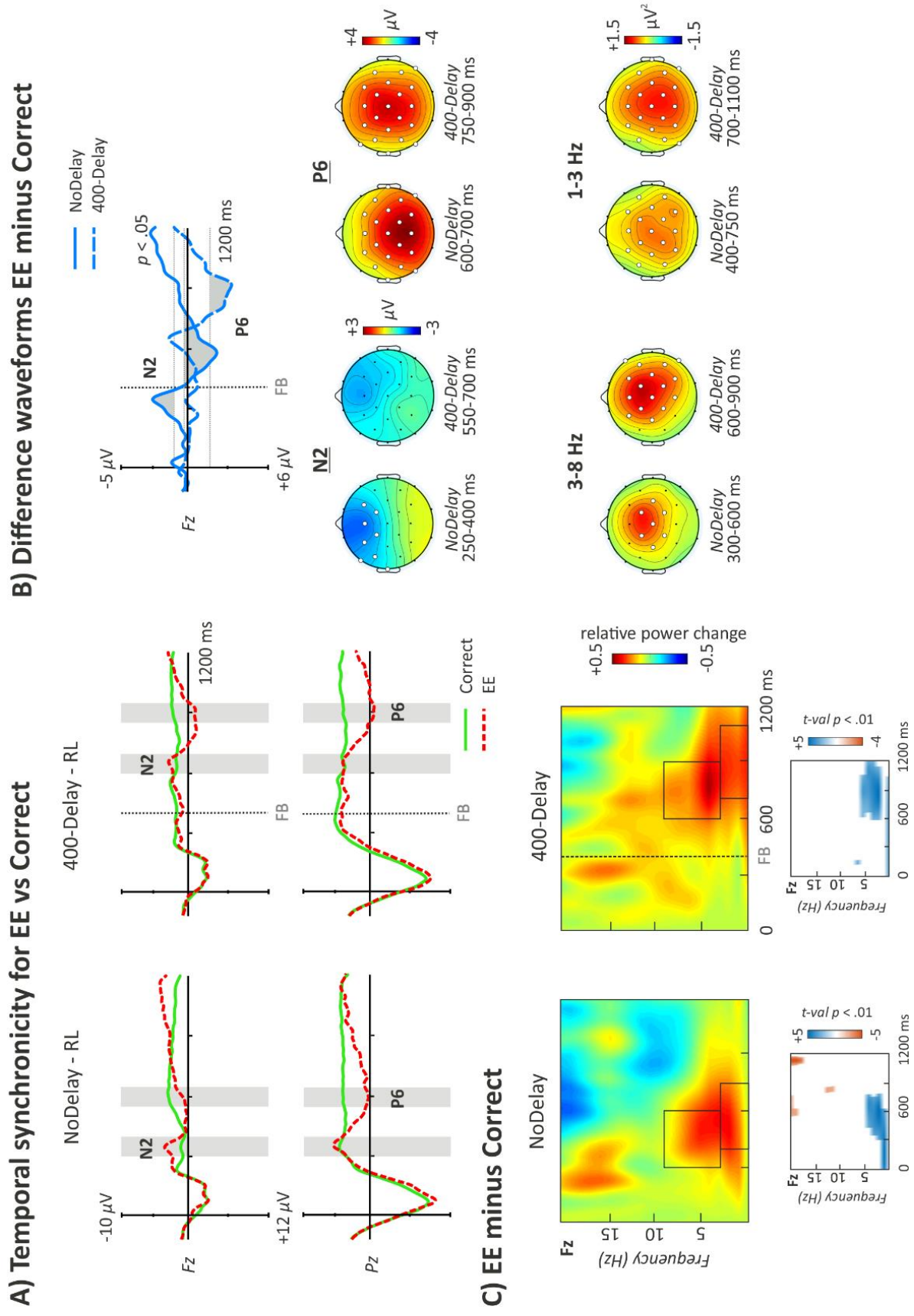

A) Temporal synchronicity for EE vs Correct
NoDelay - RL
400-Delay - RL
N2
P6
FB
1200 ms
-10 µV
+12 µV
Fz
Pz
Correct
EE
B) Difference waveforms EE minus Correct
NoDelay
400-Delay
p < .05
-5 µV
+6 µV
P6
N2
NoDelay 250-400 ms
400-Delay 550-700 ms
NoDelay 600-700 ms
400-Delay 750-900 ms
3-8 Hz
1-3 Hz
NoDelay 300-600 ms
400-Delay 600-900 ms
NoDelay 400-750 ms
400-Delay 700-1100 ms
µV
µV²
C) EE minus Correct
relative power change
+0.5
-0.5
400-Delay
NoDelay
Frequency (Hz)
t-val p < .01
600
0

Figure 7. ERP and time frequency results for Experiment 3. A) Response-locked grand average waveforms at frontal (Fz) and parietal (Pz) locations for Correct responses (green line) and EE (dotted red line) for both A-O delay conditions. **B)** Difference in waveforms for the contrasts EE minus Correct at Fz for both experimental conditions (No-Delay -blue line-/400-Delay -dotted blue line-). The N2/P600 complex waveform was registered during both conditions but shifted forwards in time to the feedback onset for the 400-Delay condition. At the bottom the topographical scalp maps depicting a significant frontal negative cluster during the No-Delay condition at the time window 250-400 ms, illustrating the N2 component. On the right the topographical maps show a significant centro-parietal positive cluster for both conditions at the P600 time window (adjusting for feedback presentation delay), confirming the presence of a significant late positivity following the insertion of EE. **C)** Changes in power spectrum with respect to baseline (100 ms period prior to the response onset) for the differences EE minus Correct at Fz for both No-Delay and 400-Delay conditions, showing an increase in theta (3-8 Hz) and delta (1-3 Hz) frequency bands, replicating our previous findings on Experiment 1. Black dotted contours show the areas where rmANOVAs were computed. At the bottom are depicted the *t*-value maps at Fz ($p < .01$). On the right the topographical scalp maps are presented showing significant clusters at the theta and delta bands. The grey vertical segments indicate the time intervals that were separately subjected to statistical analysis -rmANOVAs- (see **Table S4** at section **7. ANNEX**). The grey horizontal segments correspond to the statistically significant cluster ($p < .05$) at the mass cluster permutation testing. Significant electrodes conforming the clusters are shown in white. EE: external errors; RL: Response locked. See also **Table S4** and **Figure S3** at section **7. ANNEX** for more details on rmANOVA results and time frequency data for Experiment 3.

Correlation between the P600 component and delta oscillatory activity (1-3 Hz)

Noteworthy, we replicated our findings at Experiment 1 when examining the relationship between the P600 component and delta power (1-3 Hz) at centro-posterior electrode locations (see **Figure 4F**), confirming a significant P600-delta activity positive correlation [both No-Delay and 400-Delay trials included: $r(36) = 0.29$, $p = .04$].

Interim conclusions Experiment 3

In Experiment 3 we replicated the findings of the first experiment regarding the appearance of the delta/P600 complex in the No-Delay condition. Interestingly, when introducing an A-O delay of 400 ms (temporal aPE) during EE, the same oscillatory pattern –although latency-dependent on the occurrence of the feedback onset- was observed.

3.1.4. DISCUSSION

The ability to monitor ourselves as agents of our actions is crucial for experiencing a self independent of the external world. In the present study, we examined how the multifaceted experience of being the agent of an action can be modulated by different types of aPE (content vs. temporal) and to which extent external cues are crucial to retrospectively infer the attribution of agency. Our findings show a specific delta/P600 component associated to external agency

attributions (content aPE, Experiment 1 & 3; see **Table 1)** whenever hand-movement incongruences were presented. Importantly, the amplitude of the delta/P600 component covaried with the subjective intensity of external agency judgements (**Figure 5A** and **5C**), but not so with the internal agency attribution category. Moreover, when introducing an A-O delay of 400 ms during content aPE (Experiment 3), a shift forward in time of the delta/P600 component was observed tagged to the feedback appearance (**Figure 7B**), highlighting the role of the external sensory information in the conscious reconstruction of agency attribution judgements. In contrast, during Experiment 2, in which we manipulated A-O temporal asynchronies eliciting temporal aPE but without introducing hand-movement incongruences (Farrer et al., 2008; Farrer, Valentin & Hupe, 2013; Sato & Yasuda, 2005), a frontal FCRP modulation associated to a was observed during delayed Correct responses. Moreover, delaying the appearance of the external feedback for SE during the 400-Delay condition was associated to the appearance of a frontal FRN-theta component.

Regarding the P600 modulation associated with external agency attribution judgements, a similar effect has been observed during language processing, generally related to structural reanalysis, repair or integration following syntactic violations or ambiguities (Brouwer, Fitz & Hoeks, 2012; Friederici, 2011). Likewise, the P600 component is usually associated with the P3b-like family components under the context updating hypothesis (Donchin & Coles, 1988), which states that the P300 may be a manifestation of the processes by which the mental model that one has of his/her available environmental information (context) is evaluated and updated whenever a relevant deviant stimulus is introduced. Importantly, the delayed appearance of the P600 effect comes to confirm its association with the retrospective aspects of the agentic experience (Moore & Haggard, 2008; Synofzik, Vosgerau & Newen, 2008), which depends on the computation and differential weighting of external (multimodal somatosensory, visual, and auditory reafferent representations), intentions (goals) and internal cues.

Noteworthy, a strong positive correlation between increased P600 amplitude and delta-band power at posterior parietal electrode locations was found, reinforcing the delta/P600 linkage (**Figure 5E**), as well as its common source and functional interpretation. Delta oscillations have been related to several cognitive functions, such as the motivational relevance of the task and the salience of the target stimulus (Knyazev, 2007, 2012), as well as with reward processing (Bernat, Nelson & Baskin-Sommers, 2015; Bernat et al., 2011; Cavanagh, 2015; Nelson et al., 2011) and commission of motor errors (Yordanova et al., 2004). Moreover, parietal delta oscillations are considered to mediate signal detection and decision-making (Başar et al., 1999; Schürmann et al., 2001), modulating the rhythmic gain of information accumulation (Lakatos et al., 2008; Schroeder

& Lakatos, 2009). These findings fit comfortably with 'two-stage' decision making models (Del Cul et al., 2009). Following this perspective, a fast, low-level non-conscious sensorimotor route allowing a rapid and rudimentary evaluation of evidence using internal cues, might be engaged providing us with the default, positive agency feeling or 'minimal self' (Gallagher, 2000). This stage would be followed by a slower, more accurate high-level conscious route devoted to integrating external information and related to evaluative and context updating processes. From this view, the delta/P600 component might reflect the engagement of the latter processing stage, during which an external judgement of agency occurs indicating the need to implement control processes and update of current mental models. In line with this interpretation, a significant positive correlation was observed between delta/P600 modulations and larger PES effects following EE, supporting the existence of a tight link between agency monitoring and the engagement of control/error-monitoring processes during the execution of a voluntary action (**Figures 5B** and **5D**). The delayed appearance of the delta/P600 component contrasts with the fast onset of the ERN component (see **Figure 4**) associated to the low-level non-conscious action monitoring route (Del Cul et al., 2009), and elicited as soon as an internal mismatch between the intended movement and the information from efference copy occurs (Nieuwenhuis et al., 2002; Rodriguez-Fornells et al., 2002; van Schie et al., 2004). Importantly, the appearance of the delta/P600 complex during the No-Delay condition was first found in Experiment 1 and afterwards replicated in Experiment 3. Most remarkably, when introducing an A-O delay of 400 ms during Experiment 3, EE continued to elicit the same oscillatory pattern although latency-dependent on the occurrence of the feedback onset. Overall, we believe current results suggest that the parietal delta/P600 activity might be elicited by content discrepancies between intended/predicted vs. actual states, reflecting judgments of external control that force the system to update the predicted mental model at work using agentic inferential processing.

Interestingly, the parietal topographical distribution of the delta/P600 component seems consistent with the role of the posterior parietal cortex (PPC) in generating and maintaining internal forward models (i.e., efference copy), detecting mismatches between intended-actual states (i.e., aPE), and monitoring voluntary actions (Andersen & Bruneo, 2002; Blakemore & Sirigu, 2003; Cui, 2014; Desmurget & Grafton, 2000). Several neuroimaging and lesion studies support this proposal, reporting PPC activations when experiencing external agency attributions following incongruent visual feedback for subject's actions (David et al., 2007; Farrer & Frith, 2002; Farrer et al., 2008; Leube et al., 2003; Schnell et al., 2007). Similarly, PPC lesions have been associated with experiencing external causal attributions of agency, for example in limb awareness deficits and alien hand syndrome patients (Assal, Schwartz & Vuilleumeier, 2007; Daprati et al., 1997; Leiguarda et al., 1993; Nightingale, 1982; Sirigu et al., 1999), as well as in anosognosia for

hemiplegia (i.e., paralysis denial), which is thought to be related to the impossibility to access the information derived from the external sensory feedback following damage to components of the frontoparietal network (Fotopoulou, 2014; Monai et al., 2020; Pacella et al., 2019). Besides, hyperactivation of the right PPC for schizophrenic patients was reported in a PET study when experiencing delusions of control (Spence et al., 1997).

In contrast to the findings reported in Experiments 1 and 3, our results in Experiment 2 showed a very distinct pattern when introducing temporal aPE (see **Figure 6**). This experiment was designed based on accumulative evidence showing a gradual decrease in the reported experienced agency as A-O temporal delay increases, with an agency threshold around 200/300 ms (Blakemore, Frith & Wolpert, 1999; Farrer et al., 2008; Farrer, Valentin & Hupe, 2013; Hara et al., 2015; Krugwasser, Harel & Salomon, 2019; Sato & Yasuda, 2005). As such, we inspected the temporal aPE introducing A-O delays of 0, 150 and 400 ms. Our results converge with previous ones showing a significant gradual reduction in the reported internal agency attribution of actions as the delay increases. Importantly, no effects on the external agency reports were encountered (in contrast to Experiment 1), highlighting the relevance of distinguishing between both agency facets (internal vs. external attribution) as already previously suggested (Pacherie, 2013; Synofzik et al., 2008).

Interestingly, correct responses executed during the 400-Delay condition elicited a frontal positive component approximately 200 ms after the feedback presentation (see **Figures 6A** and **6C**), which topographically resembles the FCRP or RewP component (Baker & Holroyd, 2011; Carlson et al., 2011; Foti et al., 2011; Holroyd et al., 2008; Walsh & Anderson, 2012), characterized by an increase in frontal delta power (Figure 6C and 6D). Several prior studies have linked the RewP/FCRP to affective processing (Gehring & Willoughby, 2002; Holroyd & Coles, 2002) and motivational states (Boksem et al., 2006; Hajcak et al., 2006) in addition to reinforcement learning signals and enhanced performance monitoring (Lange, Leue & Beauducel, 2012; Threadgill & Gable, 2016). The appearance of the FCRP during temporal aPE seems to point out to a reactivation of monitoring processes under temporal uncertainty, in which a diminished feeling of ownership and internal agency is present, and attention to external confirmatory cues is increased to check the appropriateness of the previous response. Additionally, this finding is also convergent with the observation of depressed phasic dopaminergic activity in midbrain dopaminergic neurons for violations of temporal prediction of rewards, which is observed when the expected reward is delayed after an action (Hollerman & Schultz, 2008). Supporting this idea, SE trials in the 400-Delay condition showed a frontally distributed negative deflection characterized by an increase of the medial-frontal theta power matching the FRN component

(Hajcak et al., 2006; Nieuwenhuis et al., 2004) (**Figure 6A**). This electrophysiological pattern is similar to the ones previously observed by Gentsch, Ullsperger and Ullsperger (2009) and Steinhauser and Kiesel (2011) following the omission of the expected feedback and insertion of technical malfunctions, respectively. This result could be interpreted as a reinstatement of the previous ERN (following the dopamine reinforcement learning view, Holroyd & Coles, 2002), maybe indicating a need to re-evaluate or re-update their internal model, although it is intriguing considering that participants already know that they have committed an error, as evidenced by the elicited ERN in all conditions. Importantly, no differences were observed for the ERN across A-O delay conditions, indicating the preservation of internal error monitoring mechanisms working independently of outcome-related information (**Figure 6A**). Therefore, it seems that A-O temporal delays act as agency cues, increasing uncertainty, affecting internal agency attribution, and increasing the chances of a possible external intervention or control (Pacherie, 2013), although not disruptive enough to produce an increase in the judgement of external agency attribution for those actions, at least in healthy population as the one tested here. In both cases, the presence of the FCRP and FRN in correct and erroneous trials with temporal aPE speaks in favor of increased uncertainty about correctness with delayed feedback and the re-activation of error monitoring processes. Further studies might be required to evaluate this hypothesis.

Some limitations regarding the design of our study and possible generalization of our results were encountered. First, the requirement of explicit reports in the form of a questionnaire to assess the participants' agency attributions might had induced a subjective bias in the results, due to its retrospective -so, interpretative- nature. Also, in Experiment 2 we introduced A-O temporal delays in a block-wise manner, which may have allowed action-effect delay adaptations influencing our results. Future studies may tackle these concerns by evaluating the internal/external agency experience at the single-trial level and introducing A-O temporal delays in a randomized manner, respectively.

Overall, the present pattern of results offers new evidence about workings of agency attribution mechanisms, distinguishing between self and externally generated actions and their underlying neurophysiological signatures, providing valuable insight regarding the integration of internal and external information, and emphasizing the importance of considering the many facets of agency and its inherent complexity.

STUDY 2

Decoding agency attribution using single trial error-related brain potentials

3.2. STUDY 2

Decoding agency attribution using single trial error-related brain potentials.

3.2.1. INTRODUCTION

In our changeable and uncertain world, the ability to monitor and evaluate our actions is crucial for self-regulation. Successful goal achievement requires the ability to distinguish between events caused by ourselves or by another agent, promoting behavioural adaptation whenever unintended action outcomes, such as errors, occur (Rabbitt, 1966). Voluntary actions are accompanied by the implicit and automatic feeling of being the authors of one's movements and their consequences, a feeling/experience known as the sense of agency (SoA). Several behavioural [Sato & Yasuda, 2005; Tsakiris et al., 2005, see David, Newen and Vogeley (2008) for a review] and neuroimaging studies [David et al., 2007; Farrer et al., 2003; Farrer & Frith, 2002, for a recent meta-analysis see Zito, Wiest and Aybek et al. (2020)] have addressed the cognitive architecture and neural basis for the SoA, highlighting the influence of prediction-outcome consistency on the SoA attribution (Frith, Blakemore & Wolpert, 2000; Synofzik, Vosgerau & Newen, 2008; Wolpert & Miall, 1996). Importantly, the ability to distinguish self vs. externally generated actions lies at the heart of the concept of agency and individual responsibility in our society, concerning important issues such as moral and legal status of actions (Haggard & Tsakiris, 2009).

External factors, such as mechanical issues, other human agents' interventions or environmental elements can be favourable or unfavourable for successful goal achievement, therefore, influencing the processing of action selection and adaptation. Intriguingly, as a result of the new improvements in biomedical and biotechnology research, we have begun to design new brain-computer interfaces (BCI) to control other bodies (e.g. avatars and robots). For example, BCI systems are being increasingly used as assistive devices in neurorehabilitation [see Tonin and Millán (2021) for a recent review on BCI for robotic devices]. Stroke patients with motor impairments may benefit from these technological developments, using BCI-based exoskeletons (Lui et al., 2017; Zhang et al., 2018), prosthetics (Rotermund, Ernst & Pawelzik, 2007), spellers (Manyakov, 2012; Margaux, 2012) or robotic systems (Bhattacharyya, Konar & Tibarewala, 2014; Rakshit, Konar & Nagar, 2016). These new interactions confront us with new challenges regarding body(ies)-mind relationships that raise important issues concerning the moral and legal status of actions. From a classical legal definition (English Habeas Corpus Act, 1675), we are fully responsible and have authority and property rights over our own body. In this sense, the actions

of our own body are attributed to our agency and we therefore have direct legal responsibility for those actions. It is, however, more questionable the extent to which our feeling of agency and legal consequences would be the same when governing different bodies, when our surrogate body might take independent decisions or carry out actions that cause errors with drastic consequences that were not planned or caused directly by you.

Several ERP studies have already reported that erroneous actions elicit distinct neural responses (family of error-related potentials, i.e., ErrP), mostly using measures of the average time course of brain EEG activity. For example, the Error-Related Negativity (ERN) (Falkenstein et al., 1990; Gehring et al., 1993; Taylor, Stern & Gehring, 2007), is a very robust and reliable negative deflection observed at frontomedial locations during response-locked averaged EEG recordings, and which is elicited immediately after an error has been committed (50-100 ms after error commission). Additionally, a similar ERN modulation has also been recorded during the observation of incorrect actions performed by another agent (i.e., 'observational' ERN), depicting a lower amplitude than the ERN for self-generated errors and a later occurrence (van Schie et al., 2004). Interesting, our group also identified recently different ErrP signatures when participants observed external errors induced in their own body (embedded in an avatar, Padrao et al., 2017; a negative modulation at 400 ms) or observing their own hands committing an error that was not their own (Gomez-Andres et al., 2022a). In this last case, a positive modulation (referred here as P600) was observed for external vs. self-errors (Gomez-Andres et al., 2022a).

Crucially for the purpose of the present article, it has been previously observed that ErrP can be reliably decoded on a single-trial basis (Chavarriaga, Sobolewski & Millán, 2014; Iturrate et al., 2015; Kim et al., 2017; Usama et al., 2021; Zander et al., 2016; see Kumar et al., 2019 for a review about decoding ErrP), thus allowing their implementation for BCI systems, decoding the users' intentions from his/her neural activity. Subsequently, the decoding of the user's perception that an error has occurred in the form of ErrP can allow the system to undergo corrective actions, for example, by preventing the erroneous action from being completely executed (Dal Seno, Matteucci & Mainardi, 2010; Ferrez & Millán, 2008; Schalk et al., 2000) and/or reducing the possibility of errors reappearing in the future through re-calibration of the system (Artusi et al., 2011; Llera et al., 2011).

Nevertheless, despite these recent additions, most of the literature in the field concerns the classification of errors against correct actions or observing external agents committing errors, rather than the classification of the agency of errors regarding my "own actions" ("mine" vs. "externally" induced error). With this purpose in mind, we applied a linear SVM classifier to previously acquired EEG data, considering for the analysis not only the latency and amplitude of

the key ErrP features namely, the ERN or the P600 components [based on Gomez-Andres et al. (2022a)], but also the topographical distribution (considering all the electrodes and time points] of the brain's response to each agency-error condition. Our main aim was to assess to which extent single-trial EEG activity contains reliable decodable information about the agentic causal attribution of errors ("mine" vs. "external"). The possibility to access this information at the level of a trial (each real action) using decoding techniques could allow in the future to monitor brain activity while interacting with external agents and circumvent some of the problems regarding moral responsibility commented above.

3.2.2. MATERIALS AND METHODS

Participants

A sample of 25 healthy participants (graduate students) were paid to participate in the original study (Gomez-Andres et al., 2022a). Two participants were excluded from the analysis due to a low number of self-generated (SE) error trials (24 and 23, respectively). The final sample consisted of 23 participants (15 females, M = 24.2 years ± 4.2, mean ± SD). All participants were naïve with respect to the aim of the experiment. The procedures of the experiment were approved by the Biomedical Research Institute of Bellvitge (IDIBELL) and Hospital Universitari de Bellvitge ethics committee (CEIC, Ref. PR254/15). Informed consent in accordance with the Declaration of Helsinki was obtained from all participants prior to the commencement of the study.

Apparatus and experimental task

An apparatus inspired by the Rubber Hand Illusion (Kalckert & Ehrsson, 2012, 2014) and the Nielsen's (1963) paradigm was built (see **Figure 1**). The experiment was performed inside a Faraday chamber with a full HD 24.5-inch monitor displaying the experimental task at 200 Hz refresh rate. The monitor was mounted on a wooden stand and adjusted to the subjects' body, oriented with an inclination of 30º on the horizontal plane (**Figure 1A**). Participants were asked to put on a pair of white latex gloves to remove any morphological cues that could affect Self-identification and to place their hands on top of the wooden stand surface (hidden from their view due to the monitor overlap), where two fixed joysticks with a button at the top were attached. After general instructions were given, the EEG cap was set up and the state of each electrode was checked. Finally, the room lights were turned off for the realization of the experiment.

The experimental task consisted of a modified version of the Eriksen Flanker task (Padrao et al., 2016; Rodriguez-Fornells, Kurzbuch & Münte, 2002) (see **Figure 1B** and Step 1 in **Figure 2**). Stimulus presentation was controlled with EPrime (Psychology Software Tools Inc., Pittsburgh,

PA). Participants were instructed to focus on a central target arrow from a visual array of three arrows allocated vertically and to respond as fast and accurately as possible with the right or left hand (i.e., thumb press on button placed on the high-end of the fixed joystick), depending on the directionality of the target arrow. The flanker arrows located above and below the target arrow were either compatible (i.e., all arrows in the same direction) or incompatible (i.e., flanker arrows pointing in the opposite direction than the target arrow) with respect to the target arrow. To optimize the number of errors, a proportion of 40-60% of compatible and incompatible trials were presented, respectively, in a randomized order.

For each trial, a pair of life-sized hands (i.e., real hand photographs -adult size- wearing the same white latex gloves as the participants') mimicked the participants' hands actions at a coherent position with respect to the participants' hands and body posture. The duration of the target presentation was fixed to 150 ms, followed by a response threshold of 1000 ms (Reaction Time, RT). At the same time the participants were responding, the 'virtual' hands provided the participants with real-time, online feedback (Observed Response) for 100 ms. A variable fixation slide (depending on the RT) appeared at the end of the trial (see **Figure 1B** for an example trial).

The experimental task was divided in two conditions: (i) a standard condition (one block of 160 trials), during which congruent feedback was presented in all cases (i.e., the 'virtual' hand response was always the same as the participant's response) and (ii) an error induction condition (two blocks of 320 trials each, 640 trials in total), where incongruent feedback was provided in 10% of the trials (64 trials in total). During the incongruent trials, when the participants pressed the button with one hand, the opposite 'virtual' hand performed the response movement, causing an external error (EE). We avoided introducing EE in incompatible trials to avoid the pre-activation of incorrect motor channels responsible for the increase of error rates in the incompatible trials (compatibility effect). Every 80 trials a block pause of 10 seconds was presented. A training phase (20 trials) was always performed before the experiment began to ensure an adequate speed-accuracy rate.

EEG recording

The electroencephalographic (EEG) signal was recorded from the scalp using sintered Ag-AgCl ring electrodes mounted in an elastic cap (Easycap, International 10–20 System locations) and located at 27 standard positions (Fp1/2, Fz, F3/4, F7/8, Fc1/2, Fc5/6, Cz, C3/4, Cp1/2, Cp5/6, T7/8, Pz, P3/4, P7/8, O1/2). Biosignals were referenced on-line to the right mastoid electrode and posteriorly re-referenced off-line to half of the signal acquired on the left mastoid electrode.

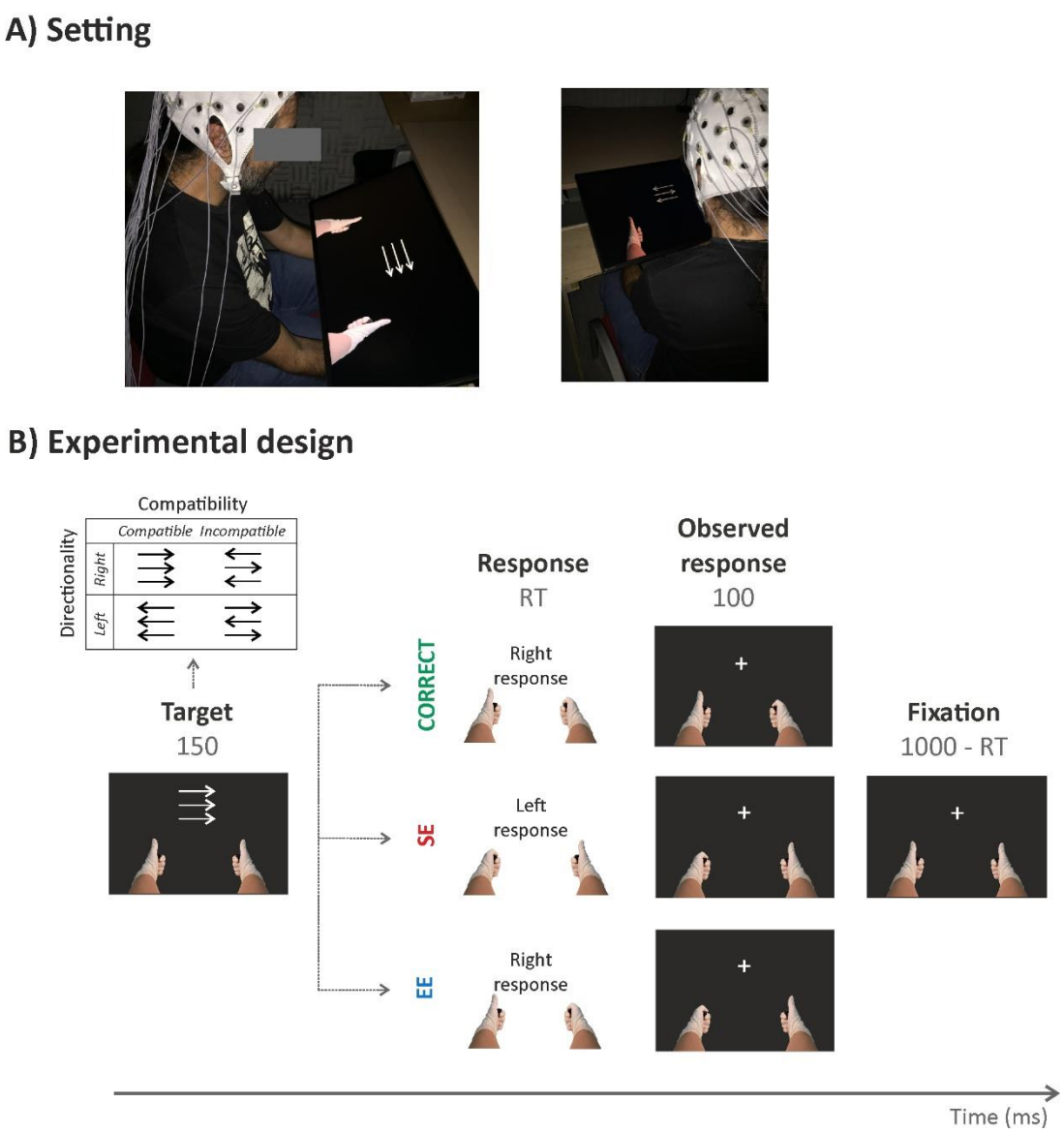


Figure 1. Experimental setting and task. A) Schematic representation of the experimental apparatus. Lateral (left) and bird-view (right) perspective of the participants' position with respect to the wooden-mounted monitor. The participant's hands are hidden underneath the monitor, in a coherent position with respect to the 'digital' hand displayed, holding the response joysticks, and the visual feedback displayed on the monitor. B) Experimental paradigm depicting a modified version of the Eriksen Flanker task (Padrao et al., 2016; Rodriguez-Fornells et al., 2002). All trials started with the target presentation (150 ms) followed by a response threshold (<1000) in which participants had to respond to the target arrow as fast as possible (RT: Reaction Times). The visual feedback (Observed Response, OR) corresponding to the virtual hand response movement was displayed for 100 ms, showing either congruent correct, congruent error or incongruent correct visual displays.

Electrode impedances were kept below 5 kΩ. For all experiments, vertical eye movements were monitored with an electrode at the infraorbital ridge of the right eye. The electrophysiological signals were filtered online with a notch-filter (50 Hz) and a high-pass filter (0.016 Hz) and digitized at a rate of 250 Hz. Participants were given instructions about how to reduce muscle artefacts by minimizing movement and to wait for a visual signal, an array of five asterisks appearing every 10 trials, to free blink for 5 seconds.

Data and statistical analysis

Behavioural measures

To inspect the behavioural effects during the task execution we computed the RT and accuracy rates for all trial types (Correct, SE and EE), exploring the compatibility effect (Danielmeier & Ullsperger, 2011; Eriksen & Schultz, 1979), in terms of RT and accuracy rates.

Also, we examined the subjective experience of agency after the standard and error induction blocks using a 7-item questionnaire (in Spanish) (see **Table 1**). After each experimental condition, participants were asked to rate their subjective experience focusing on the experience of Agency attribution, which partially depended on the sensorimotor input resulting from the action performed (Haggard, 2017). The questionnaire was designed to address internal (Q1 and Q2, i.e., 'Most of the time, the movements of the digital hands seemed to be my own movements' and Q2) and external attribution of actions (Q3 and Q4, i.e., 'It sometimes seemed as if the errors were not caused by me'). Additionally, two control questions (Q5 and Q6, i.e., 'It seemed as if I had more than two hands') and one item addressing the Sense of Ownership (SoO, Q7: 'I felt as if the digital hands were my hands') were also included. Participants were asked to rate their level of agreement with these 7 statements using a 7-level Likert-type response, ranging from "strongly disagree" (1) to "strongly agree" (7). Wilcoxon test (pairwise comparisons) were employed for testing the possible differences regarding the participants' scores on the questionnaire, with the significance alpha level adjusted to multiple comparisons.

EEG data

EEG analyses (Step 2 in **Figure 2**) were conducted using routines taken from the ERPLAB toolbox V6.1.4 (Lopez-Calderon & Luck, 2014) and custom routines from MATLAB (The MathWorks, Inc. Natick, MA). A high-pass filter of 0.1 Hz (second order Butterworth filter, 12-40 dB) was applied to the raw EEG data. Epochs of -350 to 1050 ms were defined, time-locked to the onset of the participant's response, and baseline-corrected to its preceding 100 ms. To perform artefact rejection, we excluded epochs with step-like artefacts when the amplitude jumps in the electro-

oculograms exceeded 25 μV (moving window = 400 ms, moving step = 10 ms) or in which activity was ±100 μV in any channel. No additional filtering was applied for the subsequent analyses.

Table 1: Item description of the subjective experience questionnaire examining agency attribution.

Subjective feeling	*Questionnaire item*	*Description*
Internal attribution	*My movements*	**Q1**. Most of the time, the movements of the digital hands seemed to be my own movements.
	Feeling of control	**Q2**. I felt I could control the movements of the digital hands most part of the time.
External attribution	*Not my movements*	**Q3.** Sometimes, the digital hands seemed to be moving by themselves.
	external errors	**Q4**. It sometimes seemed as if the errors were not caused by me.
Control	*Influence*	**Q5.** Sometimes I felt as if the movements of the digital hands were influencing my own movements.
	More than 2 hands	**Q6.** It seemed as if I had more than two hands.
	My hands (SoO)	**Q7**. I felt as if the digital hands were my hands.

SoO: Sense of ownership.

To study the multichannel evoked potentials, we visualized the superimposed activity of all the electrodes at once using a Butterfly plot, and the Global Mean Field Power (GMFP) was calculated (Hamburger & Van der Burgt, 1991; Lehmann & Skrandies, 1980; Skrandies, 1990). The GMFP corresponds to the standard deviation of the activity across all the Butterfly plots and has been shown as a reference-independent descriptor of the potential field (Skrandies, 1990).

Decoding

The aim of this work was to decode, at a single-trial level, the time-series of the ERPs to be able to separate the different defined conditions (Correct vs SE vs EE), and to study at which time points these three conditions were separable by a supervised learning model such as a Support Vector Machine (SVM) classifier. This decoding procedure was conducted using scikit-learn routines from Python (Pedregosa et al., 2011), and visualization of the results was done using custom routines from MATLAB. See **Figure 2** for a pipeline on the classification algorithm.

To ensure the predictive performance of the decoding procedure, the number of trials per condition was balanced (Step 3 in **Figure 2**). Since the number of EE trials was set beforehand, the limiting condition was the SE. In that sense, the participant who had less SE trials had 31, therefore, we considered 31 trials per condition and subject. For the participants who had more than 31 trials for any condition, we randomly selected 31, resulting in a total of 93 ERPs for each subject. The 70% of these trials (N_{train} = 65) were used to train the classifier, and the 30% of the trials (N_{test} = 28) were used in the test-phase (Step 4 in **Figure 2**). The proportion of trials of each class/condition (33%) was maintained in both subsets.

A SVM classifier with a linear kernel was considered for the decoding procedure (Step 5 in **Figure 2**). For each subject and time point, a single multiclass SVM was fitted using the training trials, and the accuracy of the model was calculated using the testing trials. For each classifier, the hyperparameter C was fine-tuned using a cross-validation method of the training subset with K = 5. The input of the classifier $SVM_{s,t}$ was the amplitude (in μV) from all the electrodes of the subject s at the time point t. To discriminate the classes, the classifier not only considered the amplitude of the EEG signal, but also the distribution of the EEG signal across the scalp.

Once the $SVM_{s,t}$ classifier was trained, the labels of the testing trials were predicted (Step 6 in **Figure 2**), and the performance of the model was calculated (Step 7 in **Figure 2**). To calculate this performance, we used the confusion matrix, which considers both the predicted and the real labels, the F1-score, and the accuracy. To consider a prediction as correct, the predicted and the real labels had to match, providing a very strict assessment of decoding. To compare the time-point accuracy against the chance performance (i.e., 0.33 because we defined three conditions), we used a non-parametric cluster-based Monte Carlo simulation analysis (Bae & Luck, 2018; Groppe, Urbach & Kutas et al., 2011; Maris & Oostenveld, 2007) with 10,000 permutations. This analysis allowed the authors to detect the clusters of contiguous time-points for which the performance of the classifier is above the chance ($p < 0.05$), and to obtain a cluster-level *t* mass. If the obtained cluster-level t mass was larger than the maximum of the Monte Carlo cluster-level t mass, we reported $p < 10^{-4}$. We rejected the null hypothesis (H0: the classifier performance was not different to chance performance), for any cluster-based *t* mass above the top 95% of the null distribution (critical *t* mass = 21.33, one-tailed, $a < 0.05$; see Bae & Luck, 2018 for further details on this analysis).

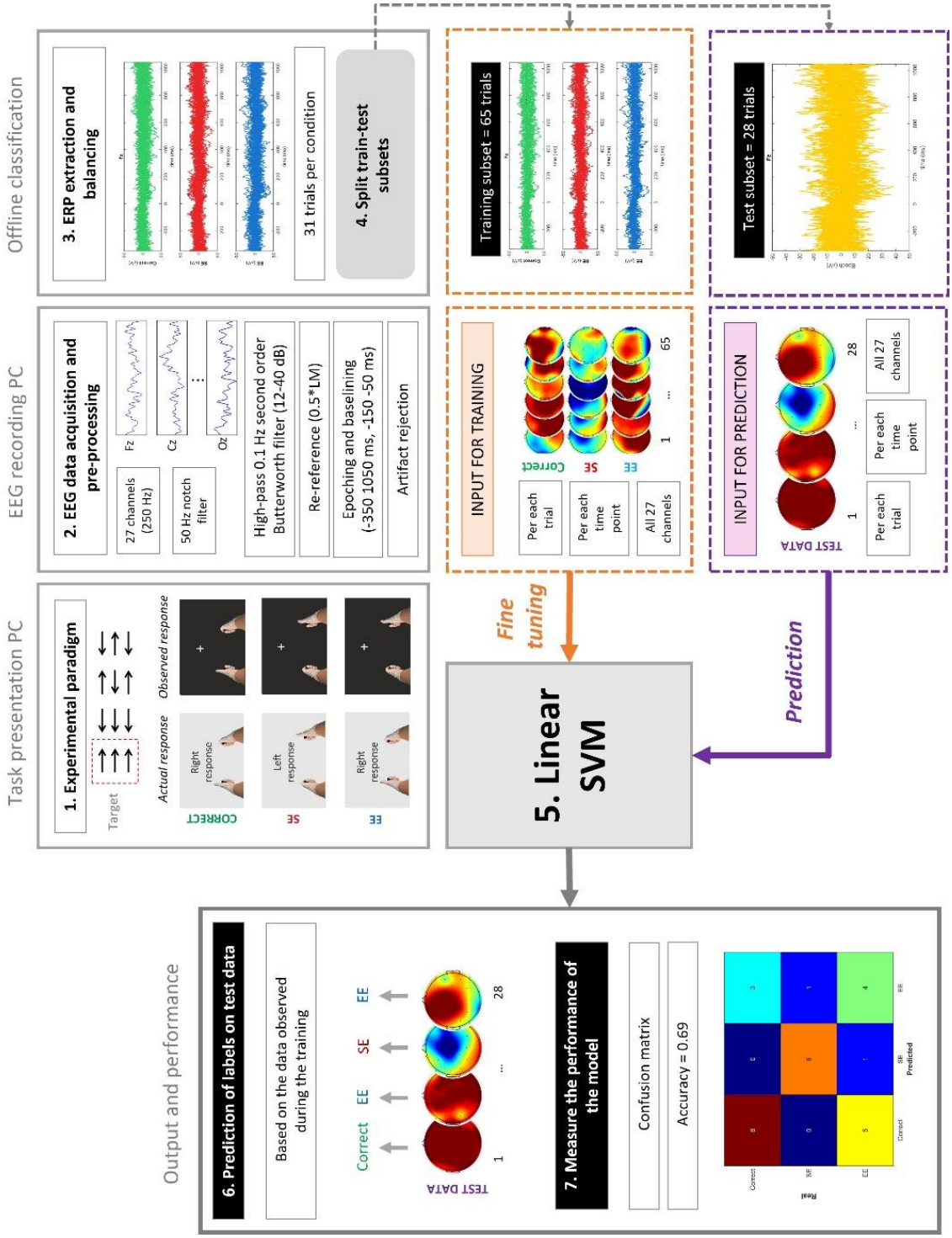


Figure 2. Classification pipeline description. EEG data previously acquired (Step 1) was pre-processed (Step 2). We extracted the ERPs and balanced the number trials across conditions (31 trials per condition) (Step 3) and split the data into the training and test subsets (Step 4). The training subset was used to fine tune the linear SVM classifier (Step 5). We then applied this fine-tuned SVM classifier to predict the labels of the test data (Step 6). Finally, we examined the output and performance of the classifiers, in terms of accuracy and confusion matrices.

