Radboud University Nijmegen

Bachelorthesis

Tracking the Sense of Agency in BCI Applications

Author:

J.-P. van Acken

0815322

j.vanacken@student.ru.nl

Supervisors:

Dr. W.F.G. Haselager

Drs. L. Roijendijk

Drs. R. Vlek

August 31, 2012

revised February 6, 2013

Abstract:

This thesis paper presents an experiment where participants were given the task to make a virtual hand on a screen execute one of two gestures, cued via audio instructions. Subjects were led to believe that they controlled the virtual hand through a brain-computer interface with a technique known as motor imagery. However, the actual hand movement was preprogrammed, making every Sense of Agency the participants reported illusory. Timing between the instruction cue and hand movement was manipulated between two blocks. After each manipulation a questionnaire was filled out, measuring the subject's Sense of Agency over the virtual hand. During the blocks an electroencephalogram (EEG) was measured. The results of this experiment were compared to the results of research by Wegner, Sparrow, and Winerman on a similar non-BCI task. The Sense of Agency rating were found to be significantly higher when compared to the non-BCI experiment. In addition this thesis holds a section on the predictive powers of several models for the Sense of Agency and concludes with the suggestion to opt for a more refined model that would actually allow to predict not only the presence of the Sense of Agency but the strength of it.

Keywords: authorship processing, brain-computer interface, comparator model, Sense of Agency, Sense of Self, Helping Hands

1 Introduction

This thesis studies the *Sense of Agency* (SA) in users that operate applications controlled via a *brain-computer interface* (BCI). First the concept of the SA will be reviewed via a look at the most common notions and models. Afterwards the basics of brain-computer interfacing will be explained and the reasons why the SA is of importance for that field will be assessed. After this theoretical groundwork I present an experiment where participants *used* an apparently working but non-functional BCI application and had to give their rating for a perceived SA. These results will then be discussed in the light of the many models for the SA[1].

While there was no consensus about how to define the SA in recent years (Gallagher, 2007), we can for the moment get by with the definition that it is *the feeling that one is the agent performing an action*. Several models do exist to describe the SA on different levels.

As a set of experiments by Wegner et al. (2004) has shown, the SA is even present – albeit only to a degree – in situations where the actual action is performed by another (partially covert) agent. To summarize, this other actor, labelled *hand helper*, moved her hands according to audio cues both the participant and the *hand helper* heard while the participant remained still. This is also known as the *Helping Hands* scenario.

BCI is an umbrella term for several techniques where "covert mental activity is measured and used directly to control a device such as a wheelchair or a computer" (van Gerven et al., 2009). That means, that while the user performs a mental task her brain activity is measured, analysed in real-time and used as control signal for a device. The device then gives feedback to the user. Control is achieved through the classification of the detected activity and the mapping of this activity to an action. Different versions of BCIs can be distinguished by looking at whether they are intrusive or not (that is if they are placed at the inside or outside of the skull), the measurement technique they use, the classifiers in

[1] During the course of this text you will encounter several sections labelled **in depth** – these are not needed to understand the main points of this thesis but serve to underline or highlight certain points of interest. They were included because I believe them to be interesting for those readers either curious or well versed in the theories surrounding the SA.

place and the devices involved in signal processing. Out of this wide array of techniques this experiment used the electroencephalogram (EEG), which is an extrusive measurement technique. The supposed modus operandi here was *imagined movement*, a form of control where the user imagines the movement of, say, a limb.

What warrants a closer look at the SA in BCI applications is the question if and how *a sense that one is the agent behind an action* can be evoked by the mediated control provided by such an interface. This is particularly interesting since the method of control for most BCI applications is a (more often than not somewhat unrelated) mental task that in turn has effect on the outside world – can such an interface, that requires no bodily action and thus takes away many of the feedback channels that we are normally used to, evoke the sense of being the agent, the actor, behind an act?

The aforementioned research by Wegner et al. (2004) indicated that the perceived agency can shift under certain conditions from another agent onto oneself. Research by Lynn, Berger, Riddle, and Morsella (2010) showed that people can be fooled into believing that they controlled some non-functional BCI. The novelty of the research presented here lies in the fact that it is a transformation of the *Helping Hands* scenario into the BCI realm. Participants had the task to get a virtual hand on a screen to perform one of two gestures, supposedly via a BCI; the movement they saw the virtual hand perform, however, was entirely pre-rendered, leaving the participants with no actual control. A comparable experiment has been run by Perez-Marcos, Slater, and Sanchez-Vives (2009) with the difference that their BCI actually worked and their virtual hand was not on a screen but embedded into a virtual reality environment.

There are two main research questions that will be delved into:

- Is the Sense of Agency higher in a BCI settings than in Wegner et al. (2004)?

I predict a higher SA by comparison, based on the idea that a BCI will be seen more like a tool than like a (co-)actor. The SA will be measured by a questionnaire identical[2] to the ones used by Wegner et al..

[2]Identical regarding the questions that Wegner et al. (2004) used to measure agency; note that the actual questionnaire (see Appendix) looks different.

- Will a manipulation of the timing between action cue and reaction have any effect? Here I predict that a closer timely connection between cue and reaction will evoke a higher Sense of Agency than a scenario with a longer delay between cue and reaction. This prediction is based on similar theories stated in Wegner and Wheatley (1999).

2 Sense of Agency

During the last years two concepts emerged, a pre-reflective as well as a reflective SA (Gallagher, 2012). Several notions for the different types of SA exist, the one best suited to give a rough, intuitive understanding in my eyes is the one by Synofzik, Vosgerau, and Newen (2008): pre-reflective SA can be seen as *Feeling of Agency* (FoA) while reflective SA – upon which one has to, as the name suggests, reflect – is labelled *Judgement of Agency* (JoA). Unless especially noted I will use the terms FoA and JoA while SA will henceforth denote the combination of FoA and JoA. The most prominent models for the FoA are derived from the so called *comparator model* by Frith (Carruthers, 2012, Frith, 2012, Gallagher, 2000); a model for the JoA has been proposed by Wegner and Wheatley (1999). Several publications also try to present a unified model that links the FoA and the JoA into one coherent picture to explain the overarching SA (Gallagher, 2012, Synofzik et al., 2008).

2.1 Judgement of Agency

Beginning at the end of last century the psychologist Daniel M. Wegner published several articles on various aspects and manipulations of the JoA. In their 1999 work Wegner and Wheatley looked for sources of the experience of *conscious will*, a term later deemed synonymous with (aspects of) the SA (see p.21, Table 1). Wegner and Wheatley defined will in terms of Hume, as "nothing but the internal impression we feel and are conscious of when we knowingly give rise to any new motion of our body, or new perception of our mind" (Wegner & Wheatley, 1999, p.580). This impression, according to the authors, arises when three preconditions are met: a thought is perceived as willed when the thought precedes the action at a proper interval (called the *priority principle*), when the thought is compatible with the reaction (*consistency principle*) and when the thought is the only apparent cause of the action (*exclusivity principle*).

The *consistency principle* is especially interesting for, say, motor learning: imagine learning how to operate a computer mouse[3]; at first you are uncertain which of your movements will result in which exact movement of the cursor on screen. You have to double check, maybe looking back and forth between the cursor and your mouse hand. A less technical example would be learning to write. When we begin we normally do not know for sure beforehand which resulting action we should consider consistent and only repetition and time will shape a consistent picture.

In depth: Critics & critiques regarding Wegner. Note that aspects of these publications – alongside a book written by Wegner in 2002, "The Illusion of Conscious Will" – have their critics. To name but a few I recap Carruthers (2010), Andersen (2006) and van Duijn and Ben (2005). This list of critics is in no way intended to be comprehensive.

As Carruthers (2010) pointed out, the model of Wegner and colleagues could not explain cases of young children, who displayed an intact SA "despite not being able to infer that their mental states cause their action." (cf. Carruthers, 2010, p.342) This clashes with the model proposed by Wegner, which states that "the ability to draw inferences about one's own mental states as the cause of one's action" (as cited by Carruthers, 2010, p.341) is needed for a SA.

van Duijn and Ben (2005) list Wegner (2003) as one example alongside other researchers that "have adopted (the experiments of Benjamin Libet) as evidence in favor of the idea that conscious will is merely *an illusion created by the brain*" (van Duijn & Ben, 2005, p.701), while they point out that "criticisms related to Libet's own interpretations of his data have accumulated, and other, less radical interpretations have been suggested that appear equally compatible with Libet's data" (ibidem, p.700).

Concerning the "illusion of conscious will" argument as a whole van Duijn and Ben, aside from highlighting the "non-persuasive empirical evidence" (ibidem, p.710), argue that Wegner (and others) made a category mistake in assuming a causal relation between neuronal activity

[3] Or: remember your first experiences with a trackball, a (multi-)touchpad, a joystick etc.

and what he calls *conscious will* (JoA, cf. Table 1, p.21). According to van Duijn and Ben saying that neuronal activity causes conscious will would be akin to saying "that H_2O molecules cause water" (van Duijn & Ben, 2005, p.707).

van Duijn and Ben see the reason for this category mistake in the "behaviorist input-output paradigm", which relies on strict causal relations. They argue, that one should adept views from Self-Organization instead, "a process that, given certain boundary conditions, gives rise to increasing order in a particular system by spontaneous synchronization of system parts, without a central executive that helps to set-up this self-organization" (van Duijn & Ben, 2005, p.704). When a differentiation between the microscopic level of neurons and the macroscopic level of mental states is made, one can proceed and see that "global macroscopic mental states are constituted by neurons on the microscopic level, while simultaneously these mental states organize the activity of the individual neurons" (p.708). With that view H_2O does not cause water, but "both water and ice are made of the same micro-level components (H_2O molecules), which lack properties such as liquidity or hardness" (ibidem).