Moreover, to analyse the confused classes/conditions at a specific time point, we used the confusion matrices at this time point to calculate a confusion metric for each pair of classes (Correct vs SE, Correct vs EE, and SE vs EE). This metric, for a given pair of classes (c1, c2), is defined as (TPc1+TPc2)/(FPc1/c2+FPc2/c1), where FPc1/c2 are the c2 trials predicted as c1 and the FPc2/c1 the c1 trials predicted as c2. This metric is between 0 and 1, and the higher the metric, the lower confusion between the pair of classes. Since this metric was calculated at each time-point, it allowed us to capture how the different pairs of classes were confused along the time.

Finally, since the GMFP captures the variability of the data at a certain time point by quantifying the amount of activity at each time point in the field considering the data from all recording electrodes, we inspected the association between the GMFP and the accuracy of the decoding procedure by means of Pearson correlation (significance level was set at $p < 0.05$). Since the times of GMFP maxima are used to determine the latencies of the relevant ERP components (Skrandies, 1990), we expected to find a positive correlation between the GMFP and the decoding accuracy driven the topographical and latency-dependent characteristics of the ErrPs.

Statistical analysis

Categorical variables were reported as absolute values, while continuous variables were reported as the mean ± either the standard deviation (SD) or the standard error of means (SEM). The normality distribution of the data was checked using the Shapiro-Francia test (function sf.test from nortest package) and visual inspection.

All statistical analyses were conducted in R (Version 3.6.0, R Core Team, 2019. https://www.R-project.org). The relationship between continuous variables was assessed using Pearson correlations. To compare groups of two factors, for the continuous normal-distributed data, two-sided, unpaired t-tests of equal variances checked by the Levene's test (R car package) were used and both the t- and p-values were reported (R stats package). For groups of more than two factors or several levels of interaction, an ANOVA for balanced designs was used, reporting the F- and *p*-values. Moreover, when necessary, p-values were corrected for multiple comparisons (p_{adj}) using the Tukey Honestly Significant Difference. Finally, for the group's comparisons, the effect sizes were reported, i.e., Cohen's *d* for the *t*-tests and η_p^2 for the ANOVA.

3.2.3. RESULTS

Behavioural results

The participant's performance was as expected for this paradigm, with a mean percentage of self-generated errors (SE) approx. of 9% ± 1 (mean ± SEM). A compatibility effect (Danielmeier &

Ullsperger, 2011; Eriksen & Schultz, 1979) was encountered, with participants responding more accurately [percentage of SE during compatible vs. incompatible trials, 13.4% ± 1.1 vs. 86.6% ± 1.1, respectively; $t(22) = 32.78$, $p < .0001$, $d = 13.8$] and faster [mean RT for correct: 282 ± 4 ms vs. 299 ± 5 ms; $t(22) = 9.73$, $p < .001$, $d = 0.8$] during compatible vs. incompatible trials. Moreover, participants also showed significantly faster RT for SE compared to correct responses [mean RT correct: 290 ± 4 ms vs. mean RT SE: 236 ± 5 ms; $t(22) = 21.7$, $p < .0001$, $d = 2.4$]. Altogether, these results indicated a correct implementation of the Eriksen flanker task.

In relation to the SoA experience, the insertion of an EE in the error induction blocks lead to an external attribution of the actions, revealing significant differences between the standard and error induction blocks for both external attribution questions: Q3 ("Sometimes, the digital hands seemed to be moving by themselves") ($Z = -3.64$, $p < .001$) and Q4 ("It sometimes seemed as if the errors were not caused by myself") ($Z = -3.64$, $p < .001$), revealing a significant external attribution judgment when introducing EE but without affecting the internal attribution ratings (Q1: $Z = -1.75$, $p = .08$; Q2: $Z = -2.26$, $p = .024$). High levels of SoO were observed for both conditions, and no other significant differences were found for any of the other control measures.

Error-related brain potentials (ERPs)

Figure 3A shows the average waveforms for Correct (green line), SE (red line) and EE (blue line) trials time-locked to the participants' response. During SE, the well-known error-related negativity (ERN) component was elicited, peaking at around 80 ms after error commission. The ERN's topography depicted a fronto-central distribution, with its maximum activity at Fz electrode locations (see **Figure 3B**). On the contrary, the insertion of EE elicited a more latter and posterior P600 component, peaking at 580 ms after the EE induction evidencing a centro-parietal topographical distribution (**Figure 3B**, EE) (Gomez-Andres et al., 2022b). Although grand average waveforms across participants are depicted, single-subject data were entered to the classification algorithm.

The multichannel evoked potential analysis (**Figure 3C**) indicates two time periods where the variability of the electrodes' activity is increased (the GMFP is higher). The first component is sharp and peaks around 95 ms, and the second one is smoother and peaks around 585 ms. In both cases, these correspond clearly to the ERN and P600 periods previously identified.

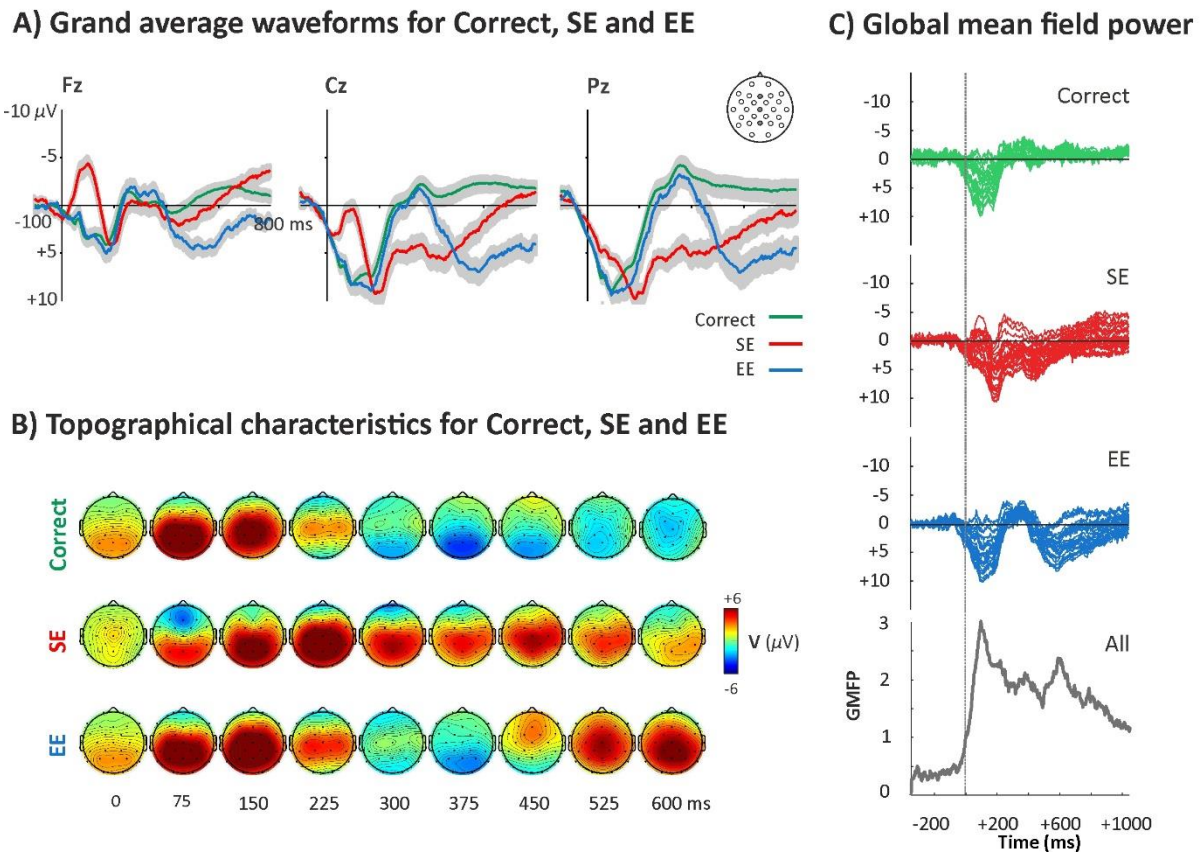


Figure 3. Event-related potentials (ERPs) and global mean field power (GMFP) measures. A) Response-locked grand-average waveforms for Correct, self-errors (SE) and external errors (EE) at Fz, Cz and Pz electrodes. In grey we show the standard deviation (SD). **B)** Topographical characteristics for Correct, SE and EE trials from 0 - 600 ms. **C)** Butterfly plots for Correct, SE and EE, and the global mean field power (GMFP) for All the electrodes at each time point.

Decoding

Cluster permutation analysis

The linear SVMs were fitted for each participant and time-point, leading to 8,050 (23 subjects x 350 time-points) classifiers. For each $SVM_{s,t}$ classifier, a confusion matrix was obtained, and the subject's mean accuracy along the time was calculated (**Figure 4**). The classifier performed better than chance when it was able to distinguish at least one class/condition over the others. During two different intervals of time this fact occurred (2 clusters): from -158 to -62 ms ($p < 10^{-4}$), and from 34 ms to the end of the ERP ($p < 10^{-4}$), being the accuracy of the first interval (0.38 ± 0.01, mean accuracy ± SEM) lower than that of the second one (0.44 ± 0.01; $t(44) = -5.90$, $p = 4.74 \times 10^{-7}$, $d = -1.74$).

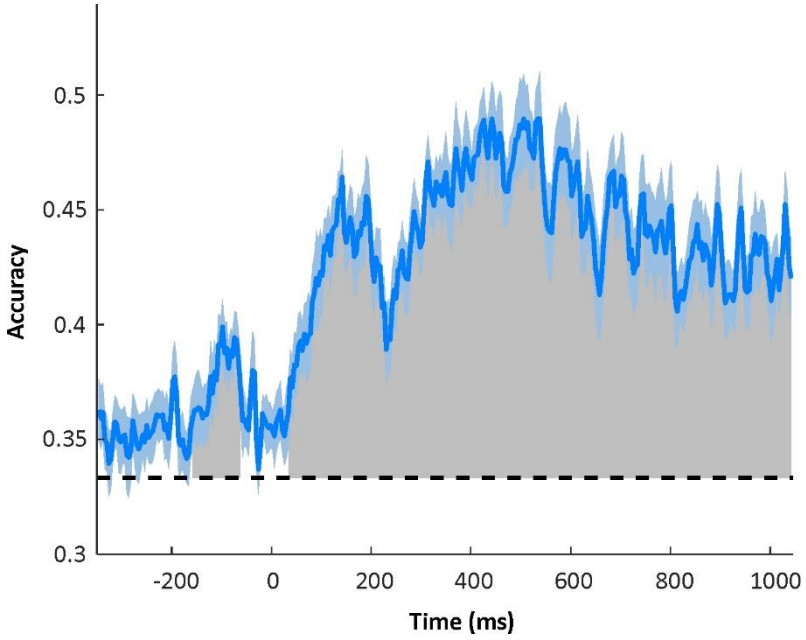


Figure 4. Classification accuracy over time. Graphical representation of the accuracy (ranging from 0 to 1) for the time period -350 ms to 1050 ms. In blue we show the mean accuracy for all participants and for every time point, and the SD (bluish shadow). In grey we highlight the presence of significant clusters after performing the non-parametric cluster-based Monte Carlo simulation analysis (Bae & Luck, 2018; Groppe et al., 2011; Maris & Oostenveld, 2007) with 10,000 permutations. Black dashed line indicates the chance level (0.33).

Confusion metric

Figure 5A shows the confusion metric along the time between all the pairs of classes. The confusion metric suggests that the significance of the first interval can be explained by the proper classification of the Correct and SE classes, i.e., the Correct vs SE, and the SE vs EE confusion metrics were increased during this period, while the Correct vs EE confusion metric was slightly increased. This first interval is followed by an interval of 100 ms where the performance of the models is not better than chance. Afterwards, during the second interval, two processes can be differentiated. During the first one (which starts at 35 ms approx.), the SE class pops-up over the rest and roughly corresponds to the ERN component time-window). During the second one, starting at 300 ms, Correct and EE classes become more distinguishable.

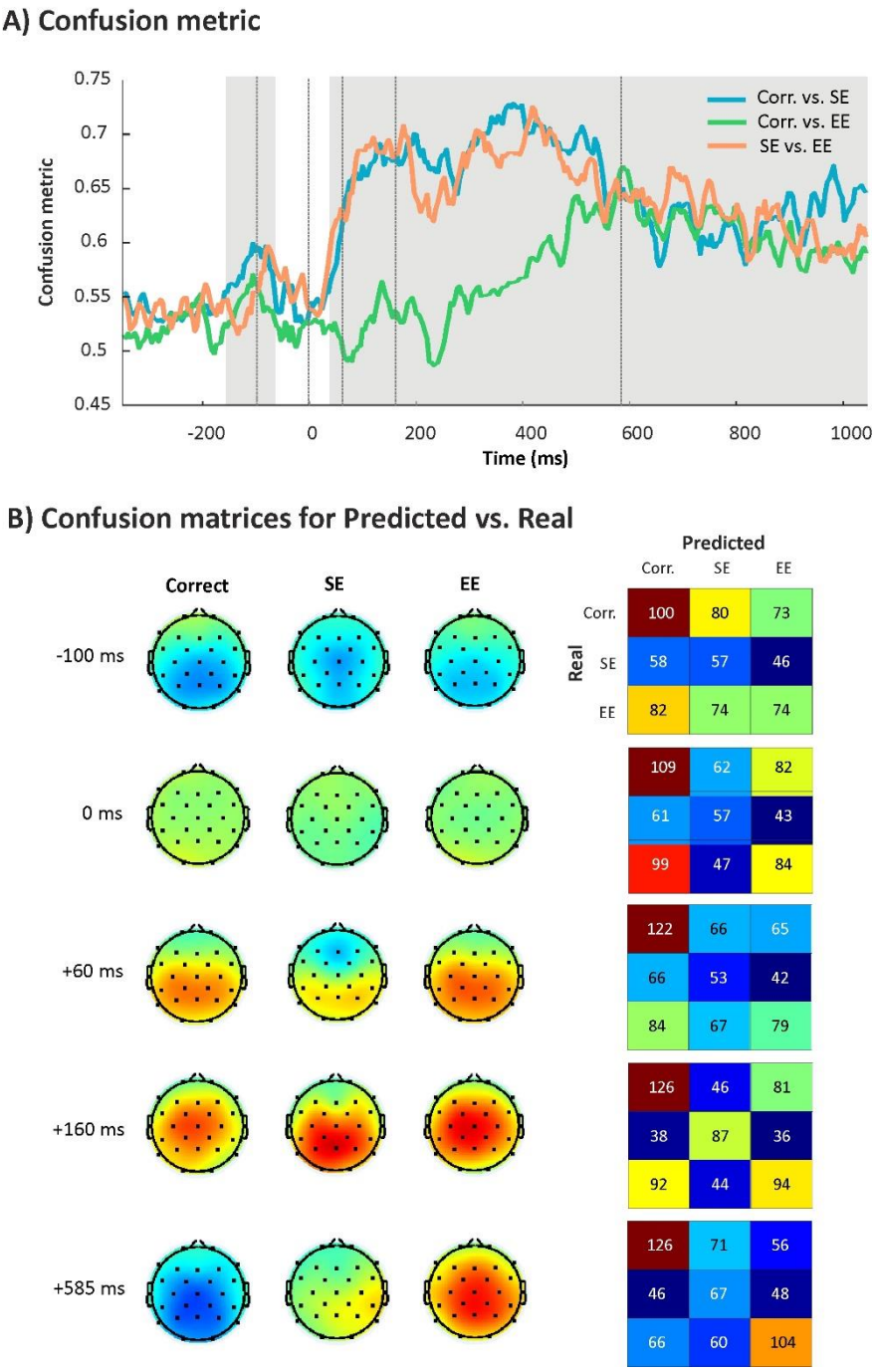


Figure 5. Confusion metric and matrices. A) Confusion metric (ranging from 0 to 1) along time for the pairs Correct vs. SE (blue line), Correct vs. EE (green line) and SE vs. EE (orange line). Grey vertical dotted lines indicate the time point were topographies and confusion matrices are depicted in B. **B)** Topographical representations of Correct, SE and EE at several time points of interest (left) and confusion matrices for Correct, SE and EE for predicted vs. real trial classifications (right).

This observation is strengthened by both the topographies and confusion matrices shown in **Figure 5B**. At 0 ms, the topographies between the three classes are very similar, which leads to classifiers that do not perform better than chance (high number of mismatches). In contrast, as the time is increased (bottom panels of **Figure 5B**), the condition topographies become more distinguishable, with a maximum differentiation between Correct and EE classes at 585 ms (corresponding to the P600 period).

Individual performance accuracy

The individual differences on the decoding performance are reported in **Table S5** (see section **7. ANNEX**), where the accuracy of the decoding procedure strongly depends on the subject (the subject accuracy, averaged during the significant intervals, ranges from 0.38 for S10, S14 and S22 to 0.53 for S06). Although individual differences are noticeable, **Figure 6** shows that, as expected, the subject's mean accuracy along time and the GMFP have a very strong correlation (r = 0.7419, p = 4.88 x 10^{-62}), indicating that the higher the GMFP, the higher the accuracy of the model.

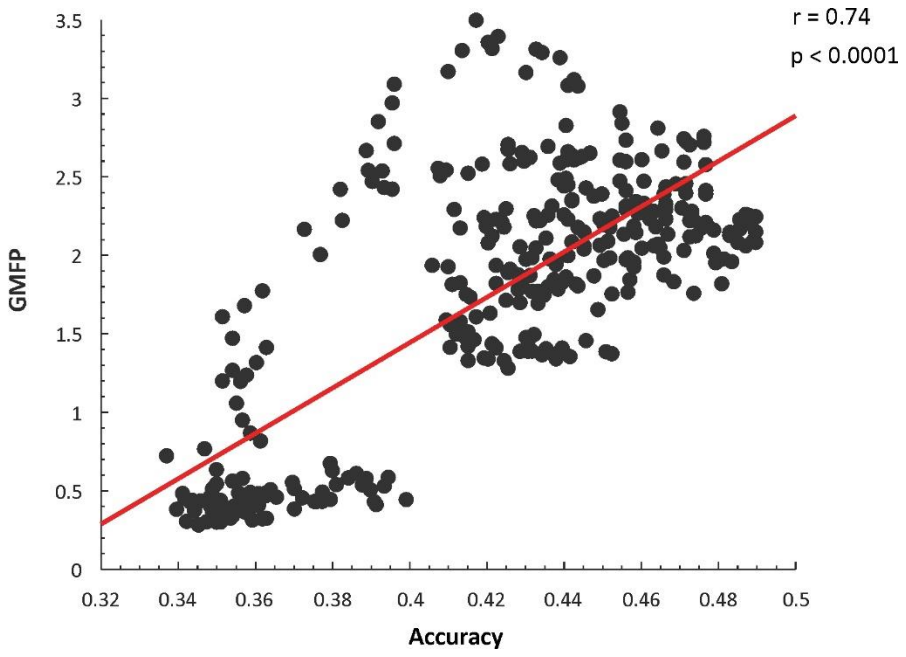


Figure 6. Correlation between global mean field power (GMFP) and accuracy of the SVM decoder. Pearson correlation for GMFP and accuracy measures depicting a significant positive association between an increase GMFP and increase in model accuracy.

F1-scores

In **Figure 7**, we analysed the mean F1-score, which indicates how well a class/condition is classified, at these two components time-windows (ERN: 95 ± 25 ms, and P600: 580 ± 50 ms). An analysis of variance model was fitted, considering the component time-windows (ERN vs P600) and the condition (Correct vs SE vs EE) factors on the F1-score. A subsequent ANOVA test indicated a main effect of Component ($F(1) = 6.06$, $p = 0.0151$, $\eta p^2 = 0.04$) and Condition ($F(2) = 4.05$, $p = 0.0197$, $\eta_p^2 = 0.06$), together with an interaction between these factors ($F(2) = 8.90$, $p = 0.0002$, $\eta_p^2 = 0.12$). Post-hoc analyses only reported significant differences between SE and EE during the elicitation of the ERN ($p_{adj} = 0.0130$).

At later stages of processing (∼580 ms – P600 period), the classification of the Correct and EE conditions improved, being their F1-score higher at P600 than at ERN (F1-score(SE) = 0.41 ± 0.02, F1-score(Correct) = 0.51 ± 0.02, and F1-score(EE) = 0.46 ± 0.02; p_{adj}(Correct) = 0.0229, p_{adj}(SE) = 0.3349, p_{adj}(EE) = 0.0266). Interestingly, the Correct condition was the best classified at the P600 latency, with significant differences between Correct and SE ($p_{adj} = 0.0029$), but not between Correct and EE ($p_{adj} = 0.3357$), and SE and EE ($p_{adj} = 0.4774$).

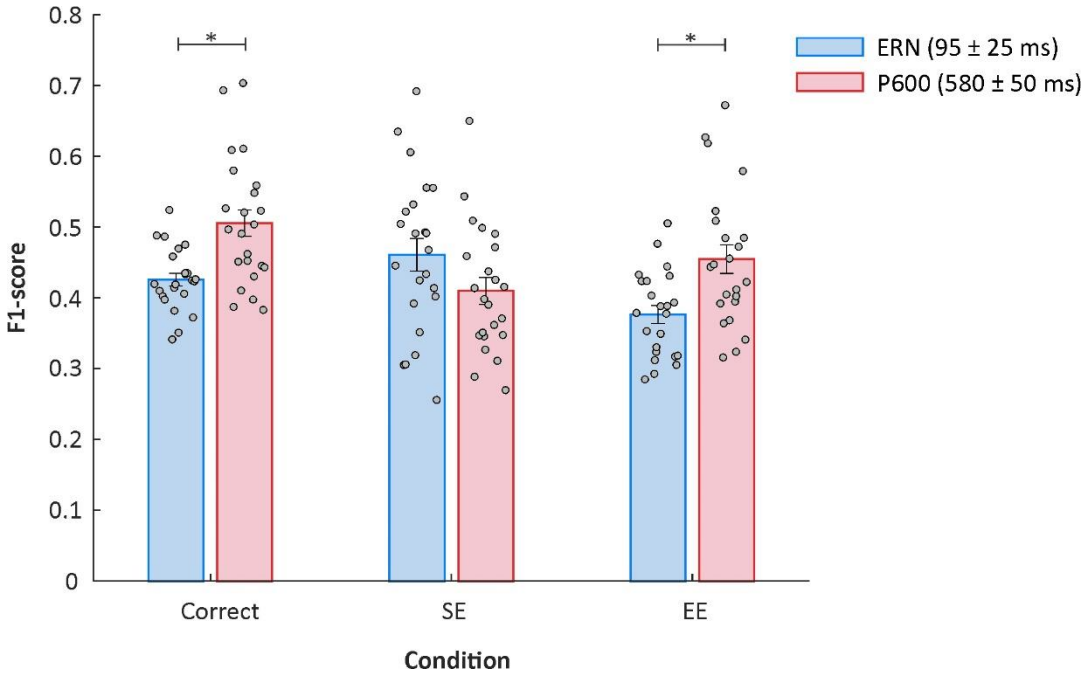


Figure 7. Mean F1-scores at the ERN and P600 time windows. Graphical representation of the F1-scores for the ERN and P600 time windows depicting the significant differences between conditions (Correct/SE/EE). During the ERN time window, significant differences were encountered between SE and EE. Significant differences between time windows were observed for Correct and EE classes. SE: self-generated errors, EE: externally generated errors.

3.2.4. DISCUSSION

In the present study, our aim was to investigate the feasibility of applying a machine learning decoder to decipher ErrPs from an EEG experimental paradigm. As previously mentioned in the introductory section, the ability to distinguish self vs. externally generated actions is a key factor influencing learning and adaptive behaviour and is crucial for agency inference. Here, we classified three types of trials, namely, Correct, SE and EE on a single-trial basis using a linear SVM model based on the ERP amplitude, latency, and topographical distribution. Significantly, our results showed that the classifier performance was better than chance from +34 ms onwards, showing two distinguishable peaks of accuracy overlapping with the latencies of the ERN during self-made errors and the P600 for externally induced errors.

As previously stated, ErrPs can be observed at the single-trial level allowing us to distinguish correct vs. erroneous actions (Chavarriaga et al., 2014; Iturrate et al., 2015; Kim et al., 2017; Usama et al., 2021; Zander et al., 2016). In the present study, the existence of two dissociable components for SE and EE was confirmed by the GMFP measures, suggesting the existence of two main ERP components, the first one peaking at 95 ms, and the second one at 585 ms, approximately (Gomez-Andres et al. 2022b). Considering the latencies of these two components, it is feasible to say that they correspond to the ERN (for SE) and P600 (for EE) components (see **Figure 3**).

Our results from the linear SVMs (fitted for each participant and time-point) evidenced a performance above chance during two clusters of time, the first one ranging from -158 to -62 ms and the second one from +34 ms onwards. When looking at the confusion metric and matrices at the different time points between all the pairs of classes, the first significant cluster was driven by the correct classification of Correct and SE classes (see **Figure 5A**). Especially interesting is the possibility of predicting the occurrence of self-errors before the actual erroneous response occurs, and to distinguish them from other action outcomes. Importantly, this predictive correct classification can be done at the singe-trial level, therefore allowing the user to potentially avoid the occurrence of the erroneous performance. The identification of this period of time previously to the response informing about its outcome (SE vs. Correct) resembles several studies that have evidenced the existence of pre-ERN modulations (in between 30-100 ms before the error) associated to on-going erroneous actions. For example, some studies have identified ERP modulations previous to the commission of an error when performing musical errors (Maidhof et al., 2009, 2013; Ruiz et al., 2009), during learning rhythmic sequences (Padrao et al., 2014) or even previous to a speech error (Möller et al., 2007). These ERP signatures registered before the actual erroneous response might index the existence of on-line error monitoring processes

implemented through neural feed-forward computations allowing the system to predict that an error might occur signaling the need to perform a corrective action beforehand. Further studies might be needed combining also information from lateralized components (e.g., LRPs) in order to evaluate the potential accuracy of pre-error signatures and the extent that this information could be temporally viable to avoid the execution of on-going errors.

Moreover, a second interval starting at +35 ms showed a good classification of the SE class (i.e., SE vs. Correct and SE vs. EE) (see **Figure 5A**), evidencing the distinguishable topographical characteristics of the ERN, while the confusion metric was maintained for the Correct vs EE. This result confirms previous evidence highlighting the utility of the ERN for decoding self-made errors, confirming its dissociable neural characteristics from correct responses (Schmidt et al., 2012; Spüler et al., 2012). Moreover, SE vs. EE could also be correctly classified at this point in time, favouring the possibility of performing rapid error corrections if the error is coming from the self. More posteriorly, at approximately +300 ms, the Correct and EE classes became more distinguishable, reflecting the P600 component appearing during the EE condition, with a maximum differentiation at 585 ms. The ability to correctly classify EE seems to be more related to later stages of processing probably indicating the need to recruit more reflective aspects of agentic processing related to higher cognitive functions (Moore & Haggard, 2008; Synofzik et al., 2008a).

Individual differences were observed in both the Agency attribution questionnaire and the model performance. One could argue that the model performance, at least on the EE condition, could correlate with the attribution of agency, but there was no significant correlation between the F1-score of the EE class during the significant intervals, and the Agency attribution score (r = 0.22, p = 0.3491). Probably, this indicates that the model performance depends on how clear and robust the ERP signal is, and not on how much agency attribution the participant has. Moreover, although individual differences existed in terms of accuracy rates, the mean accuracy along time and the GMFP had a very strong correlation (see Figure 6), indicating that the higher the GMFP, the higher the accuracy of the model.

Not surprisingly, when examining the F1-scores, the SE condition was the best classified at the ERN time window. This observation goes in line with previous electrophysiological studies that reported the ERN as a key component to detect the self-generated errors (Falkenstein et al., 1990; Gehring et al., 1993; Taylor et al., 2007). Probably, both the polarity (negativity) and the localization (frontal electrodes) of the activity make the ERN easy to detect by a classifier focused on the proper latencies (∼95 ms). In the same way, this too early component does not permit the model to classify the Correct condition better than the EE condition. In fact, the classification of

the Correct and EE conditions improves along the time, i.e., their F1-score at P600 is higher than that of at ERN. Interestingly, the Correct condition was the best classified at the P600 latency, with significant differences between Correct vs. SE but, not between Correct vs. EE and SE vs. EE. These results indicate that the EE condition is more distinguishable by later components such as P600, which goes in-line with previous studies that reported a centro-parietal large positive waves (P600) produced by erroneous or incongruent conditions (Pijnacker et al., 2012).

To sum up, in the present study we were able to correctly classify, at the single-trial level, Correct, SE and EE ERPs by means of a linear SVM classifier based on the latency and topographical characteristics of several ErrPs. Importantly, the model indicates that SE condition is the best classified condition during the frontal ERN time window, while the Correct and EE conditions were more accurately classified at a later time window (from 400 ms onwards), corresponding to the parietal P600 component. We believe the present results provide crucial new evidence on the importance of agency attribution of our actions and the actions supposedly governed by our mind (on external agents or surrogated bodies). We encountered that single-trial EEG activity contains decodable information about our capacity to accurately monitor our actions, both the ones we initiated and the ones that are imposed to us. This research has also potential value for further research applications in the near future, not only for understanding the origin of our intentions to act, actions and the sense of agency but also for the potential implications it might have regarding moral responsibility over our supposed actions.

STUDY 3

The role of the anterior insular cortex in self-monitoring: A novel study protocol with electrical stimulation mapping and functional magnetic resonance imaging

This study corresponds to:

Gomez-Andres A, Cunillera T, Rico I, Naval-Baudin P, Camins A, Fernandez-Coello A, Gabarrós A, Rodriguez-Fornells A. The role of the anterior insular cortex in self-monitoring: A novel study protocol with electrical stimulation mapping and functional magnetic resonance imaging. *Cortex.* 2022 Oct 4;157:231-244. https://doi.org/10.1016/j.cortex.2022.09.008

3.3. STUDY 3

The role of the anterior insular cortex in self-monitoring: A novel study protocol with electrical stimulation mapping and functional magnetic resonance imaging.

3.3.1. INTRODUCTION

For our daily life, it is crucial that we become aware of the consequences of our actions, errors and limitations that force us to adapt our behaviour and strategies. In clinical settings, reduced conscious perception of errors has been associated with poor insight of neurological sequelae (O'Keeffe et al., 2004), which may lead to worse functional outcome and poor compliance with rehabilitation (Ownsworth & Clare, 2006). The ability to monitor our actions is mediated by the coordinated activity of several brain regions, including the mPFC, especially the ACC and SMA, the thalamus and, as we will highlight in this study, the aIC (Holroyd & Coles 2002; Ito et al., 2003; Ullsperger & von Cramon 2001; Wessel, 2012). Dysfunction of the insular cortex and its interconnected regions are thought to be core features of many psychiatric and neurological disorders (Goodkind et al., 2015; Namkung, Kim & Sawa, 2017). In the present study, we propose a new multimodal protocol to explore the functional role of the aIC in self-monitoring in patients undergoing awake brain surgery for tumour resection.

Described by Reil in 1809, the insula (Latin for 'island') lies folded deep within the lateral sulcus of each hemisphere hidden below parts of the frontal, parietal, and temporal lobes. The insular cortex constitutes an anatomical integration hub which receives direct thalamic and horizontal cortical afferent projections carrying information from all sensory modalities and reciprocally connects to an extensive network of cortical and subcortical brain regions (such as limbic regions, the ACC, orbitofrontal and medial prefrontal cortices) serving sensory, emotional, motivational, and cognitive functions (Gogolla, 2017). Interestingly, a unique feature of the aIC of humans and a few other species (i.e., great apes, elephants, and some cetaceans) is the presence of clusters of large spindle-shaped neurons among the pyramidal neurons in layer 5, called von Economo neurons (Allman et al., 2005; Nimchinsky et al., 1999). While the precise function of these type of cells is not known, fMRI studies have shown that these type of neurons are selectively destroyed in disorders characterized by loss of emotional awareness and self-consciousness, such as frontotemporal dementia (Seeley et al., 2006; Sturm et al., 2006), schizophrenia (White et al., 2010) and autism (Minshew & Keller, 2010; Monk et al., 2009), which has led several researchers to postulate their involvement in empathy, self-awareness, and self-monitoring (Seeley et al., 2007).

Recently, it has been proposed that error awareness can be conceived as a decision process, in which the available evidence from different sources that an error has occurred (i.e., mPFC, autonomic responses, proprioceptive and other sensory inputs) is accumulated until a decision threshold is reached (Steinhauser & Yeung, 2010; Ullsperger et al., 2010; Wessel, Danielmeier & Ullsperger, 2011; Wessel, 2012). It has been proposed that this integration of the different salient signals reflecting a deviation from the predicted outcomes (i.e., prediction errors) may be processed in the insular cortex (Craig & Craig, 2009; Ham et al., 2013), particularly at the aIC, conforming the actual input to the self-monitoring network signaling the need for cognitive control and behavioural adaptation (Menon & Uddin, 2010). Previous clinical evidence and neuroimaging studies using fMRI provide converging evidence of the role of the aIC in error awareness. It has been reported that the aIC is consistently more activated for errors compared to correct responses (Ham et al., 2013), especially for consciously perceived compared to unperceived errors (Hester et al., 2005; Klein et al., 2007), as well as during negative vs positive informative feedback (Ullsperger & von Cramon, 2003). Relatedly, the aIC has been found to be involved in self-awareness and self-attribution of actions. For example, in a task where participants were instructed to view self-portrait and pictures of unknown faces, this comparison yielded a greater activation of aIC (Kircher et al., 2000; Sugiura et al., 2000). Moreover, grater activity in the aIC has been found when responding to visual feedback resulting directly from the subject's actions while driving a virtual race car around a track compared to unrelated, random feedback (Farrer & Frith, 2002) as well as reacting to auditory feedback coming from oneself compared to randomly timed feedback during a motor task (Blakemore, Rees & Frith., 1998).

ESM has been the gold standard technique for identifying essential sensory and motor cortices as well as relevant language areas in patients undergoing tumour resection (Duffau, 2008; Ojemann, 1983; Penfield & Roberts, 2014), also for single-case designs (Rofes et al., 2017; Rojas et al., 2021; Sierpowska et al., 2015). Executive aspects of self-monitoring abilities have been recently explored using ESM, for example, using the Stroop paradigm (Stroop, 1935). Wager et al. (2013) found evidence of the functional role of the ACC in performance monitoring when ESM was applied, and Puglisi et al. (2019) reported that ESM at the subcortical level over white matter sites below the inferior and middle frontal gyri, anterior to the insula and over the putamen, led to impairments in the task performance (i.e., colour-word inversion and latency increases). However, the literature concerning the functional mapping of the insular cortex for more complex cognitive functions such as self-monitoring and error awareness is rather scarce (Erez et al., 2021). In the seminal paper of Penfield and Faulk (1955), ESM of the insula elicit a number of autonomic responses such as gastrointestinal sensations, swallowing, chewing and heart rate increases, highlighting its role for interoception and pointing out to a contribution of this region in the

conscious detection of salient events, which capture the subject's attention and awareness in a similar way as errors do (Craig & Craig, 2009; Craig, 2011; see Mazzola, Mauguiere & Isnard, 2019 for a review). More recently, Bastin et al. (2016) used intracerebral electroencephalography in a group of epileptic patients during a Stop-signal paradigm; recordings at the aIC showed a rapid increase of electrophysiological activity in this region when an error was committed as well as a feedforward influence from aIC onto ACC and, subsequently, onto the SMA, providing direct evidence for a key role of aIC within the saliency-error-monitoring networks.

Both self-awareness and error monitoring abilities are core aspects of metacognition, understood as the cognitive processes that are linked with activation of "thinking about one's own thinking", by which individuals can reflect upon (monitor) their own internal mental states and apply their knowledge to evaluate and regulate (control) their own mental states (Nelson et al., 1999). Clinical findings further support the role of the aIC in self-monitoring/metacognitive processing, as is the case for patients with AHP, who present unawareness of motor deficits related to hemiplegia and are commonly associated to lesions in insular regions (Karnath et al., 2005; Vocat et al., 2010). Patients suffering from schizophrenia, especially those reporting delusions of control have also been associated to functional and morphological abnormalities in the insular cortex (Crespo-Facorro et al., 2000; Moran et al., 2013; Wylie & Tregellas, 2010).

In the present study, we propose a multimodal protocol for exploring the role of the aIC in error monitoring for self-made actions combining ESM, fMRI and neuropsychological assessment. With this purpose in mind and, due to the inherent difficulty of measuring errors in simple reaction time tasks, mainly due to the small frequency and randomness distribution, we created a situation during which we challenged self-monitoring abilities introducing random feedback informing about the correctness of the response emitted. Based on the exposed above regarding the role of aIC in self-monitoring, our main hypothesis was that ESM applied over the aIC regions would disrupt the ability to correctly detect the incongruences between the patients' action/response and the subsequent feedback appearing on the screen. As feedback correctness was random (50% chance of providing a correct information), we created a situation in which the patient was interrogated about the accuracy of the presented feedback, requiring access to her internal self-monitoring/metacognitive process. This allowed us to increase the chance of stimulating the insular region while presenting action-outcome incongruences and to obtain a larger number of trials for the evaluation of self-monitoring abilities. Moreover, we expected to find relevant and significant brain activations involving the aIC when introducing incongruent feedback during the fMRI acquisition at the relevant contrasts. Our results revealed to be quite promising regarding the usefulness of our multimodal approach in studying the implication of the aIC region in self-

monitoring and error processing, highlighting the importance of assessing the insular function during tumour resection and functional imaging allowing a better understanding and preservation of the aIC functional role in self-monitoring.

3.3.2. MATERIALS AND METHODS

Clinical case description

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

MM, a 57-year-old, right-handed woman, Spanish-Catalan bilingual was admitted to the Neurosurgery Department of the Bellvitge Academic Hospital after incidental diagnosis of a left fronto-insular lesion (diffuse astrocytoma, WHO, II) (see **Figure 3A**). Due to the lesion location, the patient was selected for the awake mapping multimodal protocol (which includes neuropsychological assessment, fMRI and ESM, in this order) to assess the aIC-related functions. The use of a multimodal protocol allows us to obtain a multidimensional perspective of the patient's state that aided adjusting the surgical procedure to approach tumour recession, benefiting in that way the patient's treatment. The study protocol was accepted by the Ethical Committee of the University Hospital of Bellvitge (reference PR075/19) in accordance with the principles of the Declaration of Helsinki and the participant signed the informed consent for the participation in the study. No part of the study analyses was pre-registered prior to the research being conducted.

Neuropsychological assessment

A comprehensive neuropsychological assessment was performed at the neurological ward of the University Hospital of Bellvitge both pre- and postoperatively. A specific neuropsychological protocol was employed assessing executive functions such as attention ["Digits -direct-" from Test Barcelona-Revisado (Peña-Casanova, 2005) and "Trail Making Test (TMT) -part A-" (Reitan, 1955)], working memory ["Digits -inverse-" from Test Barcelona-Revisado (Peña-Casanova, 2005), "Letters and Numbers" and "Arithmetic" subtests from the Wechsler Adult Intelligence Scale-IV -WAIS-IV- (Spanish Edition; Wechsler, 2008)], inhibition [Stroop Task (Stroop, 1935)], mental flexibility ["TMT -part B-" (Reitan, 1955)], set shifting abilities [Wisconsin Card Sorting Test –WCST- (Grant & Berg, 1948)] and verbal fluency (semantic and phonological verbal fluency). Moreover, language function was also examined in terms of production and naming [Boston

Naming Test –BNT- from the Boston Diagnostic Aphasia Examination -BDAE- (Goodglass & Kaplan, 2001)] and verbal comprehension [Token Test (De Renzi & Faglioni, 1978)]. Importantly, two tasks tackling more specific insular-related functioning were included namely, the Empathy scale [Interpersonal Reactivity Index -IRI- (Davis, 1983), Spanish version] and the Emotion recognition test [Ekman 60 faces Test (Ekman & Friesen, 1976)]. Mood was also assessed using the Hamilton Anxiety Rating Scale –HARS- (Hamilton, 1959) and the Hamilton Depression Rating Scale –HDRS- (Hamilton, 1960).

Standardized tests from the WAIS-IV were interpreted by using Spanish normative data from the WAIS-IV Spanish Edition (Wechsler, 2008). Additionally, Digits, verbal fluencies, TMT (parts A and B), Stroop task, BNT, Token Test were corrected using Neuronorma Spanish normative data and corrected by age and years of education (Peña-Casanova et al., 2009a; Peña-Casanova et al., 2009b). In both cases, impairment was defined as a scaled score ≤6 (see **Table 1**).