As for the causal relations that Wegner proposes: they are recapped by Andersen (2006). The following is given: first of all Wegner makes two differentiations according to Andersen, the first of which being:

- there is *empirical will* and *phenomenological will*
- observed action/behaviour counts as empirical will
- intentions and the "I did that" feeling count as phenomenological will

The second differentiation runs along those lines:

- there is will as a *feeling* and as a *causal force*
- the feeling of doing something counts as will as a feeling
- the connection of "what the mind has decided to do with the bodily motions needed to do it" counts as causal force

With this given Wegner arguments like this (Andersen, 2006, p.11):

Assumption 1 will should be understood as feeling

Assumption 2 causes are events; causes are not properties of feelings

Conclusion feelings of will cannot be causes

Note that, as Andersen points out, for this conclusion to hold one needs to make another assumption, namely "the implicit premise that feelings cannot be events" (ibidem). This yields a problem – to quote from Andersen

> "This premise is true only if feelings are not physically instantiated. Neural events can be causes. Only if feelings are completely incorporeal, having no neural activity associated with them in any fashion, would it be acceptable to say that feelings cannot be events, and so cannot serve as causes. (. . .) If we disallow this treatment of anything conscious as incorporeal, the conclusion of epiphenomenality cannot be drawn from the evidence." (Andersen, 2006, p.11f)

Andersen argues that Wegner assumes "conscious will is supposed to affect (. . .) the unconscious neural processes involved in action" (ibidem, p.9), in other words: the phenomenological will is supposed to affect the empirical will. Following up on his differentiation between empirical and phenomenological will Wegner looks for separate neural pathways for both, suspecting that – like a little lamp – there would be a neuronal area "that flashes in accompaniment to voluntary action" and that this flash would signal an action to be perceived as consciously willed. However, according to Wegner "no such thing has yet been found, and is unlikely ever to be found" (Andersen, 2006, p.7). Andersen strongly objects the conclusion Wegner drew based on this, that the separate neural pathways for experience of will and for action would mean that one of these would be illusory or causally ineffective. Andersen points to the dorsal and ventral visual streams in perception as objection, where the first "processes the visual guidance of movements (the how)" (Kolb & Whishaw, 2006, p.284) while the latter "processes the visual perception of objects (the what)" (ibidem) – separate neural pathways, none of which illusory or causally ineffective with respect to visual perception.

Let me close this segment with what Andersen said about Wegner: "his problems are artifacts of the causal representations he used, of putting a multi-level, complex causal system into a linear, single-level representation." (Andersen, 2006, p.13)

Now that the underlying theory and the three principles of *consistency*, *priority* and *exclusivity* have been explained we will look at some more research involving Wegner and colleagues on the JoA in later years before the FoA will be discussed in detail.

Wegner, Fuller, and Sparrow tried to discern which factors play a role when associating our own actions either with ourselves or projecting them onto others. They did the latter by letting their participants partake in what is called *Facilitated Movement*, "a popular but discredited technique in which communication-impaired clients are helped at keyboards by facilitators who brace the clients' hands while they type"(Wegner et al., 2003, p.5). The idea behind that technique was that the facilitators/participants would feel very small amounts of movement from the client/confederate and guide their hand accordingly without influencing it. Wegner et al. report that "the simple assumption that (another agent) *could* contribute was sufficient to undermine the (original agent's) own thoughts as causal candidates and instead encourage attribution of the actions to the (other agent)" (Wegner et al., 2003, p.16).

Aarts, Custers, and Wegner found that the priming of effect information – essentially granting pre-knowledge about the effect of a certain (motor) action – enhanced the experienced *authorship*. Authorship is used synonymously by Aarts et al. with what is called JoA here. Their results indicated that "we may experience authorship because the mere thought of the possible effect informs us that the subsequent execution of a motor program may produce the corresponding effect—whether we truly caused it or not" (Aarts et al., 2005, p.454).

Aarts, Wegner, and Dijksterhuis (2006) took a look (again) at the priming of effect information and confirmed their earlier findings; the novelty here lay in other primes they introduced: they tested for self-primes. They did so by subliminal priming of the word "I" or an unrelated word[4], but could only report a significant effect in conditions where no effect prime was given and only for participants that classified as dysphoric[5].

So much for the research into the JoA, let us now take an in depth look at the research that went into the FoA.

[4] Since this experiment took place in the Netherlands they actually primed "ik", Dutch for "I", and used "de" as unrelated term, meaning "the" in Dutch.

[5] From dysphoria, opposite of euphoria.

2.2 Feeling of Agency

In a review of different conceptions of Self the philosopher Shaun Gallagher (2000) took a look at two specific aspects of the enigma that we call Self, the *minimal self* and the *narrative self*. The *minimal self*, according to Gallagher is "phenomenologically (...) a consciousness of oneself as an immediate subject of experience, unextended in time" while the *narrative self* is "a more or less coherent self (or self-image) that is constituted with a past and a future in the various stories that we and others tell about ourselves." (Gallagher, 2000, p.15) The *minimal self* can be decomposed again, two aspects of it being the Sense of Ownership (SO), that is "the sense that I am the one undergoing an experience", and the SA, being "the sense that I am the one that is causing or generating an action" (ibidem). The differentiation between the two concepts of JoA & FoA had not been made at this time, later publications (Gallagher, 2007, 2012) suggest that Gallagher was talking about the FoA, even so he labelled it SA. According to Gallagher these aspects are indistinguishable in everyday cases of willed action but can be distinguished when looking at *involuntary action*.

Imagine being pushed over. In that case you can be certain that it is your body that rapidly approaches the ground (you experience ownership – and pain, should you hit a sufficiently hard surface). But since it was not you that hurled yourself towards the earth but some bystander you experience no FoA for that particular action.

The relationship between the three concepts of Self, FoA and SO according to Gallagher is that both the FoA and the SO are parts of the Sense of Self[6].

One model to describe the FoA is the *comparator model* by Frith, for a recent commentary see Frith (2012). There are a number of different incarnations of the *comparator model*, I will specifically focus on the one seen in Synofzik et al. (2008).

Concerning the model of Synofzik et al. (2008) (Figure 1) it is noteworthy that this

[6] More precise: parts of an aspect of the Sense of Self, the *minimal self*. On a site note: while Gallagher (2000) made a link between the FoA and the notion of the *minimal self*, Synofzik et al. point to another work of Gallagher where he apparently follows up on this train of thought and makes a link between the JoA and the *narrative self*, thus linking two aspects of the Self to the two concepts of the SA.

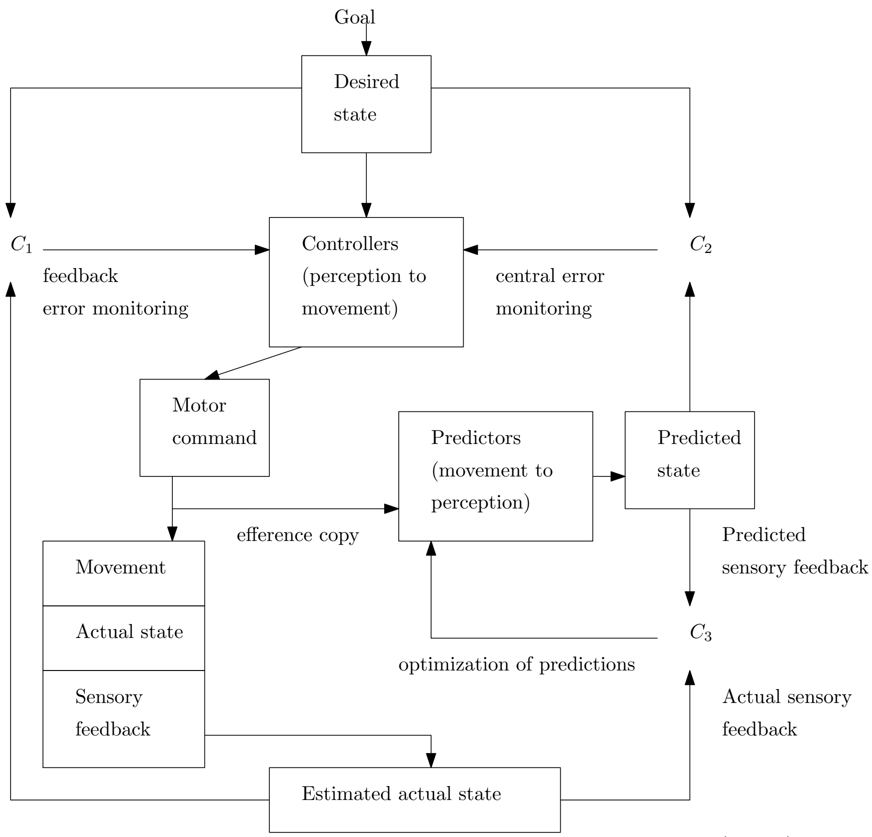


Figure 1. The comparator model as displayed in Synofzik et al. (2008). C_2 and C_3 are responsible for aspects of the FoA according to the corresponding text.

depiction of the *comparator model* actually has two comparators influencing agency – C_2 and C_3 – and three comparators in total while most others only have two. According to Synofzik et al. and their sources C_2 "evokes a sense of being in control" (Synofzik et al., 2008, p.221), given that desired and predicted states match. Comparator C_3 "allows to self-attribute sensory events" (ibidem), given that the predicted sensory feedback matches the actual sensory feedback. One term that needs explanation here is the term *efference*. *Efference* denotes an impulse exciting the central nervous system (CNS) en route to the peripheral nervous system where it will ultimately reach a muscle, a limb or another effector. *Afference* denotes an impulse caused by a stimulus entering the CNS. An *afference* caused by an *efference* via effectors and receptors is called *reafference* (von Holst & Mittelstaedt, 1950). The term *efference copy* thus denotes a – not further established – copy of an impulse to be used internally (as in: within the CNS) for predictive purposes while the original impulse travels on.