Electrical stimulation mapping (ESM)

MM underwent awake surgery for tumour removal by the senior neurosurgeon A.G. at the Department of Neurosurgery of the University Hospital of Bellvitge. A left frontotemporal craniotomy was performed, and the Sylvian fissure was opened allowing the exposure of the aIC. ESM was performed using an Ojemann cortical stimulator (Radionics, Inc.) under asleep-awake-asleep surgery, following the methodology described previously by Ojemann, Ojemann, Lettich, and Berger (2008) and our previous work (Fernández-Coello et al., 2016; Havas et al., 2015; Sierpowska et al., 2013, 2018). The interelectrode distance of the bipolar forceps was 5 mm. The stimulator delivered a biphasic current with a pulse frequency of 60 Hz and a single-pulse phase duration of 1 sec. The duration of each stimulation train was 3 sec. The current amplitude was progressively increased by 0.5 mA, beginning at 1 mA, until the desired responses were observed. During the ESM procedure, the same cortical area was never stimulated twice in succession to avoid seizures, and between each set of 2 stimulations we always performed a control trial without applying electrical current.

Firstly, the primary motor and sensory cortices were mapped. A site was considered positive for motor function when the stimulation elicited involuntary muscle contraction or speech arrest while the patient was counting. On the other hand, the primary sensory cortex was determined by sensory disturbances perceived by the patient upon electric stimulation. Due to the tumour location involving fronto-opercular regions, language mapping was also performed using a home-made simplified version of the picture-naming task, composed by 60 black and white drawings

selected from a standard stimuli database (Snodgrass & Vanderwart, 1980; Havas et al., 2015) both in Spanish and Catalan (see the functional map obtained in **Figures 3B** and **3C**).

Experimental paradigm: Stroop interference task with feedback

Once the aIC was exposed, the experimental paradigm was performed. A modified version of the Stroop task (Stroop, 1935) was designed and implemented both before and after surgery, intraoperatively and during presurgical fMRI (see **Figures 1A** and **1B**). The task was adapted for each modality (ESM vs. fMRI) in terms of timing although the same set of stimuli were employed. The Stimulus presentation was controlled with EPrime (Psychology Software Tools Inc., Pittsburgh, PA). In general terms, the task consisted of signaling the colour in which a word is printed ignoring the word itself. When the word is a colour word printed in a mismatched colour ink (i.e., *RED* printed in blue), an interference effect occurs resulting in a slower, more difficult, and error-prone condition relative to a control condition (i.e., *XXXXX* printed in blue). The performance cost in the incongruent condition is called the Stroop effect.

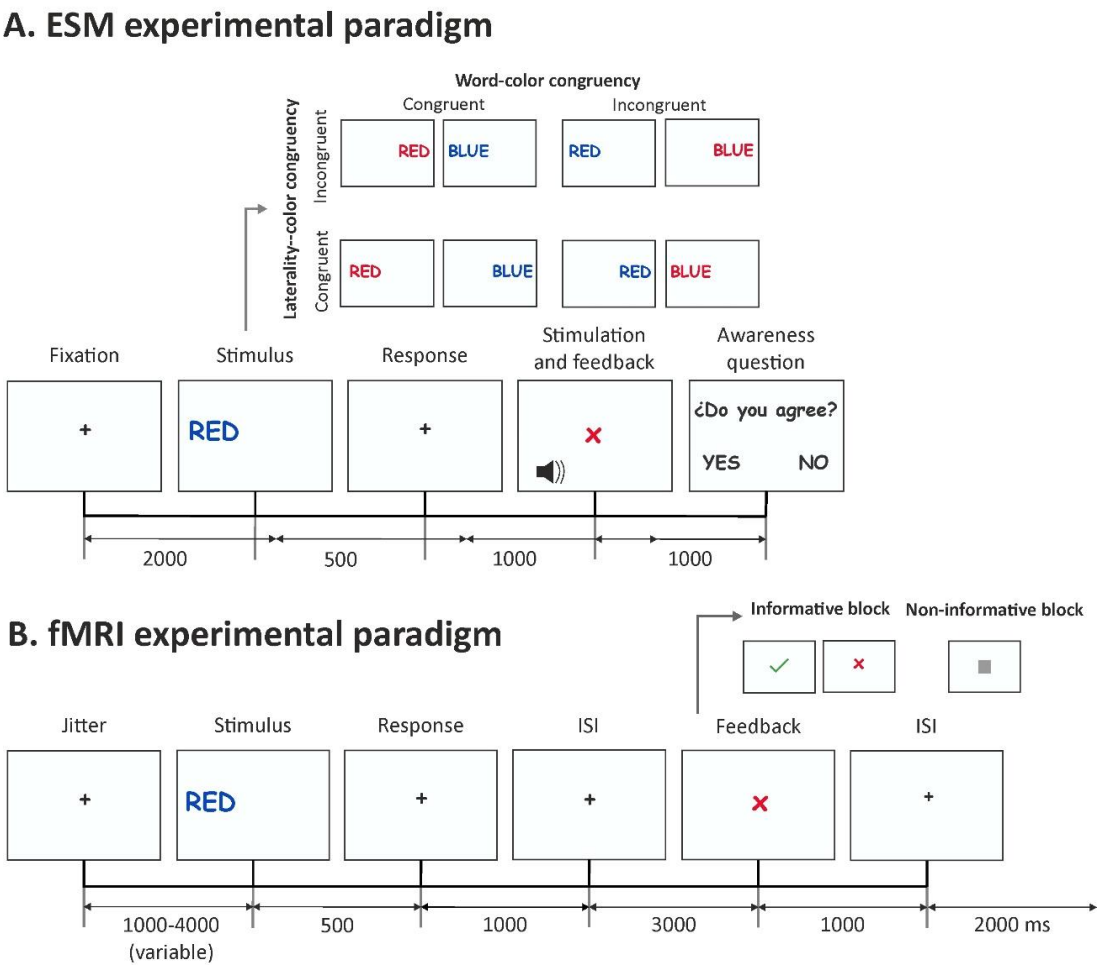


Figure 1. Experimental paradigm: Stroop interference task with feedback. A) Experimental task employed during the electrical stimulation mapping (ESM) procedure. The task always started with a fixation slide (2000 ms). The target stimulus appeared for 500 ms, which varied in terms of the laterality-colour congruency and word-colour congruency variables (8 trial types in total) The patient was instructed to respond as fast as possible to the ink colour ignoring the word by pressing either left (blue) or right (red). After the response, the feedback slide appeared for 1000 ms (either congruent or incongruent with respect to the patient's performance), during which the electrical stimulation was applied signalled by a sound to inform the neurosurgical team. After the feedback + stimulation presentation, the monitoring question was presented. **B)** Experimental task employed during the functional magnetic resonance imaging (fMRI) acquisition. Each trial started with a variable fixation slide (1000-4000 ms), followed by the target stimulus presentation (500 ms). Once the patient responded, an ISI of 3000 ms appeared followed by the feedback presentation (1000 ms). During the informative blocks, the feedback (either a red cross or green tick) appeared in a random manner (50% probability). For the non-informative blocks, neutral feedback (grey square) was always presented.

The patient was instructed to press a button with either their left or right thumb (see **Figure 1A**) whenever the target stimulus was printed in red or blue, respectively, ignoring the word. Additionally, we increased the difficulty of the task by adding a laterality variable which involved placing the target stimulus at the middle-left or middle-right location on the screen, generating either congruent or incongruent trials regarding laterality-correct response. Therefore, 8 different targets were presented based on the word-colour and laterality congruence condition (32 trials in total), balanced across the task (4 trials for each word-colour and laterality-congruence trial type). After the stimulus presentation (500 ms), the patient had 1000 ms to respond (reaction time -RT). The experimental paradigm is accessible at https://doi.org/10.34810/data212.

To be able to investigate whether the stimulation of the aIC resulted in a disruption of the ability to self-monitor committed errors, we needed to be able to anticipate the occurrence of the desired event to synchronize it to the electrical stimulation. Because the commission of errors by the participant could not be anticipated, we provided we provided feedback in a random manner [i.e., either positive or negative (green tick or red cross)], which could be either congruent or incongruent with respect to the patient's performance. In 50% of the cases, positive feedback was provided and distributed equally across the different trial types. Finally, the patient had to say whether she agreed with the feedback she had been given or not. The feedback presentation was synchronized with the electrical stimulation to evaluate the effect of the electrical impulse on performance monitoring. As shown in the trial example in **Figure 2**, a cortical site was considered functionally relevant for error/performance monitoring whenever the electrical stimulation resulted in a failure in detecting whether the feedback provided was incongruent with her

performance. To do so, the experimenter monitored online the patients' performance via an additional monitor which displayed the correctness or not of the response and the feedback that was given for each trial. The cortical locations of the self-monitoring sites detected during the ESM procedure were transferred to an arbitrary grid (see **Figure 3C**).

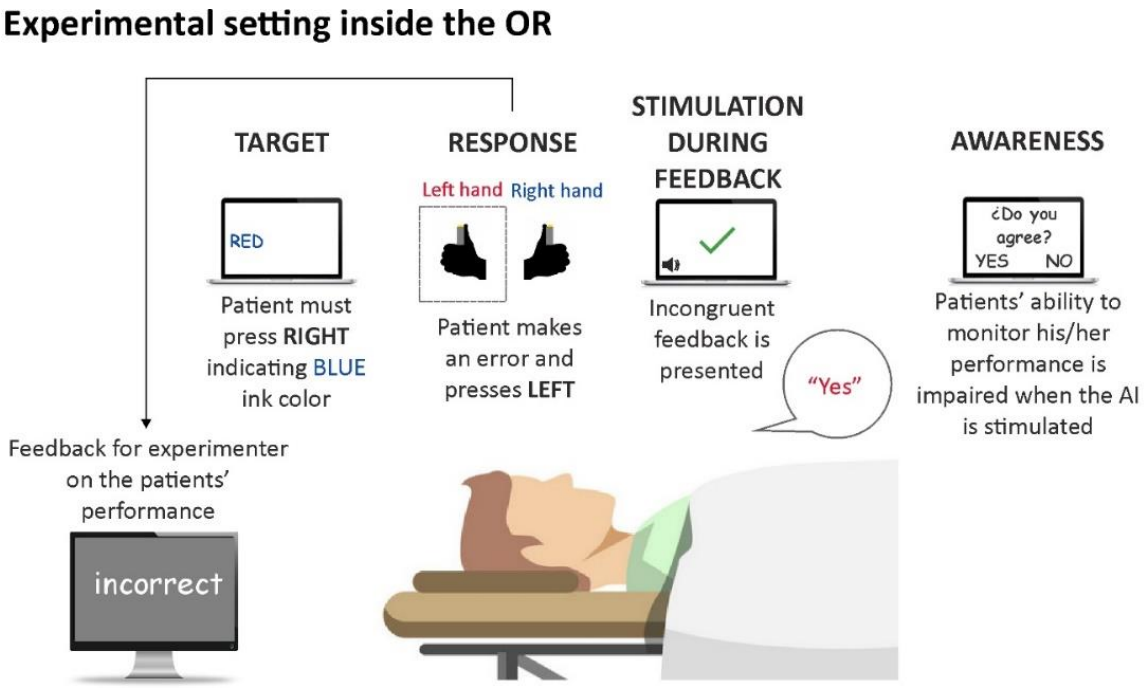


Figure 2. Experimental set-up inside the operating room (OR). Once the patient was awakened and the resection reached the insular cortex, the experimental task was performed. A computer screen was placed at the eye-level of the patient, displaying the Stroop modified paradigm. A second computer screen informing the experimenter about the patients' performance and the feedback presented was also available only for the experimenter eyes. The electrical stimulation was cued by a sound informing the neurosurgical team the feedback slide was appearing and therefore the stimulation had to be performed. Once the response and feedback had occurred, the experimenter determined whether the stimulation had impaired or not the patient's monitoring ability. If 3 or more non-consecutive stimulations over the same cortical site resulted in a monitoring impairment, that cortical point was considered to be functionally relevant.

fMRI

Patient MM was scanned before surgery in a 1.5 T MRI Philips Igenia system at University Hospital of Bellvitge for presurgical assessment supporting intraoperative neuronavigation. Functional images were acquired in the axial plane using a single-shot T2-weighted gradient-echo EPI sequence with a 3000 msec repetition time (TR), 50

msec echo time (TE) and 90º flip angle (FA). Each volume consisted of 4 mm thick slices with no inter-slice gap; voxel size = 3.59 x 3.59 x 4 mm^3; FOV = 230 mm; size of acquisition matrix 64 x 64. In addition to the functional images, a high-resolution sagittal T1-weighted image (slice thickness = 2 mm; no inter-slice gap; number of slices = 180; TR = 7.8 msec; TE = 3.8 msec; flip angle = 8 º; matrix = 250 x 250; FOV = 250 mm; voxel size = 1 x 1 x 2 mm^3) was also acquired.

The Stroop experimental paradigm was adapted during the fMRI acquisition (see **Figure 1B**). The stimuli duration was adapted as follows: First, a variable fixation slide –intertrial interval (ITI)- (jittering 1000-4000 ms) was presented followed by the stimulus presentation (500 ms), to which the patient had 1000 ms to respond. A fixation slide –interstimulus interval (ISI)- of 3000 ms was then presented, followed by the feedback (either informative or non-informative, depending on the block), during 1000 ms. Two main conditions were defined: a) Informative feedback blocks, during which informative feedback was provided in a random manner (i.e., either positive -green tick- or negative -red cross-, 50% positive/negative) and b) non-informative feedback blocks, during which neutral feedback was always delivered (i.e., grey square) (see **Figure 1B**). A total of 16 blocks were performed (8 Informative and 8 Non-Informative), each one consisting of 24 trials. The total duration of the task was 16 mins. As in the task version used during the ESM procedure, the patient was instructed to press a button with either their left or right thumb whenever the target stimulus was printed in red or blue, respectively, ignoring the word and a laterality variable was also included.

In order to examine the brain activity related to congruent vs incongruent vs neutral feedback, we performed an event-related design grouping the trials in the following manner: a) Congruent trials (29 events of interest), trials of the Informative blocks for which a congruent feedback was presented in relation to the participant performance (i.e., correct response followed by correct feedback or erroneous response followed by error feedback), b) Incongruent trial (15 events of interest), trials corresponding to an incongruent feedback (i.e., correct response followed by error feedback or erroneous response followed by correct feedback) and c) Neutral trials (48 events of interest), corresponding to all trials from the non-informative blocks.

The fMRI pre-processing and statistical analysis was performed with SPM12 (The Wellcome Trust Centre for Neuroimaging, London, UK). Image pre-processing included realignment, slice timing, segmentation, normalization and smoothing with an 8 mm gaussian kernel. Unified segmentation (Ashburner & Friston, 2005) with medium regularization was applied. A General Linear Model contrastive analysis was performed. Motion parameters extracted from the realignment were included as regressors of no interest. Statistical parametric maps were obtained for the following

contrasts: a) Congruent vs Neutral, b) Incongruent vs Neutral and c) Incongruent vs Congruent feedback. These contrasts are reported at an uncorrected level of $p < .01$.

Region of interest (ROI) analyses using an uncorrected corrected $\alpha = 0.05$ with an extent threshold = 20 continuous voxels were performed for the left insular cortex and the ACC (bilateral) based on the WFU Pickatlas toolbox (Maldjian et al., 2003).

3.3.3. RESULTS

Electrical stimulation mapping (ESM)

Functionally, the ESM over the motor and sensory cortices revealed two speech arrest sites (red tags) at the opercular region, and two sensitive points on the somatosensory regions corresponding to the palate (blue tag -1-) and oropharyngeal and nasopharyngeal sensations (blue tag -2-) (see **Figures 3B** and **3C**). Surgically, a total resection of the tumour involving the left Broca region was performed, after obtaining no positive functional points at his level neither cortically nor subcortically. Moreover, although two positive sites were encountered (see red tags at **Figures 3B** and **3C**), a subtotal resection of the tumour tissue involving the left motor operculum portion was reached in order to decrease the probability of tumour recurrence (**Figure 3A**).

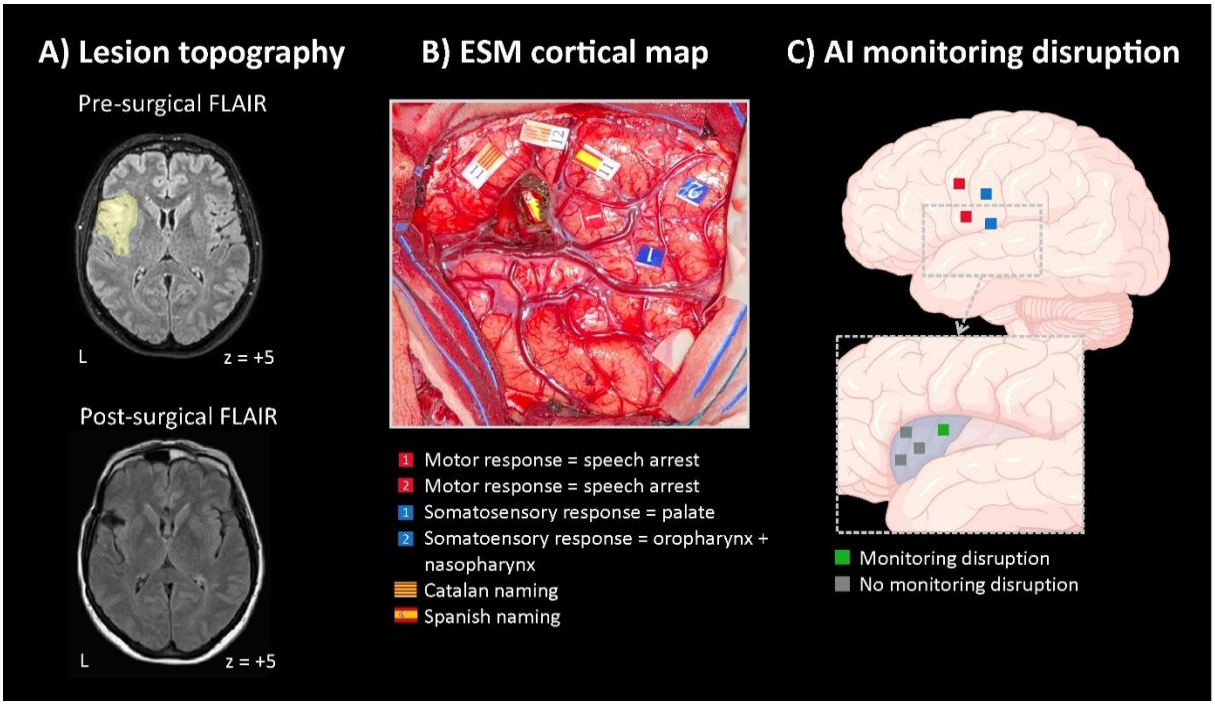


Figure 3. Lesion topography and ESM results. A) Pre and post-surgical FLAIR acquisitions depicting the tumoral tissue in light yellow and the post-surgical cavity.

B) ESM intraoperative photograph depicting the ESM results at the left fronto-temporal region. In red we see 2 functional motor sites related to speech arrest. In blue, 2 somatosensory points were located corresponding to palate sensations (top) and oropharynx + nasopharynx sensations (bottom). Language-related cortical sites are also depicted for Spanish and Catalan naming impairment. **C)** Anatomical map of the cortical surface (top) and insular region (down) depicting positive functional sites (green) and negative sites (grey) for self-monitoring abilities tested with the Stroop experimental paradigm.

Following the cortical ESM, the opening of the sylvian fissure was performed to enable the tumoral resection involving the insular region. Once the aIC was exposed, the ESM was performed. Out of the 32 trials of the experimental task, 10 trials were performed while electrically stimulating the aIC synchronized with the feedback appearance. One functional point for self-monitoring was found at the most posterior part of the aIC (see **Figure 3C**). More specifically, 3 non-consecutive stimulations over the abovementioned cortical site resulted in an inability to correctly state whether the feedback provided was congruent with their actual response. Two of these trials were characterized by erroneous responses followed by the presentation of incongruent feedback (i.e., positive feedback), to which the patient responded '*Yes*' during the awareness question revealing her inability to correctly monitor and compare her actual response to the feedback presented. In the case of reaching a third trial for which a monitoring failure was detected while stimulating, a congruent feedback after an error was presented (i.e., negative feedback) followed by an incorrect awareness assertion by the patient (i.e., MM responded '*No*' to a congruent feedback). A total surgical resection was performed of the tumour tissue involving the aIC boundary with deep inferior fronto-occipital fasciculus (IFOF) boundary.

fMRI results

Whole brain analysis

The neural networks underlying performance monitoring were assessed by an event-related analysis time-locked to the feedback onset. Results of the whole brain analysis are shown on **Figure 4A**. During the Congruent vs Neutral contrast, significant clusters of activations were mainly found at the superior frontal gyrus -SFG-, inferior frontal gyrus -IFG, SMA and ACC (**Figure 4A**). When looking at the Incongruence effect (Incongruent vs Neutral feedback contrast), similar regions including the ACC, SMA and IFG were found to be activated. Crucially for the purpose of the current study, the bilateral insular cortex was also found to be significantly engaged during the incongruent feedback (**Figure 4A**). Finally, when computing the contrast Incongruent vs Congruent feedback to isolate the incongruence effect, significant clusters were encountered at

the ACC, bilateral insular cortices (both anterior and posterior regions), superior temporal gyrus -STG- (bilaterally), and subcortical structures such as the left putamen and right globus pallidus (see **Figure 4A**). Patient data accessible at https://doi.org/10.34810/data212.

ROI analysis

ROI analysis of the left insular cortex for the contrast Incongruent vs Congruent feedback revealed a significant cluster of activation at the left posterior insula (cluster size = 43 voxels, peak *t*-score = 2.48, MNI_{XYZ} = -32, −26, and 16; **Figure 4B**).

Moreover, ROI analysis at the ACC also revealed significant clusters of activity mostly left lateralized (left ACC: cluster size = 121 voxels, peak *t*-score = 2.44, MNI_{XYZ} = -12, 46, and 2; right ACC: cluster size = 58 voxels, peak *t*-score = 2.05, MNI_{XYZ} = 6, 38, and 2; **Figure 4B**). Patient data accessible at https://doi.org/10.34810/data212.

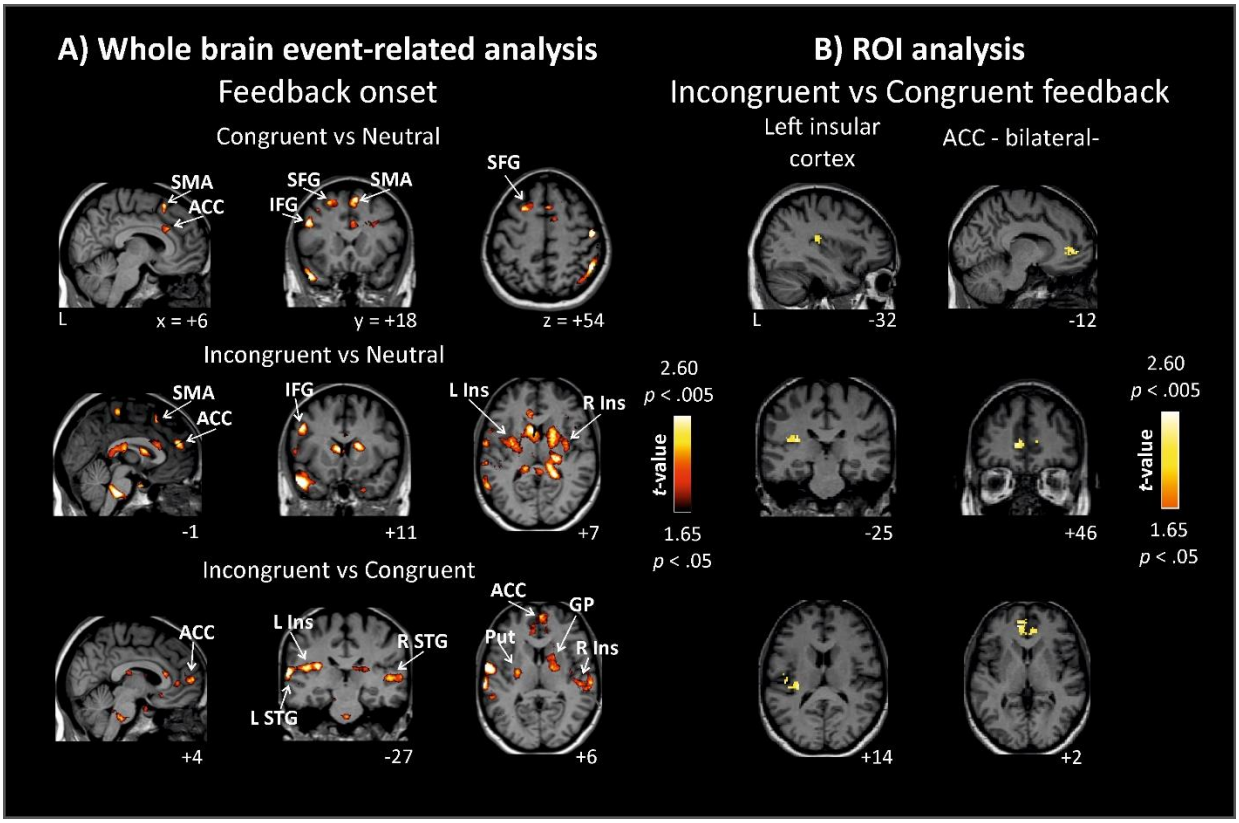


Figure 4. Functional magnetic resonance imaging results. A) Whole brain event-related analysis tagged to the feedback onset for the contrasts of interest Congruent vs Neutral, Incongruent vs Neutral and Incongruent vs Congruent feedback. **B)** ROI analysis for the contrast Incongruent vs Congruent feedback for the left insular cortex and anterior cingulate cortex (ACC) –bilateral- Results are shown normalized to MNI space at an uncorrected level of $p < .005$.

Cognitive outcome

Pre-operatively, the patients' neuropsychological profile was within the normal score range in attention and executive functioning, language, semantic knowledge, and social cognition (see **Table 1**). She did not report any difficulties in motor or sensory abilities. Moreover, the performance in the Stroop experimental paradigm was also within the normal range, obtaining a 100% accuracy on the execution and feedback awareness assessment.

At the post-surgical assessment performed 11 months after the surgery, a significant decline in some cognitive domains was observed (**Table 1**), most probably due to the resection at the level of the left motor opercular region (see **Figure 3A**). A decrease in digit span (direct) and verbal fluency, both for the semantic and phonological cue, was encountered, most probably influenced by the articulatory difficulties. A slower visuomotor processing speed revealed by a significant increase in the time needed to complete the TMT-A and the Letters and Numbers test were observed. Reduced cognitive flexibility assessed with the TMT-B and set-shifting abilities assessed using the WCST, showing a larger number of efficient errors and the impossibility of achieving any category, were also impaired after surgery. Notably, the performance on the Stroop task (standardized version) revealed a large reduction in the number of items completed during the colour-word interference condition, pointing out to an increased difficulty in her inhibitory capacity, although it might be influenced by the verbal fluency and dysarthria difficulties the patient presented at post-op after the resection. Language production abilities were preserved, although more difficulties with complex verbal comprehension were recorded.

Social cognition aspects were also assessed, evidencing more difficulties in Perspective taking and Fantasy items of the IRI scale (empathy) after surgery. Moreover, emotion recognition difficulties (Ekman test) were observed both before and after surgery, especially for negative emotions. Importantly, the patient presented a significant increase in depressive symptoms after surgery as revealed by the HDRS.

3.3.4. DISCUSSION

Very few studies addressing self-monitoring and error processing have been performed using ESM during awake brain surgery. Here, we present a new multimodal protocol proposal assessing the functional role of the aIC in performance monitoring for self-made actions. We provide ESM, fMRI and neuropsychological information data for a single patient undergoing awake brain surgery for tumour removal involving the aIC region. For this purpose, we designed a modified

Table 1: Neuropsychological assessment at baseline and post-surgery.

Function (test)	Baseline		Post-surgery		Comments
	R.S.	S.S.	R.S.	S.S.	
Attention and executive functioning					
A. Attention and verbal WM					
- Digit span -direct-	6	11	4	6	Significant decrease in attention efficiency.
- Digit span -inverse-	6	15	4	10	Slight reduction in verbal WM but preserved.
- Letters and Numbers	11	12	5	6	Slower visuomotor processing speed.
- Arithmetic	14	12	14	12	
- TMT -part A-	30 sec	12	180 sec	2	Slower visuomotor processing speed.
B. Inhibition					Significant reduction in colour-word condition (susceptibility to interference and inhibitory ability). Might be influenced by motor/articulatory difficulties.
- Stroop Test	36	10	7	3	
C. Mental flexibility					Reduced cognitive flexibility.
- TMT -part B-	53 sec	14	241 sec	5	
D. Set shifting					
- WCST					
• Correct responses	84	Preserved	76	Preserved	Increased difficulties in set-switching.
• Categories achieved	6	Preserved	0	Impaired	
• Perseverative errors	9	Preserved	18	Preserved	
• Efficient errors	23	Preserved	34	Impaired	
E. Verbal fluency					
- Semantic cue (animals)	21	10	9	4	Significant impairment in verbal fluency.
- Phonological cue ("p")	15	10	6	5	
Language					
A. Production and naming					*55/60 with phonological cue, 1 latency. **52/60 with phonological cue, 1 semantic paraphasia, 1 phonemic paraphasia, 5 circumlocutions, 2 switching errors, 10 latencies.
- BNT	52/60*	10	45/60**	8	
B. Verbal comprehension					Slight reduction in verbal comprehension
- Token Test	36/36	12	32.5/36	8	
Social cognition					
A. Empathy					
- IRI					
– Perspective taking	30	Preserved	13	Impaired	More difficulties in *Perspective taking* and *Fantasy* items were observed at post-op.
– Fantasy	26	Preserved	13	Impaired	
– Empathic concern	34	Preserved	28	Preserved	
– Personal concern	19	Preserved	18	Preserved	
B. Emotion recognition					
- Ekman 60 faces Test Total	36/60	Impaired	38/60	Impaired	Lower scores were obtained for negative emotions, both during pre-op and post-op, with global scores showing an impairment in this test.
• Anger	0/10	Impaired	8/10	Preserved	
• Disgust	9/10	Preserved	5/10	Impaired	
• Fear	1/10	Impaired	2/10	Impaired	
• Happiness	10/10	Preserved	10/10	Preserved	
• Sadness	6/10	Preserved	5/10	Impaired	
• Surprise	10/10	Preserved	8/10	Preserved	
Mood					
A. Anxiety					After surgery, the patient presented a significant increase in depressive symptoms.
– HARS	11	Preserved	14	Preserved	
B. Depression					
– HDRS	3	Preserved	15	Impaired	

R.S.: raw score; S.S.: scalar score; TMT: Trail Making Test; BNT: WCST: Wisconsin Card Sorting Test; Boston Naming Test; IRI: Interpersonal Reactivity Index; HARS: Hamilton Anxiety Rating Scale; HDRS: Hamilton Depression Rating Scale.

version of the Stroop task presenting the patient with either congruent or incongruent feedback with respect to her performance synchronized with the stimulation of the aIC sites exposed.

As previously mentioned on introductory paragraphs, ESM is the gold standard technique for locating and identifying essential motor, sensory and other cognitive functions (i.e., language related functions) in patients undergoing awake brain tumour resection (Ojemann, 1983; Penfield & Roberts, 1959), allowing to maximize the extent of the tumour resection and the preservation of functionally relevant brain sites. Among the various cognitive domains ESM is applied, executive functions and more complex cognitive abilities, such as performance monitoring and error awareness, have been less explored due to the inherent complexity of these functions. Here, we explored the functional role of the aIC in performance monitoring for self-made actions by synchronizing the ESM with the presentation of either congruent or incongruent feedback followed by a monitoring assessment as to whether the patient agreed or not with the feedback provided. Our results revealed one relevant point at the most posterior portion of the aIC related to self-monitoring. Specifically, 3 stimulations over the same cortical point of the aIC resulted in an inability to correctly monitor her performance (see **Figure 3B** and **3C**). Interestingly, all three trials resulting in a monitoring failure corresponded to erroneous responses, 2/3 when incongruent feedback was presented and 1/3 after congruent feedback.

Previous literature has pointed out the key role of the insula in self-related processing (David, 2012; Sperduti et al., 2011). Together with other brain regions such as the mPFC and the ACC, the insula, particularly the aIC, has been largely found to be involved in self-recognition, self-monitoring and discriminating the self from non-self-sources of sensory input (Blakemore et al., 1998; Farrer & Frith, 2002; Kircher et al., 2000; Sugiura et al., 2000). Self-awareness is an essential component of metacognition, through which we can adjust our beliefs of the world - the modulation of the self- by monitoring and controlling our behaviour. For example, patients with fronto-insular stroke, behavioural variant frontotemporal dementia and Alzheimer's disease show a lack of self-awareness, insight, and self-monitoring, which translate into anosognosia and daily behavioural impairments (Hebscher et al., 2016; Hebscher & Gilboa, 2016; Rosen et al., 2014; Shany-Ur et al., 2016; Sunderaraman & Cosentino, 2017), leading to worse functional outcome and poor compliance with rehabilitation (Ownsworth & Clare, 2006). One of the most prominent theories (Blakemore & Frith, 2003; Frith et al., 2000) claims that self-monitoring in healthy people is based on a comparator process that determines deviations between the predicted and actual consequences of physical or mental actions. When predicted and actual consequences match, the observed outcome is experienced as coming from the self. On the other hand, other authors have proposed that self-monitoring is usually based on a more direct comparison between the

intention underlying an action and its observed outcome (Franck et al., 2001; Fourneret et al., 2001; Jeannerod, 1999).

Furthermore, a line of evidence suggests an involvement of the aIC in performance monitoring related to its role in error awareness (Klein et al. 2007; Klein, Ullsperger, & Danielmeier, 2013). Direct evidence on the role of the aIC within the error monitoring network has been recently provided (Bastin et al., 2017), showing a broadband increase in neuronal population response at the aIC during erroneous NoGo trials in a group of epileptic patients undergoing intracerebral electroencephalography. Recent models regarding error monitoring suggest the existence of a detection threshold that error signals might need to surpass in order to be detected by our system (Steinhauser & Yeung, 2010; Ullsperger et al., 2010; Wessel, Danielmeier & Ullsperger, 2011; Wessel, 2012). Interestingly, several authors have proposed that these signals provide multimodal information to be integrated in the insula conforming the input to the error monitoring network and reflecting an error-awareness signal (Ham et al., 2013; Klein et al., 2013; Sridharan, Levitin & Menon, 2008; Ullsperger et al., 2010).

Consistently with the abovementioned studies, our results support the functional relevance of the aIC in error monitoring/processing indicated by the inability to correctly detect and compare the participants' response/performance to the feedback provided when inducing a virtual lesion on aIC regions, specifically during self-made errors, although caution with this interpretation should be taken due to the reduced number of trials used to evaluate performance monitoring. Hence, these results seem to indicate that the aIC might play a relevant role in monitoring deviances from the predicted action consequences during self-made actions, informing the system than an unexpected outcome/prediction error has occurred. This interpretation fits well with clinical populations showing difficulties in the detection of discrepancies between predicted vs. actual outcomes, such as the case for AHP patients, commonly associated to lesions at the aIC region (Karnath, Baier & Nägele, 2005; Vocat et al., 2010). Moreover, neuroimaging studies in patients with schizophrenia, who commonly present difficulties in discriminating between self-generated and externally generated sources of sensory stimuli, have consistently reported both structural and functional abnormalities of the insula (Makris et al., 2006; Moran et al., 2013; Palaniyappan et al., 2013; Saze et al., 2007; Takahashi et al., 2004, 2005).

Our event-related fMRI results for the contrasts of interest (Incongruent vs Congruent and Incongruent vs Neutral feedback) support ESM findings by revealing significant clusters of activation at the most posterior portion of the aIC (bilaterally) when the patient was presented with incongruent feedback (**Figure 4A).** Importantly, other error-related regions such as the ACC

and basal ganglia nuclei such as the putamen (bilaterally) and right globus pallidus were significantly activated during the presentation of incongruent feedback. These regions have been previously related to monitoring and outcome prediction of ongoing events (Botvinick et al. 2001; Mathalon et al. 2003). Several authors suggest that the aIC involvement in self-monitoring and error awareness may pertain to its role in the salience network (Seeley et al., 2007; Menon & Uddin 2010), which involves the mPFC, ACC, IFG, amygdala, IPL, thalamus aIC and other brainstem nuclei (Seeley et al., 2007) and responding to behaviourally relevant events. Functional connectivity analysis of aIC during error awareness confirms its coordinated activity with distant brain regions of the salience network, presumably to amplify the neural salience-signal and motivational properties of the detected error (Deen et al., 2011; Dosenbach et al., 2008). Therefore, it might be the case that when encountering events that deviate from expectations (i.e., prediction errors), as in our case, inserting incongruent feedback, the basal ganglia relays information to the ACC via the thalamus and communicates with the executive control network, signaling the need for increased cognitive control and monitoring (Ham et al. 2013; Kennerley et al. 2006). Interestingly, as previously mentioned on introductory paragraphs, the human insula contains von Economo neurons, whose large axons could provide a neuronal basis for rapid signal communication between aIC and ACC, as well as with other brain networks (Menon et al., 2020).

Several limitations were encountered during this project. On one hand, the limited availability of ESM data might hinder the possible generalization of our findings, therefore more patients must be recruited and tested with this protocol. Moreover, the cognitive decline in the patient's performance at the articulatory level, probably related to the resection at the left motor opercular region, might have influenced her performance on other cognitive domains dependent on verbal articulation outputs. Also, a sample of healthy individuals could also be recruited to perform the fMRI experimental paradigm favouring the generalization of our results. Therefore, the current protocol should be tested with other similar patients with brain tumours involving the aIC region to properly assess the benefits of this multimodal protocol in evaluating aIC function, allowing for example stricter corrections regarding the significance of the fMRI results.

In conclusion, the present study offers new insight regarding the exploration of the aIC in patients undergoing awake brain surgery to ensure a preservation of the integrity of the self-awareness/monitoring network. We provide novel insight regarding the functional role of the aIC in performance monitoring, following previous findings reporting aIC responses to error trials and incongruent feedback processing (Bastin et al., 2017; Ham et al., 2013). A better understanding of the aIC's role during self-attributed outcomes may help shed light on feedback/error processing abnormalities reported in several neuropsychiatric disorders

associated with functional and structural abnormalities of the aIC (Diener et al., 2012; Hatton et al., 2012; Naqvi & Bechara, 2009; Palaniyappan & Liddle, 2012; Shepherd et al., 2012), for example, patients with schizophrenia, AHP, major depression, and/or drug addiction, who show difficulties in self-monitoring as well as an abnormal sense of agency for their thoughts or actions (Eshel and Roiser, 2010; Karnath et al., 2005; Vocat & Vuilleumeier, 2010; Ziauddeen and Murray, 2010).

GENERAL DISCUSSION

4. GENERAL DISCUSSION

As human beings, we are not only conscious of the world surrounding us but also of the fact that "I" am the one initiating and controlling my actions in the outside world. The feeling of being in control of our actions is a fundamental aspect of our human experience. The experimental work included in the present thesis aimed at extending the actual knowledge on the agency attribution mechanisms, capitalizing on A-O modulations while inspecting the neural correlates and subjective experiences across a series of experimental settings. Although the findings of each study have been discussed in detail in the previous chapters, I will now provide an overview of the insights gathered from this work and attempt to relate them more generally to prior knowledge.

4.1. Agency attribution in the face of error

External agency attribution judgements: Neurophysiological signatures associated to agency prediction errors (aPE)

As previously exposed on Chapter 1, there is an abundant corpus of knowledge demonstrating that A-O inconsistencies, either in terms of temporal synchronicity or spatial consistency, significantly reduces the SoA and increases the external attribution for the actions performed (Blakemore et al., 1998; Ebert & Wegner, 2010; Farrer et al., 2003, 2008, 2013; Fourneret & Jeannerod, 1998; Haggard et al., 2002; Hon et al., 2013; Kawabe, 2013; Kannape et al., 2010; Knoblich & Kircher, 2004; Kühn et al., 2011; MacDonald & Paus, 2003; Nielsen, 1963; Synofzik et al., 2006; Sato & Yasuda, 2005; Tsakiris et al., 2005, 2006; Wen et al., 2015, see **Table 4**). Drawing upon these premises and the lack of current knowledge about the neural correlates underlying the feeling of 'mineness', in **Study 1** we conducted three EEG experiments using a modified version of the flanker task inducing aPE. To do so, we developed a new experimental set-up allowing the illusion of ownership and agency over a pair of digital hands while recording the EEG activity in a group of healthy participants. Taking advantage of its high temporal resolution, we explored ERPs and time-frequency data and its association with the participants' agency attribution reports. Based on previous evidence indicating that A-O content consistency and temporal contiguity are required to infer the cause of an effect (i.e., '*the agent*') (Hume, 1739/1888, p. 293; Sato and Yasuda, 2005), we predicted a modulation of internal vs. eternal agency attribution judgements when introducing A-O incongruences.

Concerning one of the main goals of this thesis, our main finding was the characterization of a neurophysiological P6-delta modulation underlying external agency attribution judgements (Experiments 1 and 3) whenever content incongruences were presented (i.e., incongruent hand response movement with respect to the participants' action). Importantly, we found a significant association between the amplitude of the delta power increase and the P6 amplitude with the subjective intensity of external agency attribution reports.