With the general models for loose FoA and JoA taken care of, it is now time to consider the unifying models for an overarching SA, springing from a combination of FoA and JoA.

In depth: Comparing the different comparator models. As mentioned in the main part of this thesis the comparator model has seen a number of incarnations. The one featured prominently so far was the one depicted in Figure 1 by Synofzik et al. (2008), but I also want to shed some light on others:

- the movement focused version seen in Gallagher (2000)
- the thought focused version seen ibidem
- the incarnation depicted in Carruthers (2012)

Gallagher presented two incarnations of the so called *comparator model* originally postulated by Frith: one on how both the FoA and the SO tie in with motor actions (Figure 2) and one how the two tie in with thoughts (Figure 3) – but note that recently Frith himself called into question if thoughts can be treated as action (Frith, 2012).

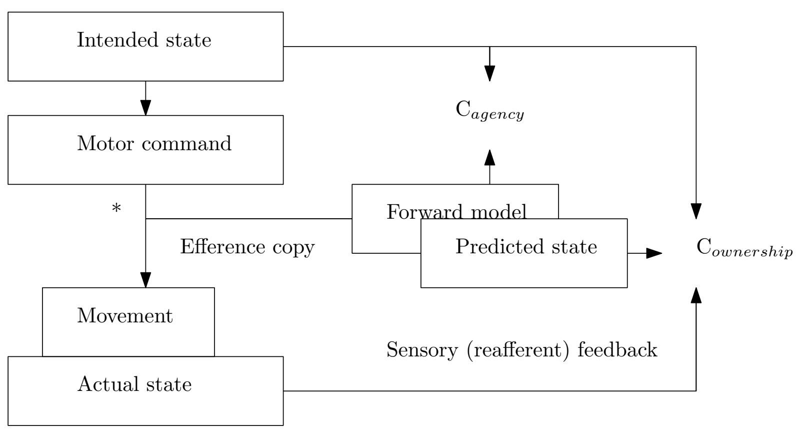


Figure 2. The comparator model for movement as seen in Gallagher (2000). C denotes a comparator, the the subscript indicates the sense that is felt when a match occurs.

In the earlier example where I asked you to imagine being pushed over we already looked into the SA and SO for motor action. To see how FoA and SO are tied in with thoughts I will reuse the train of thought presented by Gallagher (2000): in the case of thought an example for experienced SO but no SA would be "phenomena such as thought insertion, hearing voices, . . . " (ibidem, p.17) as they can be found in, e.g., schizophrenia.

Note the differences between the two incarnations: in the movement focused version Gallagher begins with the intended state, moves on to a motor command, resulting in movement followed by the actual state. In the thought focused version intention leads to thought generation, which in turn leads to what he calls actual stream of consciousness.

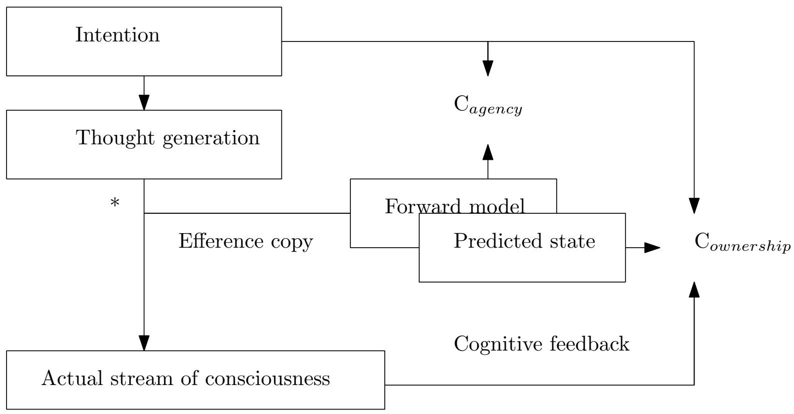


Figure 3. The comparator model for thought as seen in Gallagher (2000). C denotes a comparator, the subscript indicates the sense that is felt when a match occurs.

Pace Gallagher, but I am somewhat uncomfortable with his use of the term *efference copy* within the model for thought since *efference* according to his source denotes "(Impulse), (. . .), die dann – (. . .) – von dort (aus dem ZNS) wieder herauskommen" (von Holst & Mittelstaedt, 1950, p.464, translated to: impulses leaving the central nervous system). The central nervous system is the composed by the brain and the spinal cord (Kolb & Whishaw, 2006, p.6), so if an *efference* is something leaving the CNS, then what place does it have in a model only concerned with thought?

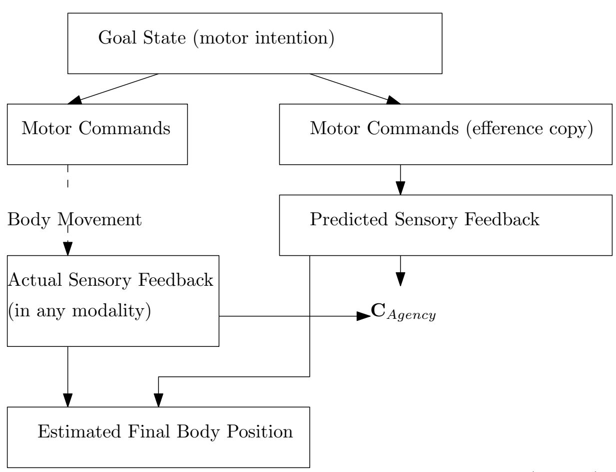


Figure 4. The comparator model as found in Carruthers (2012), extended by some explanatory text fragments taken from the corresponding text.

Another version of the model can be found in Carruthers (2012). Note that this particular incarnation of the *comparator model* (Figure 4) postulates that the FoA comes from a comparison between *actual sensory feedback* and *predicted sensory feedback* with no mentioning of the SO, while the incarnation seen in Gallagher (2000) proposes that agency comes from a comparison between the efference copy and the *intended* state. What Carruthers labels agency rather equals SO in the *comparator model* incarnation seen in Gallagher (2000); cf. Figure 2, page 12.

2.3 Unified Sense of Agency

Synofzik et al. (2008) see the *comparator model* as unfit to explain SA on two grounds: first they reference studies where a SA was felt even though the relevant comparators should have reported a mismatch. They argue that while this inability of the model to explain findings could be explained away "in terms of a bias or insensitivity in the comparator processing" (Synofzik et al., 2008, p.223) the explanatory load it could shoulder would be small and it would limit the contribution of the comparator to SA. On a second ground Synofzik et al. state that pure *motor efference* and *afference* fall short as explanans for particular cases that would need either an extension of the model to support multiple sensory feedback or a generalisation to "unspecific efferent-afferent congruencies" (ibidem, p.224) .

Synofzik et al. underline this by pointing towards neuroimaging studies that strengthen their position. They go on stating that according to them the *comparator model* can be left aside completely for some – not all – cases, which might be explained through congruency between intention and effects arriving at the comparator[7]. As notable exceptions for explanation through simple congruency they name motor learning and perceptual learning.

With that they introduce their *two-step account*, differentiating between two versions of SA, the *Feeling of Agency* and the *Judgement of Agency*, and offer a framework for the interaction between the two (Figure 5). A novel addition of this model was the idea that

[7]While this may appear to be reminiscent of the *consistency principle* it actually is not the same thing. The principle deals with the JoA while the comparator model handles the FoA.

not only the SA is comprised of two parts, but these two also influence each other while being based on different contributors. The contributors named by Synofzik et al. (2008) resemble the ones proposed by the *comparator model* for the FoA; perceptual feed forward and feedback elements. As for the JoA Synofzik et al. may not list the three principles of Wegner and Wheatley, but one can reason about said principles based upon the contributors named by Synofzik et al..

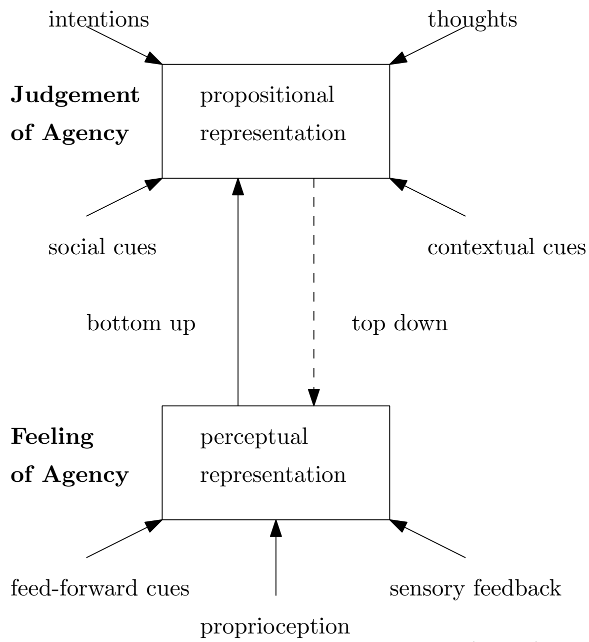


Figure 5. The two-step account as seen in Synofzik et al. (2008).

The ideas that came forth in the two-step account did not come out of thin air, ideas and notions hinting in the same general directions can also be found with Bayne and Pacherie (2007), who sketch an integrated model, as well as with Gallagher (2007). When discussing "possibilities for explaining the pathological loss of the sense of agency" (p.355f.) Gallagher rejected pure top-down, bottom-up or intentional theories in favour of the possibility of multiple (integrated) aspects.

Introducing another model to this discussion, Gallagher (2012) proposed a link of the FoA and JoA with the *three stage intentional cascade* approach of Pacherie. I will briefly sum up the main points of that approach based on a more recent work (Pacherie, 2008). As the name suggest there are three (cascading) stages of intentions: distal intentions (D-intentions, also called future / F-intentions at other places (Gallagher, 2012)), proximal or P-intentions and motor or M-intentions, each with their own up- and downstream (Figure 6).

Take note, that "all three levels of intentions coexist, each exerting its own form of control over the action. Yet the relation between the intentions at the three levels is not merely one of coexistence. They form an intentional cascade, with D-intentions causally generating[8] P-intentions and P-intentions causally generating in turn M-intentions" (Pacherie, 2008, p.188).

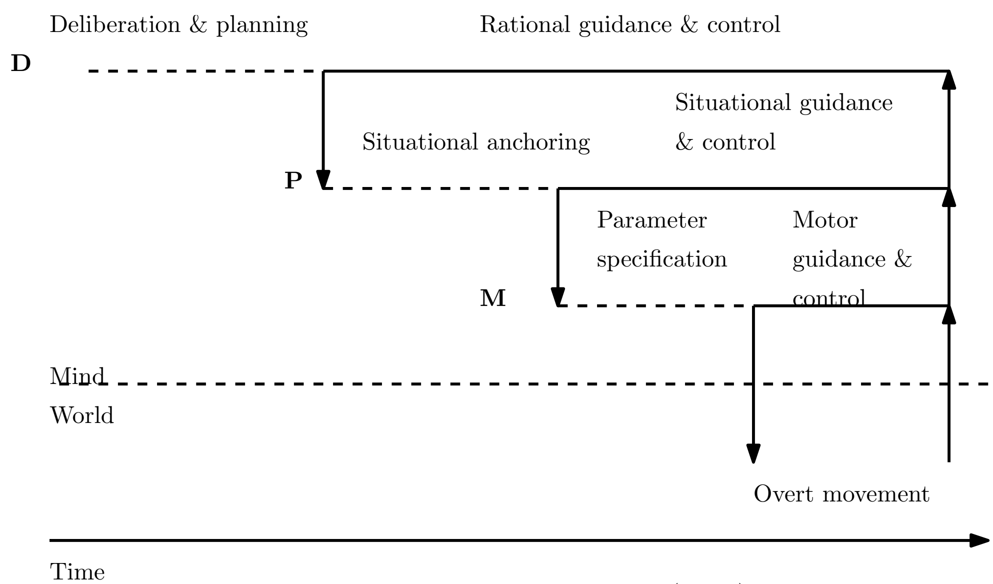


Figure 6. The intentional cascade as depicted in Pacherie (2008)

While Figure 6 only gives a rough idea, to grasp the full extent one can replace every level of intentions with one slightly modified instance of the *comparator model*[9], the relations in all instances stay the same, it is only that the nodes represent different concepts in each instance. For a look at this more elaborate version please cf. Figure 7.

What Gallagher now went on to do is suggesting that JoA can be generated *prospectively* "and that as such it correlates with what Pacherie calls D-intentions[10] and in some cases, with P-intentions" (Gallagher, 2012, p.19). He continued by pointing out that even though

[8]In case the section on the critic on Wegner has been read: I actually wonder what Andersen would have to say on this notion of causality.

[9]The incarnation that can be seen in, e.g., Synofzik et al. (2008), cf. Figure 1 here.

[10]Originally called F-intention by Gallagher here.

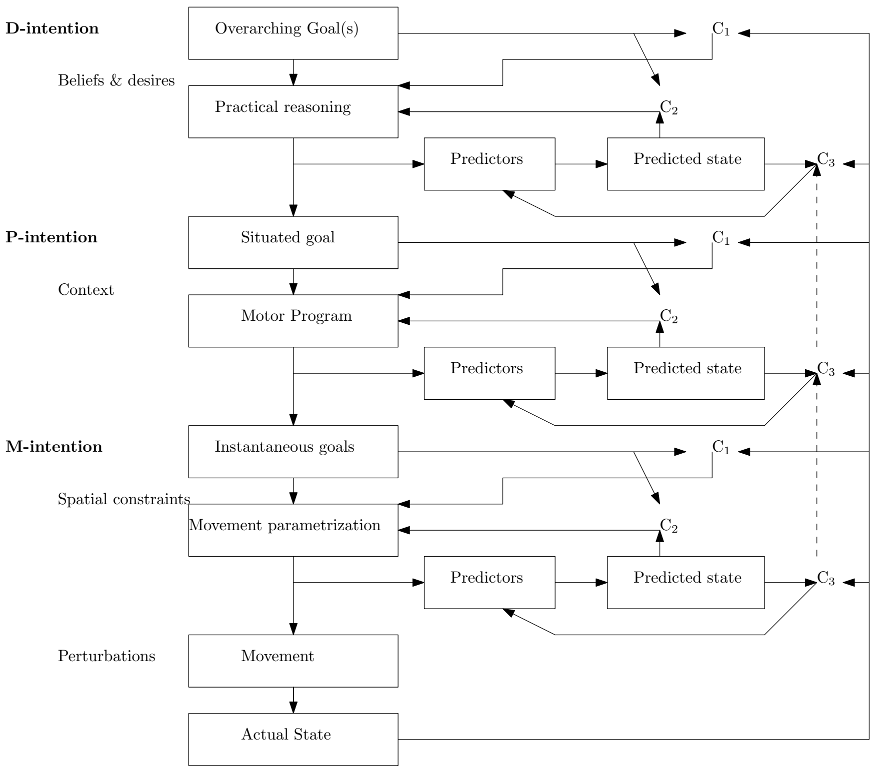


Figure 7. The complete model of Pacherie's intentional cascade as seen in Pacherie (2008) with comparator indications akin to the notions of Synofzik et al. (2008).

the two intentions add something to the JoA they are "neither a necessary nor a sufficient condition" (ibidem) for the JoA. Where he disagreed with Pacherie is on the point that P-intentions would be directly inherited from D-intentions. He suggested that P-intentions would not necessarily need a foregoing D-intention to exist.

In another argument it is suggested that it is even possible – in cases of quasi-automated behaviour for instance – to have D-intentions and M-intentions but no connecting P-intentions; as an example of which Gallagher mentions driving to work in a car: there is the D-intention to take the car rather than going by train, but the driving itself is seen as M-intentions. Gallagher lists the functions of P-intentions from Pacherie here – initiating, unfolding, guiding and monitoring. He agrees with her on the idea that "some of these, or some parts of these functions, may be taken over by M-intentions" (Gallagher, 2012, p.21), possibly explaining how one can drive (safely) from A to B without killing or injuring countless beings in the process if there would be no initiation, guidance or monitoring of the sensorimotor representations that involve actions like avoiding things or hitting the breaks. For cases that involve no D-intentions (and thus only consists of P- and M-intentions) Gallagher brings up examples that he himself believes to stem from habitual practice[11]. Over the course of his paper Gallagher pointed out several suggestions that would add an arrow or two to the the *intentional cascade* model such that D-intentions can influence M-intentions (seemingly without having to take the route along P-intentions) and he concludes that "we can identify at least five different contributors to the sense of agency" (ibidem, p.29). As these five he lists the following:

- Basic efferent motor-control processes that generate a first aspect the FoA
- Pre-reflective perceptual monitoring, making up a second aspect of the FoA
- Forming D-intentions, contributing to the JoA
- Conscious action monitoring (P-intentions), contributing to a more specific JoA
- Retrospective attribution, contributing to or reinforcing JoA

The novel aspect of this five-item list lies in its third item; this *prospective* notion for a contribution is – to the extend of my knowledge – not postulated that explicitly anywhere earlier. One could argue that it has been hinted at by other authors, the priming of effect information by Aarts et al. (2005) comes to mind.

In depth: Critics & critiques regarding Pacherie. Like I have done with the theories of Wegner I will now present to you a recent critique on the *intentional cascade* model by Pacherie, written by Uithol, Burnston, and Haselager (2012). They argue that the notion of *intentions* – "a discrete state that causes an action" (Uithol et al., 2012, Abstract) – is a *folk* one and "deeply incompatible with the dynamic organization of the prefrontal cortex, the agreed upon neural locus of the causation and control of actions" (ibidem). In short, Uithol et al. bring up the following points:

- the folk notion of *intentions* assumes them to be discrete and rather static
- a review of neuroscientific data suggests that the neurol locus is organized dynamically

[11] What I would like to add here is that one could argue the same way for *implementation intentions*: like habitual responses they need some trigger, they are stable over time and they do not need a new D-intention every time they occur.

- the folk notion and the the neuroscientific viewpoint are incompatible

With that as groundwork Uithol et al. suggest important things:

- the term *intentions* should be used only as the folk term that it is
- for (neuro-)scientific debate one should rather talk about *action control.* Note that this does not solve the problem of *thought as a form of action*, brought up earlier here and within Frith (2012).

They close by providing a dynamic model for *action control.* All this is related to the *intentional cascade* in such a way that Uithol et al. provide an alternative model, consistent with recent findings within neuroscience. The *action control* they propose shows functional similarities with the notion of *M-intentions* of Pacherie, while Uithol et al. do not agree with either the *intentional cascade* model, nor the notions of D- or P-intentions.

Another notion that developed over the last few years was the integration of what Gallagher (2011) called *interaction theory* into the SA discussions: after hinting at it in an earlier publication (Gallagher, 2006, Tsakiris, Schütz-Bosbach, & Gallagher, 2007)[12] the idea is made more explicit on a later date. He suggests that the SA and related concepts are "constituted in interaction and in communicative and narrative practices" (Gallagher, 2011, p.69); or, as he said earlier: "just as when two people dance the tango, something dynamic is created that neither one could create on their own" (ibidem, p.65f.), pointing to some research on early (prenatal) development research that suggest that "we are in the tango before we even know it" (ibidem, p.66) due to interaction with the maternal body.