We interpreted the P6 component as a P3-like component following the *context updating hypothesis* (Donchin & Coles, 1988), previously mentioned during the introductory section (see subsection ***1.4.3.2. Context updating: P300 component***). This theory states that the P3-family components index brain activity related to the revision of the mental representation induced by incoming stimuli, after the initial sensory processing takes place (Donchin, 1981, see **Figure 24**). During this evaluation of the mental representation, if no stimulus attribute mismatch is detected, the current mental model or "schema" of the stimulus context is maintained (only sensory evoked potentials are recorded -N100, P200, N200-), whether if a new stimulus is detected, an attentionally-driven process undergoes the 'updating' of the stimulus representation giving rise to a P300 potential (Donchin et al., 1986).

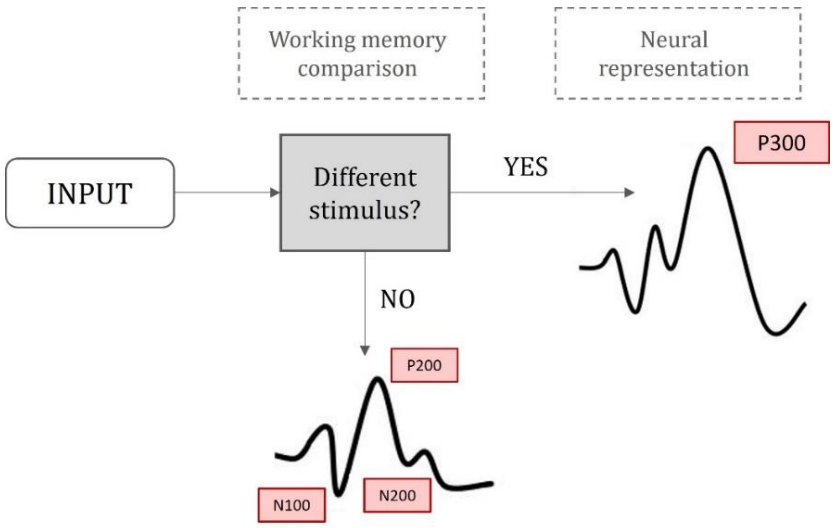


Figure 24. Schematic illustration of the P300 context-updating model. Stimuli enter the processing system, and a memory comparison process is engaged that ascertains whether the current stimulus is either the same or not as the previous stimulus or not. If the incoming stimulus is the same, the neural model of the stimulus environment is unchanged, and sensory evoked potentials (N100, P200, N200) are obtained after signal averaging. If the incoming stimulus is not the same and the subject allocates attentional resources to the target, the neural representation of the stimulus environment is changed or updated, such that a P300 potential is generated in addition to the sensory evoked potentials. *Adapted from Polich (2003).*

Following this theory, our findings suggest that whenever an external error was introduced, a violation of the predicted mental context of the sensorimotor consequences occurred, resulting in the external agency attribution for the actions performed. A fast process of revision and reanalysis of the predicted A-O models initiates, starting at about 400 ms after the incongruence is detected, and reaching its maximum at about 600 ms, as manifested in the P600 oscillation.

Moreover, another hypothesis regarding P3-like deflections propose that these neural signatures reflect the neuromodulatory effect of the locus coeruleus (LC) norepinephrine (NE) system in the neocortex (Nieuwenhuis et al., 2005), suggesting that they indicate a LC-mediated enhancement of neural responsivity in the cortex to task-relevant stimuli. This hypothesis is supported both by similarities between the target areas of NE projections, reported P3 sources/generators, the latency of the LC-NE effects (100–200 ms post-discharge) (Aston-Jones et al., 1986; Berridge & Waterhouse, 2003; Foote et al., 1983), consistent with the latency of the P3, as well as by the condition to which they are both sensitive to, including motivational significance, probability of occurrence, and attention allocation. In a latter study, Nieuwenhuis, De Geus and Aston-Jones (2011) proposed that the *LC-P3 hypothesis* and the *context updating hypothesis* were complementary to each other, with the *LC-P3 hypothesis* providing a mechanistic account to the more conceptual *context updating hypothesis*. Our results fit well with the *LC-P3-context updating* explanation, as the insertion of EE constitute an unexpected and salient event which most probably might have engaged a phasic NE response indicating the need to attend to the environment to learn from it (Bouret & Sara, 2004).

Although not directly related to the study of agency attribution, previous ERP studies looking at unexpected response outcomes have reported similar complex waveform patterns, in which an initial negativity is followed by a later positivity, namely an N2/P3 complex (Wessel, 2012; Wessel & Aron, 2013). Wessel and Aron (2013) directly compared two kinds of unexpected events and showed that the ERN/Pe complex after action errors and the N2/P3 complex after unexpected action outcomes are interconnected brain components, which might be sustained by a common mechanistic basis leading to PES effects, attentional orienting and context updating. Following this line of evidence, we suggest the appearance of the N2/P6 compound might index an explicit and conscious judgement of external agency attribution, reflecting an external causal attribution of action outcome and indicating the need to implement control processes requiring the update of the current mental model (i.e., task set reconfiguration).

Importantly, the delayed appearance of the P600 effect comes to confirm its association with the retrospective aspects of the agentic experience (Moore & Haggard, 2008; Synofzik et al., 2008a), which depends on the computation and differential weighting of external postdictive cues (multimodal somatosensory, visual and auditory reafferent representations), intentions (goals) and internal predictive cues. This interpretation also fit comfortably with 'two-stage' models such as the one proposed by Del Cul et al. (2009) for decision-making. Firstly, a fast, non-conscious sensorimotor route allowing rapid and rudimentary evaluation of evidence would be responsible for the processing of internal cues, providing us with the *FoA* or 'minimal self' (Gallagher, 2000), followed by a slower, more accurate conscious detection route associated to external/postdictive information and related to judgement/evaluative processes and context updating. In line with this proposal, we suggest that the P600 component indexes the occurrence of aPE, reflecting a metacognitive *JoA* resulting in an external causal attribution of the action indicating the need to implement control processes and an update of current mental models. Besides, the delayed appearance of the P600 component contrasts with the fast onset of the ERN component, associated to the non-conscious action monitoring route (Del Cul et al., 2009) and elicited as soon as an internal mismatch between the intended movement and the predictive/efference copy information occurred (Rodriguez-Fornells et al., 2002; van Schie et al., 2004; Niuwenhuis et al., 2002) (see **Figure 25**).

Remarkably, the parietal topographical distribution of the P600 is consistent with the role of the PPC in generating and maintaining internal forward models of the future body state (integrating visual, somatosensory and efference copy signal), and in detecting mismatches between the desired and the actual movement to aid in monitoring voluntary actions (Andersen & Buneo, 2002; Blakemore & Sirigu, 2003; Cui, 2014; Desmurget & Grafton, 2000).

In another interesting study using an ERP-Virtual Reality paradigm (1PP) (Padrao et al., 2016), the introduction of EE by the avatar diluted the SoA in healthy participants. EE triggered a delayed negativity which topographically resembled the N400 parietal component associated to semantic or conceptual violations (Kutas & Federmeier, 2011). These authors proposed the N400 reflected a violation in the process of understanding "our own body movement semantics". It could be the case that although the same level of SoA and SoO was reported in our experimental setting for the congruent block during Experiment 1 of **Study 1** (both ratings corresponding to 'strongly agree' for the items addressing SoA and SoO), as well as the same degree of SoA reduction/external agency attribution when introducing EE as in Padrao et al. (2016), a more enhanced sense of embodiment was provided during the 1PP virtual immersion due to the influence of multimodal

reafferent information provided, unavailable in physical reality (Kilteni, Groten & Slater, 2012), resulting in the presence of the N400 deflection as a disembodiment indicator.

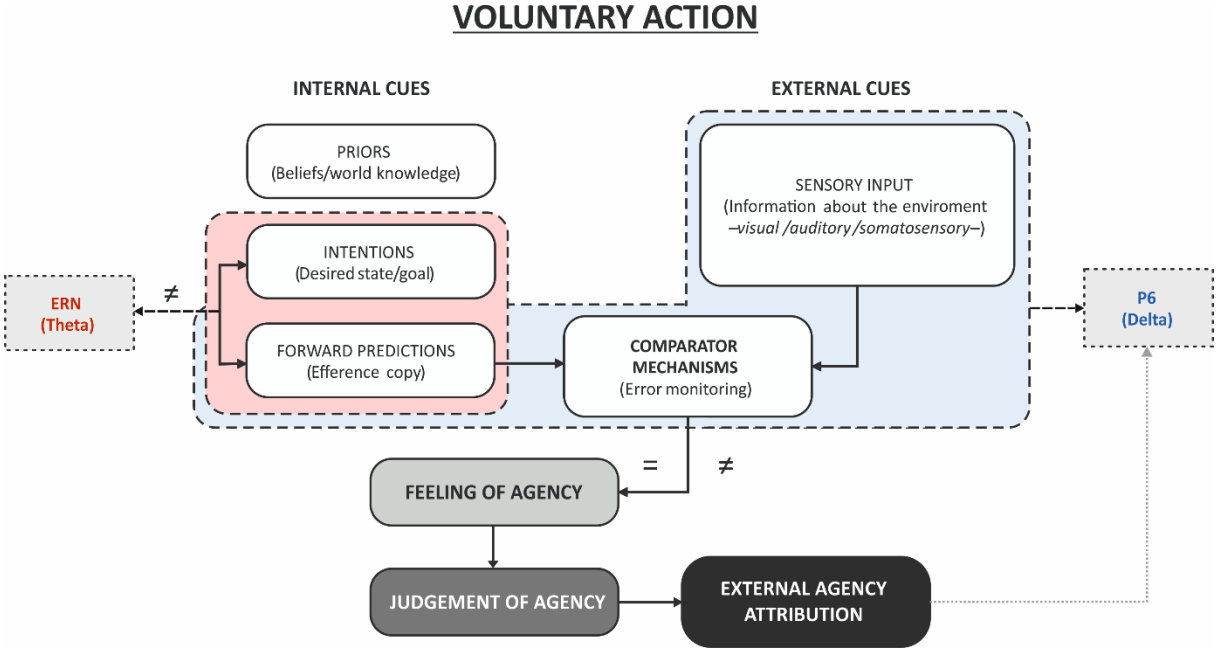


Figure 25. A summarized integrative account for our results on Study 1. Sense of Agency (SoA) is depicted. The SoA arises based on the integration of both internal and external cues. On one hand, internal cues such as prior beliefs, intentions to act and forward predictions are computed. By means of comparator mechanisms, discrepancies between intended actions and efference copies might elicit the ERN component, signaling the need for fast compensatory error correction mechanisms. Depending on the environment and context, predictions are compared to the sensory inputs coming from the external world, resulting in a positive/negative feeling of agency depending on the congruence or not of these comparisons. A higher-level cognitive judgement of agency may follow the pre-reflective feeling of being the agent, influenced by other background information about the external world. In the present study, when an incongruence was introduced between the predicted and the actual sensory information (External errors), a P6 component appeared signaling the need for revision and belief updating associated to an increased external agency attribution judgement. Based on *Synofzyk et al., (2008a)* and *Wolpert and Miall (1996).*

The prominent role of retrospective aspects of outcome attribution in the determination of the attribution of agency was reinforced by our findings in Experiment 3 of **Study 1**, when observing a shift forward in time of the delta/P600 component tagged to the feedback appearance, clearly highlighting the role of the external sensory input in agency attribution.

Brain oscillatory activity focusing on low-frequency responses (1-15 Hz range) was also examined during **Study 1** in relation to action monitoring and error detection processes (Bernat et al., 2015;

Cavanagh et al., 2010; Cavanagh & Frank, 2014; Cohen, 2011; Luu et al., 2004). Our results evidenced a very distinctive pattern of activity characterized by a two-stage oscillatory pattern depicting an increase in early frontal theta (3-8 Hz) followed by a more latter posterior delta (1-3 Hz) activity, which was in line with our ERP results depicting a distinguishable frontal-N2 parietal-P6 complex.

Theta-band oscillations (3-8 Hz) in the frontal network seem to be responsible for transmitting internal signals of error (i.e., prediction errors), and for providing the system with relevant information about the need to implement performance and behavioural adjustments at an earlier stage of processing, as reflected here by the modulation of the N2 component (Cavanagh, Cohen & Allen, 2009). In **Study 1**, although participants had responded correctly, the occurrence of an externally imposed error might have triggered control mechanisms by which neural networks supporting performance monitoring and cognitive control communicate with each other to guide behaviour, resulting in an increase of theta activity.

Noteworthy, we also encountered a strong positive correlation between increased delta power and P6 amplitude at posterior electrode locations, reinforcing the P6-delta linkage. Notably, similar to what we observed for the P6 component, a clear association between delta power increase and external agency attribution judgements was found, further supporting our proposal regarding its functional involvement during agency attribution. Delta oscillations have been related to several cognitive functions, such as the motivational relevance of the task and the salience of the target stimulus (Knyazev, 2007, 2012), as well as with reward processing (Bernat et al., 2011, 2015; Cavanagh, 2015; Knyazev, 2007; Nelson et al., 2011) and commission of motor errors (Yordanova et al., 2004). Moreover, parietal delta oscillations are considered to mediate signal detection and decision-making (Başar et al., 1999; Schürmann et al., 2001), modulating the rhythmic gain of information accumulation (Lakatos et al., 2008; Schroeder & Lakatos, 2009). We suggest that parietal delta-band activity underlying the experience of external agency attribution in **Study 1** might be revealing the conscious reflection on the external agency judgement produced by the discrepancy between internal and external cues (i.e., model mismatch), informing the system about the need to re-evaluate the predicted mental model allowing an updating of the agentic inferential processing.

Temporal synchronicity and agency attribution judgements

Temporal contiguity discrepancies between action and its related effects, such as shifting the sensory reafferent information with respect to the participant's forward predictions, have also been found to induce a gradual decrease in the reported SoA as A-O temporal delay increases, with

an agency threshold around 150 ms (Asai & Tanno, 2007; Ebert & Wegner, 2010; Farrer et al., 2008; Franck et al., 2001; Hara et al., 2015; Imaizumi & Asai, 2015; Kawabe et al., 2013; Kühn et al., 2013; Sato & Yasuda, 2005; Wen et al., 2015). Following our results on Experiment 1 (**Study 1**) showing a P6-delta association as a neural marker for external agency attribution, we performed Experiment 2 with the aim of investigating the influence of A-O temporal delays on agency attribution, by introducing delays of 0, 150 and 400 ms between the participant's response (i.e., button press/hand movement) and the visual effect (i.e., digital button press/hand movement) displayed. Our results on Experiment 2 are in line with previous evidence showing a significant gradual reduction in the reported internal agency attribution as well as SoO as the delay increased. Nevertheless, no significant increases in external agency attribution judgements were observed.

As previously mentioned during the introductory section, introducing A-O temporal delays might not be as impactful as one might think. It has been proposed that decreases in internal agency attribution following temporal delay (Ebert & Wegner, 2010; Farrer et al., 2013; Hon et al., 2013; Kawabe, 2013; Kühn et al., 2011; MacDonald & Paus, 2003; Sato & Yasuda, 2005; Tsakiris et al., 2006; Wen et al., 2015) might be due to the influence of other high-order cognitive processes such as motor control or response bias arising from graded conditions and response options (Wen, 2019). For example, it has been shown that participants still report a causal relationship between their actions and the perceived consequences even when confronted with temporal delays ~1 sec (Farrer et al., 2008). Our results in Experiment 2 seem to go in line with this assumption, showing a gradual decrease in the internal attribution of actions as A-O delay increased and engagement of outcome monitoring processes (FCRP and FRN modulations for correct and SE, respectively), but without eliciting a sufficient disruption in the agentic inference to reach an external agency judgement. Therefore, our results seem to indicate that sensory predictions for external agency might be more flexible and adaptable to temporal aPE than to content aPE. It might be the case that although A-O mismatches in the temporal domain influence the FoA and/or internal agency judgements, other cognitive instances might still indicate a correct goal achievement of the intended action, therefore preserving the self-attribution, contrary to what we observed during content or spatial aPE.

4.2. Decoding agency attribution

In recent years, brain computer interfaces (BCI) technologies have seen very rapid developments displaying a wide range of potential applications in the near future. In essence, BCI systems record the brain activity (i.e., neural signals) of the user as inputs, process and transforms them into outputs to control an external device. BCIs have a lot of potential in restorative and clinical uses (Glannon, 2014) for example, by allowing the communication in paralyzed patients (Mullin, 2017) or enabling control of prosthetic devices in patients with motor disabilities or amputees (Bowsher et al., 2016). Also, they may proof useful for rehabilitative and neuropsychiatric treatment by providing neurofeedback (Lim et al. 2012; Sreedharan et al. 2013).

EEG signals are widely used to translate the subject's brain activity into action with the purpose of controlling external devices, due to its non-invasive, high temporal resolution and low cost (Lopez-Larraz et al., 2018). One of the main challenges of BCI systems can be their non-optimal performance, sometimes leading to erroneous interpretations of the user's intentions which might end up in the execution of incorrect actions with undesirable consequences. Relevantly for our purposes, specific neurophysiological signals have been recorded in response to awareness of self-performed errors, observed mistakes performed by another agent or during conflict monitoring. These brain markers have been termed error-related potentials (ErrPs). More importantly, it has been previously reported that ErrPs can be reliably decoded on a single-trial basis (Chavarriaga & Millán, 2010; Chavarriaga et al., 2014; Iturrate et al., 2015; Kim et al., 2017; Usama et al., 2021; Zander et al., 2016). Based on this evidence, we designed **Study 2** with the aim of decoding ErrPs from our previously acquired EEG data from **Study 1** (Experiment 1) on a single-trial basis. To do so, we applied a support vector machine (SVM) decoder to classify self-made errors, externally generated errors and correct responses based on the topographical distribution (considering all the electrodes and time points) of the brain's response to each condition. To explore the neurophysiological distinctions between the brain's responses to these error conditions, we used time domain data to compare the latency, amplitude and topographical distribution of the ERN (self-made errors) and P6 (external errors) components. To do so, we employed a SVM linear multiclass classification algorithm. In general terms, an SVM analysis typically consists of three phases: (i) a training phase, during which a portion of the data (i.e., training set) is used to build a classifier or hyperplane discriminating between classes by fitting model parameters; (ii) a validation phase, during which another sub-set of the data (i.e., validation set) is used for making adjustments to the classifier by tuning its hyperparameters by means of a cross-validation approach; and (iii) a testing phase, in which the classifier attempts to predict the

new set of data (i.e., testing set) into the specified classes, determining its performance by comparing its class predictions to the actual classes (Miller, Sacched & Gotlieb, 2020).

Our results yielded significant findings, showing a performance better than chance from +34 ms onwards, with overlapping peaks of accuracy with the ERN and P6 time windows. Interestingly, a pre-response correct classification could be done at the singe-trial level for SE vs. Correct trials, allowing the user to potentially avoid the occurrence of the erroneous performance. Numerous studies have been performed with the aim of distinguishing errors and correct actions by means of decoding ErrPs using EEG at a single-trial level, showing variations in classification accuracy. For example, Kim et al. (2017) employed a robotic arm setting during which users observed it making either corrector erroneous movements, achieving classification rates of 95% in the error class and 86% in the correct class. Chavarriaga and Millán (2010) further incorporated ErrPs during an automatic cursor's movement system to correct erroneous movements by detecting ErrPs, reporting mean classification rates of 75.8% for the error class and 63.2% for the correct class. Also, some studies have investigated differences in ErrPs in relation to the error source, for example, distinguishing between i) response-ErrPs, elicited during error detection, ii) feedback-ErrPs, caused when the user is informed about an error which he/she has not consciously perceived, iii) observation-ErrP, when the user observed an error committed by the robot/another agent or iv) execution-ErrPs occurring when the machine fails to execute the intended command of the human agent (Ferrez & Millán, 2008; Kim et al., 2017; Spüler & Niethammer, 2015).

Nevertheless, despite these recent investigations, most of the studies addressing error classification only focus on distinguishing erroneous vs. correct actions, without taking into consideration the agentic nature of the erroneous actions. As previously discussed on the introdutory section (see **Figure 11**), the existence of two distinct error monitoring subsystems running in parallel to the forward computations has been proposed (Charles et al., 2013; Logan & Crump, 2010): i) an internal error-monitoring loop, in charge of detecting deviations between intended vs. predicted actions and implementing fast error-correction or compensatory mechanisms and ii) an external error-monitoring loop, responsible for evaluating predicted vs. actual state mismatches. Based on this perspective, dissociable ErrP were recorded as functional indexes of these loops (**Study 1**), as the ERN and P6 components associated to internal/predictive and external/postdictive cues. Therefore, based on their distinct topographical characteristics and given the fact that they appear at different processing stages (before and after the feedback, respectively), it is possible to correctly decipher and classify distinct erroneous actions based on their agency, as our findings in **Study 2** reveal. We concluded that the implementation of decoding

algorithms for classifying the agency of errors might benefit the correct implementation of BCI systems, distinguishing between self and externally generated errors, affecting the processing of action selection, adaptation and subsequent goal achievement, as well as potentially detecting pre-error signatures to avoid the execution of on-going errors.

4.3. The functional role of the anterior insular cortex (aIC) in self-monitoring

Being able to monitor our own performance and the outcome of our actions is fundamental for adapting our behaviour in a changing environment. During this monitoring process, we are constantly comparing our internal/forward predictions with the actual outcomes giving rise to prediction errors whenever a mismatch is detected during this comparison process. From a clinical point of view, the preservation of self-monitoring abilities is of crucial importance, as an impaired conscious perception of errors has been associated with poor insight in consequences of neurological conditions (O'Keeffe et al., 2004). One of the brain regions that have been related to self-monitoring abilities is the insular cortex, especially its anterior portion (Holroyd & Coles 2002; Ullsperger & von Cramon 2001; Wessel, 2012, see **Figure 26**). The aim of **Study 2** was to investigate the functional role of the aIC in self-monitoring and error awareness in a patient undergoing awake brain surgery for tumour resection. To do so we designed a multimodal approach combining neuropsychological assessment, ESM and fMRI. We employed a modified version of the Stroop task (Stroop, 1935) to induce A-O incongruences by introducing random feedback which could either congruent or incongruent with the patient's performance. Our results revealed new direct evidence of the involvement of the aIC in monitoring performance, showing increased difficulties in detecting A-O mismatches when stimulating a cortical site located at the most posterior part of the aIC as well as significant activations at this region during outcome incongruences for self-made actions during the fMRI acquisition.

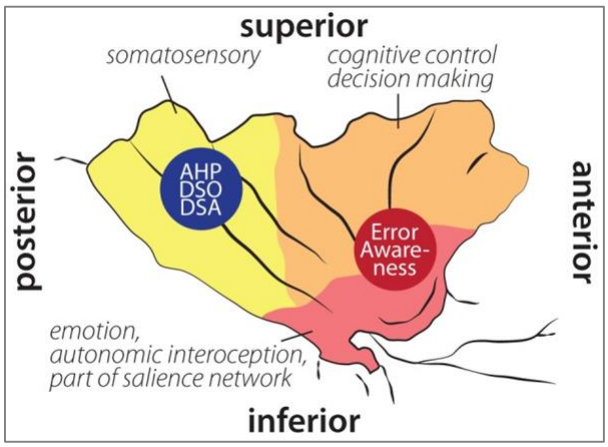


Figure 26. Schematic illustration of insular cortex involvement in error awareness, anosognosia for hemiplegia (AHP), disturbed sense of ownership (DSO) and disturbed

sense of agency (DSA) overlaid on schematic drawing of functional areas within the insula according to *Deen et al. (2011*). *Extracted from Klein et al. (2013).*

Recent findings report significant activations at the aIC at a wide range of conditions and behaviours (see **Figure 27**), such as interoceptive stimuli (Critchley et al., 2004) like pain (Kong et al., 2006), body awareness (Farrer & Frith, 2002; Farrer et al., 2003), emotional awareness including maternal and romantic love (Bartels & Zeki, 2004; Leibenluft et al., 2004), time perception (Coull, 2004) and attention (Mason et al., 2007; Weissman et al., 2006). Moreover, several studies have pointed out the key role of the insula in self-related processing (David, 2012; Sperduti et al., 2011). Together with other brain regions such as the mPFC and the ACC, the insula, particularly the aIC, has been largely found to be involved in self attribution and self vs. other distinctions (Blakemore et al., 1998; Farrer & Frith, 2002; Kircher et al., 2000; Sugiura et al., 2000). For example, hearing distorted feedback of one's own voice increases the response in the left insula, compared to hearing feedback of a different person's voice (McGuire et al., 1996). Also, during the realization of a motor task, the left insula was shown to respond to auditory feedback coming from the self, compared to randomly timed feedback (Blakemore et al., 1998). Bilateral activations of the aIC region have also been reported when presented with feedback resulting directly from the participant's action, compared to unrelated, random feedback (Farrer & Frith, 2002). Following this pattern of evidence, it has been proposed that the aIC might be a key central hub for self-awareness, "*providing a neural substrate that instantiates all subjective feelings from the body and feelings of emotion in the immediate present (now)"* (Craig & Craig, 2009, p. 65).

Furthermore, there is an alternative theory that suggests an involvement of the aIC in performance monitoring related to its role in error awareness (Klein et al. 2007; Klein, Ullsperger & Danielmeier, 2013). This theory is in line with fMRI studies showing that the aIC is consistently activated for consciously perceived errors compared to unperceived errors (Klein et al., 2007; Hester et al., 2005). Direct evidence on the role of the aIC within the error monitoring network has been recently provided (Bastin et al., 2017) by recording intracerebral electrophysiological signals from the aIC in epileptic patients while performing a Stop-signal paradigm (i.e., participants had to rapidly press a button after GO cues and try to withhold their prepared response after the STOP cues). Their results indicate that the aIC rapidly detects and conveys error signals to dorsomedial PFC, which might use this input to adapt behaviour following inappropriate actions (Bastin et al., 2017). Current views suggest that error awareness can be explained by an accumulating evidence account (Ullsperger et al., 2010; Vocat et al., 2010; Wessel, Danielmeier & Ullsperger, 2011; Wessel, 2012). This account describes the accumulation of

evidence for an error from very different sources (posterior MFC activity, autonomic responses, proprioceptive and other sensory input that deviates from expectation).

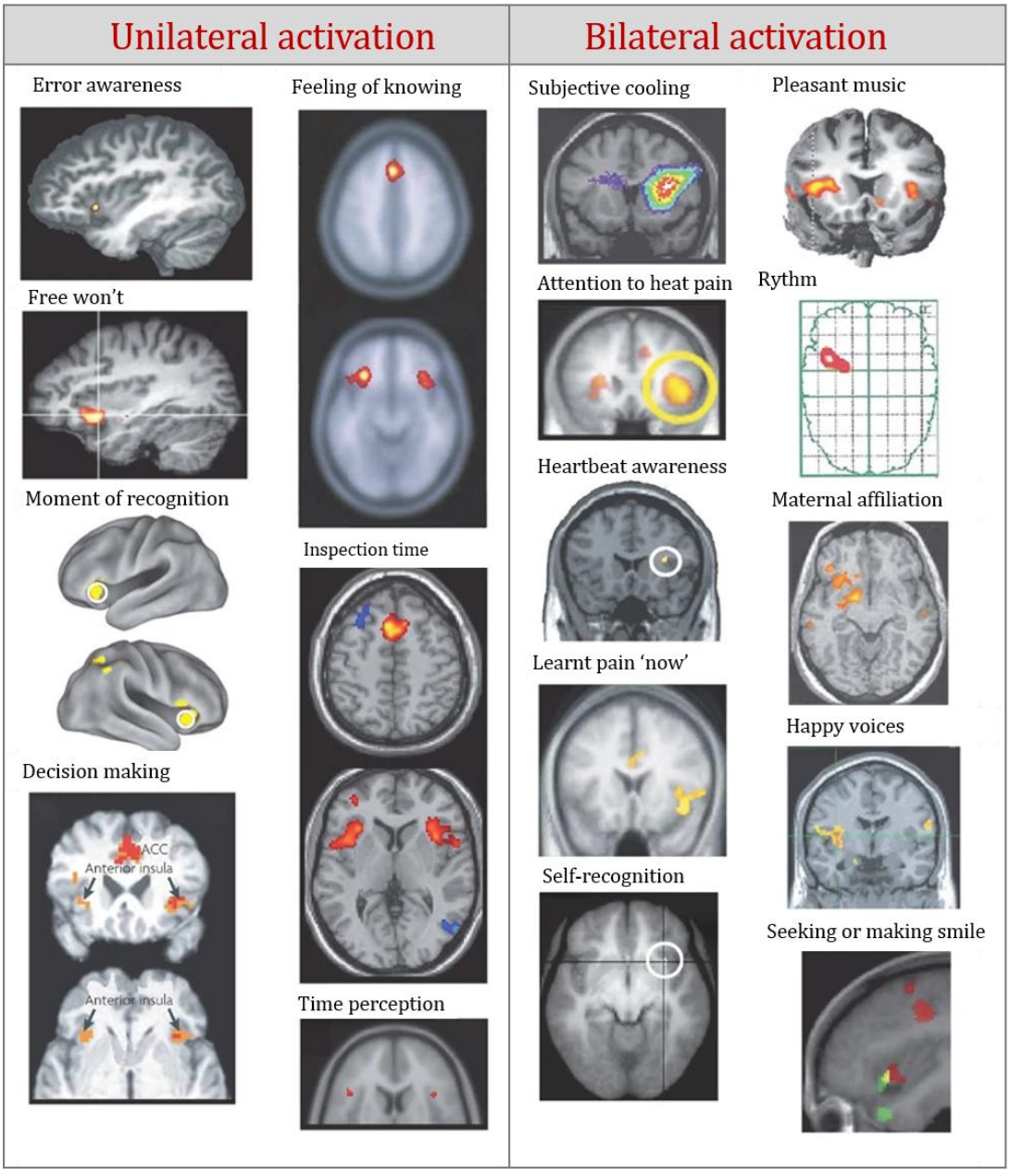


Figure 27. A summary of imaging results showing activation of the anterior insular cortex (aIC). Activation of the left aIC was reported in mothers viewing photos of their own child; greater activation of the left than of the right aIC was associated with both maternal and romantic love; activation of the left aIC was reported while subjects were either seeing or making a smile; activation of the left aIC was found while subjects attended to happy voices; activation of the left aIC was associated with hearing pleasant music; selective activation of the left aIC was observed in subjects experiencing joy; and selective activation of the left aIC in females was found that correlated with self-reported orgasm ratings. ACC: anterior cingulate cortex. *Adapted from Craig and Craig (2009).*

Consistently with the abovementioned studies, our results reported in **Study 3** support the functional relevance of the aIC in error processing indicated by the inability to correctly detect and compare the participant response/performance to the feedback provided, specifically during self-made errors.

As previously discussed, a plausible explanation for our results might be that when inducing a virtual lesion onto the aIC, a failure in detecting the mismatch between the participants expected outcome and the actual outcome might occur due to a failure in the monitoring/comparator mechanisms in charge of detecting the presence of these prediction errors. As previously exposed on the introductory section, the *Comparator model* (Frith, 1992; Wolpert et al., 1995, 2001) is based on the idea that our system contains internal models representing aspects of our own body and the external world. Accordingly, the forward model uses a copy of the current motor command (i.e., efference copy) to build up predictions about the future state of our system and the sensory consequences of our actions, which are matched against the actual sensory feedback enabling the distinction between self-generated and external generated signals (by attenuating the sensory effects of self-produced movements) (Blakemore et al., 1998, 2000; Sato & Yasuda, 2005; Sato, 2008). From this perspective, our results seem to point out to the aIC as a relevant brain region involved in this monitoring process, informing the system that an unexpected outcome/prediction error has occurred. Interestingly, as depicted in **Figure 28**, it has been proposed that the integration of the different salient signals reflecting a deviation from the predicted outcomes (i.e., prediction errors) may be processed in the insular cortex (Craig & Craig, 2009; Ham et al., 2013), conforming the input to the error monitoring network and reflecting an error-awareness signal (Ham et al., 2013; Klein et al., 2013; Sridharan et al., 2008; Ullsperger et al., 2010). Furthermore, the integrative nature of the insular cortex, due to its anatomical connectivity pattern receiving direct thalamic and horizontal cortical afferents and reciprocally connecting to an extensive network of cortical and subcortical brain regions, might represent the cortical basis for awareness carrying the mental representations of the sentient self through the integration of salient events at each moment of time (Craig & Craig, 2009, see **Figure 28**).

Moreover, our results during the event-related fMRI analysis revealed an interesting pattern of activity. Significant clusters of activation for the contrasts of interest (i.e., Incongruent vs. Congruent feedback and Incongruent vs Neutral feedback) were found at the most posterior portion of the aIC (bilaterally), supporting our findings during the ESM procedure as well as with previous literature. Together with the aIC, significant clusters at other error monitoring regions were found, such as the ACC, putamen (bilaterally) and right globus pallidus. Of special interest was the conjoint activation of the aIC and ACC, which has been repeatedly found in the literature

due to their large functional and anatomical connectivity (Taylor, Seminowicz & Davis, 2009) and their prominent role within the *salience network* (Seeley et al., 2007). The *salience network* main function is to identify relevant internal and/or external stimuli in order to guide our behaviour (Seeley et al., 2007). Additionally, the *salience network* shows a high degree of overlap with activity patterns found during performance monitoring [see Klein et al. (2007)] and is also strongly connected with the *LC/NE system* previously described (Aston-Jones et al. 1986, 1991). Consequently, we propose that the insertion of A-O mismatches in terms of incongruent feedback might entail the recruitment of monitoring regions, such as the basal ganglia, where a decreased mesencephalic dopaminergic activity is transmitted throughout the ACC (via thalamus) to the mPFC regions (Holroyd & Coles, 2002; Nieuwenhuis et al., 2004). These signals conveyed in the mPFC help the organism to detect potential cognitive conflicts arising from previous expectations and unexpected outcomes, enhancing action monitoring and control processes, and allowing believe updating (Botvinick, Cohen & Carter, 2004; Holroyd & Coles, 2002; Padrao et al., 2014; Ridderinkhof et al., 2004).

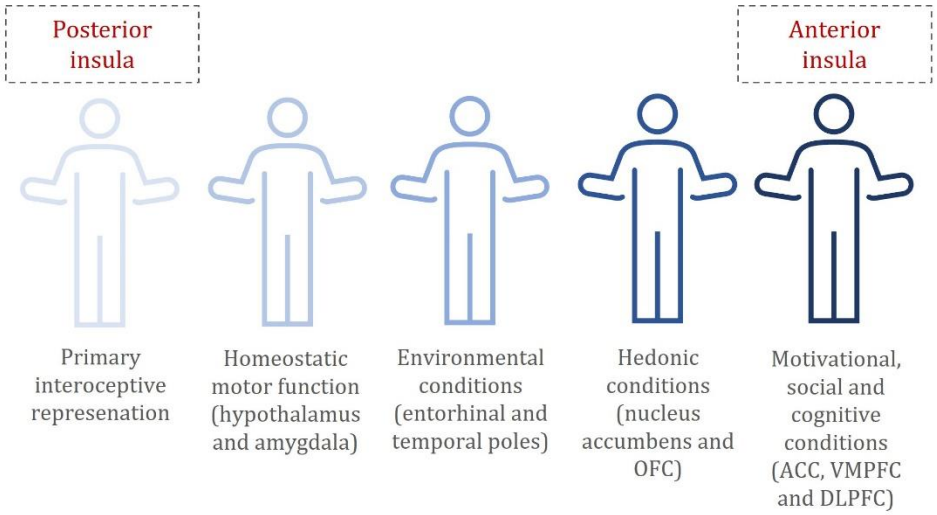


Figure 28. Integration of salient activity, progressing from the posterior insula (left) to the anterior insula (right). The integration successively includes homeostatic, environmental, hedonic, motivational, social and cognitive activity to produce a 'global emotional moment', which represents the sentient self at one moment of time. OFC orbitofrontal cortex; ACC: anterior cingulate cortex; VmPFC: ventromedial prefrontal cortex; DLPFC: dorsolateral prefrontal cortex. *Adapted from Craig and Craig (2009).*

Moreover, the aIC and the ACC share a unique feature at the neuronal level, they contain a special class of neurons, termed von Economo neurons (VENs), which possess distinctive anatomical and

functional features (Nimchinsky et al. 1999). VENs are present in all the great apes, such as chimpanzees, bonobos, gorillas, and orangutans, as well as in elephants and some cetaceans (Allman et al., 2005; Nimchinsky et al., 1999). Interestingly, several clinical populations that course with self-awareness deficits show damage to or abnormal activation or development of the aIC regions, as well as degeneration of VENs (Seeley et al., 2006). For example, some patients with frontotemporal dementia (FTD) show degeneration at fronto-insular and cingulate cortices affecting VENs and display a significant loss of self-awareness behaviours and a loss of emotional self-awareness and of others (Sturm et al., 2006). Other clinical conditions characterized by self-awareness dysfunction such as AHP (Karnath et al., 2005; Vocat et al., 2010) and schizophrenia (Crespo-Facorro et al., 2000; Moran et al., 2013; Wylie & Tregellas, 2010) have also been associated to strokes in insular regions and morphological al abnormalities in the insular cortices, respectively.

Furthermore, many neuroimaging studies addressing emotional awareness report joint activation of the aIC and the ACC in participants experiencing emotional feelings, such as maternal and romantic love, anger, fear, sadness, happiness, sexual arousal, disgust, aversion, unfairness, inequity, indignation, uncertainty, disbelief, social exclusion, trust, empathy or beauty (Craig & Craig, 2009). Based on this evidence, it has been postulated from the A*ffective-control framework* that conflict and cognitive control might be dependent on emotion (Dreisbach & Fischer, 2015; Inzlicht et al., 2015; Saunders et al., 2017). More specifically, this account proposes that conflict or errors elicit an affective reaction which drives changes in behavioural adaptation, updating the task representations. Like this, the occurrence of an unexpected event such as A-O incongruences might trigger an affective response detected by the aIC and ACC driving control adaptation (Steenbergen, Band & Hommel, 2009).

All in all, I believe the encountered results in **Study 3** support the importance of evaluating self-monitoring abilities, promoting the preservation of metacognitive abilities aiding in their functional recovery, and in preventing loss of insight and deficit awareness after surgery. Although there are some limitations to be considered, I believe our work might help elucidate the functional implication of the aIC in self-monitoring and error processing, highlighting the importance of evaluating the aIC functional properties during awake brain surgery and functional neuroimaging promoting the preservation of metacognitive structures. A better understanding of the aIC's role during self-attributed outcomes may help shed light on feedback/error processing abnormalities reported in several neuropsychiatric disorders associated with functional and structural abnormalities of the aIC. for example, patients with schizophrenia, AHP, major depression, and/or drug addiction, who show difficulties in self-monitoring as well as an

abnormal SoA for their thoughts or actions (Diener et al., 2012; Eshel & Roiser, 2010; Hatton et al., 2012; Karnath et al., 2005; Naqvi & Bechara, 2009; Palaniyappan & Liddle, 2012; Shepherd et al., 2012; Vocat & Vuilleumeier, 2010; Ziauddeen & Murray, 2010).

4.4. Limitations and future directions

Despite the empirical contribution of the current findings, there are some limitations that must be considered to understand the scope of these findings and the future research directions:

a) Overall, due to the contextual circumstances in relation to Covid-19, data collection during approximately 1 year was not available. Therefore, due to the delay on data acquisition, we could not conclude the analysis and redaction of what would have been Study 4 of this thesis. Nevertheless, some preliminary data are shown on point **d)** of this subsection.

b) Some limitations regarding the design of our study and possible generalization of our results were encountered in **Study 1**. On the one hand, the use of an explicit questionnaire to examine the SoA experience may have been a source of bias given the need to retrospectively assess the experiences and feelings. Explicit judgements of agency are rare in everyday life, as we usually experience a clear feeling (or *flow*) of agency during everyday actions, where no evaluation or judgement is required (Haggard, 2017). Given these limitations, it has been proposed that the SoA should be examined with implicit measures. Implicit measures aim to capture this *'mineness'* feeling without requiring people to explicitly think about agency or control. Nevertheless, I believe that the information provided in Study 1 through the EEG analysis in terms of ERP and time frequency measures allow us to have direct, online, and non-biased information of the underlying mechanisms and processes that are being engaged in the different conditions. Because of the complex nature of the agentic experience, I believe that by employing a correlational approach we can overcome these limitations assessing the neural correlates of self-made vs. externally triggered actions and their relation to other implicit (i.e., RT and PES) and explicit (i.e., questionnaire) measures of SoA.

c) Also in **Study 1**, we introduced A-O temporal delays in a block-wise manner, which may have allowed A-O delay adaptations influencing our results. Our aim when designing Experiment 2 (Study 1) was to address the differences in agency attribution and processing when introducing different A-O delays ranging from 0 (self-attribution), 150 ms (self-attribution threshold) and 400 ms (external attribution) based on previous literature (Farrer et al., 2008, 2013; Hara et al., 2015; Krugwasser et al., 2019; Sato & Yasuda, 2005). Because we wanted to examine both Correct and SE trials at the different delays, we employed a block design to obtain a minimum number of error trials per delay condition (approximately 10% of total

trials are SE at the Eriksen Flanker task, therefore we expected to have 16 trials per block for each condition, 32 in total for No-Delay, 150-Delay and 400-Delay). Nonetheless, to surpass these limitations, future studies may tackle this concern by assessing the influence of temporal aPE in a randomized manner, evaluating the internal/external agency experience at the single-trial level. In fact, we have now conducted a new set of experiments using a modified version of our experimental paradigm in Experiment 2 (Study 1) (Gomez-Andres et al., 2022c), employing a randomized design and introducing both temporal anticipations and delays at ±50/±100/±150/±200/±250 ms, while acquiring EEG data as well as agency attribution reports at the single-trial level.

d) Moreover, as commented above, we recently performed another experiment during which instead of generating an aPE by incorporating incongruent visual feedback, we introduced externally generated errors in the motor/proprioceptive modality by means of a newly developed device allowing passive movements (Gomez-Andres et al., 2022c) (see **Figure 29A**). Based on our previous results (Study 1), we expected to encounter a modulation of agency attribution judgements when introducing EE in terms of external agency attribution judgments, accompanied by a dissociable neural pattern indicating that an aPE has occurred. To preview our results, during the insertion of passive EE, we encountered an N2-P3 complex waveform depicting a frontal and centro-parietal distribution, respectively (see **Figure 29B**). Interestingly, a significant association between the mean amplitude of the N2 component and external attribution judgements was observed (**Figure 29C**). These results point out to an engagement of conflict monitoring mechanisms whenever a passive EE movement was introduced, showing an N2-conflict signal triggering a control adaptation-P3 modulation that specifies a need for changes in attention (Allport, 1987; Botvinick et al., 2001; Miller & Cohen, 2001; Norman & Shallice, 1986). Nevertheless, further exploration and analysis of our data will be performed to ensure a correct interpretation of our results.

e) As previously exposed on section **1.3. Dysfunctions of the self**, several clinical and neuropsychiatric conditions have been related to disturbances in SoA processing. For example, patients with schizophrenia suffering from delusions of control seem to have specific difficulties with forward predictive mechanisms which, as I discussed on previous chapters, are crucial for experiencing a coherent sense of self. Also, patients with AHP, who are unaware of their motor impairments after a stroke involving the aIC, right PMC or the inferior frontal gyrus (IFG) (brain regions related to motor initiation, preparation, and monitoring) (Berti et al., 2005; Fotopoulou et al., 2010; Kortte et al., 2015), also show difficulties in the comparator mechanisms and/or prediction updating. As such, on future studies we would like to examine

the neural signatures elicited in these clinical populations when encountering aPE as well as their subjective experience of agency attribution, which might help elucidate what are the neural and cognitive processes underlying these clinical syndromes.

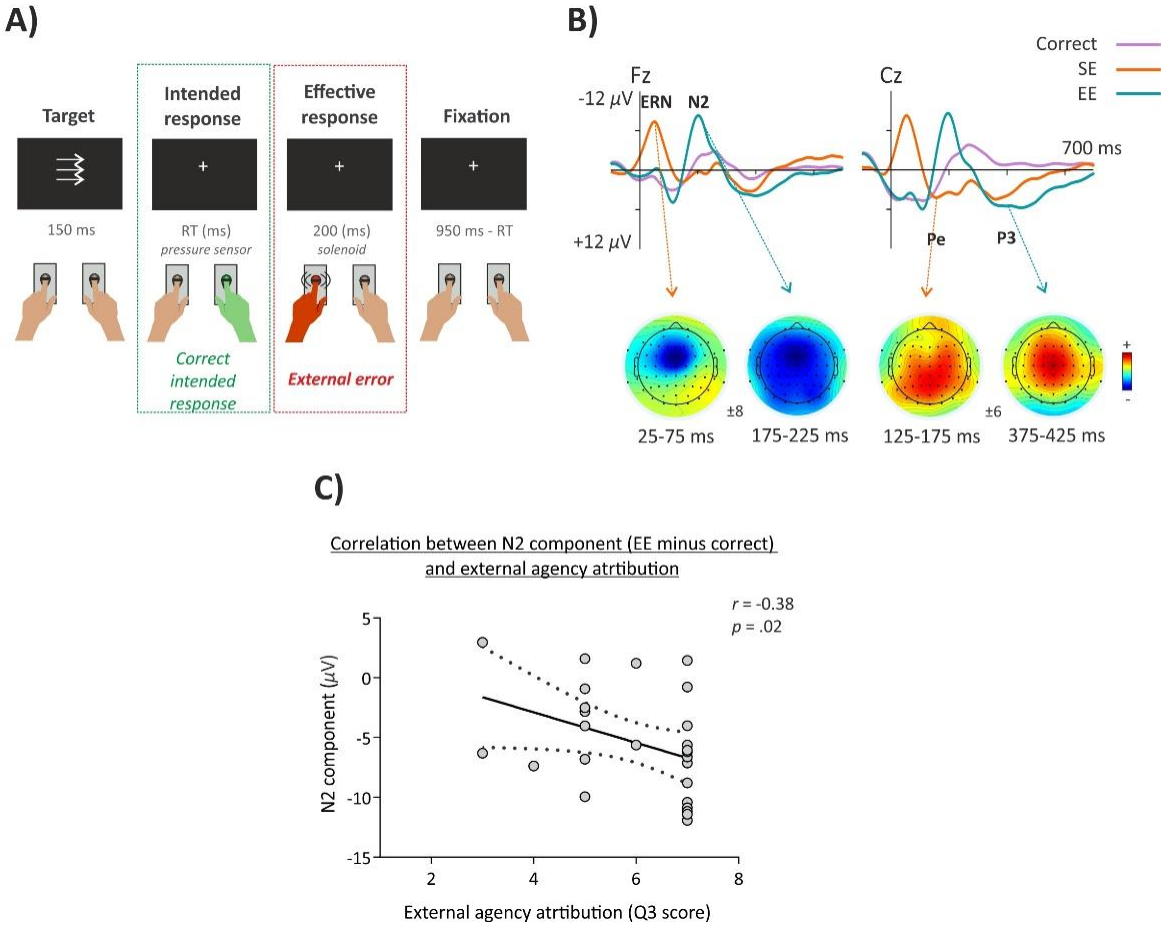


Figure 29. Preliminary results for Gomez-Andres et al. (2022c). A) Experimental task performed during the Error induction block. The target was presented for 150 ms, which could be either compatible or incompatible, both with left and right modalities. A pressure sensor located on each of the response devices tracked the onset of the intended response (reaction time -RT-). During this block, in 10% of compatible trials, an external error (EE) was introduced when a correct intended response was detected (i.e., pressure sensor), eliciting a passive movement of the contralateral finger (i.e., solenoid). After the response (200 ms), a fixation slide appeared for a variable duration of 950 ms minus the RT. **B)** Grand average waveforms for all trial types are depicted at Fz and Cz electrode locations. At frontal locations, the ERN and N2 components can be observed, elicited for SE and EE, respectively. At more centro-parietal sites, the Pe and P3 components were registered, for SE and EE, respectively. Topographical maps for voltage -V- (μV) are also depicted at the time windows of the ERN (25-75 ms), N2 (175-225 ms), Pe (125-175 ms) and P3 (375-425 ms). **C)** Correlation between the N2 component (difference external errors -EE- minus correct) at Fz electrode and the subjective reports for external agency attribution judgements (Q3).

f) Regarding **Study 3**, the limited availability of ESM data might hinder the possible generalization of our findings. During the realization of the ESM procedure, other neuropsychological protocols addressing language production were also applied at the cortical level to functionally locate relevant brain areas for Spanish and Catalan naming. As a result, and due to an increased level of fatigue reported by the patient while inside the OR, a limited number of trials could be performed during the Stroop experimental paradigm (total of the 32 trials, 10 ESM trials). Therefore, on future studies, a larger number of ESM trials should be performed, favouring the generalization of our results. Also, a larger sample of patients undergoing awake brain surgery protocols for tumour resection involving the insular regions must be recruited.

g) Moreover, the postoperative neuropsychological assessment during **Study 3** revealed a cognitive decline in the patient's performance at the articulatory level, probably related to the resection at the left motor opercular region which might have influenced her performance on other cognitive domains dependent on verbal articulation outputs. Therefore, the current protocol should be tested with other patients with brain tumours involving the aIC region to properly assess the benefits of this multimodal protocol in evaluating aIC function, maybe also including patients with brain tumours located on the right hemisphere, to allow the functional comparison between the left and right insular functional relevance.

GENERAL CONCLUSIONS

5. GENERAL CONSLUSIONS

5.1. Conclusions

The SoA is a central aspect of human self-consciousness which lies at the heart of the concept of personal responsibility in our social context. However, alongside its conceptual importance in the philosophical discourse, impairments in volition have also prompted the scientific investigation of the psychological processes and neurobiology of the SoA. In the present thesis, I have described specific neural markers associated to the internal and external agency attribution mechanisms underlying the experience of agency and its disruption, by means of a series of modified experimental paradigms, including the Eriksen flanker task and the Stroop task, using EEG, decoding algorithms and ESM. In Study 1, our EEG and behavioural results revealed that during content aPE (i.e., incongruent hand movements), a parietal-P6 oscillation peaking at about 580 ms related to an enhancement of posterior delta (1-3 Hz) activity (Experiments 1 and 3) was elicited. Importantly, both the P6 component and delta activity were strongly correlated with an increase in external agency attribution judgements. On the other hand, temporal aPE (Experiment 2) were associated to a decrease in internal agency attribution, accompanied by the elicitation of outcome monitoring modulations such as the FCRP and FRN for correct and SE, respectively. These results might be of great usefulness for understanding disorders of agency and action awareness, such as delusions of control in schizophrenia or AHP. Further studies employing this same experimental paradigm may help elucidate the relationship between agency attribution and the error monitoring system on the construction of a non-veridical motor awareness displayed by these groups of patients. Importantly, we emphasize the need to explore not only the anatomical basis but also the neurophysiological markers associated to self/other monitoring, as well as the multifaceted nature of the experience of agency.

Following our results form Study 1, we designed Study 2 with the aim of evaluating the feasibility of employing a SVM classifier to accurately disentangle self-agency errors from other-agency errors from the EEG signal at a single-trial level. Our results in Study 2 resulted in a satisfactorily classification of correct, SE and EE trials at a single-trial basis, giving rise to high performance rates, especially at the ERN and P6 time windows. These results offer a new perspective on how to distinguish self vs. externally generated errors providing new potential implementations on BCI systems. By gaining more insight regarding the integration of internal/predictive and external/sensory information, this new evidence might aid in the development of adaptation and error-correcting mechanisms. If an internal/self-made error is detected, this information can be used to perform online corrections of the ongoing actions. On the other hand, if externally

generated erroneous actions are detected, this information can be used for adapting the system to external/environmental factors influencing goal achievement.

Finally, Study 3 allowed us to investigate the role of the aIC in self-monitoring abilities by employing a modified version of the Stroop paradigm inducing A-O incongruences in terms of incongruent feedback. Our results revealed a cortical site located at the aIC associated to monitoring impairment, resulting in an inability to correctly detect A-O discrepancies when disrupting this cortical site during an ESM protocol. Moreover, the pattern of activation resulting from the fMRI event-related analysis confirmed the involvement of this region, amongst other performance monitoring-related brain regions. I believe our work might help elucidate the functional implication of the aIC in monitoring self-made actions, evidencing the importance of assessing the aIC during awake brain surgery protocols for tumour resections involving this region, promoting the preservation of self-awareness and monitoring abilities.

In sum, this novel research provides new insight regarding the neural workings of the attribution of agency, bridging the gap between the phenomenological experience of 'mineness' and its underlying brain mechanisms as well as new evidence on how agency can be monitored in real-time supporting the development of BCI systems useful for developing new assistive applications. These findings will hopefully pave the path for future basic and clinical research concerning agency and its disruption, which might help elucidate what are the neural and cognitive processes underlying certain neurological and psychopathological conditions offering diagnostic and therapeutical potential.

REFERENCES

6. REFERENCES

Allman, J. M., Watson, K. K., Tetreault, N. A., & Hakeem, A. Y. (2005). Intuition and autism: a possible role for Von Economo neurons. *Trends in cognitive sciences*, *9*(8), 367-373. https://doi.org/10.1016/j.tics.2005.06.008

Allport, A. (1987). Selection for action: Some behavioural and neurophysiological considerations of attention and action. *Perspectives on perception and action*, *15*, 395-419.

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, R. A., & Buneo, C. A. (2002). Intentional maps in posterior parietal cortex. *Annual review of neuroscience*, *25*(1), 189-220. https://doi.org/10.1146/annurev.neuro.25.112701.142922

Andersen, R. A., & Zipser, D. (1988). The role of the posterior parietal cortex in coordinate transformations for visual-motor integration. *Canadian journal of physiology and pharmacology*, *66*(4), 488–501. https://doi.org/10.1139/y88-078

Andreasen, N. C., & Flaum, M. (1991). Schizophrenia: the characteristic symptoms. *Schizophrenia bulletin*, *17*(1), 27–49. https://doi.org/10.1093/schbul/17.1.27

Antusch, S., Aarts, H., & Custers, R. (2019). The role of intentional strength in shaping the sense of agency. *Frontiers in Psychology*, 1124. https://psycnet.apa.org/doi/10.3389/fpsyg.2019.01124

Appelros, P., Karlsson, G. M., & Hennerdal, S. (2007). Anosognosia versus unilateral neglect. Coexistence and their relations to age, stroke severity, lesion site and cognition. *European Journal of Neurology*, *14*(1), 54-59. https://doi.org/10.1111/j.1468-1331.2006.01544.x

Armel, K. C., & Ramachandran, V. S. (2003). Projecting sensations to external objects: evidence from skin conductance response. *Proceedings. Biological sciences*, *270*(1523), 1499–1506. https://doi.org/10.1098/rspb.2003.2364

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in cognitive sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Artusi, X., Niazi, I. K., Lucas, M. F., & Farina, D. (2011). Accuracy of a BCI based on movement-related and error potentials. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2011*, 3688–3691. https://doi.org/10.1109/IEMBS.2011.6090624

Asai, T., & Tanno, Y. (2007). The relationship between the sense of self-agency and schizotypal personality traits. *Journal of motor behaviour*, *39*(3), 162–168. https://doi.org/10.3200/JMBR.39.3.162-168

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *Neuroimage*, *26*(3), 839-851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Assal, F., Schwartz, S., & Vuilleumier, P. (2007). Moving with or without will: functional neural correlates of alien hand syndrome. *Annals of neurology*, *62*(3), 301–306. https://doi.org/10.1002/ana.21173

Aston-Jones, G., Ennis, M., Pieribone, V. A., Nickell, W. T., & Shipley, M. T. (1986). The brain nucleus locus coeruleus: restricted afferent control of a broad efferent network. *Science (New York, N.Y.)*, *234*(4777), 734–737. https://doi.org/10.1126/science.3775363

Babik, I., Cunha, A. B., Ross, S. M., Logan, S. W., Galloway, J. C., & Lobo, M. A. (2019). Means-end problem solving in infancy: Development, emergence of intentionality, and transfer of knowledge. *Developmental psychobiology*, *61*(2), 191–202. https://doi.org/10.1002/dev.21798

Bach, P., Gunter, T. C., Knoblich, G., Prinz, W., & Friederici, A. D. (2009). N400-like negativities in action perception reflect the activation of two components of an action representation. *Social neuroscience*, *4*(3), 212–232. https://doi.org/10.1080/17470910802362546

Bae, G. Y., & Luck, S. J. (2018). Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *38*(2), 409–422. https://doi.org/10.1523/JNEUROSCI.2860-17.2017

Baier, B., & Karnath, H. O. (2008). Tight link between our sense of limb ownership and self-awareness of actions. *Stroke*, *39*(2), 486–488. https://doi.org/10.1161/STROKEAHA.107.495606

Baker, T. E., & Holroyd, C. B. (2011). Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biological psychology,* 87(1), 25-34. https://doi.org/10.1016/j.biopsycho.2011.01.010

Balconi, M., & Vitaloni, S. (2014). Dorsolateral pFC and the representation of the incorrect use of an object: the transcranial direct current stimulation effect on N400 for visual and linguistic stimuli. *Journal of cognitive neuroscience*, *26*(2), 305–318. https://doi.org/10.1162/jocn_a_00500

Ball, T., Demandt, E., Mutschler, I., Neitzel, E., Mehring, C., Vogt, K., Aertsen, A., & Schulze-Bonhage, A. (2008). Movement related activity in the high gamma range of the human EEG. *NeuroImage*, *41*(2), 302–310. https://doi.org/10.1016/j.neuroimage.2008.02.032

Ball, T., Schreiber, A., Feige, B., Wagner, M., Lücking, C. H., & Kristeva-Feige, R. (1999). The role of higher-order motor areas in voluntary movement as revealed by high-resolution EEG and fMRI. *NeuroImage*, *10*(6), 682–694. https://doi.org/10.1006/nimg.1999.0507

Balslev, D., Nielsen, F. A., Lund, T. E., Law, I., & Paulson, O. B. (2006). Similar brain networks for detecting visuo-motor and visuo-proprioceptive synchrony. *NeuroImage*, *31*(1), 308–312. https://doi.org/10.1016/j.neuroimage.2005.11.037

Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, *110*(31), 12846-12851. https://doi.org/10.1073/pnas.1306779110

Bard, K. A., Todd, B. K., Bernier, C., Love, J., & Leavens, D. A. (2006). Self-awareness in human and chimpanzee infants: What is measured and what is meant by the mark and mirror test? *Infancy*, *9*(2), 191-219. https://psycnet.apa.org/doi/10.1207/s15327078in0902_6

Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. *NeuroImage*, *21*(3), 1155–1166. https://doi.org/10.1016/j.neuroimage.2003.11.003

Barraquer-Bordas, L. (1974). Afasias, Apraxias, Agnosias. Toray, Barcelona.

Başar, E., Başar-Eroğlu, C., Karakaş, S., & Schürmann, M. (1999). Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG?. *Neuroscience letters*, *259*(3), 165-168. https://doi.org/10.1016/s0304-3940(98)00934-3

Başar-Eroglu, C., & Başar, E. (1991). A compound P300-40 Hz response of the cat hippocampus. *The International journal of neuroscience*, *60*(3-4), 227–237. https://doi.org/10.3109/00207459109167035

Bäss, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology*, *70*(2), 137–143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Bastin, J., Deman, P., David, O., Gueguen, M., Benis, D., Minotti, L., ... & Jerbi, K. (2016). Direct recordings from human anterior insula reveal its leading role within the error-monitoring network. *Cerebral Cortex*, bhv352. https://doi.org/10.1093/cercor/bhv352

Bastos, A. M., Briggs, F., Alitto, H. J., Mangun, G. R., & Usrey, W. M. (2014). Simultaneous recordings from the primary visual cortex and lateral geniculate nucleus reveal rhythmic interactions and a cortical source for gamma-band oscillations. *Journal of Neuroscience*, *34*(22), 7639-7644. https://doi.org/10.1523/JNEUROSCI.4216-13.2014

Bates, A. T., Kiehl, K. A., Laurens, K. R., & Liddle, P. F. (2002). Error-related negativity and correct response negativity in schizophrenia. *Clinical neurophysiology*, *113*(9), 1454-1463. https://doi.org/10.1016/S1388-2457(02)00154-2

Bauer, M., Stenner, M. P., Friston, K. J., & Dolan, R. J. (2014). Attentional modulation of alpha/beta and gamma oscillations reflect functionally distinct processes. *Journal of Neuroscience*, *34*(48), 16117-16125. https://doi.org/10.1523/JNEUROSCI.3474-13.2014

Bays, P. M., Flanagan, J. R., & Wolpert, D. M. (2006). Attenuation of self-generated tactile sensations is predictive, not postdictive. *PLoS biology*, *4*(2), e28. https://doi.org/10.1371/journal.pbio.0040028

Bays, P. M., Wolpert, D. M., & Flanagan, J. R. (2005). Perception of the consequences of self-action is temporally tuned and event driven. *Current biology: CB*, *15*(12), 1125–1128. https://doi.org/10.1016/j.cub.2005.05.023

Bednark, J. G., & Franz, E. A. (2014). Agency attribution: Event-related potentials and outcome monitoring. *Experimental Brain Research*, *232*(4), 1117-1126. https://doi.org/10.1007/s00221-014-3821-4

Bednark, J. G., Reynolds, J. N., Stafford, T., Redgrave, P., & Franz, E. A. (2013). Creating a movement heuristic for voluntary action: electrophysiological correlates of movement-outcome learning. *Cortex*, *49*(3), 771-780. https://doi.org/10.1016/j.cortex.2011.12.005

Bellebaum, C., & Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience*, *27*(7), 1823-1835. https://doi.org/10.1111/j.1460-9568.2008.06138.x

Berlad, I., & Pratt, H. (1995). P300 in response to the subject's own name. *Electroencephalography and clinical neurophysiology*, *96*(5), 472–474. https://doi.org/10.1016/0168-5597(95)00116-a

Bernat, E. M., Nelson, L. D., & Baskin-Sommers, A. R. (2015). Time-frequency theta and delta measures index separable components of feedback processing in a gambling task. *Psychophysiology*, *52*(5), 626–637. https://doi.org/10.1111/psyp.12390

Bernat, E. M., Nelson, L. D., Steele, V. R., Gehring, W. J., & Patrick, C. J. (2011). Externalizing psychopathology and gain–loss feedback in a simulated gambling task: Dissociable components of brain response revealed by time-frequency analysis. *Journal of abnormal psychology*, *120*(2), 352. https://doi.org/10.1037/a0022124

Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus-noradrenergic system: modulation of behavioural state and state-dependent cognitive processes. *Brain research. Brain research reviews*, *42*(1), 33–84. https://doi.org/10.1016/s0165-0173(03)00143-7

Berti, A., Bottini, G., Gandola, M., Pia, L., Smania, N., Stracciari, A., ... & Paulesu, E. (2005). Shared cortical anatomy for motor awareness and motor control. *Science*, *309*(5733), 488-491. https://doi.org/10.1126/science.1110625

Bhattacharyya, S., Konar, A., & Tibarewala, D. N. (2014). Motor imagery, P300 and error-related EEG-based robot arm movement control for rehabilitation purpose. *Medical & biological engineering & computing*, *52*(12), 1007-1017. https://doi.org/10.1007/s11517-013-1123-9

Biran, I., Giovannetti, T., Buxbaum, L., & Chatterjee, A. (2006). The alien hand syndrome: What makes the alien hand alien? *Cognitive Neuropsychology*, *23*(4), 563-582. https://doi.org/10.1080/02643290500180282

Bisiach, E., Vallar, G., Perani, D., Papagno, C., & Berti, A. (1986). Unawareness of disease following lesions of the right hemisphere: anosognosia for hemiplegia and anosognosia for hemianopia. *Neuropsychologia*, *24*(4), 471-482. https://doi.org/10.1016/0028-3932(86)90092-8

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature neuroscience*, *1*(7), 635–640. https://doi.org/10.1038/2870

Blakemore, S. J., Rees, G., & Frith, C. D. (1998). How do we predict the consequences of our actions? A functional imaging study. *Neuropsychologia*, *36*(6), 521–529. https://doi.org/10.1016/s0028-3932(97)00145-0

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999a). Spatio-temporal prediction modulates the perception of Self-produced stimuli. *Journal of cognitive neuroscience*, *11*(5), 551-559. https://doi.org/10.1162/089892999563607

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1999b). The cerebellum contributes to somatosensory cortical activity during self-produced tactile stimulation. *NeuroImage*, *10*(4), 448–459. https://doi.org/10.1006/nimg.1999.0478

Blakemore, S. J., Smith, J., Steel, R., Johnstone, E. C., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychological medicine*, *30*(5), 1131-1139. https://doi.org/10.1017/S0033291799002676

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in cognitive sciences*, *6*(6), 237–242. https://doi.org/10.1016/s1364-6613(02)01907-1

Blakemore, S. J., & Frith, C. (2003). Self-awareness and action. *Current opinion in neurobiology*, *13*(2), 219-224. https://doi.org/10.1016/s0959-4388(03)00043-6

Blakemore, S. J., & Sirigu, A. (2003). Action prediction in the cerebellum and in the parietal lobe. *Experimental Brain Research*, *153*(2), 239-245. https://doi.org/10.1007/s00221-003-1597-z

Boksem, M. A., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological psychology*, 72(2), 123-132. https://doi.org/10.1016/j.biopsycho.2005.08.007

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, *35*(8), 3478-3484. https://psycnet.apa.org/doi/10.1523/JNEUROSCI.0797-14.2015

Borhani, K., Beck, B., & Haggard, P. (2017). Choosing, Doing, and Controlling: Implicit Sense of Agency Over Somatosensory Events. *Psychological science*, *28*(7), 882–893. https://doi.org/10.1177/0956797617697693

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel 'touch that eyes see. *Nature*, *391*(6669), 756-756. https://doi.org/10.1038/35784

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review*, *108*(3), 624. https://psycnet.apa.org/doi/10.1037/0033-295X.108.3.624

Botvinick M. (2004). Neuroscience. Probing the neural basis of body ownership. *Science (New York, N.Y.)*, *305*(5685), 782–783. https://doi.org/10.1126/science.1101836

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in cognitive sciences*, *8*(12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Botvinick M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, affective & behavioural neuroscience*, *7*(4), 356–366. https://doi.org/10.3758/cabn.7.4.356

Bouret, S., & Sara, S. J. (2004). Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. *The European journal of neuroscience*, *20*(3), 791–802. https://doi.org/10.1111/j.1460-9568.2004.03526.x

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin*, *138*(3), 389–414. https://doi.org/10.1037/a0026450

Braun, N., Emkes, R., Thorne, J. D., & Debener, S. (2016). Embodied neurofeedback with an anthropomorphic robotic hand. *Scientific reports*, *6*, 37696. https://doi.org/10.1038/srep37696

Braun, N., Thorne, J. D., Hildebrandt, H., & Debener, S. (2014). Interplay of agency and ownership: the intentional binding and rubber hand illusion paradigm combined. *PloS one*, *9*(11), e111967. https://doi.org/10.1371/journal.pone.0111967

Brooks, J. X., & Cullen, K. E. (2013). The primate cerebellum selectively encodes unexpected self-motion. *Current biology: CB*, *23*(11), 947–955. https://doi.org/10.1016/j.cub.2013.04.029

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain research*, *1446*, 127-143. https://doi.org/10.1016/j.brainres.2012.01.055

Buchholz, V. N., David, N., Sengelmann, M., & Engel, A. K. (2019). Belief of agency changes dynamics in sensorimotor networks. *Scientific reports*, *9*(1), 1995. https://doi.org/10.1038/s41598-018-37912-w

Buehner, M. J., & Humphreys, G. R. (2009). Causal binding of actions to their effects. *Psychological science*, *20*(10), 1221–1228. https://doi.org/10.1111/j.1467-9280.2009.02435.x

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE transactions on medical imaging*, *18*(1), 32-42. https://doi.org/10.1109/42.750253

Burnod, Y., Baraduc, P., Battaglia-Mayer, A., Guigon, E., Koechlin, E., Ferraina, S., Lacquaniti, F., & Caminiti, R. (1999). Parieto-frontal coding of reaching: an integrated framework. *Experimental brain research*, *129*(3), 325–346. https://doi.org/10.1007/s002210050902

Buzsáki, G., & Watson, B. O. (2012). Brain rhythms and neural syntax: implications for efficient coding of cognitive content and neuropsychiatric disease. *Dialogues in clinical neuroscience*, *14*(4), 345–367. https://doi.org/10.31887/DCNS.2012.14.4/gbuzsaki

Cao, L., Veniero, D., Thut, G., & Gross, J. (2017). Role of the Cerebellum in Adaptation to Delayed Action Effects. *Current biology: CB*, *27*(16), 2442–2451.e3. https://doi.org/10.1016/j.cub.2017.06.074

Carhart-Harris, R. L., Muthukumaraswamy, S., Roseman, L., Kaelen, M., Droog, W., Murphy, K., Tagliazucchi, E., Schenberg, E. E., Nest, T., Orban, C., Leech, R., Williams, L. T., Williams, T. M., Bolstridge, M., Sessa, B., McGonigle, J., Sereno, M. I., Nichols, D., Hellyer, P. J., Hobden, P., … Nutt, D. J. (2016). Neural correlates of the LSD experience revealed by multimodal neuroimaging. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4853–4858. https://doi.org/10.1073/pnas.1518377113

Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical

activity: a combined ERP and fMRI study. *Neuroimage,* 57(4), 1608-1616. http://dx.doi.org/10.1016/j.neuroimage.2011.05.037

Carruthers G. (2012). The case for the comparator model as an explanation of the sense of agency and its breakdowns. *Consciousness and cognition*, *21*(1), 30–58. https://doi.org/10.1016/j.concog.2010.08.005

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science (New York, N.Y.)*, *280*(5364), 747–749. https://doi.org/10.1126/science.280.5364.747

Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. (2016). Coercion Changes the Sense of Agency in the Human Brain. *Current biology: CB*, *26*(5), 585–592. https://doi.org/10.1016/j.cub.2015.12.067

Cavanagh J. F. (2015). Cortical delta activity reflects reward prediction error and related behavioural adjustments, but at different times. *NeuroImage*, *110*, 205–216. https://doi.org/10.1016/j.neuroimage.2015.02.007

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences*, *18*(8), 414-421. https://doi.org/10.1016/j.tics.2014.04.012

Cavanagh, J. F., Cohen, M. X., & Allen, J. J. (2009). Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. *Journal of Neuroscience*, *29*(1), 98-105. https://doi.org/10.1523/jneurosci.4137-08.2009

Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. (2010). Frontal theta links prediction errors to behavioural adaptation in reinforcement learning. *Neuroimage*, *49*(4), 3198-3209. https://doi.org/10.1016/j.neuroimage.2009.11.080

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford Paperbacks.

Chambon, V., Moore, J. W., & Haggard, P. (2015). TMS stimulation over the inferior parietal cortex disrupts prospective sense of agency. *Brain Structure and Function*, *220*(6), 3627-3639. https://doi.org/10.1007/s00429-014-0878-6

Chambon, V., Wenke, D., Fleming, S. M., Prinz, W., & Haggard, P. (2013). An online neural substrate for sense of agency. *Cerebral Cortex*, *23*(5), 1031-1037. https://doi.org/10.1093/cercor/bhs059

Chaminade, T., & Decety, J. (2002). Leader or follower? Involvement of the inferior parietal lobule in agency. *Neuroreport*, *13*(15), 1975–1978. https://doi.org/10.1097/00001756-200210280-00029

Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage*, *73*, 80-94. https://doi.org/10.1016/j.neuroimage.2013.01.054

Chavarriaga, R., Sobolewski, A., & Millán, J. D. R. (2014). Errare machinale est: the use of error-related potentials in brain-machine interfaces. *Frontiers in neuroscience*, 208. https://doi.org/10.3389/fnins.2014.00208

Cheyne, D., Bells, S., Ferrari, P., Gaetz, W., & Bostan, A. C. (2008). Self-paced movements induce high-frequency gamma oscillations in primary motor cortex. *Neuroimage*, *42*(1), 332-342. https://doi.org/10.1016/j.neuroimage.2008.04.178

Choi, H., & Scholl, B. J. (2006). Perceiving causality after the fact: postdiction in the temporal dynamics of causal perception. *Perception*, *35*(3), 385–399. https://doi.org/10.1068/p5462

Christensen, M. S., Lundbye-Jensen, J., Geertsen, S. S., Petersen, T. H., Paulson, O. B., & Nielsen, J. B. (2007). Premotor cortex modulates somatosensory cortex during voluntary movements without proprioceptive feedback. *Nature neuroscience*, *10*(4), 417–419. https://doi.org/10.1038/nn1873

Christoff, K., Cosmelli, D., Legrand, D., & Thompson, E. (2011). Specifying the self for cognitive neuroscience. *Trends in cognitive sciences*, *15*(3), 104–112. https://doi.org/10.1016/j.tics.2011.01.001

Clark A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioural and brain sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Cocchini, G., Beschin, N., Fotopoulou, A., & Della Sala, S. (2010). Explicit and implicit anosognosia or upper limb motor impairment. *Neuropsychologia*, *48*(5), 1489-1494. https://doi.org/10.1016/j.neuropsychologia.2010.01.019

Cohen, M. X. (2011). Error-related medial frontal theta activity predicts cingulate-related structural connectivity. *Neuroimage*, *55*(3), 1373-1383. https://doi.org/10.1016/j.neuroimage.2010.12.072

Cohen, M. X. (2014). A neural microcircuit for cognitive conflict detection and signaling. *Trends in neurosciences*, *37*(9), 480-490. https://doi.org/10.1016/j.tins.2014.06.004

Cohen, M. X., & Ridderinkhof, K. R. (2013). EEG source reconstruction reveals frontal-parietal dynamics of spatial conflict processing. *PloS one*, *8*(2), e57293. https://doi.org/10.1371/journal.pone.0057293

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual review of neuroscience*, *22*, 319–349. https://doi.org/10.1146/annurev.neuro.22.1.319

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology*, *206*(4), 515-530. https://doi.org/10.1007/s00213-009-1561-0

Corlett, P. R., Honey, G. D., & Fletcher, P. C. (2007). From prediction error to psychosis: ketamine as a pharmacological model of delusions. *Journal of psychopharmacology (Oxford, England)*, *21*(3), 238–252. https://doi.org/10.1177/0269881107077716

Corlett, P. R., Honey, G. D., Krystal, J. H., & Fletcher, P. C. (2011). Glutamatergic model psychoses: prediction error, learning, and inference. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, *36*(1), 294–315. https://doi.org/10.1038/npp.2010.163

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R., Shanks, D. R., Robbins, T. W., ... & Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain*, *130*(9), 2387-2400. https://doi.org/10.1093/brain/awm173

Coull J. T. (2004). fMRI studies of temporal attention: allocating attention within, or towards, time. *Brain research. Cognitive brain research*, *21*(2), 216–226. https://doi.org/10.1016/j.cogbrainres.2004.02.011

Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. *Current opinion in neurobiology*, *13*(4), 500-505. https://doi.org/10.1016/S0959-4388(03)00090-4

Craig, A. D., & Craig, A. D. (2009). How do you feel--now? The anterior insula and human awareness. *Nature reviews neuroscience*, *10*(1). https://doi.org/10.1038/nrn2555

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature reviews. Neuroscience*, *9*(8), 587–600. https://doi.org/10.1038/nrn2457

Crespo-Facorro, B., Kim, J. J., Andreasen, N. C., O'Leary, D. S., Bockholt, H. J., & Magnotta, V. (2000). Insular cortex abnormalities in schizophrenia: a structural magnetic resonance imaging study of first-episode patients. *Schizophrenia research*, *46*(1), 35-43. https://doi.org/10.1016/S0920-9964(00)00028-1

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature neuroscience*, *7*(2), 189-195. https://doi.org/10.1038/nn1176

Cui, H. (2014). From intention to action: hierarchical sensorimotor transformation in the posterior parietal cortex. *eNeuro*, *1*(1). https://dx.doi.org/10.1523%2FENEURO.0017-14.2014

Dal Seno, B., Matteucci, M., & Mainardi, L. (2010). Online detection of P300 and error potentials in a BCI speller. *Computational intelligence and neuroscience*, *2010*. https://dx.doi.org/10.1155%2F2010%2F307254

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Houghton Mifflin Harcourt.

Damen, T. G., Van Baaren, R. B., Brass, M., Aarts, H., & Dijksterhuis, A. (2015). Put your plan into action: The influence of action plans on agency and responsibility. *Journal of personality and social psychology*, *108*(6), 850. https://psycnet.apa.org/doi/10.1037/pspa0000024

Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. *Frontiers in psychology*, *2*, 233. https://doi.org/10.3389/fpsyg.2011.00233

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., & Jeannerod, M. (1997). Looking for the agent: an investigation into consciousness of action and Self-consciousness in schizophrenic patients. *Cognition*, *65*(1), 71-86. https://doi.org/10.1016/s0010-0277(97)00039-5

David N. (2012). New frontiers in the neuroscience of the sense of agency. *Frontiers in human neuroscience*, *6*, 161. https://doi.org/10.3389/fnhum.2012.00161

David, N., Cohen, M. X., Newen, A., Bewernick, B. H., Shah, N. J., Fink, G. R., & Vogeley, K. (2007). The extrastriate cortex distinguishes between the consequences of one's own and others' behaviour. *Neuroimage*, *36*(3), 1004-1014. https://doi.org/10.1016/j.neuroimage.2007.03.030

David, N., Newen, A., & Vogeley, K. (2008). The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and cognition*, *17*(2), 523–534. https://doi.org/10.1016/j.concog.2008.03.004

David, N., Skoruppa, S., Gulberti, A., Schultz, J., & Engel, A. K. (2016). The sense of Agency is more sensitive to manipulations of outcome than movement-related feedback irrespective of sensory modality. *PloS one*, *11*(8), e0161156. https://doi.org/10.1371/journal.pone.0161156

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113. https://psycnet.apa.org/doi/10.1037/0022-3514.44.1.113

de Bruijn, E. R., de Lange, F. P., von Cramon, D. Y., & Ullsperger, M. (2009). When errors are rewarding. *Journal of Neuroscience*, *29*(39), 12183-12186. https://doi.org/10.1523/JNEUROSCI.1751-09.2009

De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. *Cortex*, *14*(1), 41-49. http://dx.doi.org/10.3233/NRE-172244

De Vignemont, F., & Fourneret, P. (2004). The sense of agency: A philosophical and empirical review of the "Who" system. *Consciousness and Cognition*, *13*(1), 1-19. https://doi.org/10.1016/S1053-8100(03)00022-9

Decety, J., Chaminade, T., Grèzes, J., & Meltzoff, A. N. (2002). A PET exploration of the neural mechanisms involved in reciprocal imitation. *NeuroImage*, *15*(1), 265–272. https://doi.org/10.1006/nimg.2001.0938

Deen, B., Pitskel, N. B., & Pelphrey, K. A. (2011). Three systems of insular functional connectivity identified with cluster analysis. *Cerebral cortex*, *21*(7), 1498-1506. https://doi.org/10.1093/cercor/bhq186

Deiber, M. P., Passingham, R. E., Colebatch, J. G., Friston, K. J., Nixon, P. D., & Frackowiak, R. S. (1991). Cortical areas and the selection of movement: a study with positron emission tomography. *Experimental brain research*, *84*(2), 393–402. https://doi.org/10.1007/BF00231461

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*, *132*(9), 2531-2540. https://doi.org/10.1093/brain/awp111

Della-Maggiore, V., Malfait, N., Ostry, D. J., & Paus, T. (2004). Stimulation of the posterior parietal cortex interferes with arm trajectory adjustments during the learning of new dynamics. *Journal of Neuroscience*, *24*(44), 9971-9976. https://doi.org/10.1523/JNEUROSCI.2833-04.2004

Demiralp, T., Ademoglu, A., Istefanopulos, Y., Başar-Eroglu, C., & Başar, E. (2001). Wavelet analysis of oddball P300. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology*, *39*(2-3), 221–227. https://doi.org/10.1016/s0167-8760(00)00143-4

Demiralp, T., Ademoglu, A., Schürmann, M., Başar-Eroglu, C., & Başar, E. (1999). Detection of P300 waves in single trials by the wavelet transform (WT). *Brain and language*, *66*(1), 108–128. https://doi.org/10.1006/brln.1998.2027

den Ouden, H. E., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in psychology*, *3*, 548. https://doi.org/10.3389/fpsyg.2012.00548

Derrfuss, J., Danielmeier, C., Klein, T. A., Fischer, A. G., & Ullsperger, M. (2021). Unbiased post-error slowing in interference tasks: A confound and a simple solution. *Behaviour Research Methods*, 1-12. https://doi.org/10.3758/s13428-021-01673-8

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in cognitive sciences*, *4*(11), 423-431. https://doi.org/10.1016/S1364-6613(00)01537-0

Desmurget, M., & Sirigu, A. (2009). A parietal-premotor network for movement intention and motor awareness. *Trends in cognitive sciences*, *13*(10), 411-419. https://doi.org/10.1016/j.tics.2009.08.001

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science (New York, N.Y.)*, *324*(5928), 811–813. https://doi.org/10.1126/science.1169896

Diedrichsen, J., Hashambhoy, Y., Rane, T., & Shadmehr, R. (2005). Neural correlates of reach errors. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *25*(43), 9919–9931. https://doi.org/10.1523/JNEUROSCI.1874-05.2005

Diener, C., Kuehner, C., Brusniak, W., Ubl, B., Wessa, M., & Flor, H. (2012). A meta-analysis of neurofunctional imaging studies of emotion and cognition in major depression. *Neuroimage*, *61*(3), 677-685. https://doi.org/10.1016/j.neuroimage.2012.04.005

Donchin, E. (1981). Surprise!... surprise? *Psychophysiology*, *18*(5), 493-513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Donchin, E. (1986). Cognitive psychophysiology and human information processing. *Psychophysiology: Systems, processes and applications*, 244-267.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioural and brain sciences*, *11*(3), 357-374. https://doi.org/10.1017/S0140525X00058027

Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. *Event-related brain potentials in man*, *349*, 411.

Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in cognitive sciences*, *12*(3), 99-105. https://doi-org.sire.ub.edu/10.1016/j.tics.2008.01.001

Dreisbach, G., & Fischer, R. (2015). Conflicts as aversive signals for control adaptation. *Current Directions in Psychological Science*, *24*(4), 255-260. https://psycnet.apa.org/doi/10.1177/0963721415569569

Duffau H. (2005). Lessons from brain mapping in surgery for low-grade glioma: insights into associations between tumour and brain plasticity. *The Lancet. Neurology*, *4*(8), 476–486. https://doi.org/10.1016/S1474-4422(05)70140-X

Duffau, H., Gatignol, P., Mandonnet, E., Capelle, L., & Taillandier, L. (2008). Intraoperative subcortical stimulation mapping of language pathways in a consecutive series of 115 patients with Grade II glioma in the left dominant hemisphere. *Journal of neurosurgery*, *109*(3), 461-471. https://doi.org/10.3171/jns/2008/109/9/0461

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1998). Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *Journal of neurophysiology*, *79*(1), 126–136. https://doi.org/10.1152/jn.1998.79.1.126

Dummer, T., Picot-Annand, A., Neal, T., & Moore, C. (2009). Movement and the rubber hand illusion. *Perception*, *38*(2), 271–280. https://doi.org/10.1068/p5921

Ebert, J. P., & Wegner, D. M. (2010). Time warp: authorship shapes the perceived timing of actions and events. *Consciousness and cognition*, *19*(1), 481–489. https://doi.org/10.1016/j.concog.2009.10.002

Ehrsson, H. H., Holmes, N. P., & Passingham, R. E. (2005). Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *25*(45), 10564–10573. https://doi.org/10.1523/JNEUROSCI.0800-05.2005

Ehrsson, H. H., Spence, C., & Passingham, R. E. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science (New York, N.Y.)*, *305*(5685), 875–877. https://doi.org/10.1126/science.1097011

Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J., & Passingham, R. E. (2007). Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(23), 9828–9833. https://doi.org/10.1073/pnas.0610011104

Eimer, M. (1998). The lateralized readiness potential as an on-line measure of central response activation processes. *Behaviour Research Methods, Instruments, & Computers*, *30*(1), 146-156. https://doi.org/10.3758/BF03209424

Ekman, P. (1976). Pictures of facial affect. *Consulting psychologists press*.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of experimental psychology. Human perception and performance*, *27*(1), 229–240. https://doi.org/10.1037//0096-1523.27.1.229

Engbert, K., & Wohlschläger, A. (2007). Intentions and expectations in temporal binding. *Consciousness and cognition*, *16*(2), 255–264. https://doi.org/10.1016/j.concog.2006.09.010

Erez, Y., Assem, M., Coelho, P., Romero-Garcia, R., Owen, M., McDonald, A., ... & Santarius, T. (2021). Intraoperative mapping of executive function using electrocorticography for patients with low-grade gliomas. *Acta Neurochirurgica*, *163*(5), 1299-1309. https://doi.org/10.1007/s00701-020-04646-6

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & psychophysics*, *25*(4), 249-263.

Eshel, N., & Roiser, J. P. (2010). Reward and punishment processing in depression. *Biological psychiatry*, *68*(2), 118-124. https://doi.org/10.1016/j.biopsych.2010.01.027

Falkenstein, M. (1990). Effects of errors in choice reaction tasks on the ERP under focused and divided attention. *Psychophysiological brain research*.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and clinical neurophysiology*, *78*(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological psychology*, *51*(2-3), 87-107. https://doi.org/10.1016/s0301-0511(99)00031-9

Farrer, C., & Frith, C. D. (2002). Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *Neuroimage*, *15*(3), 596-603. https://doi.org/10.1006/nimg.2001.1009

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: a positron emission tomography study. *NeuroImage*, *18*(2), 324–333. https://doi.org/10.1016/s1053-8119(02)00041-1

Farrer, C., & Franck, N. (2007). Self-monitoring in schizophrenia. *Current Psychiatry Reviews*, *3*(4), 243-251.

Farrer, C., Bouchereau, M., Jeannerod, M., & Franck, N. (2008). Effect of distorted visual feedback on the sense of Agency. *Behavioural neurology*, *19*(1, 2), 53-57. https://doi.org/10.1155/2008/425267

Farrer, C., Frey, S. H., Van Horn, J. D., Tunik, E., Turk, D., Inati, S., & Grafton, S. T. (2008). The angular gyrus computes action awareness representations. *Cerebral cortex*, *18*(2), 254-261. https://doi.org/10.1093/cercor/bhm050

Farrer, C., Valentin, G., & Hupé, J. M. (2013). The time windows of the sense of agency. *Consciousness and cognition*, *22*(4), 1431-1441. https://doi.org/10.1016/j.concog.2013.09.010

Feinberg I. (1978). Efference copy and corollary discharge: implications for thinking and its disorders. *Schizophrenia bulletin*, *4*(4), 636–640. https://doi.org/10.1093/schbul/4.4.636

Fernández-Coello, A., Havas, V., Juncadella, M., Sierpowska, J., Rodríguez-Fornells, A., & Gabarrós, A. (2016). Age of language acquisition and cortical language organization in multilingual patients undergoing awake brain mapping. *Journal of neurosurgery*, *126*(6), 1912-1923. https://doi.org/10.3171/2016.5.JNS152791

Ferrez, P. W., & Millán, J. D. R. (2008). Error-related EEG potentials generated during simulated brain–computer interaction. *IEEE transactions on biomedical engineering*, *55*(3), 923-929. https://doi.org/10.1109/tbme.2007.908083

Fink, G. R., Marshall, J. C., Halligan, P. W., Frith, C. D., Driver, J., Frackowiak, R. S., & Dolan, R. J. (1999). The neural consequences of conflict between intention and the senses. *Brain: a journal of neurology*, *122 (Pt 3)*, 497–512. https://doi.org/10.1093/brain/122.3.497

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*(5614), 1898-1902. https://doi.org/10.1126/science.1077349

Fisher C. M. (2000). Alien hand phenomena: a review with the addition of six personal cases. *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques*, *27*(3), 192–203. https://doi.org/10.1017/s0317167100000834

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48-58. https://doi.org/10.1038/nrn2536

Flor, H., Diers, M., & Andoh, J. (2013). The neural basis of phantom limb pain. *Trends in cognitive sciences*, *17*(7), 307-308. https://doi.org/10.1016/j.tics.2013.04.007

Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., & Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (area F4). *Journal of neurophysiology*, *76*(1), 141–157. https://doi.org/10.1152/jn.1996.76.1.141

Foote, S. L., Bloom, F. E., & Aston-Jones, G. (1983). Nucleus locus ceruleus: new evidence of anatomical and physiological specificity. *Physiological reviews*, *63*(3), 844–914. https://doi.org/10.1152/physrev.1983.63.3.844

Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. *Human brain mapping*, 32(12), 2207-2216. https://doi.org/10.1002%2Fhbm.21182

Foti, D., Kotov, R., Bromet, E., & Hajcak, G. (2012). Beyond the broken error-related negativity: functional and diagnostic correlates of error processing in psychosis. *Biological psychiatry*, *71*(10), 864–872. https://doi.org/10.1016/j.biopsych.2012.01.007

Fotopoulou A. K. (2012). Illusions and delusions in anosognosia for hemiplegia: from motor predictions to prior beliefs. *Brain: a journal of neurology*, *135*(Pt 5), 1344–1346. https://doi.org/10.1093/brain/aws094

Fotopoulou, A. (2014). Time to get rid of the 'Modular'in neuropsychology: A unified theory of anosognosia as aberrant predictive coding. *Journal of neuropsychology*, *8*(1), 1-19. https://doi.org/10.1111/jnp.12010

Fotopoulou, A., Pernigo, S., Maeda, R., Rudd, A., & Kopelman, M. A. (2010). Implicit awareness in anosognosia for hemiplegia: unconscious interference without conscious re-representation. *Brain*, *133*(12), 3564-3577. https://doi.org/10.1093/brain/awq233

Fotopoulou, A., Rudd, A., Holmes, P., & Kopelman, M. (2009). Self-observation reinstates motor awareness in anosognosia for hemiplegia. *Neuropsychologia*, *47*(5), 1256-1260. https://doi.org/10.1016/j.neuropsychologia.2009.01.018

Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, *36*(11), 1133–1140. https://doi.org/10.1016/s0028-3932(98)00006-2

Fourneret, P., Franck, N., Slachevsky, A., & Jeannerod, M. (2001). Self-monitoring in schizophrenia revisited. *Neuroreport*, *12*(6), 1203–1208. https://doi.org/10.1097/00001756-200105080-00030

Fourneret, P., Paillard, J., Lamarre, Y., Cole, J., & Jeannerod, M. (2002). Lack of conscious recognition of one's own actions in a haptically deafferented patient. *Neuroreport,* 13, 541–547. https://doi.org/10.1097/00001756-200203250-00036

Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Daléry, J., d'Amato, T., & Jeannerod, M. (2001). Defective recognition of one's own actions in patients with schizophrenia. *The American journal of psychiatry*, *158*(3), 454–459. https://doi.org/10.1176/appi.ajp.158.3.454

Fried, I., Katz, A., McCarthy, G., Sass, K. J., Williamson, P., Spencer, S. S., & Spencer, D. D. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *11*(11), 3656–3666. https://doi.org/10.1523/JNEUROSCI.11-11-03656.1991

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological reviews*, *91*(4), 1357-1392. https://doi.org/10.1152/physrev.00006.2011

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and biobehavioural reviews*, *25*(4), 355–373. https://doi.org/10.1016/s0149-7634(01)00019-7

Friston K. (2010). The free-energy principle: a unified brain theory? *Nature reviews. Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston K. (2012). Prediction, perception and agency. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology*, *83*(2), 248–252. https://doi.org/10.1016/j.ijpsycho.2011.11.014

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Frith, C. (2005). The neural basis of hallucinations and delusions. *Comptes rendus biologies*, *328*(2), 169-175. https://doi.org/10.1016/j.crvi.2004.10.012

Frith, C. D. (1992/2014). *The cognitive neuropsychology of schizophrenia*. Psychology press.

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *355*(1404), 1771-1788. http://dx.doi.org/10.1098/rstb.2000.0734

Fukushima, H., Goto, Y., Maeda, T., Kato, M., & Umeda, S. (2013). Neural substrates for judgment of self-agency in ambiguous situations. *PloS one*, *8*(8), e72267. https://doi.org/10.1371/journal.pone.0072267

Gaetz, W., Liu, C., Zhu, H., Bloy, L., & Roberts, T. P. (2013). Evidence for a motor gamma-band network governing response interference. *NeuroImage*, *74*, 245–253. https://doi.org/10.1016/j.neuroimage.2013.02.013

Gallagher I., I (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, *4*(1), 14–21. https://doi.org/10.1016/s1364-6613(99)01417-5

Gallagher, S. (2004). Neurocognitive models of schizophrenia: a neurophenomenological critique. *Psychopathology*, *37*(1), 8-19. https://doi.org/10.1159/000077014

Gallagher, S. (2007). The natural philosophy of agency. *Philosophy Compass*, *2*(2), 347-357. https://doi.org/10.1111/j.1747-9991.2007.00067.x

Gallop G. G., Jr (1970). Chimpanzees: self-recognition. *Science (New York, N.Y.)*, *167*(3914), 86–87. https://doi.org/10.1126/science.167.3914.86

Gallup, G. G., Jr., Anderson, J. R., & Shillito, D. J. (2002). The mirror test. In M. Bekoff, C. Allen, & G. M. Burghardt (Eds.), *The cognitive animal: Empirical and theoretical perspectives on animal cognition* (pp. 325–333). MIT Press.

Gao, J. H., Parsons, L. M., Bower, J. M., Xiong, J., Li, J., & Fox, P. T. (1996). Cerebellum implicated in sensory acquisition and discrimination rather than motor control. *Science (New York, N.Y.)*, *272*(5261), 545–547. https://doi.org/10.1126/science.272.5261.545

Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *21*(23), 9430–9437. https://doi.org/10.1523/JNEUROSCI.21-23-09430.2001

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science (New York, N.Y.)*, *295*(5563), 2279–2282. https://doi.org/10.1126/science.1066893

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological science*, *4*(6), 385-390. https://psycnet.apa.org/doi/10.1111/j.1467-9280.1993.tb00586.x

Gellman, R., Gibson, A. R., & Houk, J. C. (1985). Inferior olivary neurons in the awake cat: detection of contact and passive body displacement. *Journal of Neurophysiology*, *54*(1), 40-60. https://doi.org/10.1152/jn.1985.54.1.40

Gentsch, A., & Schütz-Bosbach, S. (2011). I did it: unconscious expectation of sensory consequences modulates the experience of self-agency and its functional signature. *Journal of cognitive neuroscience*, *23*(12), 3817-3828. https://psycnet.apa.org/doi/10.1162/jocn_a_00012

Gentsch, A., Ullsperger, P., & Ullsperger, M. (2009). Dissociable medial frontal negativities from a common monitoring system for Self-and externally caused failure of goal achievement. *Neuroimage*, *47*(4), 2023-2030. https://doi.org/10.1016/j.neuroimage.2009.05.064

Georgieff, N., & Jeannerod, M. (1998). Beyond consciousness of external reality: a "who" system for consciousness of action and self-consciousness. *Consciousness and cognition*, *7*(3), 465–477. https://doi.org/10.1006/ccog.1998.0367

Gergely, G., & Watson, J. S. (1999). Early socio-emotional development: Contingency perception and the social-biofeedback model. *Early social cognition: Understanding others in the first months of life*, *60*, 101-136.

Gialanella, B., Monguzzi, V., Santoro, R., & Rocchi, S. (2005). Functional recovery after hemiplegia in patients with neglect: the rehabilitative role of anosognosia. *Stroke*, *36*(12), 2687-2690. https://doi.org/10.1161/01.STR.0000189627.27562.c0

Giovannetti, T., Buxbaum, L. J., Biran, I., & Chatterjee, A. (2005). Reduced endogenous control in alien hand syndrome: evidence from naturalistic action. *Neuropsychologia*, *43*(1), 75–88. https://doi.org/10.1016/j.neuropsychologia.2004.06.017

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision

making. *Cerebral cortex (New York, N.Y. : 1991)*, *19*(2), 483–495. https://doi.org/10.1093/cercor/bhn098

Glazer, J. E., Kelley, N. J., Pornpattananangkul, N., Mittal, V. A., & Nusslock, R. (2018). Beyond the FRN: Broadening the time-course of EEG and ERP components implicated in reward processing. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology*, *132*(Pt B), 184–202. https://doi.org/10.1016/j.ijpsycho.2018.02.002

Gogolla, N. (2017). The insular cortex. *Current Biology*, *27*(12), R580-R586. https://doi.org/10.1016/j.cub.2017.05.010

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, *30*, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Goldberg, G., Mayer, N. H., & Toglia, J. U. (1981). Medial frontal cortex infarction and the alien hand sign. *Archives of neurology*, *38*(11), 683–686. https://doi.org/10.1001/archneur.1981.00510110043004

Goldstein, K. (1908). Zur Lehre der motorischen apraxie. *Journal fur Psychologie und Neurologie*, *11*, 169-87.

Gomez-Andres, A., Cunillera, T., Rico, I., Naval-Baudin, P., Camins, A., Fernandez-Coello, A., … Rodriguez-Fornells, A. (in press). The role of the anterior insular cortex in self-monitoring: A novel study protocol with electrical stimulation mapping and functional magnetic resonance imaging. *Cortex.*

Gomez-Andres, A., Cunillera, T., Cucurell, D., & Rodriguez-Fornells, A. (2022a). "The complex nature of agency attribution": Neurophysiological signatures associated to monitoring self vs. external erroneous actions. Submitted for publication.

Gomez-Andres, A., Cerda-Company, X., Cucurell, D., Cunillera, T., & Rodriguez-Fornells, A. (2022b). Decoding Agency Attribution using single trial error-related brain potentials. Submitted for publication.

Gomez-Andres, A., Ballestero, M., Cucurell, D., Cunillera, T., & Rodriguez-Fornells, A. (2022c). Modulating the Sense of Agency: Motor incongruence and intentionality. In preparation

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders*. Philadelphia: Lea & Febiger.

Goodkind, M. S., Sturm, V. E., Ascher, E. A., Shdo, S. M., Miller, B. L., Rankin, K. P., & Levenson, R. W. (2015). Emotion recognition in frontotemporal dementia and Alzheimer's disease: A new film-based assessment. *Emotion*, *15*(4), 416. https://dx.doi.org/10.1037%2Fa0039261

Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., Ortega, B. N., Zaiko, Y. V., Roach, E. L., Korgaonkar, M. S., Grieve, S. M., Galatzer-Levy, I., Fox, P. T., & Etkin, A.

(2015). Identification of a common neurobiological substrate for mental illness. *JAMA psychiatry*, *72*(4), 305–315. https://doi.org/10.1001/jamapsychiatry.2014.2206

Grant, D. A., & Berg, E. (1948). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of experimental psychology*, *38*(4), 404. https://psycnet.apa.org/doi/10.1037/h0059831

Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *Journal of experimental social psychology*, *40*(2), 216-224. https://psycnet.apa.org/doi/10.1016/S0022-1031(03)00092-1

Graziano, M. S., & Gross, C. G. (1993). A bimodal map of space: somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Experimental brain research*, *97*(1), 96–109. https://doi.org/10.1007/BF00228820

Graziano, M. S., & Gross, C. G. (1998). Spatial maps for the control of movement. *Current opinion in neurobiology*, *8*(2), 195–201. https://doi.org/10.1016/s0959-4388(98)80140-2

Graziano, M. S., Cooke, D. F., & Taylor, C. S. (2000). Coding the location of the arm by sight. *Science (New York, N.Y.)*, *290*(5497), 1782–1786. https://doi.org/10.1126/science.290.5497.1782

Grent-'t-Jong, T., Gross, J., Goense, J., Wibral, M., Gajwani, R., Gumley, A. I., Lawrie, S. M., Schwannauer, M., Schultze-Lutter, F., Navarro Schröder, T., Koethe, D., Leweke, F. M., Singer, W., & Uhlhaas, P. J. (2018). Resting-state gamma-band power alterations in schizophrenia reveal E/I-balance abnormalities across illness-stages. *eLife*, *7*, e37799. https://doi.org/10.7554/eLife.37799

Greville, W. J., & Buehner, M. J. (2010). Temporal predictability facilitates causal learning. *Journal of experimental psychology. General*, *139*(4), 756–771. https://doi.org/10.1037/a0020976

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*(12), 1711-1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Gulbinaite, R., van Rijn, H., & Cohen, M. X. (2014). Fronto-parietal network oscillations reveal relationship between working memory capacity and cognitive control. *Frontiers in human neuroscience*, *8*, 761. https://doi.org/10.3389/fnhum.2014.00761

Haans, A., Ijsselsteijn, W. A., & de Kort, Y. A. (2008). The effect of similarities in skin texture and hand shape on perceived ownership of a fake limb. *Body image*, *5*(4), 389–394. https://doi.org/10.1016/j.bodyim.2008.04.003

Haggard P. (2005). Conscious intention and motor cognition. *Trends in cognitive sciences*, *9*(6), 290–295. https://doi.org/10.1016/j.tics.2005.04.012

Haggard P. (2008). Human volition: towards a neuroscience of will. *Nature reviews. Neuroscience*, *9*(12), 934–946. https://doi.org/10.1038/nrn2497

Haggard P. (2017). Sense of agency in the human brain. *Nature reviews. Neuroscience*, *18*(4), 196–207. https://doi.org/10.1038/nrn.2017.14

Haggard, P., & Clark, S. (2003). Intentional action: conscious experience and neural prediction. *Consciousness and cognition*, *12*(4), 695–707. https://doi.org/10.1016/s1053-8100(03)00052-7

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental brain research*, *126*(1), 128–133. https://doi.org/10.1007/s002210050722

Haggard, P., & Magno, E. (1999). Localising awareness of action with transcranial magnetic stimulation. *Experimental brain research*, *127*(1), 102–107. https://doi.org/10.1007/s002210050778

Haggard, P., & Tsakiris, M. (2009). The experience of agency: Feelings, judgments, and responsibility. *Current Directions in Psychological Science*, *18*(4), 242-246. https://psycnet.apa.org/doi/10.1111/j.1467-8721.2009.01644.x

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature neuroscience*, *5*(4), 382-385. https://doi.org/10.1038/nn827

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological psychology*, *71*(2), 148-154. https://doi.org/10.1016/j.biopsycho.2005.04.001

Ham, T., Leff, A., de Boissezon, X., Joffe, A., & Sharp, D. J. (2013). Cognitive control and the salience network: an investigation of error processing and effective connectivity. *Journal of Neuroscience*, *33*(16), 7091-7098. https://dx.doi.org/10.1523%2FJNEUROSCI.4692-12.2013

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & biological engineering & computing*, *32*(1), 35-42. https://doi.org/10.1007/bf02512476

Hamburger, H. L. (1991). Global field power measurement versus classical method in the determination of the latency of evoked potential components. *Brain topography*, *3*(3), 391-396. https://doi.org/10.1007/BF01129642

Hamilton, M. (1959). The assessment of anxiety states by rating. The British journal of medical psychology, 32(1), 50–55. https://doi.org/10.1111/j.2044-8341.1959.tb00467.x

Hamilton, M. (1960). A rating scale for depression. *Journal of neurology, neurosurgery, and psychiatry*, 23(1), 56–62. https://doi.org/10.1136/jnnp.23.1.56

Hanley, M. A., Jensen, M. P., Ehde, D. M., Hoffman, A. J., Patterson, D. R., & Robinson, L. R. (2004). Psychosocial predictors of long-term adjustment to lower-limb amputation and phantom limb pain. *Disability and rehabilitation*, *26*(14-15), 882–893. https://doi.org/10.1080/09638280410001708896

Hänsel, A., Lenggenhager, B., von Känel, R., Curatolo, M., & Blanke, O. (2011). Seeing and identifying with a virtual body decreases pain perception. *European journal of pain (London, England)*, *15*(8), 874–879. https://doi.org/10.1016/j.ejpain.2011.03.013

Happé, F., & Frith, U. (2014). Annual research review: Towards a developmental neuroscience of atypical social cognition. *Journal of Child Psychology and Psychiatry*, *55*(6), 553-577. https://doi.org/10.1111/jcpp.12162

Hara, M., Nabae, H., Yamamoto, A., & Higuchi, T. (2015). A novel rubber hand illusion paradigm allowing active Self-touch with variable force feedback controlled by a haptic device. *IEEE Transactions on Human-Machine Systems*, *46*(1), 78-87. https://doi.org/10.1109/THMS.2015.2487499

Hatton, S. N., Lagopoulos, J., Hermens, D. F., Naismith, S. L., Bennett, M. R., & Hickie, I. B. (2012). Correlating anterior insula gray matter volume changes in young people with clinical and neurocognitive outcomes: an MRI study. *BMC psychiatry*, *12*(1), 1-10. https://dx.doi.org/10.1186%2F1471-244X-12-45

Hauf, P., & Försterling, F. (Eds.). (2007). *Making minds: the shaping of human minds through social context* (Vol. 4). John Benjamins Publishing.

Havas, V., Gabarrós, A., Juncadella, M., Rifa-Ros, X., Plans, G., Acebes, J. J., ... & Rodríguez-Fornells, A. (2015). Electrical stimulation mapping of nouns and verbs in Broca's area. *Brain and Language*, *145*, 53-63. https://doi.org/10.1016/j.bandl.2015.04.005

Hebscher, M. & Gilboa, A. (2016) A boost of confidence: The role of the ventromedial prefrontal cortex in memory, decision-making, and schemas. *Neuropsychologia*, 90, 46–58, https://doi.org/10.1016/j.neuropsychologia.2016.05.003

Hebscher, M., Barkan-Abramski, M., Goldsmith, M., Aharon-Peretz, J. & Gilboa, A. (2016) Memory, Decision-Making, and the Ventromedial Prefrontal Cortex (vmPFC): The Roles of Subcallosal and Posterior Orbitofrontal Cortices in Monitoring and Control Processes. *Cerebral Cortex,* 26, 4590–4601, https://doi.org/10.1093/cercor/bhv220

Heinrichs-Graham, E., Hoburg, J. M., & Wilson, T. W. (2018). The peak frequency of motor-related gamma oscillations is modulated by response competition. *NeuroImage*, *165*, 27–34. https://doi.org/10.1016/j.neuroimage.2017.09.059

Hester, R., Foxe, J. J., Molholm, S., Shpaner, M., & Garavan, H. (2005). Neural mechanisms involved in error processing: a comparison of errors made with and without

awareness. *Neuroimage*, *27*(3), 602-608. https://doi.org/10.1016/j.neuroimage.2005.04.035

Hikosaka, O., Sakai, K., Miyauchi, S., Takino, R., Sasaki, Y., & Putz, B. (1996). Activation of human presupplementary motor area in learning of sequential procedures: a functional MRI study. *Journal of neurophysiology*, *76*(1), 617-621. https://doi.org/10.1152/jn.1996.76.1.617

Hoechstetter, K., Berg, P., & Scherg, M. (2010). BESA research tutorial 4: distributed source imaging. *BESA Research Tutorial*, 1-29.

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature neuroscience*, *1*(4), 304-309. https://doi.org/10.1038/1124

Holmes, N. P., Snijders, H. J., & Spence, C. (2006). Reaching with alien limbs: visual exposure to prosthetic hands in a mirror biases proprioception without accompanying illusions of ownership. *Perception & psychophysics*, *68*(4), 685–701. https://doi.org/10.3758/bf03208768

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological review*, *109*(4), 679. https://psycnet.apa.org/doi/10.1037/0033-295X.109.4.679

Holroyd, C. B., & Coles, M. G. (2008). Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behaviour. *Cortex; a journal devoted to the study of the nervous system and behaviour*, *44*(5), 548–559. https://doi.org/10.1016/j.cortex.2007.08.013

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G., & Cohen, J. D. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature neuroscience*, *7*(5), 497-498. https://doi.org/10.1038/nn1238

Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: Sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology*, *45*(5), 688-697. http://dx.doi.org/10.1111/j.1469-8986.2008.00668.x

Holroyd, C. B., Yeung, N., Coles, M. G., & Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General*, *134*(2), 163. https://psycnet.apa.org/doi/10.1037/0096-3445.134.2.163

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning. *The Behavioural and brain sciences*, *24*(5), 849–937. https://doi.org/10.1017/s0140525x01000103

Hon, N., Poh, J. H., & Soon, C. S. (2013). Preoccupied minds feel less control: sense of agency is modulated by cognitive load. *Consciousness and cognition*, *22*(2), 556–561. https://doi.org/10.1016/j.concog.2013.03.004

Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: the role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological bulletin*, *139*(1), 133–151. https://doi.org/10.1037/a0028566

Hume, D. (1896). *A treatise of human nature*. Clarendon Press.

Hummelsheim, H., Bianchetti, M., Wiesendanger, M., & Wiesendanger, R. (1988). Sensory inputs to the agranular motor fields: a comparison between precentral, supplementary-motor and premotor areas in the monkey. *Experimental brain research*, *69*(2), 289–298. https://doi.org/10.1007/BF00247574

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M. C., Mazziotta, J. C., & Rizzolatti, G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(24), 13995–13999. https://doi.org/10.1073/pnas.241474598

Ikeda, A., Lüders, H. O., Burgess, R. C., & Shibasaki, H. (1992). Movement-related potentials recorded from supplementary motor area and primary motor area: role of supplementary motor area in voluntary movements. *Brain*, *115*(4), 1017-1043. https://doi.org/10.1093/brain/115.4.1017

Imaizumi, S., & Asai, T. (2015). Dissociation of agency and body ownership following visuomotor temporal recalibration. *Frontiers in integrative neuroscience*, *9*, 35. https://doi.org/10.3389/fnint.2015.00035

Imaizumi, S., & Tanno, Y. (2019). Intentional binding coincides with explicit sense of agency. *Consciousness and cognition*, *67*, 1–15. https://doi.org/10.1016/j.concog.2018.11.005

Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive control. *Trends in cognitive sciences*, *19*(3), 126–132. https://doi.org/10.1016/j.tics.2015.01.004

Isabella, S., Ferrari, P., Jobst, C., Cheyne, J. A., & Cheyne, D. (2015). Complementary roles of cortical oscillations in automatic and controlled processing during rapid serial tasks. *Neuroimage*, *118*, 268-281. https://doi.org/10.1016/j.neuroimage.2015.05.081

Ito M. (2006). Cerebellar circuitry as a neuronal machine. *Progress in neurobiology*, *78*(3-5), 272–303. https://doi.org/10.1016/j.pneurobio.2006.02.006

Ito, S., Stuphorn, V., Brown, J. W., & Schall, J. D. (2003). Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science*, *302*(5642), 120-122. https://doi.org/10.1126/science.1087847

Iturrate, I., Chavarriaga, R., Montesano, L., Minguez, J., & Millán, J. D. R. (2015). Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control. *Scientific reports*, *5*(1), 1-10. https://doi.org/10.1038/srep13893

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *24*(34), 7540–7548. https://doi.org/10.1523/JNEUROSCI.1921-04.2004

James, W., Burkhardt, F., Bowers, F., & Skrupskelis, I. K. (1890). *The principles of psychology* (Vol. 1, No. 2). London: Macmillan.

Jeannerod, M. (1999). The 25th Bartlett Lecture: To act or not to act: Perspectives on the representation of actions. *The Quarterly Journal of Experimental Psychology Section A*, *52*(1), 1-29. https://doi.org/10.1080%2F713755803

Jeannerod, M., & Pacherie, E. (2004). Agency, simulation and self-identification. *Mind & language*, *19*(2), 113-146. https://doi.org/10.1111/J.1468-0017.2004.00251.X

Jehkonen, M., Laihosalo, M., & Kettunen, J. (2006). Anosognosia after stroke: assessment, occurrence, subtypes and impact on functional outcome reviewed. *Acta Neurologica Scandinavica*, *114*(5), 293-306. https://doi.org/10.1111/j.1600-0404.2006.00723.x

Jékely, G., Godfrey-Smith, P., & Keijzer, F. (2021). Reafference and the origin of the self in early nervous system evolution. *Philosophical Transactions of the Royal Society B*, *376*(1821), 20190764. https://doi.org/10.1098/rstb.2019.0764

Jovanovic, B., & Schwarzer, G. (2007). Infant perception of the relative relevance of different manual actions. *European Journal of Developmental Psychology*, *4*(1), 111-125. https://doi.org/10.1080/17405620601033186

Kalaska, J. F., Scott, S. H., Cisek, P., & Sergio, L. E. (1997). Cortical control of reaching movements. *Current opinion in neurobiology*, *7*(6), 849-859. https://doi.org/10.1016/S0959-4388(97)80146-8

Kalckert, A., & Ehrsson, H. H. (2012). Moving a rubber hand that feels like your own: a dissociation of ownership and Agency. *Frontiers in human neuroscience*, *6*, 40. https://doi.org/10.3389/fnhum.2012.00040

Kalckert, A., & Ehrsson, H. H. (2014). The moving rubber hand illusion revisited: Comparing movements and visuotactile stimulation to induce illusory ownership. *Consciousness and cognition*, *26*, 117-132. https://doi.org/10.1016/j.concog.2014.02.003

Kang, S. Y., Im, C. H., Shim, M., Nahab, F. B., Park, J., Kim, D. W., Kakareka, J., Miletta, N., & Hallett, M. (2015). Brain Networks Responsible for Sense of Agency: An EEG Study. *PloS one*, *10*(8), e0135261. https://doi.org/10.1371/journal.pone.0135261

Kannape, O. A., Schwabe, L., Tadi, T., & Blanke, O. (2010). The limits of agency in walking humans. *Neuropsychologia*, *48*(6), 1628–1636. https://doi.org/10.1016/j.neuropsychologia.2010.02.005

Kant, I. 1787/1929. *Critique of Pure Reason*. Translated by N. Kemp Smith. New York: MacMillan.

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *American journal of Psychiatry*, *160*(1), 13-23. https://doi.org/10.1176/appi.ajp.160.1.13

Karakaş, S., Erzengin, O. U., & Başar, E. (2000). A new strategy involving multiple cognitive paradigms demonstrates that ERP components are determined by the superposition of oscillatory responses. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, *111*(10), 1719–1732. https://doi.org/10.1016/s1388-2457(00)00418-1

Karnath, H. O., Baier, B., & Nägele, T. (2005). Awareness of the functioning of one's own limbs mediated by the insular cortex? *Journal of Neuroscience*, *25*(31), 7134-7138. https://dx.doi.org/10.1523%2FJNEUROSCI.1590-05.2005

Kawabe T. (2013). Inferring sense of agency from the quantitative aspect of action outcome. *Consciousness and cognition*, *22*(2), 407–412. https://doi.org/10.1016/j.concog.2013.01.006

Kelso J. (2016). On the Self-Organizing Origins of Agency: (Trends in Cognitive Sciences 20, 490-499; June 14, 2016). *Trends in cognitive sciences*, *20*(11), 868. https://doi.org/10.1016/j.tics.2016.08.011

Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J., & Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. *Nature neuroscience*, *9*(7), 940-947. https://doi.org/10.1038/nn1724

Kerns, J. G., Cohen, J. D., MacDonald, A. W., 3rd, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science (New York, N.Y.)*, *303*(5660), 1023–1026. https://doi.org/10.1126/science.1089910

Kew, J. J., Halligan, P. W., Marshall, J. C., Passingham, R. E., Rothwell, J. C., Ridding, M. C., ... & Brooks, D. J. (1997). Abnormal access of axial vibrotactile input to deafferented somatosensory cortex in human upper limb amputees. *Journal of Neurophysiology*, *77*(5), 2753-2764. https://doi.org/10.1152/jn.1997.77.5.2753

Kim, S. K., Kirchner, E. A., Stefes, A., & Kirchner, F. (2017). Intrinsic interactive reinforcement learning–Using error-related potentials for real world human-robot interaction. *Scientific reports*, *7*(1), 1-16. https://doi.org/10.1038/s41598-017-17682-7

Kircher, T. T., Senior, C., Phillips, M. L., Benson, P. J., Bullmore, E. T., Brammer, M., ... & David, A. S. (2000). Towards a functional neuroanatomy of self processing: effects of faces and words. *Cognitive Brain Research*, *10*(1-2), 133-144. https://doi.org/10.1016/s0926-6410(00)00036-7

Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D. Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *Neuroimage*, *34*(4), 1774-1781. http://dx.doi.org/10.1016/j.neuroimage.2006.11.014

Klein, T. A., Ullsperger, M., & Danielmeier, C. (2013). Error awareness and the insula: links to neurological and psychiatric diseases. *Frontiers in human neuroscience*, *7*, 14. https://doi.org/10.3389/fnhum.2013.00014

Knoblich, G., & Kircher, T. T. (2004). Deceiving oneself about being in control: conscious detection of changes in visuomotor coupling. *Journal of experimental psychology. Human perception and performance*, *30*(4), 657–666. https://doi.org/10.1037/0096-1523.30.4.657

Knoblich, G., & Sebanz, N. (2005). Agency in the face of error. *Trends in cognitive sciences*, *9*(6), 259–261. https://doi.org/10.1016/j.tics.2005.04.006

Knyazev G. G. (2007). Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neuroscience and biobehavioural reviews*, *31*(3), 377–395. https://doi.org/10.1016/j.neubiorev.2006.10.004

Knyazev G. G. (2012). EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neuroscience and biobehavioural reviews*, *36*(1), 677–695. https://doi.org/10.1016/j.neubiorev.2011.10.002

Koch, C. (2012). *Consciousness: Confessions of a romantic reductionist*. MIT press.

Kolev, V., Demiralp, T., Yordanova, J., Ademoglu, A., & Isoglu-Alkaç, U. (1997). Time-frequency analysis reveals multiple functional components during oddball P300. *Neuroreport*, *8*(8), 2061–2065. https://doi.org/10.1097/00001756-199705260-00050

Kometer, M., Pokorny, T., Seifritz, E., & Volleinweider, F. X. (2015). Psilocybin-induced spiritual experiences and insightfulness are associated with synchronization of neuronal oscillations. *Psychopharmacology*, *232*(19), 3663–3676. https://doi.org/10.1007/s00213-015-4026-7

Kong, J., White, N. S., Kwong, K. K., Vangel, M. G., Rosman, I. S., Gracely, R. H., & Gollub, R. L. (2006). Using fMRI to dissociate sensory encoding from cognitive evaluation of heat pain intensity. *Human brain mapping*, *27*(9), 715–721. https://doi.org/10.1002/hbm.20213

Kontaris, I., Wiggett, A. J., & Downing, P. E. (2009). Dissociation of extrastriate body and biological-motion selective areas by manipulation of visual-motor congruency. *Neuropsychologia*, *47*(14), 3118–3124. https://doi.org/10.1016/j.neuropsychologia.2009.07.012

Kooijman, C. M., Dijkstra, P. U., Geertzen, J., Elzinga, A., & van der Schans, C. P. (2000). Phantom pain and phantom sensations in upper limb amputees: an epidemiological study. *Pain*, *87*(1), 33–41. https://doi.org/10.1016/S0304-3959(00)00264-5

Kortte, K. B., McWhorter, J. W., Pawlak, M. A., Slentz, J., Sur, S., & Hillis, A. E. (2015). Anosognosia for hemiplegia: The contributory role of right inferior frontal gyrus. *Neuropsychology*, *29*(3), 421–432. https://doi.org/10.1037/neu0000135

Krämer, U. M., Cunillera, T., Camara, E., Marco-Pallarés, J., Cucurell, D., Nager, W., ... & Münte, T. F. (2007). The impact of catechol-O-methyltransferase and dopamine D4 receptor genotypes on neurophysiological markers of performance monitoring. *Journal of Neuroscience*, *27*(51), 14190-14198. https://doi.org/10.1523/jneurosci.4229-07.2007

Krugwasser, A. R., Harel, E. V., & Salomon, R. (2019). The boundaries of the self: The sense of agency across different sensorimotor aspects. *Journal of vision*, *19*(4), 14-14. https://doi.org/10.1167/19.4.14

Kühn, S., & Brass, M. (2010). The cognitive representation of intending not to act: Evidence for specific non-action-effect binding. *Cognition*, *117*(1), 9–16. https://doi.org/10.1016/j.cognition.2010.06.006

Kühn, S., Brass, M., & Haggard, P. (2013). Feeling in control: Neural correlates of experience of agency. *Cortex; a journal devoted to the study of the nervous system and behaviour*, *49*(7), 1935–1942. https://doi.org/10.1016/j.cortex.2012.09.002

Kühn, S., Nenchev, I., Haggard, P., Brass, M., Gallinat, J., & Voss, M. (2011). Whodunnit? Electrophysiological correlates of agency judgements. *PloS one*, *6*(12), e28657. https://doi.org/10.1371/journal.pone.0028657

Kumar, A., Gao, L., Pirogova, E., & Fang, Q. (2019). A review of error-related potential-based brain–computer interfaces for motor impaired people. *IEEE Access*, *7*, 142451-142466. https://doi.org/10.1109/ACCESS.2019.2944067

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science (New York, N.Y.)*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*(5872), 110-113. https://doi.org/10.1126/science.1154735

Langer, K. G., & Levine, D. N. (2014). Babinski, J.(1914). Contribution to the Study of the Mental Disorders in Hemiplegia of Organic Cerebral Origin (Anosognosia). Translated by KG Langer & DN Levine: Translated from the original Contribution à l'Étude des Troubles Mentaux dans l'Hémiplégie Organique Cérébrale (Anosognosie). *Cortex: A Journal Devoted to the Study of the Nervous System and Behaviour*. https://doi.org/10.1016/j.cortex.2014.04.019

Larson, M. J., Clayson, P. E., & Clawson, A. (2014). Making sense of all the conflict: a theoretical review and critique of conflict-related ERPs. *International journal of psychophysiology*, *93*(3), 283-297. https://doi.org/10.1016/j.ijpsycho.2014.06.007

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2007). Manipulating the experienced onset of intention after action execution. *Journal of cognitive neuroscience*, *19*(1), 81–90. https://doi.org/10.1162/jocn.2007.19.1.81

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to intention. *Science (New York, N.Y.)*, *303*(5661), 1208–1210. https://doi.org/10.1126/science.1090973

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and clinical neurophysiology*, *48*(6), 609–621. https://doi.org/10.1016/0013-4694(80)90419-8

Leibenluft, E., Gobbini, M. I., Harrison, T., & Haxby, J. V. (2004). Mothers' neural activation in response to pictures of their children and other children. *Biological psychiatry*, *56*(4), 225–232. https://doi.org/10.1016/j.biopsych.2004.05.017

Leiguarda, R., Starkstein, S., Nogues, M., Berthier, M., & Arbelaiz, R. (1993). Paroxysmal alien hand syndrome. *Journal of Neurology, Neurosurgery & Psychiatry*, *56*(7), 788-792. https://doi.org/10.1136/jnnp.56.7.788

Leube, D. T., Knoblich, G., Erb, M., & Kircher, T. T. (2003a). Observing one's hand become anarchic: An fMRI study of action identification. *Consciousness and cognition*, *12*(4), 597-608. https://doi.org/10.1016/S1053-8100(03)00079-5

Leube, D. T., Knoblich, G., Erb, M., Grodd, W., Bartels, M., & Kircher, T. T. (2003b). The neural correlates of perceiving one's own movements. *NeuroImage*, *20*(4), 2084–2090. https://doi.org/10.1016/j.neuroimage.2003.07.033

Li, P., Jia, S., Feng, T., Liu, Q., Suo, T., & Li, H. (2010). The influence of the diffusion of responsibility effect on outcome evaluations: electrophysiological evidence from an ERP study. *Neuroimage*, *52*(4), 1727-1733. https://doi.org/10.1016/j.neuroimage.2010.04.275

Libet B. (1985). Mediation of slow-inhibitory postsynaptic potentials. *Nature*, *313*(5998), 161–162. https://doi.org/10.1038/313161b0

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain: a journal of neurology*, *106 (Pt 3)*, 623–642. https://doi.org/10.1093/brain/106.3.623

Linden D. E. (2005). The p300: where in the brain is it produced and what does it tell us? *The Neuroscientist: a review journal bringing neurobiology, neurology and psychiatry*, *11*(6), 563–576. https://doi.org/10.1177/1073858405280524

Lindner, A., Thier, P., Kircher, T. T., Haarmeier, T., & Leube, D. T. (2005). Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. *Current biology: CB*, *15*(12), 1119–1124. https://doi.org/10.1016/j.cub.2005.05.049

Liu, D. X., Wu, X., Du, W., Wang, C., Chen, C., & Xu, T. (2017). Deep spatial-temporal model for rehabilitation gait: Optimal trajectory generation for knee joint of lower-limb exoskeleton. *Assembly Automation*. https://doi.org/10.1108/AA-11-2016-155

Llera, A., van Gerven, M. A., Gómez, V., Jensen, O., & Kappen, H. J. (2011). On the use of interaction error potentials for adaptive brain computer interfaces. *Neural network : the official journal of the International Neural Network Society*, *24*(10), 1120–1127. https://doi.org/10.1016/j.neunet.2011.05.006

Logan, G. D., & Crump, M. J. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science (New York, N.Y.)*, *330*(6004), 683–686. https://doi.org/10.1126/science.1190483

Longo, M. R., Betti, V., Aglioti, S. M., & Haggard, P. (2009). Visually induced analgesia: seeing the body reduces pain. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *29*(39), 12125–12130. https://doi.org/10.1523/JNEUROSCI.3072-09.2009

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Luck, S. J., & Kappenman, E. S. (Eds.). (2011). *The Oxford handbook of event-related potential components*. Oxford university press.