When comparing the *two-step account* of Synofzik et al. (2008) with the *comparator model* Carruthers (2012) came to the conclusion that as long as there is no hypothesis "as to how weights are assigned to each agency cue" in the model of Synofzik et al. "this comes at the price of apparent unfalsifiability"(Carruthers, 2012, p.45). Because of that

[12] "Freely willed action is something accomplished in the world, in situations that motivate embedded reflection, and amongst the things that I reach for and the people that I affect" (Gallagher, 2006, p.123); "(...) the same processes that support a minimal self-awareness of embodied action, contribute to the resonance systems that support our perception and understanding of another person's action" (Tsakiris et al., 2007, p.658)

	Gallagher	Bayne and Pacherie (2007)	Synofzik et al. (2008)	Wegner (2003)
Concept I	SA(1), pre-reflective / first-order SA	agentive experience, comparator based	Feeling of Agency	
Concept II	SA(2), reflective / high order SA	agentive judgement, narrator based	Judgement of Agency	conscious will
Umbrella term		agentive (self-) awareness	Sense of Agency	*

Table 1
*Some of the different terms in use throughout literature for two concepts of SA; * = according to how I read Bayne & Pacherie they hint that the notion of conscious will is also used as umbrella term by Wegner since they relate it with agentive awareness on p.478f.*

Carruthers opts for the *comparator model*, while noting that the model needs to be somewhat changed to explain certain cases. One thing that the model does not yet account for and that seems to be of special interest to Carruthers is that "sometimes representations of actual and predicted sensory feedback in the visual mode seems to matter more than such representations in non-visual modes" (ibidem, p.44). One remarkable example of such a case is watching a ventriloquist; one is normally able to hear that it's not the puppet speaking because the source of the sound does not match with the position of the puppet. But the mouth of the puppet moves while the mouth of the puppeteer seems closed. The situation leaves us with two possible agents: the puppet and the puppeteer. The auditory information points towards the puppeteer, the visual information points towards the puppet. In almost all cases we attribute the role of the speaker to the puppet, based on the visual information.

As a take-home message on the several models of unified SA (Pacherie, 2008, Synofzik et al., 2008) it can be said that Pacherie (2008) (Figure 7, p.18) can be seen as repeated instances of the *comparator model* of Synofzik et al. (2008) (Figure 1, p.10) and that the two-step account of Synofzik et al. (2008) (Figure 5, p.16) suffers from its *magic components*. I use the term *magic* as defined by the *Jargon file*[13] "as yet unexplained, or too complicated to explain". A *magic component*, simply put, thus can be seen as an unexplained component. The *magic components* here primarily surround the weight assignment for the different cues, for Synofzik et al. do not give a hypothesis as to how this is achieved (as Carruthers (2012) points out).

[13] Jargon File 4.4.7, as available on February 6, 2013 via `http://catb.org/~esr/jargon/html/index.html`; for a book version cf. Raymond (1996).

In depth: A possible combination of models. One can go one step further and knit these models into one another, thus combining

- the *two-step account* of Synofzik et al.
- the *three-stage intentions* model by Pacherie
- the *authorship indicators* out of Wegner et al. (2004)
- and the five contributors to JoA & FoA mentioned by Gallagher (2012)

The result is depicted in Figure 8, a combination that to my knowledge has not been made before (at least not explicitly).

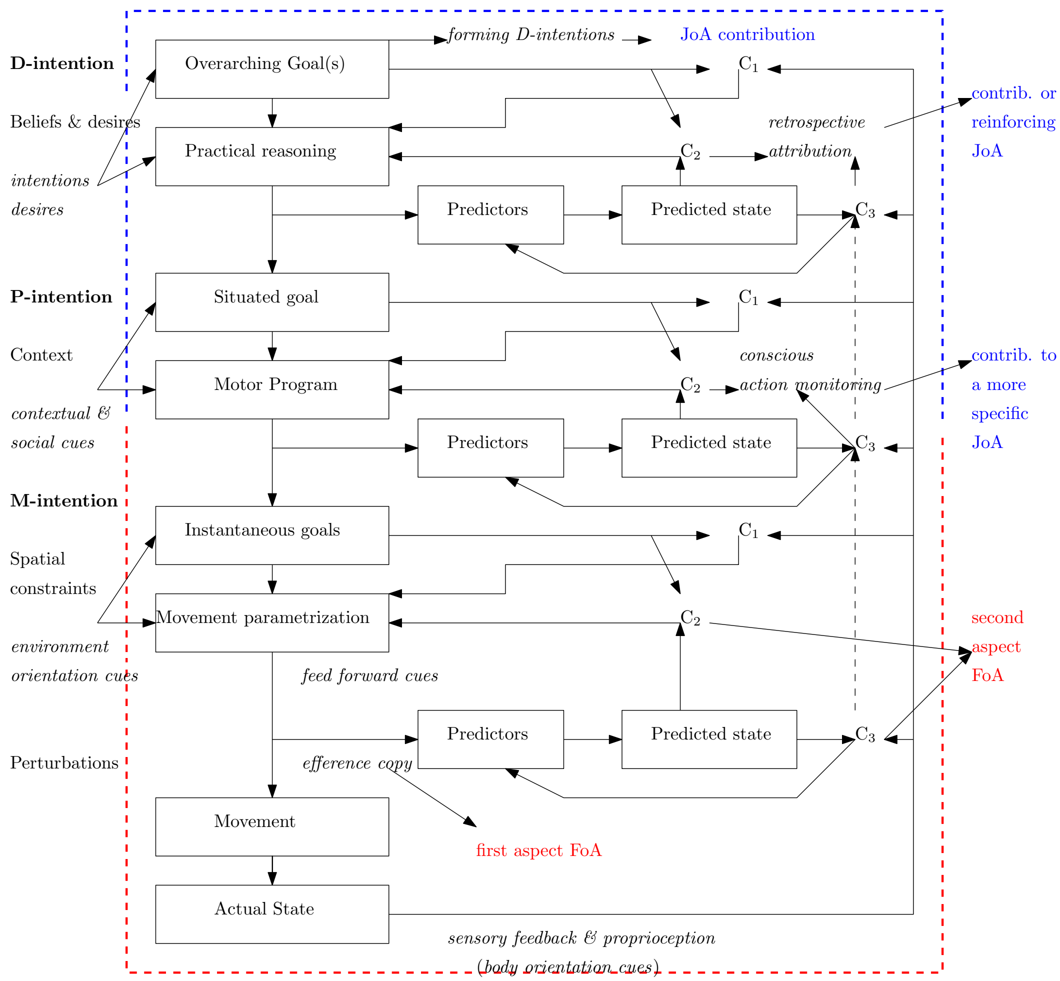


Figure 8. A combination of several models for the SA.

The model of Pacherie (2008) is used as groundwork here. The model of Synofzik et al. (2008) is not immediately obvious. Looking back at it (Figure 5) we see the distinction between

JoA and FoA, indicated in Figure 8 only by the dotted lines surrounding the Pacherie bits. The JoA is indicated by the blue area, while the FoA is depicted in red. The reason for me to color P-intentions ambiguously lies in the reasoning of Gallagher (2012): on the one hand P-intentions can be omitted in certain cases that require only D- and M-intentions (the driving example), on the other hand he suggests that some P-intentions would not necessarily need a foregoing D-intention to exist. The bottom-up route of Synofzik et al. is akin to the rightmost line of arrows in Figure 8, the top-down route can be found in the leftmost line of arrows.

The *contributors* suggested by Gallagher are depicted in red and blue for better visibility and consistent with the aspect of the SA they contribute to.

The *authorship indicators* listed by Wegner et al. (2004) have only partially been integrated, both *action consequences* and *action relevant thought* are missing in my attempt for I was unsure where to place them.

3 Brain-Computer Interfaces

BCIs can employ different measurements techniques, the key differences being the temporal and spatial resolutions of these techniques as well as the distinction between invasive and non-invasive methods. The non-invasive measurement technique used here, the electroencephalogram (EEG), has the advantage of a high temporal resolution while the spatial resolution is rather bad in comparison to other BCI methods. For a comparison of spatial and temporal resolution in different BCI techniques see van Gerven et al. (2009).

EEG is suited to detect fast changes, however it is not optimal to determine the precise origin of the measured changes. One way to control certain types of BCIs is via *imagined movement*, also known as *motor imagery*. This works by measuring the motorcortex for activation, utilizing the fact that imagining a body movement generates brain activity in this area very close to that of actual movement. The practical upshot of this is that *imagined movement* is especially useful for patients with neurodegenerative diseases, missing limbs or in a paralysed condition. Patients can use this method of control – to a certain extent – even if they are unable to actually move the related limb. *Motor imagery* is used to control all sorts of different devices, from spelling devices (Blankertz et al., 2006) to artificial limbs (Neuper, Müller-Putz, Scherer, & Pfurtscheller, 2006).

To transform the brain signals into control commands for a device several steps are necessary. First some preprocessing and feature extraction has to take place "to transform measured brain signals such that the signal-to-noise ratio is maximised" (van Gerven et al., 2009, section 5), the results of which will be fed into a classifier that tries to match the input data to some form of output command to be send to the device.

3.1 BCI & Agency

The experiment by Lynn et al. (2010) has shown that it is possible to generate illusory intent for BCI applications – participants reported that they deliberately caused the movement of an object on a screen after being tasked to try moving it as often as possible[14] even though the movement they saw during the experiment was completely pre-rendered and allowed for no interaction.

To explain this illusory intent the theories of Wegner and Wheatley (1999) can be used: the object moved after participants allegedly began "emitting the intention of moving the line", in line with the *priority principle*[15]. The object traversed the screen in a way that the participants had been led to expect through their briefing and appeared to do so consistent with their prior knowledge of the "BCI" of Lynn et al., thus the *consistency principle* was satisfied. The participants were the only visible actors, satisfying the *exclusivity principle*.