Ludeman, D. A., Farrar, N., Riesgo, A., Paps, J., & Leys, S. P. (2014). Evolutionary origins of sensation in metazoans: functional evidence for a new sensory organ in sponges. *BMC evolutionary biology*, *14*, 3. https://doi.org/10.1186/1471-2148-14-3

Luu, P., & Tucker, D. M. (2001). Regulating action: alternating activation of midline frontal and motor cortical networks. *Clinical neurophysiology: official journal of the International*

Federation of Clinical Neurophysiology, *112*(7), 1295–1306. https://doi.org/10.1016/s1388-2457(01)00559-4

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clinical neurophysiology*, *115*(8), 1821-1835. https://doi.org/10.1016/j.clinph.2004.03.031

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological science*, *14*(1), 47–53. https://doi.org/10.1111/1467-9280.01417

MacDonald, P. A., & Paus, T. (2003). The role of parietal cortex in awareness of self-generated movements: a transcranial magnetic stimulation study. *Cerebral cortex (New York, N.Y. : 1991)*, *13*(9), 962–967. https://doi.org/10.1093/cercor/13.9.962

Maidhof C. (2013). Error monitoring in musicians. *Frontiers in human neuroscience*, *7*, 401. https://doi.org/10.3389/fnhum.2013.00401

Maidhof, C., Rieger, M., Prinz, W., & Koelsch, S. (2009). Nobody is perfect: ERP effects prior to performance errors in musicians indicate fast monitoring processes. *PloS one*, *4*(4), e5032. https://doi.org/10.1371/journal.pone.0005032

Makin, T. R., Scholz, J., Filippini, N., Henderson Slater, D., Tracey, I., & Johansen-Berg, H. (2013). Phantom pain is associated with preserved structure and function in the former hand area. *Nature communications*, *4*, 1570. https://doi.org/10.1038/ncomms2571

Makris, N., Goldstein, J. M., Kennedy, D., Hodge, S. M., Caviness, V. S., Faraone, S. V., Tsuang, M. T., & Seidman, L. J. (2006). Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia research*, *83*(2-3), 155–171. https://doi-org.sire.ub.edu/10.1016/j.schres.2005.11.020

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*, *19*(3), 1233-1239. https://doi.org/10.1016/s1053-8119(03)00169-1

Manoach, D. S., & Agam, Y. (2013). Neural markers of errors as endophenotypes in neuropsychiatric disorders. *Frontiers in human neuroscience*, *7*, 350. https://doi.org/10.3389/fnhum.2013.00350

Manyakov, N. V., Combaz, A., Chumerin, N., Robben, A., Vliet, M. V., & Hulle, M. M. V. (2012). Feasibility of error-related potential detection as novelty detection problem in P300 mind spelling. In *International Conference on Artificial Intelligence and Soft Computing* (pp. 293-301). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29350-4-35

Marcel, A. (2003). Awareness and Ownership of Action. *Agency and self-awareness: Issues in philosophy and psychology*, *2*, 48.

Marcel, A. J., Tegnér, R., & Nimmo-Smith, I. (2004). Anosognosia for plegia: specificity, extension, partiality and disunity of bodily unawareness. *Cortex*, *40*(1), 19-40. https://doi.org/10.1016/S0010-9452(08)70919-5

Marco-Pallares, J., Camara, E., Münte, T. F., & Rodríguez-Fornells, A. (2008). Neural mechanisms underlying adaptive actions after slips. *Journal of cognitive neuroscience*, *20*(9), 1595-1610. https://psycnet.apa.org/doi/10.1162/jocn.2008.20117

Marco-Pallares, J., Cucurell, D., Cunillera, T., García, R., Andrés-Pueyo, A., Münte, T. F., & Rodríguez-Fornells, A. (2008). Human oscillatory activity associated to reward processing in a gambling task. *Neuropsychologia*, *46*(1), 241–248. https://doi.org/10.1016/j.neuropsychologia.2007.07.016

Margaux, P., Emmanuel, M., Sébastien, D., Olivier, B., & Jérémie, M. (2012). Objective and subjective evaluation of online error correction during P300-based spelling. *Advances in Human-Computer Interaction*, *2012*. https://doi.org/10.1155/2012/578295

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, *164*(1), 177-190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martini, M., Perez-Marcos, D., & Sanchez-Vives, M. V. (2014). Modulation of pain threshold by virtual body ownership. *European journal of pain (London, England)*, *18*(7), 1040–1048. https://doi.org/10.1002/j.1532-2149.2014.00451.x

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science (New York, N.Y.)*, *315*(5810), 393–395. https://doi.org/10.1126/science.1131295

Mathalon, D. H., Whitfield, S. L., & Ford, J. M. (2003). Anatomy of an error: ERP and fMRI. *Biological psychology*, *64*(1-2), 119-141. https://doi.org/10.1016/S0301-0511(03)00105-4

Matsuzawa, M., Matsuo, K., Sugio, T., Kato, C., & Nakai, T. (2005). Temporal relationship between action and visual outcome modulates brain activation: an fMRI study. *Magnetic resonance in medical sciences: MRMS: an official journal of Japan Society of Magnetic Resonance in Medicine*, *4*(3), 115–121. https://doi.org/10.2463/mrms.4.115

Mazzola, L., Mauguiere, F., & Isnard, J. (2019). Functional mapping of the human insula: Data from electrical stimulations. *Revue neurologique*, *175*(3), 150-156.

McCarthy, G., & Donchin, E. (1981). A metric for thought: a comparison of P300 latency and reaction time. *Science (New York, N.Y.)*, *211*(4477), 77–80. https://doi.org/10.1126/science.7444452

McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., Frackowiak, R. S., & Frith, C. D. (1996). The neural correlates of inner speech and auditory verbal imagery in schizophrenia:

relationship to auditory verbal hallucinations. *The British journal of psychiatry: the journal of mental science*, *169*(2), 148–159. https://doi.org/10.1192/bjp.169.2.148

Meehl, P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American psychologist*, *17*(12), 827. https://psycnet.apa.org/doi/10.1037/h0041029

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain structure and function*, *214*(5-6), 655-667. https://doi.org/10.1007/s00429-010-0262-0

Menon, V., Gallardo, G., Pinsk, M. A., Nguyen, V. D., Li, J. R., Cai, W., & Wassermann, D. (2020). Microstructural organization of human insula is linked to its macrofunctional circuitry and predicts cognitive control. *eLife*, *9*, e53470. https://doi.org/10.7554/eLife.53470

Menzer, F., Brooks, A., Halje, P., Faller, C., Vetterli, M., & Blanke, O. (2010). Feeling in control of your footsteps: Conscious gait monitoring and the auditory consequences of footsteps. *Cognitive neuroscience*, *1*(3), 184–192. https://doi.org/10.1080/17588921003743581

Metzinger, T. (2004). *Being no one: The self-model theory of subjectivity*. mit Press.

Miall R. C. (1998). The cerebellum, predictive control and motor coordination. *Novartis Foundation symposium*, *218*, 272–290. https://doi.org/10.1002/9780470515563.ch15

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural networks*, *9*(8), 1265-1279. https://doi.org/10.1016/s0893-6080(96)00035-4

Miele, D. B., Wager, T. D., Mitchell, J. P., & Metcalfe, J. (2011). Dissociating neural correlates of action monitoring and metacognition of agency. *Journal of cognitive neuroscience*, *23*(11), 3620–3636. https://doi.org/10.1162/jocn_a_00052

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological bulletin*, *82*(2), 213. https://psycnet.apa.org/doi/10.1037/h0076486

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Miller, C. H., Sacchet, M. D., & Gotlib, I. H. (2020). Support vector machines and affective science. *Emotion Review*, *12*(4), 297-308. https://doi.org/10.1177%2F1754073920930784

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. *Journal of cognitive neuroscience*, *9*(6), 788-798. https://doi.org/10.1162/jocn.1997.9.6.788

Minshew, N. J., & Keller, T. A. (2010). The nature of brain dysfunction in autism: functional brain imaging studies. *Current opinion in neurology*, *23*(2), 124–130. https://doi.org/10.1097/WCO.0b013e32833782d4

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research*, *6*(1), 57-77. https://doi.org/10.1016/0166-4328(82)90081-X

Miyazaki, M., & Hiraki, K. (2006). Delayed intermodal contingency affects young children's recognition of their current self. *Child development*, *77*(3), 736–750. https://doi.org/10.1111/j.1467-8624.2006.00900.x

Molinaro, N., Monsalve, I. F., & Lizarazu, M. (2016). Is there a common oscillatory brain mechanism for producing and predicting language? *Language, Cognition and Neuroscience*, *31*(1), 145-158. http://dx.doi.org/10.1080/23273798.2015.1077978

Möller, J., Jansma, B. M., Rodriguez-Fornells, A., & Münte, T. F. (2007). What the brain does before the tongue slips. *Cerebral cortex (New York, N.Y.: 1991)*, *17*(5), 1173–1178. https://doi.org/10.1093/cercor/bhl028

Monai, E., Bernocchi, F., Bisio, M., Bisogno, A. L., Salvalaggio, A., & Corbetta, M. (2020). Multiple network disconnection in anosognosia for hemiplegia. *Frontiers in Systems Neuroscience*, *14*, 21. https://doi.org/10.3389/fnsys.2020.00021

Monat, J. P. (2017). The emergence of humanity's self-awareness. *Futures*, *86*, 27-35. https://doi.org/10.1016/j.futures.2016.08.002

Monk, C. S., Peltier, S. J., Wiggins, J. L., Weng, S. J., Carrasco, M., Risi, S., & Lord, C. (2009). Abnormalities of intrinsic functional connectivity in autism spectrum disorders. *NeuroImage*, *47*(2), 764–772. https://doi.org/10.1016/j.neuroimage.2009.04.069

Moore, J. W., & Fletcher, P. C. (2012). Sense of Agency in health and disease: a review of cue integration approaches. *Consciousness and cognition*, *21*(1), 59-68. https://doi.org/10.1016/j.concog.2011.08.010

Moore, J. W., & Haggard, P. (2010). Intentional binding and higher order agency experience. *Consciousness and cognition*, *19*(1), 490-491. https://doi.org/10.1016/j.concog.2009.11.007

Moore, J. W., Turner, D. C., Corlett, P. R., Arana, F. S., Morgan, H. L., Absalom, A. R., ... & Fletcher, P. C. (2011). Ketamine administration in healthy volunteers reproduces aberrant agency experiences associated with schizophrenia. *Cognitive neuropsychiatry*, *16*(4), 364-381. https://doi.org/10.1080/13546805.2010.546074

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and cognition*, *18*(4), 1056–1064. https://doi.org/10.1016/j.concog.2009.05.004

Moore, J., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and cognition*, *17*(1), 136-144. https://doi.org/10.1016/j.concog.2006.12.004

Moran, L. V., Tagamets, M. A., Sampath, H., O'Donnell, A., Stein, E. A., Kochunov, P., & Hong, L. E. (2013). Disruption of anterior insula modulation of large-scale brain networks in schizophrenia. *Biological psychiatry*, *74*(6), 467-474. https://doi.org/10.1016/j.biopsych.2013.02.029

Moritz, S., & Woodward, T. S. (2006). Metacognitive control over false memories: a key determinant of delusional thinking. *Current psychiatry reports*, *8*(3), 184-190. https://doi.org/10.1007/s11920-006-0022-2

Moro, V., Pernigo, S., Zapparoli, P., Cordioli, Z., & Aglioti, S. M. (2011). Phenomenology and neural correlates of implicit and emergent motor awareness in patients with anosognosia for hemiplegia. *Behavioural brain research*, *225*(1), 259-269. https://doi.org/10.1016/j.bbr.2011.07.010

Mort, D. J., Malhotra, P., Mannan, S. K., Rorden, C., Pambakian, A., Kennard, C., & Husain, M. (2003). The anatomy of visual neglect. *Brain: a journal of neurology*, *126*(Pt 9), 1986–1997. https://doi.org/10.1093/brain/awg200

Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H., & Acuna, C. (1975). Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. *Journal of neurophysiology*, *38*(4), 871-908. https://doi.org/10.1152/jn.1975.38.4.871

Mulert, C., Pogarell, O., Juckel, G., Rujescu, D., Giegling, I., Rupp, D., Mavrogiorgou, P., Bussfeld, P., Gallinat, J., Möller, H. J., & Hegerl, U. (2004). The neural basis of the P300 potential. Focus on the time-course of the underlying cortical generators. *European archives of psychiatry and clinical neuroscience*, *254*(3), 190–198. https://doi.org/10.1007/s00406-004-0469-2

Muthukumaraswamy S. D. (2010). Functional properties of human primary motor cortex gamma oscillations. *Journal of neurophysiology*, *104*(5), 2873–2885. https://doi.org/10.1152/jn.00607.2010

Nachev, P., Wydell, H., O'neill, K., Husain, M., & Kennard, C. (2007). The role of the pre-supplementary motor area in the control of action. *Neuroimage*, *36*, T155-T163. https://doi.org/10.1016/j.neuroimage.2007.03.034

Nahab, F. B., Kundu, P., Gallea, C., Kakareka, J., Pursley, R., Pohida, T., Miletta, N., Friedman, J., & Hallett, M. (2011). The neural processes underlying self-agency. *Cerebral cortex (New York, N.Y.: 1991)*, *21*(1), 48–55. https://doi.org/10.1093/cercor/bhq059

Namkung, H., Kim, S. H., & Sawa, A. (2017). The insula: an underestimated brain area in clinical neuroscience, psychiatry, and neurology. *Trends in neurosciences*, *40*(4), 200-207. https://doi.org/10.1016/j.tins.2017.02.002

Naqvi, N. H., & Bechara, A. (2009). The hidden island of addiction: the insula. *Trends in neurosciences*, *32*(1), 56-67. https://doi.org/10.1016/j.tins.2008.09.009

Nelson, L. D., Patrick, C. J., Collins, P., Lang, A. R., & Bernat, E. M. (2011). Alcohol impairs brain reactivity to explicit loss feedback. *Psychopharmacology*, *218*(2), 419. https://doi.org/10.1007/s00213-011-2323-3

Nelson, T. O., Stuart, R. B., Howard, C., & Crowley, M. (1999). Metacognition and clinical psychology: A preliminary framework for research and practice. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *6*(2), 73-79. https://doi.org/10.1002/(SICI)1099-0879(199905)6:2%3C73::AID-CPP187%3E3.0.CO;2-7

Newen, A., & Vogeley, K. (2003). Self-representation: searching for a neural signature of self-consciousness. *Consciousness and cognition*, *12*(4), 529–543. https://doi.org/10.1016/s1053-8100(03)00080-1

Nielsen, T. I. (1963). Volition: A new experimental approach. *Scandinavian journal of psychology*, *4*(1), 225-230. https://doi.org/10.1111/j.1467-9450.1963.tb01326.x

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology*, *38*(5), 752–760.

Nieuwenhuis, S., Ridderinkhof, K. R., Talsma, D., Coles, M. G., Holroyd, C. B., Kok, A., & Van der Molen, M. W. (2002). A computational account of altered error processing in older age: dopamine and the error-related negativity. *Cognitive, Affective, & Behavioural Neuroscience*, *2*(1), 19-36. https://doi.org/10.3758/CABN.2.1.19

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioural Reviews*, *28*(4), 441-448. https://doi.org/10.1016/j.neubiorev.2004.05.003

Nieuwenhuis, S., Yeung, N., Holroyd, C. B., Schurger, A., & Cohen, J. D. (2004). Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cerebral Cortex*, *14*(7), 741-747. https://doi.org/10.1093/cercor/bhh034

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological bulletin*, *131*(4), 510–532. https://doi.org/10.1037/0033-2909.131.4.510

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*(2), 162–175. https://doi.org/10.1111/j.1469-8986.2010.01057.x

Nigbur, R., Ivanova, G., & Stürmer, B. (2011). Theta power as a marker for cognitive interference. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, *122*(11), 2185–2194. https://doi.org/10.1016/j.clinph.2011.03.030

Nightingale, S. (1982). Somatoparaphrenia: a case report. *Cortex*, *18*(3), 463-467. https://doi.org/10.1016/s0010-9452(82)80043-9

Nimchinsky, E. A., Gilissen, E., Allman, J. M., Perl, D. P., Erwin, J. M., & Hof, P. R. (1999). A neuronal morphologic type unique to humans and great apes. *Proceedings of the National Academy of Sciences*, *96*(9), 5268-5273. https://doi.org/10.1073/pnas.96.9.5268

Norman, D. A., & Shallice, T. (1986). Attention to action. In *Consciousness and self-regulation* (pp. 1-18). Springer, Boston, MA.

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

Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: an orienting account. *Cognition*, *111*(2), 275–279. https://doi.org/10.1016/j.cognition.2009.02.002

Ojemann, G. A. (1983). Brain organization for language from the perspective of electrical stimulation mapping. *Behavioural and Brain Sciences*, *6*(2), 189-206. https://psycnet.apa.org/doi/10.1017/S0140525X00015491

O'Keeffe, F. M., Dockree, P. M., & Robertson, I. H. (2004). Poor insight in traumatic brain injury mediated by impaired error processing?: Evidence from electrodermal activity. *Cognitive Brain Research*, *22*(1), 101-112. https://doi.org/10.1016/j.cogbrainres.2004.07.012

Orfei, M. D., Robinson, R. G., Prigatano, G. P., Starkstein, S., Rüsch, N., Bria, P., ... & Spalletta, G. (2007). Anosognosia for hemiplegia after stroke is a multifaceted phenomenon: a systematic review of the literature. *Brain*, *130*(12), 3075-3090. https://doi.org/10.1093/brain/awm106

Ownsworth, T., & Clare, L. (2006). The association between awareness deficits and rehabilitation outcome following acquired brain injury. *Clinical psychology review*, *26*(6), 783-795. https://doi.org/10.1016/j.cpr.2006.05.003

Pacella, V., Foulon, C., Jenkinson, P. M., Scandola, M., Bertagnoli, S., Avesani, R., ... & De Schotten, M. T. (2019). Anosognosia for hemiplegia as a tripartite disconnection syndrome. *elife*, *8*, e46075. https://dx.doi.org/10.7554%2FeLife.46075

Pacherie, E. (2001). Agency lost and found: a commentary on Spence. *Philosophy, Psychiatry, & Psychology*, *8*(2), 173-176. https://doi.org/10.1353/ppp.2001.0013

Pacherie, E. (2013). Sense of Agency: Many Facets, Multiple Sources. In Metcalfe, J., & Terrace, H. S. (Eds.), *Agency and joint attention* (pp. 321-331). Oxford University Press.

Pacherie, E., Green, M., & Bayne, T. (2006). Phenomenology and delusions: Who put the 'alien'in alien control? *Consciousness and Cognition*, *15*(3), 566-577. https://doi.org/10.1016/j.concog.2005.11.008

Padrao, G., Gonzalez-Franco, M., Sanchez-Vives, M. V., Slater, M., & Rodriguez-Fornells, A. (2016). Violating body movement semantics: Neural signatures of Self-generated and external-generated errors. *Neuroimage*, *124*, 147-156. https://doi.org/10.1016/j.neuroimage.2015.08.022

Padrao, G., Penhune, V., de Diego-Balaguer, R., Marco-Pallares, J., & Rodriguez-Fornells, A. (2014). ERP evidence of adaptive changes in error processing and attentional control during rhythm synchronization learning. *NeuroImage*, *100*, 460-470. https://doi.org/10.1016/j.neuroimage.2014.06.034

Palaniyappan, L., & Liddle, P. F. (2012). Does the salience network play a cardinal role in psychosis? An emerging hypothesis of insular dysfunction. *Journal of Psychiatry & Neuroscience*. https://doi.org/10.1503/jpn.100176

Palaniyappan, L., Simmonite, M., White, T. P., Liddle, E. B., & Liddle, P. F. (2013). Neural primacy of the salience processing system in schizophrenia. *Neuron*, *79*(4), 814–828. https://doi.org/10.1016/j.neuron.2013.06.027

Palmer, C. E., Davare, M., & Kilner, J. M. (2016). Physiological and Perceptual Sensory Attenuation Have Different Underlying Neurophysiological Correlates. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *36*(42), 10803–10812. https://doi.org/10.1523/JNEUROSCI.1694-16.2016

Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of psychophysiology*, *18*(1), 49-65. https://doi.org/10.1016/0167-8760(84)90014-X

Pedersen, P. M., Jørgensen, H. S., Nakayama, H., Raaschou, H. O., & Olsen, T. S. (1996). Orientation in the acute and chronic stroke patient: impact on ADL and social activities. The Copenhagen Stroke Study. *Archives of physical medicine and rehabilitation*, *77*(4), 336-339. https://doi.org/10.1016/S0003-9993(96)90080-5

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830.

Peled, A., Ritsner, M., Hirschmann, S., Geva, A. B., & Modai, I. (2000). Touch feel illusion in schizophrenic patients. *Biological psychiatry*, *48*(11), 1105-1108. https://doi.org/10.1016/S0006-3223(00)00947-1

Peña-Casanova, J. (2005). *Programa integrado de exploración neuropsicológica: TRB: Test Barcelona Revisado*. Elsevier España.

Peña-Casanova, J., Quinones-Ubeda, S., Gramunt-Fombuena, N., Quintana-Aparicio, M., Aguilar, M., Badenes, D., ... & NEURONORMA Study Team. (2009a). Spanish Multicenter Normative Studies (NEURONORMA Project): norms for verbal fluency tests. *Archives of Clinical Neuropsychology*, *24*(4), 395-411. https://doi.org/10.1093/arclin/acp042

Peña-Casanova, J., Quiñones-Úbeda, S., Quintana-Aparicio, M., Aguilar, M., Badenes, D., Molinuevo, J. L., ... & Blesa, R. (2009b). Spanish Multicenter Normative Studies (NEURONORMA Project): norms for verbal span, visuospatial span, letter and number sequencing, trail making test, and symbol digit modalities test. *Archives of Clinical Neuropsychology*, *24*(4), 321-341. https://doi.org/10.1093/arclin/acp038

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*(4), 389-443. https://psycnet.apa.org/doi/10.1093/brain/60.4.389

Penfield, W., & Faulk Jr, M. E. (1955). The insula: further observations on its function. *Brain*, *78*(4), 445-470. https://doi.org/10.1093/brain/78.4.445

Penfield, W., & Roberts, L. (2014). *Speech and brain mechanisms*. Princeton University Press.

Petkova, V. I., Björnsdotter, M., Gentile, G., Jonsson, T., Li, T. Q., & Ehrsson, H. H. (2011). From part- to whole-body ownership in the multisensory brain. *Current biology: CB*, *21*(13), 1118–1122. https://doi.org/10.1016/j.cub.2011.05.022

Pfister, R., Kiesel, A., & Hoffmann, J. (2011). Learning at any rate: Action–effect learning for stimulus-based actions. *Psychological Research*, *75*(1), 61-65. https://doi.org/10.1007/s00426-010-0288-1

Pfurtscheller, G., Graimann, B., Huggins, J. E., Levine, S. P., & Schuh, L. A. (2003). Spatiotemporal patterns of beta desynchronization and gamma synchronization in corticographic data during self-paced movement. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, *114*(7), 1226–1236. https://doi.org/10.1016/s1388-2457(03)00067-1

Pia, L., Neppi-Modona, M., Ricci, R., & Berti, A. (2004). The anatomy of anosognosia for hemiplegia: a meta-analysis. *Cortex; a journal devoted to the study of the nervous system and behaviour*, *40*(2), 367–377. https://doi.org/10.1016/s0010-9452(08)70131-x

Pijnacker, J., Geurts, B., Van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, *48*(10), 2940-2951. https://doi-org.sire.ub.edu/10.1016/j.neuropsychologia.2010.06.003

Pisella, L., Gréa, H., Tilikete, C., Vighetto, A., Desmurget, M., Rode, G., Boisson, D., & Rossetti, Y. (2000). An 'automatic pilot' for the hand in human posterior parietal cortex: toward reinterpreting optic ataxia. *Nature neuroscience*, *3*(7), 729–736. https://doi.org/10.1038/76694

Polich, J. (2003). Theoretical overview of P3a and P3b. *Detection of change*, 83-98.

Polich J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Posada, A., Franck, N., Augier, S., Georgieff, N., & Jeannerod, M. (2007). Altered processing of sensorimotor feedback in schizophrenia. *Comptes rendus biologies*, *330*(5), 382–388. https://doi.org/10.1016/j.crvi.2007.02.003

Prinz, W. (1997). Perception and action planning. *European journal of cognitive psychology*, *9*(2), 129-154. https://doi.org/10.1080/713752551

Proverbio, A. M., Crotti, N., Manfredi, M., Adorni, R., & Zani, A. (2012). Who needs a referee? How incorrect basketball actions are automatically detected by basketball players' brain. *Scientific reports*, *2*, 883. https://doi.org/10.1038/srep00883

Puglisi, G., Howells, H., Sciortino, T., Leonetti, A., Rossi, M., Conti Nibali, M., ... & Bello, L. (2019). Frontal pathways in cognitive control: direct evidence from intraoperative stimulation and diffusion tractography. *Brain*, *142*(8), 2451-2465. https://doi.org/10.1093/brain/awz178

Pyasik, M., Burin, D., & Pia, L. (2018). On the relation between body ownership and sense of agency: A link at the level of sensory-related signals. *Acta Psychologica*, *185*, 219-228. https://doi.org/10.1016/j.actpsy.2018.03.001

Rabbitt P. M. (1966). Errors and error correction in choice-response tasks. *Journal of experimental psychology*, *71*(2), 264–272. https://doi.org/10.1037/h0022853

Rabbitt, P. (2002). Consciousness is slower than you think. *The Quarterly Journal of Experimental Psychology Section A*, *55*(4), 1081-1092. https://doi.org/10.1080/02724980244000080

Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology*, *29*(4), 727-743. https://psycnet.apa.org/doi/10.1080/14640747708400645

Rakshit, A., Konar, A., & Nagar, A. K. (2020). A hybrid brain-computer interface for closed-loop position control of a robot arm. *IEEE/CAA Journal of Automatica Sinica*, *7*(5), 1344-1360. https://doi.org/10.1109/JAS.2020.1003336

Ramachandran, V. S., & Hirstein, W. (1998). The perception of phantom limbs. The D. O. Hebb lecture. *Brain: a journal of neurology*, *121 (Pt 9)*, 1603–1630. https://doi.org/10.1093/brain/121.9.1603

Ramachandran, V. S., & Rogers-Ramachandran, D. (1996). Synaesthesia in phantom limbs induced with mirrors. *Proceedings. Biological sciences*, *263*(1369), 377–386. https://doi.org/10.1098/rspb.1996.0058

Ramachandran, V. S., Rogers-Ramachandran, D., & Stewart, M. (1992). Perceptual correlates of massive cortical reorganization. *Science (New York, N.Y.)*, *258*(5085), 1159–1160. https://doi.org/10.1126/science.1439826

Ramnani N. (2006). The primate cortico-cerebellar system: anatomy and function. *Nature reviews. Neuroscience*, *7*(7), 511–522. https://doi.org/10.1038/nrn1953

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Reitan, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of consulting psychology*, *19*(5), 393. https://doi.org/10.1037/h0044509

Renes, R. A., van der Weiden, A., Prikken, M., Kahn, R. S., Aarts, H., & van Haren, N. E. (2015). Abnormalities in the experience of self-agency in schizophrenia: A replication study. *Schizophrenia Research*, *164*(1-3), 210-213. https://doi.org/10.1016/j.schres.2015.03.015

Ridderinkhof K. R. (2002). Micro- and macro-adjustments of task set: activation and suppression in conflict tasks. *Psychological research*, *66*(4), 312–323. https://doi.org/10.1007/s00426-002-0104-7

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*(5695), 443-447. https://doi.org/10.1126/science.1100301

Ritterband-Rosenbaum, A., Nielsen, J. B., & Christensen, M. S. (2014). Sense of agency is related to gamma band coupling in an inferior parietal-preSMA circuitry. *Frontiers in human neuroscience*, *8*, 510. https://doi.org/10.3389/fnhum.2014.00510

Rizzolatti, G., Fogassi, L., & Gallese, V. (1997). Parietal cortex: from sight to action. *Current opinion in neurobiology*, *7*(4), 562-567. https://doi.org/10.1016/S0959-4388(97)80037-2

Rochat P. (1998). Self-perception and action in infancy. *Experimental brain research*, *123*(1-2), 102–109. https://doi.org/10.1007/s002210050550

Rochat P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and cognition*, *12*(4), 717–731. https://doi.org/10.1016/s1053-8100(03)00081-3

Rochat, P., & Striano, T. (1999). Emerging self-exploration by 2-month-old infants. *Developmental Science*, *2*(2), 206-218.

Rodriguez-Fornells, A., Kurzbuch, A. R., & Münte, T. F. (2002). Time course of error detection and correction in humans: neurophysiological evidence. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *22*(22), 9990–9996. https://doi.org/10.1523/JNEUROSCI.22-22-09990.2002

Rofes, A., Spena, G., Talacchi, A., Santini, B., Miozzo, A., & Miceli, G. (2017). Mapping nouns and finite verbs in left hemisphere tumors: a direct electrical stimulation study. Neurocase, 23(2), 105-113. https://doi.org/10.1080/13554794.2017.1307418

Rojas, P. H., Sivaraju, A., Quraishi, I. H., Vanderlind, M., Rofes, A., Połczynska-Bletsos, M. M. ... & Benjamin, C. F. (2021). Electrical cortical stimulation can impair production of the alphabet without impairing counting. Epilepsy & behaviour reports, 15, 100433. https://doi.org/10.1016/j.ebr.2021.100433

Romano, D., & Maravita, A. (2014). The visual size of one′ s own hand modulates pain anticipation and perception. *Neuropsychologia*, *57*, 93-100. https://doi.org/10.1016/j.neuropsychologia.2014.03.002

Rosen, H. J., Alcantar, O., Zakrzewski, J., Shimamura, A. P., Neuhaus, J., & Miller, B. L. (2014). Metacognition in the behavioural variant of frontotemporal dementia and Alzheimer's disease. *Neuropsychology*, *28*(3), 436. https://doi.org/10.1037/neu0000012

Rotermund, D., Ernst, U. A., & Pawelzik, K. R. (2006). Towards on-line adaptation of neuro-prostheses with neuronal evaluation signals. *Biological cybernetics*, *95*(3), 243-257. https://doi.org/10.1007/s00422-006-0083-7

Roussel, C., Hughes, G., & Waszak, F. (2014). Action prediction modulates both neurophysiological and psychophysical indices of sensory attenuation. *Frontiers in human neuroscience*, *8*, 115. https://doi.org/10.3389/fnhum.2014.00115

Ruchsow, M., Grothe, J., Spitzer, M., & Kiefer, M. (2002). Human anterior cingulate cortex is activated by negative feedback: evidence from event-related potentials in a guessing task. *Neuroscience letters*, *325*(3), 203-206. https://doi.org/10.1016/s0304-3940(02)00288-4

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., & Rushworth, M. F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *28*(51), 13775–13785. https://doi.org/10.1523/JNEUROSCI.3541-08.2008

Ruiz, M. H., Jabusch, H. C., & Altenmüller, E. (2009). Detecting wrong notes in advance: neuronal correlates of error monitoring in pianists. *Cerebral cortex*, *19*(11), 2625-2639. https://doi.org/10.1093/cercor/bhp021

Ruvolo, P., Messinger, D., & Movellan, J. (2015). Infants Time Their Smiles to Make Their Moms Smile. *PloS one*, *10*(9), e0136492. https://doi.org/10.1371/journal.pone.0136492

Sala, C. M. S. D. (1998). Disentangling the alien and anarchic hand. *Cognitive neuropsychiatry*, *3*(3), 191-207. https://psycnet.apa.org/doi/10.1080/135468098396143

Salomon, R., Malach, R., & Lamy, D. (2009). Involvement of the intrinsic/default system in movement-related self recognition. *PloS one*, *4*(10), e7527. https://doi.org/10.1371/journal.pone.0007527

San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in human neuroscience*, *6*, 304. https://doi.org/10.3389/fnhum.2012.00304

Sato A. (2009). Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition*, *110*(1), 74–83. https://doi.org/10.1016/j.cognition.2008.10.011

Sato, A., & Yasuda, A. (2005). Illusion of sense of Self-Agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of Self-Agency, but not the sense of Self-ownership. *Cognition*, *94*(3), 241-255. https://doi.org/10.1016/j.cognition.2004.04.003

Saunders, B., Lin, H., Milyavskaya, M., & Inzlicht, M. (2017). The emotive nature of conflict monitoring in the medial prefrontal cortex. *International Journal of Psychophysiology*, *119*, 31-40. https://psycnet.apa.org/doi/10.1016/j.ijpsycho.2017.01.004

Saze, T., Hirao, K., Namiki, C., Fukuyama, H., Hayashi, T., & Murai, T. (2007). Insular volume reduction in schizophrenia. *European archives of psychiatry and clinical neuroscience*, *257*(8), 473–479. https://doi-org.sire.ub.edu/10.1007/s00406-007-0750-2

Scepkowski, L. A., & Cronin-Golomb, A. (2003). The alien hand: cases, categorizations, and anatomical correlates. *Behavioural and cognitive neuroscience reviews*, *2*(4), 261–277. https://doi.org/10.1177/1534582303260119

Schaefer, M., Heinze, H. J., & Galazky, I. (2010). Alien hand syndrome: neural correlates of movements without conscious will. *PLoS One*, *5*(12), e15010. https://doi.org/10.1371/journal.pone.0015010

Schalk, G., Wolpaw, J. R., McFarland, D. J., & Pfurtscheller, G. (2000). EEG-based communication: presence of an error potential. *Clinical neurophysiology*, *111*(12), 2138-2144. https://doi.org/10.1016/s1388-2457(00)00457-0

Scheffers, M. K., & Coles, M. G. (2000). Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 141. https://psycnet.apa.org/doi/10.1037/0096-1523.26.1.141

Schlerf, J., Ivry, R. B., & Diedrichsen, J. (2012). Encoding of sensory prediction errors in the human cerebellum. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *32*(14), 4913–4922. https://doi.org/10.1523/JNEUROSCI.4504-11.2012

Schmidt, N. M., Blankertz, B., & Treder, M. S. (2012). Online detection of error-related potentials boosts the performance of mental typewriters. *BMC neuroscience*, *13*(1), 1-13. https://doi.org/10.1186/1471-2202-13-19

Schneider, K. (1959). *Clinical psychopathology*. Grune & Stratton.

Schnell, K., Heekeren, K., Schnitker, R., Daumann, J., Weber, J., Heßelmann, V., ... & Gouzoulis-Mayfrank, E. (2007). An fMRI approach to particularize the frontoparietal network for visuomotor action monitoring: detection of incongruence between test subjects' actions and resulting perceptions. *NeuroImage*, *34*(1), 332-341. https://doi.org/10.1016/j.neuroimage.2006.08.027

Schnell, K., Heekeren, K., Schnitker, R., Daumann, J., Weber, J., Hesselmann, V., Möller-Hartmann, W., Thron, A., & Gouzoulis-Mayfrank, E. (2007). An fMRI approach to particularize the frontoparietal network for visuomotor action monitoring: Detection of incongruence between test subjects' actions and resulting perceptions. *NeuroImage*, *34*(1), 332–341. https://doi.org/10.1016/j.neuroimage.2006.08.027

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in neurosciences*, *32*(1), 9-18. https://doi.org/10.1016/j.tins.2008.09.012

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual review of neuroscience*, *23*(1), 473-500. https://doi.org/10.1146/annurev.neuro.23.1.473

Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(42), E2904–E2913. https://doi.org/10.1073/pnas.1210467109

Schürmann, M., Başar-Eroglu, C., Kolev, V., & Başar, E. (2001). Delta responses and cognitive processing: single-trial evaluations of human visual P300. *International Journal of Psychophysiology*, *39*(2-3), 229-239. https://doi.org/10.1016/s0167-8760(00)00144-6

Searle, J. R., & Willis, S. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge university press.

Seeley, W. W., Carlin, D. A., Allman, J. M., Macedo, M. N., Bush, C., Miller, B. L., & DeArmond, S. J. (2006). Early frontotemporal dementia targets neurons unique to apes and humans. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *60*(6), 660-667. https://doi.org/10.1002/ana.21055

Seeley, W. W., Allman, J. M., Carlin, D. A., Crawford, R. K., Macedo, M. N., Greicius, M. D., ... & Miller, B. L. (2007). Divergent social functioning in behavioural variant frontotemporal dementia and Alzheimer disease: reciprocal networks and neuronal evolution. *Alzheimer Disease & Associated Disorders*, *21*(4), S50-S57. https://doi.org/10.1097/wad.0b013e31815c0f14

Seghezzi, S., Giannini, G., & Zapparoli, L. (2019). Neurofunctional correlates of body-ownership and sense of agency: A meta-analytical account of self-consciousness. *Cortex; a journal devoted to the study of the nervous system and behaviour*, *121*, 169–178. https://doi.org/10.1016/j.cortex.2019.08.018

Shanks, D. R., Lopez, F. J., Darby, R. J., & Dickinson, A. (1996). *Distinguishing associative and probabilistic contrast theories of human contingency judgment*. Academic Press.