This last point can be criticised on the grounds that the reactions of a BCI can not only be attributed to only the user but to a number of different agents, impacting the performance of any given BCI: we can assume errors in measurement, preprocessing or feature extraction, leading to faulty predictions, in turn generating wrong output – cf. van Gerven et al. (2009)

[14] To be clear: that is not an imagined movement task. One possible imagined movement version of this would have asked the participants to imagine moving their right hand to move the object to the right and vice versa for the left.

[15] Note that Lynn et al. state that "it is difficult to determine the exact moment at which a given participant begins to doubt the effectiveness of the BCI." Given that they did not measure anything anyway – if they did their paper does not indicate it in any way – it was impossible for their setting to determine whether or not their participants actually embarked on the task given to them or the moment that they stopped doing so.

– we could even point at bad wiring, faulty chipsets or badly programmed software and may end up with BCI-technicians, programmers or even manufacturers as possible guilty parties[16] (Grübler, 2011).

The entire responsibility discussion gets even more material if *intelligent devices* are thrown into the mix. Exemplary ID are "BCI-driven prostheses, exoskeletons or wheelchairs (equipped) with environment-sensing, obstacle-avoidance (or) path-finding capabilities." (Haselager, 2012) Here the actions of the devices are more likely to be outside of the direct control of the user (depending on the actual implementation). A noteworthy example by Haselager goes as follows:

> Imagine Fred in his 200 pound BCI driven semi-intelligent[17], semi-autonomous [18] wheelchair and hitting Ken with serious consequences. Even though Fred may have wanted to drive his wheelchair in a specific direction, his performance of the mental task may not have produced the required brain states, or he may have produced the brain states but the BCI failed. In both cases, (without Fred knowing), the ID might have taken over control implicitly, on the basis of its own interpretation of the situation and/or on the basis of its background assumptions about its impression of Fred's general intentions. So in even in cases where Fred is going where he wanted to go, he may not have done it, although it might feel[19] to him that he did.

I have briefly mentioned that one can theoretically view a BCI as either a (co-)actor or as a tool. To see the BCI as a tool it is sufficient for the the BCI to be reliable, since that basically is the common ground for useful tools: you may use them for one thing and one thing only but they really help you. Think of a hammer. A hammer is a great tool for driving nails into walls and it is very reliable at that. To see the BCI as agent

[16] See Grübler (2011) for a proposed idea on how to overcome this so called responsibility gap by introducing a set of rules for BCI usage.
[17] Parts of it can adapt (for example through *machine learning* algorithms).
[18] Shared control scenario, the exact moments when & where who is in control depend on the implementation.
[19] If that *feeling* is FoA or rather JoA is debatable.

one has to know that many of the underlying algorithms mentioned before (like the one for classification) can be specified in an adaptive manner. By adaptive I mean that their internal parameters may change over time. This means that their output may change, even if the input stays the same, thus yielding another influence within the BCI cycle. If your hammer would randomly shake whenever it *felt* like it and would only once in a while actually get that nail into the wall reliably one would not perceive the hammer as a tool any longer.

Aarts et al. (2005) have shown that priming can enhance perceived authorship[20]; combining this knowledge with the usage of plausible *authorship indicators* (Wegner et al., 2004) should enable a SA for a non-functional BCI scenario. Lynn et al. (2010) *have* shown that a non-functional BCI can generate SA but did only indirectly measure it while Wegner et al. (2004) provided a questionnaire.

The experiment presented here will compare the SA for an EEG based BCI scenario with the SA reported by Wegner et al. (2004) for the *Helping Hands* illusion. This thesis has not been written with the idea in mind of contributing to the hunt for the neural correlates or of SA. Additionally it stands to wonder if such a complex concept like the SA even has one particular locus – for a discussion about a related problem for intentions cf. Uithol et al. (2012) and the previous *in depth* section *Critique & critics regarding Pacherie*, for details about the mentioned link between SA and intentions cf. Gallagher (2012).

Wegner et al. (2004) were asking about a form of the JoA with their questionnaire (cf. the *Vicarious Agency* section in Wegner et al. (2004)); in the illusion to intend research by Lynn et al. (2010) the experimenters did not directly inquire about any agency but they point towards Wegner's research, implying that they were looking into JoA as well. Obhi and Hall (2011) state that "real-life agency experience is often quite different from agency as it is studied in the laboratory[21]" (p.663), the latter being the JoA. This does not imply that the FoA has not been looked into by researchers; for an overview about some research

[20]Not very surprising if one keeps the three principles of conscious will in mind. (Wegner, 2003)

[21]The notion of real-life and laboratory agency is one that I do not particularly agree with; the important point the authors make is that most researchers tend to look at the JoA while the FoA is not explored as thoroughly.

done regarding the FoA cf. Gallagher (2012).

3.2 Learning & the Sense of Agency

As mentioned before the mental task and the resulting action are not always what one would consider natural in BCI settings. Granted, there are examples of natural mappings. Imagine some form of control where an object has to move either to the left or to the right. An example of a natural mapping would be to link object movement to the left with imagined movement of ones left foot and vice versa. In other cases of 2-way decision problems one can still employ such imagined movement tasks, but if the resulting actions have nothing to do with right or left or spatial orientation in general then this particular link might need some learning effort before it feels natural. In addition to mapping there is the issue of feedback in mediated tasks, depending on the nature of the tool/medium there is a lot of feedback that does never get to the user.

The special issue of learning involved within *motor imagery* is this: first one has to learn *how* this imagination task actually works. Consider the task to perform imagined up and down movement of your right hand. One first has to learn what exactly it is that one has to imagine. Do you try and imagine seeing just *a* moving hand? Or do you try and imagine actuating related muscles? Do you try to imagine the feeling of a moving hand? Or should you rather try a combination of the aforementioned? Depending on the feedback provided by the BCI this is a tough question to answer. Furthermore one has to learn the relation between the mental task and the resulting action. There is even another learner involved: as mentioned before BCIs can use adaptive algorithms, which are able to adjust their parameters at runtime. While the user thus tries to learn how to operate a novel BCI, parts of the BCI try to learn how to attune as good as possible with their user. To pick up the tango example made by Gallagher (2011) again: while you try to lead your lady through the dance and have to think how to best convey your intentions your lady is constantly trying to learn how you initiate certain steps. So what has happened if you step on your lady's toes after a number of error-free steps?

- Did you try to initiate in a novel way?
- Were you unfocused, unintentionally initiating something?
- Did you mix up the established initiators?
- Were your intentions misread?

The error here can thus lie either with the user or the BCI but – unlike a dance partner – one cannot normally ask the interface what went wrong or who was responsible for the most recent mistake. Granted, it is possible to build systems that provide the user with a constant (and somewhat delayed) error analysis, but this would pretty much be like a pilot not piloting his plane but piloting the cockpit, focused solely on gauges and instruments instead of flying.

As for an example of feedback, consider steering a motorboat: unlike a car, turning the steering device one way does not lead to an immediate change of direction but with practice one can handle a boat. Since a BCI needs more then just a snapshot of data but rather a (small) window of data from time t_0 to t_1 to make accurate predictions BCI applications also have a certain reaction time Δt. Using such applications can feel much like handling a boat. Once one is used to that, however, one can anticipate the delayed reaction and has no trouble linking the reaction to the prior action. In conclusion this leads me to assume that faster feedback may help in improving the learning rate and possibly also the perceived SA as intentions and resulting actions become more consistent.

4 Methods

4.1 Subjects

The experimental data was gathered from 6 students of the Radboud University Nijmegen (3 male, 3 female, 22.3 years of age on average, all but one right handed), who participated for credits relevant to their study where applicable, otherwise for free. Regarding BCI experiments four of the participants reported to be novices while the remaining two reported earlier participation in BCI experiments.

4.2 Procedure

Upon entering the participants where seated in front of a 17" TFT screen, where the EEG cap was fitted. Caps with 64 channels were used; in addition to that we used four electrodes to keep the eye movement under surveillance (EOG, electroocculogram). Two more pairs were used to record muscle activity in their lower arms (EMG, electromyogram) to control for any overt movement. While connecting the cables the participants where asked whether or not they had any experience regarding EEG experiments. Once fully connected the participants received written instructions (cf. Appendix) while the time they spent reading was used to start the needed computer programs. After that their hands were placed close to a button box and both the box and their hands were covered with a towel. The written instructions were left with the participants.

The setup used in this experiment was an Apple Intel iMac computer with the Mac OS X 10.7 operating system. Running on that machine was the Matlab[22] program by Mathworks, version 2010a, incorporating the toolboxes BrainStream[23] and Psychtoolbox[24]. Using these, two small animations were integrated; one being a virtual hand moving from a resting state to the okay gesture and back, the second being a hand moving from a resting state to the thumbs-up gesture and back. See the Appendix for a depiction. These animations were linked to small audio clips of a human voice, saying either the words "thumbs-up" or "okay"[25].

Akin to the consistent and inconsistent preview conditions seen at Wegner et al. (2004) two consistent and two inconsistent films could be created. A consistent film would use matching audio cues and hand movements, while the inconsistent version would use mismatching versions (say, the "okay" audio clip but the "thumbs-up" animation). With modulations in the time between voice onset and movement onset two different conditions –

[22]To my knowledge the code written for this experiment should also run fine under the free software alternative Octave – `http://www.gnu.org/software/octave/`

[23]`http://www.brainstream.nu`

[24]`http://psychtoolbox.org/HomePage`, (Brainard, 1997)

[25]Since the whole experiment was taken in Dutch they actually did not hear "thumbs-up" but "duim omhoog", but for the reader's convenience I will refer to the English terms.

akin to what Wegner et al. (2004) called *early preview* and *preview* – could be done. Some pretesting led us to choose a 5.5 seconds delay for the *early preview* condition and 2.5 seconds for the *normal preview* condition. All in all, consistency ∗ timing ∗ movement, that makes 8 different clips.

Each session for a participant consisted of 60 trials, 30 instances playing an "okay"-clip, 30 playing a "thumbs-up"-clip; 6 of the 60 were error conditions, that is the audio preview and the video gesture did not match, simulating an error on the user and/or BCI side. To proceed to a new trial the buttons on the button box had to be used. The 60 trials per session were displayed in one of two random orders. Every participant had to perform one session in the *early preview* and one in the *normal preview* condition; the order of which was counterbalanced over participants.

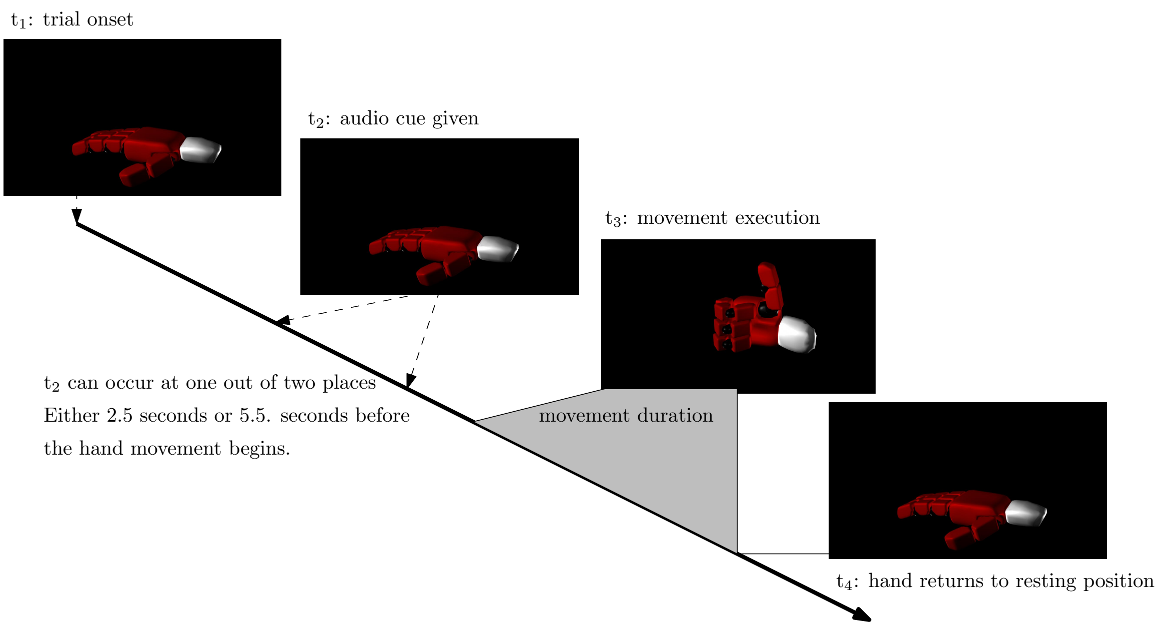


Figure 9. Depiction of a single trial. Trials are initiate by pushing a button on the button box. Note that the arrows are not drawn to scale.

After the participants had read the instructions they were asked whether they understood them or not. Their understanding was put to the test with a one-trial test run after which they were asked which hand they imagined to move. Participants then went through the first of two session, answered the questionnaire and were asked how their session went.

Following up on that they went to a second session, were faced with a questionnaire again and asked how the last session went. As a final act the EEG cap and all electrodes connected to their skin were removed. In the debriefing sessions held immediately afterwards they were first asked if they had noticed anything in particular – to check whether they would call the bluff of control or not – and subsequently were told that the entire BCI part was basically an elaborate illusion and that they had no control about the hand they saw on screen. To see this in a less verbatim way please cf. Figure 10. For the questionnaire used here as well as the questionnaire used by Wegner et al. (2004) please cf. the Appendix; differences and design choices will also be explained there.

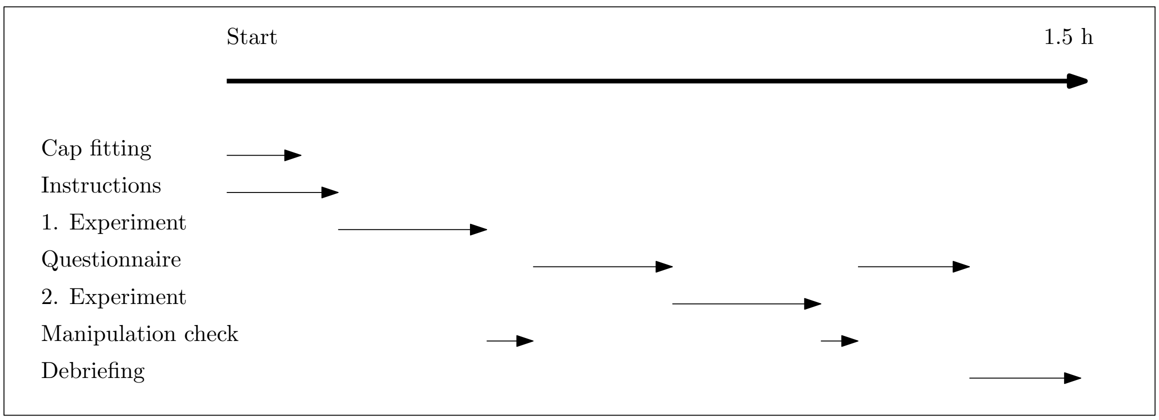


Figure 10. The timeline of the experiment. Note that the arrows are not drawn to scale.

The gathered data has been analysed with SPSS. I ran one-sample t-tests to compare the acquired agency ratings with the ones reported in Wegner et al. (2004). Additionally several paired samples t-tests have been conducted to look for effects of the time manipulation[26].

An initial preprocessing of the EEG data was conducted by PhD student Linsey Roijendijk of the Donders Institute for Brain Cognition and Behavior Nijmegen, using Matlab, extended by the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). As a follow up a preliminary analysis of the EEG data gathered here was done. What one typically observes in imagined movement tasks involving imagined right hand movement and imagined left hand movement is a difference between the right and left hemisphere. This is

[26] We removed all t-test statistics that were reported in the previous version of the thesis, because the assumptions for the tests were not met.

known as lateralization; imagined movement of the left hand is followed by a decrease in the μ- and β-band power (event related decrease, ERD) in the right hemisphere and imagined right hand movement is followed by an ERD in the left hemisphere. The question that were of special interest given the data were:

1. Do the signals of the subjects show the typical lateralization between left and right hand imagined movement in the μ- and β-band?

2. Are there differences in the lateralization between the early and the late timing conditions?

For an in depth description of the procedure please see the Appendix.

5 Results

The questionnaire data gathered in the experiment is depicted in Table 2 below. In the remainder of the text the participants will be addressed as P1 - P6 whenever there is something to be said about one specific participant. In reaction to the debriefing it turned out that four of the six participants reported that they thought to have controlled the BCI. P2 reported to have had slight doubts, based on the fact that almost dozing off seemed not to worsen the performance of the seemingly EEG-controlled hand on screen. After debriefing P5 announced to have seen through the illusion halfway through the first session; however: looking at the data of P5 suggests that this answer was likely motivated by the circumstance that P5 had just been told to have been tricked and may not have wanted to appear like a fool.

To be able to compare findings with the data reported by Wegner et al. (2004) I computed an *Agency* variable the same way Wegner et al. did. I report M = 4.33, SD = 0.875 for the *early preview* condition[27] and M = 5.0, SD = 0.316 for the *normal preview* condition[28]. For details on this process and my reasons for doing so, see the Appendix.

As for the comparison of early preview and normal preview condition in my experiment

[27]Addendum February 6, 2013: In the previous version of this text incorrect numbers were given for the mean and standard deviation of the agency variable in the *early preview* condition.

[28]Addendum February 6, 2013: We removed all t-test statistics that were reported in the previous version of the thesis, because the assumptions for the tests were not met.

I tentatively suggest that there is a trend visible, seeing how every normal preview measurement is as least as high ranked as its early preview counterpart, three of the six are actually higher.

More subjects would be needed to generalise these findings. However, this suggested trend is in line with the results of (Wegner & Wheatley, 1999).

Participant	P1		P2		P3	
Condition	in time	early	early	in time	in time	early
Control	5	5	3	5	5	4
Cons. Will	5	5	5	5	4	5
Agency	5	5	4	5	4.5	4.5
Looks	7	6	2	1	2	3
Feel	6	6	4	4	2	3
Bother	1	1	1	1	2	3
$\text{Growth}_{control}$	5	3	5	5	5	5
$\text{Constant}_{control}$	5	5	3	4	4	3
Brain	5	3	5	4	5	5
EEG	3	3	4	5	2	5
Participant	P4		P5		P6	
Condition	early	in time	in time	early	early	in time
Control	4	5	5	3	6	6
Cons. Will	4	5	5	3	5	5
Agency	4	5	5	3	5.5	5.5
Looks	1	2	4	3	3	2
Feel	2	1	5	3	3	5
Bother	5	2	2	5	1	3
$\text{Growth}_{control}$	4	3	3	4	3	3
$\text{Constant}_{control}$	2	1	3	3	5	3
Brain	3	2	5	4	5	3
EEG	3	4	5	4	4	3

Table 2
Answers given by participants on a 7-point scale; for the questionnaire in use see the Appendix, for the Agency variable (which is not accessed directly via the questionnaire) cf. the text

5.1 EEG data

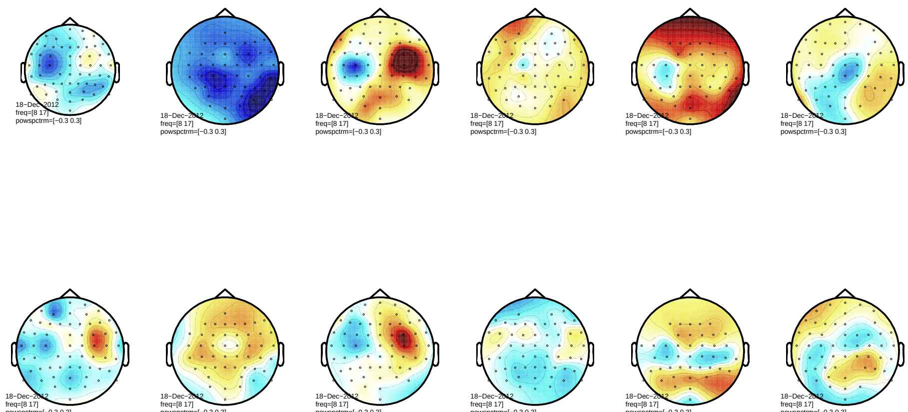


Figure 11. Topographical plots of the alpha modulation of all participants; upper row depicts results from the natural mapping condition, lower depicts unnatural mapping condition data. Frequency range: 8-17 Hz, spectrum [-0.3, 0.3].