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgement of causality by human subjects. *The Quarterly Journal of Experimental Psychology*, *41*(2), 139-159. https://doi.org/10.1080%2F14640748908401189

Shany-Ur, T., Lin, N., Rosen, H. J., Sollberger, M., Miller, B. L., & Rankin, K. P. (2014). Self-awareness in neurodegenerative disease relies on neural structures mediating reward-driven attention. *Brain*, *137*(8), 2368-2381. https://doi.org/10.1093/brain/awu161

Shepherd, A. M., Matheson, S. L., Laurens, K. R., Carr, V. J., & Green, M. J. (2012). Systematic meta-analysis of insula volume in schizophrenia. *Biological psychiatry*, *72*(9), 775-784. https://doi.org/10.1016/j.biopsych.2012.04.020

Shergill, S. S., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2003). Two eyes for an eye: the neuroscience of force escalation. *Science (New York, N.Y.)*, *301*(5630), 187. https://doi.org/10.1126/science.1085327

Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2005). Evidence for sensory prediction deficits in schizophrenia. *American Journal of Psychiatry*, *162*(12), 2384-2386. https://doi.org/10.1176/appi.ajp.162.12.2384

Sherman, R. A., Arena, J. G., Sherman, C. J., & Ernst, J. L. (1989). The mystery of phantom pain: growing evidence for psychophysiological mechanisms. *Biofeedback and self-regulation*, *14*(4), 267-280. https://doi.org/10.1007/BF00999118

Sherman, R. A., Sherman, C. J., & Parker, L. (1984). Chronic phantom and stump pain among American veterans: results of a survey. *Pain*, *18*(1), 83–95. https://doi.org/10.1016/0304-3959(84)90128-3

Shima, K., & Tanji, J. (1998). Role for cingulate motor area cells in voluntary movement selection based on reward. *Science (New York, N.Y.)*, *282*(5392), 1335–1338. https://doi.org/10.1126/science.282.5392.1335

Sierpowska, J., Gabarrós, A., Ripollés, P., Juncadella, M., Castañer, S., Camins, Á., ... & Rodríguez-Fornells, A. (2013). Intraoperative electrical stimulation of language switching in two bilingual patients. *Neuropsychologia*, *51*(13), 2882-2892. https://doi.org/10.1016/j.neuropsychologia.2013.09.003

Sierpowska, J., Gabarrós, A., Fernandez-Coello, A., Camins, À., Castañer, S., Juncadella, M. ... & Rodríguez-Fornells, A. (2015). Morphological derivation overflow as a result of disruption of the left frontal aslant white matter tract. Brain and Language, 142, 54-64. https://doi.org/10.1016/j.bandl.2015.01.005

Sierpowska, J., Fernandez-Coello, A., Gomez-Andres, A., Camins, À., Castaner, S., Juncadella, M., ... & Rodriguez-Fornells, A. (2018). Involvement of the middle frontal gyrus in language switching as revealed by electrical stimulation mapping and functional magnetic resonance imaging in bilingual brain tumour patients. *Cortex*, *99*, 78-92. https://doi.org/10.1016/j.cortex.2017.10.017

Singer, W. (2001). Consciousness and the binding problem. *Annals of the New York Academy of Sciences*, *929*(1), 123-146. https://doi.org/10.1111/j.1749-6632.2001.tb05712.x

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., & Haggard, P. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nature neuroscience*, *7*(1), 80–84. https://doi.org/10.1038/nn1160

Sirigu, A., Daprati, E., Pradat-Diehl, P., Franck, N., & Jeannerod, M. (1999). Perception of Self-generated movement following left parietal lesion. *Brain*, *122*(10), 1867-1874. https://doi.org/10.1093/brain/122.10.1867

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: an electrophysiological investigation. *Psychophysiology*, *40*(1), 160–164. https://doi.org/10.1111/1469-8986.00016

Skrandies W. (1990). Global field power and topographic similarity. *Brain topography*, *3*(1), 137–141. https://doi.org/10.1007/BF01128870

Slachevsky, A., Pillon, B., Fourneret, P., Pradat-Diehl, P., Jeannerod, M., & Dubois, B. (2001). Preserved adjustment but impaired awareness in a sensory-motor conflict following prefrontal lesions. *Journal of cognitive neuroscience*, *13*(3), 332–340. https://doi.org/10.1162/08989290151137386

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, *6*(2), 174. https://psycnet.apa.org/doi/10.1037/0278-7393.6.2.174

Spence, S. A., Brooks, D. J., Hirsch, S. R., Liddle, P. F., Meehan, J., & Grasby, P. M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain: a journal of neurology*, *120*(11), 1997-2011. https://psycnet.apa.org/doi/10.1093/brain/120.11.1997

Spencer, K. M., & Polich, J. (1999). Poststimulus EEG spectral analysis and P300: attention, task, and probability. *Psychophysiology*, *36*(2), 220–232.

Spengler, S., von Cramon, D. Y., & Brass, M. (2009). Was it me or was it you? How the sense of agency originates from ideomotor learning revealed by fMRI. *Neuroimage*, *46*(1), 290-298. https://doi.org/10.1016/j.neuroimage.2009.01.047

Sperduti, M., Delaveau, P., Fossati, P., & Nadel, J. (2011). Different brain structures related to self- and external-agency attribution: a brief review and meta-analysis. *Brain Structure and Function*, *216*(2), 151-157. https://doi.org/10.1007/s00429-010-0298-1

Sperry R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of comparative and physiological psychology*, *43*(6), 482–489. https://doi.org/10.1037/h0055479

Spüler, M., Bensch, M., Kleih, S., Rosenstiel, W., Bogdan, M., & Kübler, A. (2012). Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI. *Clinical Neurophysiology*, *123*(7), 1328-1337. https://doi.org/10.1016/j.clinph.2011.11.082

Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences*, *105*(34), 12569-12574. https://doi.org/10.1073/pnas.0800005105Ullspeger

Stampfer, H. G., & Başar, E. (1985). Does frequency analysis lead to better understanding of human event related potentials. *The International journal of neuroscience*, *26*(3-4), 181–196. https://doi.org/10.3109/00207458508985616

Stanton, T. R., & Spence, C. (2020). The Influence of Auditory Cues on Bodily and Movement Perception. *Frontiers in psychology*, *10*, 3001. https://doi.org/10.3389/fpsyg.2019.03001

Steenbergen, H. V., Band, G. P., & Hommel, B. (2009). Reward counteracts conflict adaptation: Evidence for a role of affect in executive control. *Psychological Science*, *20*(12), 1473-1477. https://doi.org/10.1111%2Fj.1467-9280.2009.02470.x

Steinhauser, M., & Kiesel, A. (2011). Performance monitoring and the causal attribution of errors. *Cognitive, Affective, & Behavioural Neuroscience*, *11*(3), 309-320. https://psycnet.apa.org/doi/10.3758/s13415-011-0033-2

Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. *Journal of Neuroscience*, *30*(46), 15643-15653. https://doi.org/10.1523/jneurosci.1899-10.2010

Streng, M. L., Popa, L. S., & Ebner, T. J. (2018). Complex Spike Wars: a New Hope. *Cerebellum (London, England)*, *17*(6), 735–746. https://doi.org/10.1007/s12311-018-0960-3

Striano, T., & Rochat, P. (2000). Emergence of Selective Social Referencing in Infancy. *Infancy: the official journal of the International Society on Infant Studies*, *1*(2), 253–264. https://doi.org/10.1207/S15327078IN0102_7

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643. https://psycnet.apa.org/doi/10.1037/h0054651

Sturm, V. E., Rosen, H. J., Allison, S., Miller, B. L., & Levenson, R. W. (2006). Self-conscious emotion deficits in frontotemporal lobar degeneration. *Brain*, *129*(9), 2508-2516. https://doi.org/10.1093/brain/awl145

Sugiura, M., Kawashima, R., Nakamura, K., Okada, K., Kato, T., Nakamura, A., ... & Fukuda, H. (2000). Passive and active recognition of one's own face. *Neuroimage*, *11*(1), 36-48. https://doi.org/10.1006/nimg.1999.0519

Sunderaraman, P., & Cosentino, S. (2017). Integrating the constructs of anosognosia and metacognition: a review of recent findings in dementia. *Current neurology and neuroscience reports*, *17*(3), 27. https://doi.org/10.1007/s11910-017-0734-1

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Synofzik, M., Vosgerau, G., & Newen, A. (2008a). I move, therefore I am: a new theoretical framework to investigate agency and ownership. *Consciousness and cognition*, *17*(2), 411–424. https://doi.org/10.1016/j.concog.2008.03.008

Synofzik, M., Vosgerau, G., & Newen, A. (2008b). Beyond the comparator model: a multifactorial two-step account of Agency. *Consciousness and cognition*, *17*(1), 219-239. https://doi.org/10.1016/j.concog.2007.03.010

Synofzik, M., Thier, P., & Lindner, A. (2006). Internalizing agency of self-action: perception of one's own hand movements depends on an adaptable prediction about the sensory action outcome. *Journal of neurophysiology*, *96*(3), 1592–1601. https://doi.org/10.1152/jn.00104.2006

Synofzik, M., Vosgerau, G., & Lindner, A. (2009). Me or not me--an optimal integration of agency cues? *Consciousness and cognition*, *18*(4), 1065–1068. https://doi.org/10.1016/j.concog.2009.07.007

Synofzik, M., Thier, P., Leube, D. T., Schlotterbeck, P., & Lindner, A. (2010). Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain*, *133*(1), 262-271. https://doi.org/10.1093/brain/awp291

Synofzik, M., & Vosgerau, G. (2012). Beyond the comparator model. *Consciousness and cognition*, *21*(1), 1–3. https://doi.org/10.1016/j.concog.2012.01.007

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: an interplay between prediction and postdiction. *Frontiers in psychology*, *4*, 127. https://doi.org/10.3389/fpsyg.2013.00127

Tajadura-Jiménez, A., Tsakiris, M., Marquardt, T., & Bianchi-Berthouze, N. (2015). Action sounds update the mental representation of arm dimension: contributions of kinaesthesia and agency. *Frontiers in psychology*, *6*, 689. https://doi.org/10.3389/fpsyg.2015.00689

Takahashi, T., Suzuki, M., Hagino, H., Zhou, S. Y., Kawasaki, Y., Nohara, S., Nakamura, K., Yamashita, I., Seto, H., & Kurachi, M. (2004). Bilateral volume reduction of the insular cortex in patients with schizophrenia: a volumetric MRI Study. *Psychiatry research*, *132*(2), 187–196. https://doi.org/10.1016/j.pscychresns.2004.11.002

Takahashi, T., Suzuki, M., Zhou, S. Y., Hagino, H., Tanino, R., Kawasaki, Y., Nohara, S., Yamashita, I., Seto, H., & Kurachi, M. (2005). Volumetric MRI study of the short and long insular cortices in schizophrenia spectrum disorders. *Psychiatry research*, *138*(3), 209–220. https://doi.org/10.1016/j.pscychresns.2005.02.004

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *Journal of Neuroscience*, *16*(13), 4240-4249. https://doi.org/10.1523/JNEUROSCI.16-13-04240.199

Taylor, S. F., Stern, E. R., & Gehring, W. J. (2007). Neural systems for error monitoring: recent findings and theoretical perspectives. *The Neuroscientist*, *13*(2), 160-172. https://doi.org/10.1177%2F1073858406298184

Taylor, K. S., Seminowicz, D. A., & Davis, K. D. (2009). Two systems of resting state connectivity between the insula and cingulate cortex. *Human brain mapping*, *30*(9), 2731–2745. https://doi.org/10.1002/hbm.20705

Tieri, G., Tidoni, E., Pavone, E. F., & Aglioti, S. M. (2015). Mere observation of body discontinuity affects perceived ownership and vicarious agency over a virtual hand. *Experimental brain research*, *233*(4), 1247–1259. https://doi.org/10.1007/s00221-015-4202-3

Tonin, L., & Millán, J. D. R. (2021). Noninvasive brain–machine interfaces for robotic devices. *Annual Review of Control, Robotics, and Autonomous Systems*, *4*, 191-214. https://dx.doi.org/10.1146/annurev-control-012720-093904

Tsakiris M. (2010). My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia*, *48*(3), 703–712. https://doi.org/10.1016/j.neuropsychologia.2009.09.034

Tsakiris, M., & Haggard, P. (2005). Experimenting with the acting self. *Cognitive neuropsychology*, *22*(3), 387–407. https://doi.org/10.1080/02643290442000158

Tsakiris, M., Costantini, M., & Haggard, P. (2008). The role of the right temporo-parietal junction in maintaining a coherent sense of one's body. *Neuropsychologia*, *46*(12), 3014–3018. https://doi.org/10.1016/j.neuropsychologia.2008.06.004

Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. *Cognition*, *96*(3), 215–231. https://doi.org/10.1016/j.cognition.2004.08.002

Tsakiris, M., Hesse, M. D., Boy, C., Haggard, P., & Fink, G. R. (2007). Neural signatures of body ownership: a sensory network for bodily self-consciousness. *Cerebral cortex (New York, N.Y. : 1991)*, *17*(10), 2235–2244. https://doi.org/10.1093/cercor/bhl131

Tsakiris, M., Longo, M. R., & Haggard, P. (2010). Having a body versus moving your body: neural signatures of agency and body-ownership. *Neuropsychologia*, *48*(9), 2740-2749. https://doi.org/10.1016/j.neuropsychologia.2010.05.021

Tseng, Y. W., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of neurophysiology*, *98*(1), 54–62. https://doi.org/10.1152/jn.00266.2007

Tzur, G., & Berger, A. (2009). Fast and slow brain rhythms in rule/expectation violation tasks: Focusing on evaluation processes by excluding motor action. *Behavioural brain research*, *198*(2), 420-428. https://psycnet.apa.org/doi/10.1016/j.bbr.2008.11.041

Ulloa J. L. (2021). The Control of Movements via Motor Gamma Oscillations. *Frontiers in human neuroscience*, *15*, 787157. https://doi.org/10.3389/fnhum.2021.787157

Ullsperger, M., & Von Cramon, D. Y. (2001). Subprocesses of performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *NeuroImage*, *14*(6), 1387-1401. https://doi.org/10.1006/nimg.2001.0935

Ullsperger, M., & Von Cramon, D. Y. (2003). Error monitoring using external feedback: specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *Journal of Neuroscience*, *23*(10), 4308-4314. https://doi.org/10.1523/jneurosci.23-10-04308.2003

Ullsperger M, Danielmeier C, Jocham G. Neurophysiology of performance monitoring and adaptive behaviour. Physiol Rev. 2014 Jan;94(1):35-79. https://doi.org/10.1152/physrev.00041.2012

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in cognitive sciences*, *18*(5), 259–267. https://doi.org/10.1016/j.tics.2014.02.009

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure and Function*, *214*(5), 629-643. http://dx.doi.org/10.1007/s00429-010-0261-1

Usama, N., Niazi, I. K., Dremstrup, K., & Jochumsen, M. (2021). Detection of Error-Related Potentials in Stroke Patients from EEG Using an Artificial Neural Network. *Sensors*, *21*(18), 6274. https://dx.doi.org/10.3390%2Fs21186274

van den Bos, E., & Jeannerod, M. (2002). Sense of body and sense of action both contribute to self-recognition. *Cognition*, *85*(2), 177–187. https://doi.org/10.1016/s0010-0277(02)00100-2

van Kemenade, B. M., Arikan, B. E., Podranski, K., Steinsträter, O., Kircher, T., & Straube, B. (2019). Distinct Roles for the Cerebellum, Angular Gyrus, and Middle Temporal Gyrus in Action-Feedback Monitoring. *Cerebral cortex (New York, N.Y.: 1991)*, *29*(4), 1520–1531. https://doi.org/10.1093/cercor/bhy048

van Schie, H. T., Mars, R. B., Coles, M. G., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature neuroscience*, *7*(5), 549-554. https://doi.org/10.1038/nn1239

Vocat, R., Staub, F., Stroppini, T., & Vuilleumier, P. (2010). Anosognosia for hemiplegia: a clinical-anatomical prospective study. *Brain*, *133*(12), 3578-3597. https://doi.org/10.1093/brain/awq297

Vocat, R., Staub, F., Stroppini, T., & Vuilleumier, P. (2010). Anosognosia for hemiplegia: a clinical-anatomical prospective study. *Brain*, *133*(12), 3578-3597. https://doi.org/10.1093/brain/awq297

von Holst, E., & Mittelstaedt, H. (1950). Das reafferenzprinzip. *Naturwissenschaften*, *37*(20), 464-476. https://doi.org/10.1007/BF00622503

Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. (2010). Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain*, *133*(10), 3104-3112. https://doi.org/10.1093/brain/awq152

Wager, M., Du Boisgueheneuc, F., Pluchon, C., Bouyer, C., Stal, V., Bataille, B., ... & Gil, R. (2013). Intraoperative monitoring of an aspect of executive functions: administration of the Stroop test in 9 adult patients during awake surgery for resection of frontal glioma. *Operative Neurosurgery*, *72*(2), ons169-ons181. https://doi.org/10.1227/NEU.0b013e31827bf1d6

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. Neuroscience and biobehavioral reviews, 36(8), 1870–1884. https://doi.org/10.1016/j.neubiorev.2012.05.008Agnew, Z., & Wise, R. J. (2008). Separate areas for mirror responses and agency within the parietal operculum. *Journal of Neuroscience*, *28*(47), 12268-12273. https://doi.org/10.1523/JNEUROSCI.2836-08.2008

Waszak, F., Cardoso-Leite, P., & Hughes, G. (2012). Action effect anticipation: neurophysiological basis and functional consequences. *Neuroscience & Biobehavioural Reviews*, *36*(2), 943-959. https://psycnet.apa.org/doi/10.1016/j.neubiorev.2011.11.004

Watanabe, H., & Taga, G. (2006). General to specific development of movement patterns and memory for contingency between actions and events in young infants. *Infant behaviour & development*, *29*(3), 402–422. https://doi.org/10.1016/j.infbeh.2006.02.001

Watanabe, H., & Taga, G. (2009). Flexibility in infant actions during arm- and leg-based learning in a mobile paradigm. *Infant behaviour & development*, *32*(1), 79–90. https://doi.org/10.1016/j.infbeh.2008.10.003

Watson, J. S., & Ramey, C. T. (1972). Reactions to response-contingent stimulation in early infancy. *Merrill-Palmer Quarterly of Behaviour and Development*, *18*(3), 219-227.

Wechsler, D. (2008). Escala de Inteligencia de Wechsler Para Adultos-IV (WAIS-IV): Manual de Aplicación y Corrección. Madrid: TEA.

Wegner D. M. (2003). The mind's best trick: how we experience conscious will. *Trends in cognitive sciences*, *7*(2), 65–69. https://doi.org/10.1016/s1364-6613(03)00002-0

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation. Sources of the experience of will. *The American psychologist*, *54*(7), 480–492. https://doi.org/10.1037//0003-066x.54.7.480

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: experiencing control over the movements of others. *Journal of personality and social psychology*, *86*(6), 838–848. https://doi.org/10.1037/0022-3514.86.6.838

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature neuroscience*, *9*(7), 971–978. https://doi.org/10.1038/nn1727

Welniarz, Q., Worbe, Y., & Gallea, C. (2021). The Forward Model: A Unifying Theory for the Role of the Cerebellum in Motor Control and Sense of Agency. *Frontiers in systems neuroscience*, *15*, 644059. https://doi.org/10.3389/fnsys.2021.644059

Wen W. (2019). Does delay in feedback diminish sense of agency? A review. *Consciousness and cognition*, *73*, 102759. https://doi.org/10.1016/j.concog.2019.05.007

Wen, W., Yamashita, A., & Asama, H. (2015). The influence of action-outcome delay and arousal on sense of agency and the intentional binding effect. *Consciousness and cognition*, *36*, 87–95. https://doi.org/10.1016/j.concog.2015.06.004

Wessel, J. R. (2012). Error awareness and the error-related negativity: evaluating the first decade of evidence. *Frontiers in human neuroscience*, *6*, 88. https://doi.org/10.3389/fnhum.2012.00088

Wessel, J. R., Danielmeier, C., & Ullsperger, M. (2011). Error awareness revisited: accumulation of multimodal evidence from central and autonomic nervous systems. *Journal of cognitive neuroscience*, *23*(10), 3021-3036. https://doi.org/10.1162/jocn.2011.21635

Wessel, J. R., & Aron, A. R. (2015). It's not too late: the onset of the frontocentral P3 indexes successful response inhibition in the stop-signal paradigm. *Psychophysiology*, *52*(4), 472–480. https://doi.org/10.1111/psyp.12374

White, T. P., Joseph, V., Francis, S. T., & Liddle, P. F. (2010). Aberrant salience network (bilateral insula and anterior cingulate cortex) connectivity during information processing in schizophrenia. *Schizophrenia research*, *123*(2-3), 105–115. https://doi.org/10.1016/j.schres.2010.07.020

Williams, D., & Happé, F. (2010). Representing intentions in self and other: studies of autism and typical development. *Developmental science*, *13*(2), 307–319. https://doi.org/10.1111/j.1467-7687.2009.00885.x

Williams, Z. M., Bush, G., Rauch, S. L., Cosgrove, G. R., & Eskandar, E. N. (2004). Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nature neuroscience*, *7*(12), 1370–1375. https://doi.org/10.1038/nn1354

Wittgenstein, L. (1958). *Philosophical investigations*. (Translated by GE Anscombe) Oxford: Blackwell.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current biology: CB*, *11*(18), R729–R732. https://doi.org/10.1016/s0960-9822(01)00432-8

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural networks: the official journal of the International Neural Network Society*, *11*(7-8), 1317–1329. https://doi.org/10.1016/s0893-6080(98)00066-5

Wolpert, D. M., & Miall, R. C. (1996). Forward Models for Physiological Motor Control. *Neural networks: the official journal of the International Neural Network Society*, *9*(8), 1265–1279. https://doi.org/10.1016/s0893-6080(96)00035-4

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880-1882. https://doi.org/10.1126/science.7569931

Womelsdorf, T., Johnston, K., Vinck, M., & Everling, S. (2010). Theta-activity in anterior cingulate cortex predicts task rules and their adjustments following errors. *Proceedings of the National Academy of Sciences*, *107*(11), 5248-5253. https://doi.org/10.1073/pnas.0906194107

Wong, H. Y. (2010). Bodily awareness and bodily action. *A companion to the Philosophy of Action. Chicester: Wiley-Blackwell*, 227-235.

Wylie, K. P., & Tregellas, J. R. (2010). The role of the insula in schizophrenia. *Schizophrenia research*, *123*(2-3), 93-104. http://dx.doi.org/10.1016/j.schres.2010.08.027

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310-1321. https://dx.doi.org/10.1098%2Frstb.2011.0416

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological review*, *111*(4), 931–959. https://doi.org/10.1037/0033-295x.111.4.939

Yomogida, Y., Sugiura, M., Sassa, Y., Wakusawa, K., Sekiguchi, A., Fukushima, A., Takeuchi, H., Horie, K., Sato, S., & Kawashima, R. (2010). The neural basis of agency: an fMRI study. *NeuroImage*, *50*(1), 198–207. https://doi.org/10.1016/j.neuroimage.2009.12.054

Yordanova, J., Devrim, M., Kolev, V., Ademoglu, A., & Demiralp, T. (2000). Multiple time-frequency components account for the complex functional reactivity of P300. *Neuroreport*, *11*(5), 1097–1103. https://doi.org/10.1097/00001756-200004070-00038

Yordanova, J., Falkenstein, M., Hohnsbein, J., & Kolev, V. (2004). Parallel systems of error processing in the brain. *NeuroImage*, *22*(2), 590-602. https://doi.org/10.1016/j.neuroimage.2004.01.040

Zander, T. O., Krol, L. R., Birbaumer, N. P., & Gramann, K. (2016). Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. *Proceedings of the National Academy of Sciences*, *113*(52), 14898-14903. https://doi.org/10.1073/pnas.1605155114

Zhang, T., & Huang, H. (2018). A lower-back robotic exoskeleton: Industrial handling augmentation used to provide spinal support. *IEEE Robotics & Automation Magazine*, *25*(2), 95-106. https://doi.org/10.1109/MRA.2018.2815083

Zhou, A., Shi, Z., Zhang, P., Liu, P., Han, W., Wu, H., Li, Q., Zuo, Q., & Xia, R. (2010). An ERP study on the effect of self-relevant possessive pronoun. *Neuroscience letters*, *480*(2), 162–166. https://doi.org/10.1016/j.neulet.2010.06.033

Ziauddeen, H., & Murray, G. K. (2010). The relevance of reward pathways for schizophrenia. *Current opinion in psychiatry*, *23*(2), 91-96. http://dx.doi.org/10.1097/YCO.0b013e328336661b

Zito, G. A., Wiest, R., & Aybek, S. (2020). Neural correlates of sense of agency in motor control: A neuroimaging meta-analysis. *PloS one*, *15*(6), e0234321. https://doi.org/10.1371/journal.pone.0234321

ANNEX

7. ANNEX

Table S1. Repeated measures ANOVA results for Experiment 1. Results for 3 x 3 x 3 repeated measures ANOVA with factors *Correctness* (3 levels: Correct, SE, EE), *Anterior–posterior electrode location* (3 levels: frontal –F3/Fz/F4-, central –C3/Cz/C4-, and parietal –P3/Pz/P4- locations), and *Lateral scalp location* (3 levels: parasagittal left –F3/C3/P3-, midline –Fz/Cz/Pz-, and parasagittal right –F4/C4/P4-) for the ERP components [error-related negativity (ERN), N2 and P600] and time frequency analyses (delta and theta frequency bands).

			Factor	*df*	*F-value*	*P-value*	*Partial eta-squared (η_p^2)*
ERPs		**ERN** (70-120 ms)	Correctness	2, 48	53.12	$p < .001$	0.69
			Anterior-Posterior location	2, 48	40.77	$p_{GG} < .001$	0.63
			Correctness x Anterior-Posterior	4, 96	20.43	$p_{GG} < .001$	0.46
			Correctness x Lateral scalp location	4, 96	33.60	$p_{GG} < .001$	0.58
			Anterior-Posterior x Lateral scalp location	4, 96	16.76	$p_{GG} < .001$	0.41
			Correctness x Anterior-Posterior x Lateral scalp location	8, 192	13.36	$p_{GG} < .001$	0.36
		N2 (300-400 ms)	Correctness	2, 48	48.87	$p < .001$	0.67
			Correctness x Anterior-Posterior	4, 96	61.40	$p_{GG} < .001$	0.72
			Correctness x Anterior-Posterior x Lateral scalp location	8, 192	9.63	$p_{GG} < .001$	0.28
		P600 (530-630 ms)	Correctness	2, 48	39.06	$p < .001$	0.67
			Correctness x Anterior-Posterior	4, 96	7.32	$p_{GG} = .001$	0.23
			Correctness x Lateral scalp location	4, 96	15.89	$p_{GG} < .001$	0.40
Time frequency	**Theta** (3-8 Hz)	50-150 ms	Correctness	2, 48	61.95	$p < .001$	0.72
			Anterior-Posterior location	2, 48	25.94	$p < .001$	0.52
			Lateral scalp location	2, 48	3.83	*p = .029*	0.14
			Correctness x Anterior-Posterior location	4, 96	6.28	$p_{GG} = .002$	0.21
			Correctness x Lateral scalp location	4, 96	3.78	$p_{GG} = .024$	0.14
			Anterior-Posterior location x Lateral scalp location	4, 96	7.32	$p_{GG} < .001$	0.23
			Correctness x Anterior-Posterior location x Lateral scalp location	8, 192	2.70	$p_{GG} = .036$	0.10
		200-400 ms	Correctness	2, 48	5.12	$p = .032$	0.18
			Anterior-Posterior location	2, 48	5.21	$p_{GG} = .014$	0.18
			Lateral scalp location	2, 48	3.86	$p_{GG} = .029$	0.14
			Correctness x Anterior-Posterior location	4, 96	17.69	$p_{GG} < .001$	0.42
			Anterior-Posterior location x Lateral scalp location	4, 96	8.06	$p_{GG} < .001$	0.25
		400-600 ms	Correctness	2, 48	19.85	$p < .001$	0.45
			Anterior-Posterior location	2, 48	10.00	$p_{GG} = .001$	0.29
			Correctness x Anterior-Posterior location	4, 96	3.63	$p_{GG} = .023$	0.13
			Anterior-Posterior x Lateral scalp location	4, 96	7.09	$p_{GG} < .001$	0.23
	Delta (1-3 Hz)	200-400 ms	Correctness	2, 48	28.30	$p < .001$	0.54
			Correctness x Anterior-Posterior location	4, 96	3.73	$p_{GG} = .033$	0.13
		400-600 ms	Correctness	2, 48	32.90	$p_{GG} < .001$	0.58
			Anterior-Posterior location	2, 48	7.66	$p_{GG} = .007$	0.24
			Lateral scalp location	2, 48	4.54	$p_{GG} = .023$	0.16
			Correctness x Anterior-Posterior location	4, 96	3.57	$p_{GG} = .041$	0.13

Table S2. Results for the subjective experience questionnaire for Experiment 2. Summary of the results on the assessment of both facets of agency explored (internal attribution and external attribution) and the control items for Experiment 2. Bonferroni-corrected Wilcoxon pairwise comparisons are shown for all 3 A-O delay conditions.

Subjective reports	*Questionnaire item*	*No-Delay vs. 150-Delay*	*150-Delay vs. 400-Delay*	*No-Delay vs. 400-Delay*
Internal attribution	Q1. My movements	n.s.	Z = -4.16 $p < .001$	Z = -4.63 $p < .001$
	Q2. Feeling of control	n.s.	Z = -4.12 $p < .001$	Z = -4.90 $p < .001$
External attribution	Q3. Not my movements	n.s.	Z = -3.65 $p < .001$	n.s.
	Q4. External errors	n.s.	n.s.	n.s.
Control	Q5. Influence	n.s.	n.s.	n.s.
	Q6. More than 2 hands	n.s.	n.s.	n.s.
	Q7. Ownership	n.s.	Z = -5.12 $p < .001$	Z = -5.23 $p < .001$

n.s.: non-significant.

Table S3: Repeated measures ANOVA results for Experiment 2. Results for 3 x 3 x 3 repeated measures ANOVA with the factors *A-O condition* (No-Delay, 150-Delay, 400-Delay), *Anterior–posterior electrode location* (Frontal, Central, Parietal), and *Lateral scalp location* (Left, Midline, Right) for Correct responses - Feedback correct-related positivity (FCRP) (time window 600-700 ms) and SE -Feedback-related negativity (FRN) (time window 525-625 ms) and time frequency results for Theta oscillatory activity (3-8 Hz) at the ERN time window (50-250 ms) are described. A 3 x 3 rmANOVA with the factors *A-O condition* (No-Delay/150-Delay/400-Delay) and *Midline electrode location* (Fz/Cz/Pz) was performed for inspecting the theta-frequency band at the FRN time window (550-850 ms) are also shown.

			Factor	***df***	***F-value***	***P-value***	***Partial eta-squared (η_p^2)***
ERPs	**FCRP** (600-700 ms)		A-O condition	2, 46	19.85	$p < .001$	0.47
			Anterior-Posterior electrode location	2, 46	4.75	$p_{GG} = .037$	0.18
			A-O condition x Lateral scalp location	4, 92	6.90	$p_{GG} < .001$	0.24
			A-O condition x Anterior-Posterior electrode location x Lateral scalp location	8, 184	4.81	$p_{GG} < .001$	0.18
	FRN (525-625 ms)		A-O condition	2, 46	9.60	$p < .001$	0.29
			Lateral scalp location	2, 46	4.72	$p_{GG} = .01$	0.17
			Lateral scalp location x Anterior-Posterior electrode location	4, 92	3.37	$p_{GG} = .033$	0.13
Time frequency	**Delta** (1-3 Hz)		Anterior-Posterior electrode location	2, 46	56.60	$p < .001$	0.80
			A-O condition x Anterior-Posterior electrode location	4, 92	2.38	$p_{GG} = .05$	0.12
	Theta (3-8 Hz)	50-250 ms	Correctness	1, 23	25.65	$p < .001$	0.54
			Anterior-Posterior electrode location	2, 46	17.02	$p_{GG} < .001$	0.44
			Correctness x Anterior-Posterior electrode location	2,46	52.79	$p_{GG} < .001$	0.71
		550-850 ms	A-O condition	2, 46	10.39	$p_{GG} = .001$	0.33

Table S4. Repeated measures ANOVA results for Experiment 3. Results for 3 x 2 x 3 x 3 repeated measures ANOVA with factors *Correctness* (Correct, SE, EE), *A-O condition* (No-Delay, 400-Delay), *Anterior–posterior electrode location* (Frontal, Central, Parietal), and *Lateral scalp location* (Left, Midline, Right).

		Factor	***df***	***F-value***	***P-value***	***Partial eta-squared (η_p^2)***
ERPs	**N2**	Anterior-Posterior location	2, 34	5.33	p_{GG} = .024	0.24
		Lateral scalp location	2, 34	4.73	p_{GG} = .027	0.22
		Correctness x Anterior-Posterior	2, 34	6.55	p_{GG} = .017	0.28
		Anterior-Posterior x Lateral scalp location	4, 68	4.05	p_{GG} = .026	0.19
	P600	Correctness	1, 17	70.26	p < .001	0.81
		Lateral scalp location	2, 34	6.95	p_{GG} = .009	0.29
		Correctness x Anterior-Posterior	2, 34	5.00	p_{GG} = .032	0.23
		Correctness x Lateral scalp location	2, 34	8.31	p_{GG} = .032	0.33
		Anterior-Posterior x Lateral scalp location	4, 68	4.54	p_{GG} = .032	0.21
		Correctness x Anterior-Posterior x Lateral scalp location	4, 68	5.38	p_{GG} = .008	0.24

Table S5. Individual results for F1-scores and classifier accuracy. Performance of the models for each subject, averaged along the significant intervals. SE: self-generated errors, EE: externally generated errors.

Subject	F1 Score			Accuracy
	Correct	SE	EE	
S01	0.50	0.48	0.39	0.46
S02	0.46	0.54	0.39	0.45
S03	0.47	0.39	0.41	0.42
S04	0.52	0.44	0.50	0.50
S05	0.53	0.43	0.51	0.49
S06	0.56	0.50	0.51	0.53
S07	0.50	0.38	0.44	0.45
S08	0.43	0.46	0.38	0.42
S09	0.55	0.41	0.38	0.45
S10	0.42	0.38	0.36	0.38
S11	0.46	0.45	0.44	0.45
S12	0.44	0.50	0.47	0.47
S13	0.43	0.41	0.36	0.40
S14	0.39	0.42	0.33	0.38
S15	0.39	0.46	0.33	0.39
S16	0.45	0.38	0.40	0.40
S17	0.47	0.40	0.43	0.44
S18	0.44	0.38	0.43	0.42
S19	0.44	0.31	0.40	0.39
S20	0.47	0.40	0.41	0.43
S21	0.50	0.40	0.45	0.45
S22	0.42	0.37	0.35	0.38
S23	0.50	0.47	0.39	0.46
Mean	**0.47**	**0.42**	**0.41**	**0.44**

Figure S1: Time-frequency analysis depicting changes in relative power at Cz for Experiment 1. Dotted contours show the areas where repeated measures ANOVA's were computed. SE: Self errors; EE: External errors.

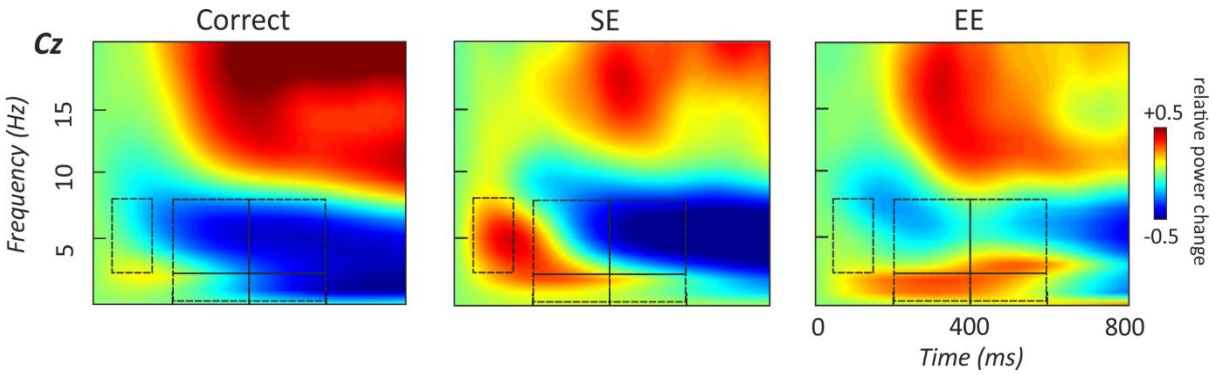


Figure S2. Time-frequency analysis depicting changes in relative power for Experiment 2. A) Changes in power for Correct and SE at the No-Delay, 150-Delay and 400-Delay A-O conditions. **B)** Changes in power for the contrasts SE minus Correct for all A-O conditions. Dotted contours show the areas where rmANOVA's were computed. SE: Self errors.

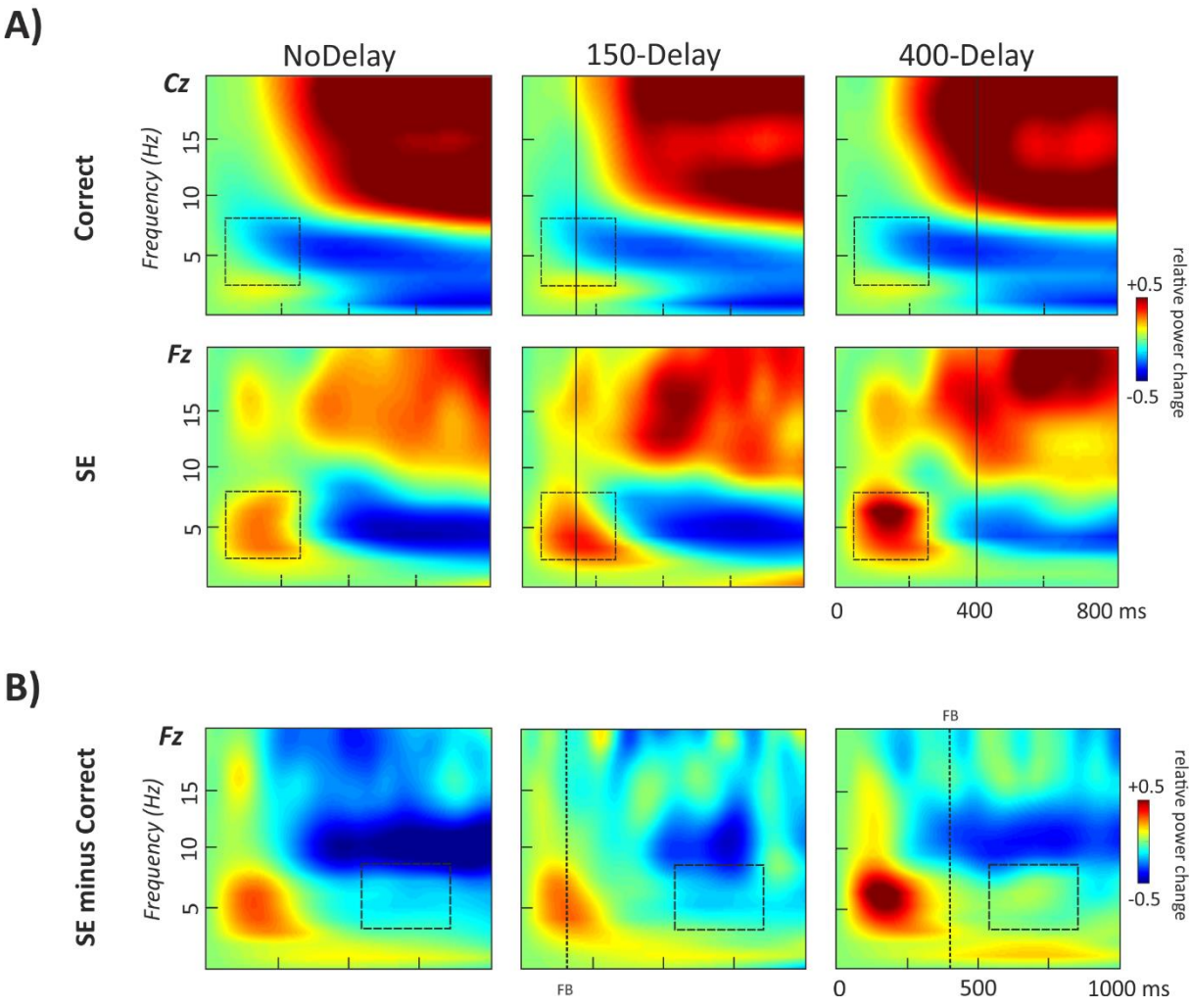
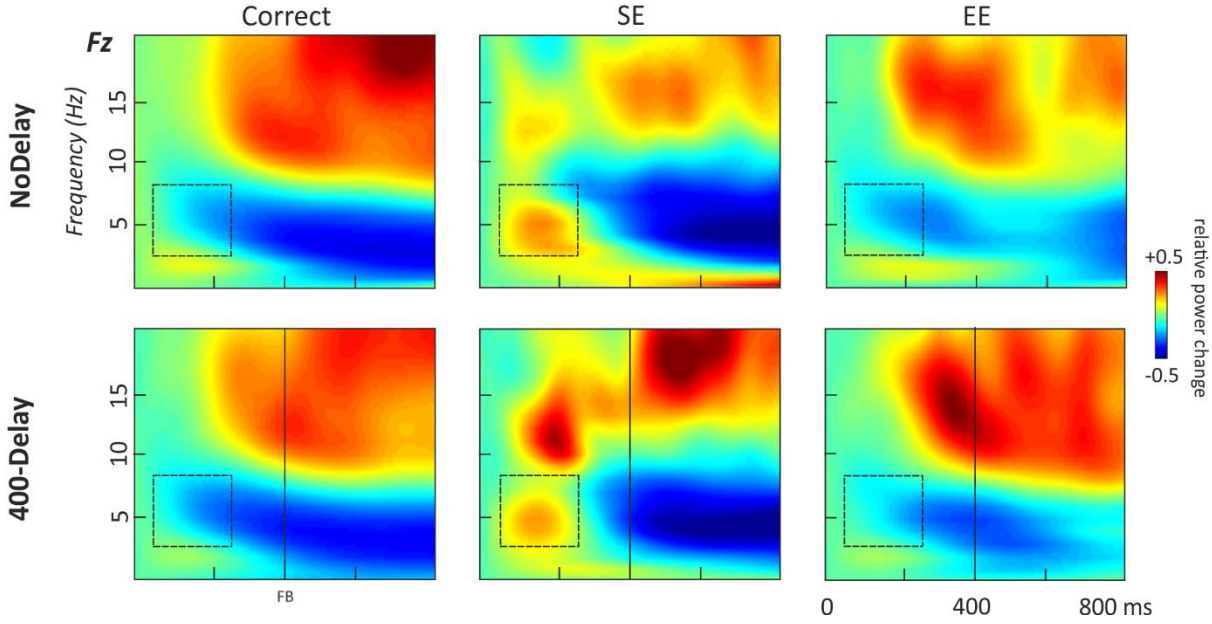


Figure S3. Time-frequency analysis depicting changes in relative power for Experiment 3. Dotted contours show the areas where repeated measures ANOVA's were computed. SE: Self errors; EE: External errors.



Doctoral program in

Brain, Cognition and Behaviour

2022