Averaged per participants and segmented per timing condition it can be seen in Figure 11 that participants P1 and P3 show a visible power difference between the left and right motor regions.

While not all participants show that difference (at least not with such a magnitude) the average over all participants (Figure 5) show what I would tentatively suggest to be a visible (if not significantly so) difference between the left and right motor areas.

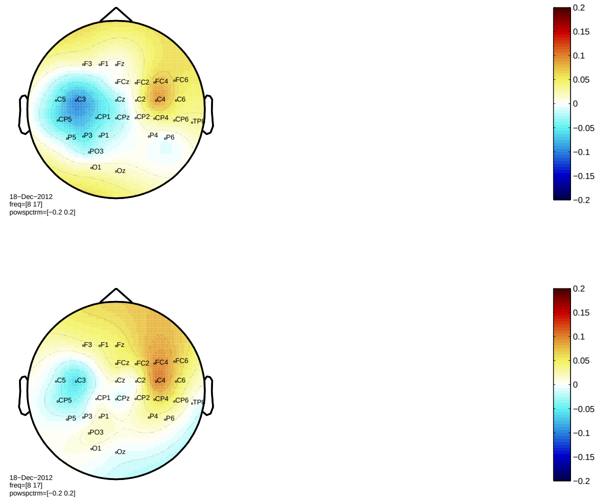


Figure 12. Average alpha modulation over all participants, split by condition; upper row shows the early timing condition, lower row shows the normal timing condition. Frequency range: 8-17 Hz, spectrum [-0.2, 0.2].

This trend continues on in the average over all participants & conditions, depicted in Figure 13, focusing around the electrodes C3 and C4.

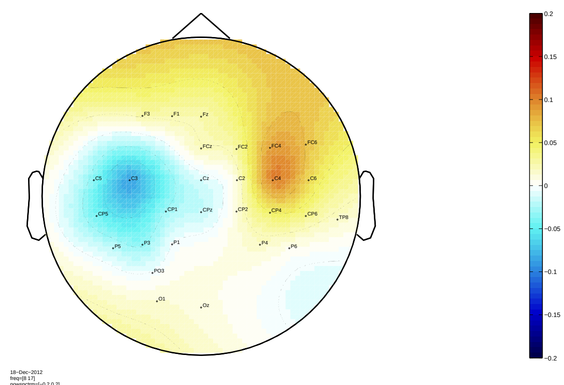


Figure 13. Average power over all participants and conditions, frequency range: 8-17 Hz, spectrum [-0.2, 0.2]

None of the statistical tests yielded significant results[29].

[29]Addendum February 6, 2013: Compared to the previous version single subject results and speculative explanations of brain patterns have been removed.

6 Discussion

Looking at the results we find that the reported SA for our BCI setting is higher than the SA reported by participants in the *Helping Hands* setting. The reasons for this can be sought after at different levels: first of all the setting of Wegner et al. (2004) is more obvious as to who is the agent as there is a person right behind the participant (almost literally breathing down his or her neck). In addition to this we assume that in our daily lives we are (hopefully) not suspecting deception around every corner, suggesting that our non-functional BCI had the benefit of the doubt.

In every setting the hand moved after the audio cue (which in turn served as start signal for the participants to begin with their imagined movement), I argue that the *priority principle* holds. As far as the participants knew they were the only actors involved during each trial, the *exclusivity principle* holds. The basis for this argument lies in the answers the participants gave to the question "Did you feel as if the EEG interpretation improved over time?", where none reported that they felt the capacities of the fictional EEG algorithm improve. While the question may not be phrased optimally[30] I would tentatively suggest that the BCI is perceived as a tool here. I base this suggestion on the following idea: a high rating on the question concerning an improving interpretation by the EEG would mean that participants agree with the notion that the EEG improved. This would hint that they perceive the BCI as (co-)actor.

Another argument in favour of the *exclusivity principle* can be grounded in the answers by participants concerning how they felt about the experiment after each session session: among the answers were phrases like "I think I did something wrong a few times, I really thought about moving my left hand but the virtual hand did the okay sign." Phrases to indicate a perceived co-authorship from the EEG would likely have gone more along the lines of "I thought about X but *it* did Y" where *it* may be replaced by BCI, EEG or a related term of your choosing, attributing some agency to *it*.

[30] This only covers an improving EEG, it is impossible to differentiate between a constant performance and a worsening performance. Even with another question added to cover for that, one would still need to look into the perceived performance of the participants.

As for the *consistency principle* one might argue that it holds here since the mapping from audio cue to imagined movement to the movements of the digital hand stayed consistent most of the time. The strength of this consistency may vary, since we look not at a single intention (from audio cue to movement of the digital hand), but at a double intention: the audio cue should trigger the intention for imagined movement while the imagined movement is done with the intent to cause movement of the virtual hand. I argue that this derived intentions are generally harder to grasp depending on their abstractness. This in turn should make it harder to perceive any consistence, while this difficulty can theoretically be removed by learning.

With all three principles accounted for – even if some might be rather weak – all sources for a JoA according to Wegner and Wheatley (1999) are present. In comparison with Wegner et al. (2004) it seems plausible to say that, by comparison, my setting allowed for stronger *exclusivity*, roughly equal or lower *priority* and – due to the, say, novelty of the task our participants had to perform – somewhat weaker *consistency* than comparable experiments reported by Wegner et al..

To explain the findings reported here with the approach of Wegner and Wheatley (1999) works rather well on the one hand since all three principles are accounted for, but to accept this would also mean that the role of two of the three principles is rather marginal compared to the third. This is an interesting point to be made, since there is – to my knowledge – no explanation or suggestion out there as to how the three principles interact with each other to yield the JoA.

The results of the time manipulation fit with earlier findings reported by Wegner and Wheatley (1999) and Wegner et al. (2004). In these earlier publications it was suggested that the perceived SA peaks if the action and reaction are reasonably far apart – according to theories this perception breaks if action and reaction occur too close to each other, to far apart or if the action occurs after the reaction. They also suggest that SA grows up to the breaking point, so we can assume that – for a BCI – that breaking point is somewhere between immediate response to a cue and 2.5 seconds while an interval of 5.5 seconds is still close enough to contribute to the SA.

6.1 Discussion of the EEG analysis

As you may recall the EEG analysis tried to give answers to two questions:

1. Do the signals of the subjects show the typical lateralization between left and right hand imagined movement in the μ- and β-band?

2. Are there differences in the lateralization between the early and the late timing conditions?

The first question can be seen as improvement regarding the experiment by Lynn et al. (2010), since they did not record any data of this kind during their experiment and could thus merely speculate whether or not their participants did what they were supposedly doing[31]

Looking at the results there are indicators – lateralization patterns – that (some) of the participants did in fact perform what they were asked to do: imagined movement. Due to the low number of participants this can not be seen as a definite answer to the question if a greater SA yields stronger lateralization patterns to more effectively drive motor imagery BCI's.

7 Different models of agency fitted against data

Since there are a number of models that feature the SA, the JoA and the FoA I will now try and see in how far they are able to predict the data acquired in this experiment.

Previously the predictive power of the model of Wegner and Wheatley (1999) has been discussed.

Synofzik et al. (2008) have suggested that certain cases could be explained solely by congruency between intentions and effects. Participants had the *intention* to move the virtual hand. The perceivable *effect* was the – delayed – movement of the virtual hand.

[31] Addendum February 6, 2013: Since "20 percent of participants do not show strong enough motor- related mu-rhythm variations for effective motor imagery BCI" (Roijendijk, 2009, p.9) what one can say based on EEG data is that if one can see lateralization patterns then the participants have done their task. If one cannot see lateralization patterns then one cannot be sure where to attribute this to.

While this sounds like a closed case on first glance, Synofzik et al. mentioned motor learning (and perceptual learning) as explicit exception and I argue that learning that *imagined movement* moves the (novel) virtual hand can be counted as motor learning. Applied to Wegner et al. (2004) congruency falls short as well, there are no intentions to move – given that the participants adhered to the instructions to keep their arms at rest – and still the hands of the *hand helper* move. This means that neither the findings of this thesis nor the findings of Wegner et al. can be explained by congruency. The former cannot be explained because the task falls into one of the categories named as explicit exception for explanations via congruency, the latter cannot be explained via congruency between intention and action for it simply is *not* congruent.

When fitting the *comparator model* against the data it must be kept in mind that the *comparator model* and the model proposed by Wegner deal with different concepts of agency; the first deals with the FoA, the latter with the JoA.

Looking at the *comparator model* as displayed in Synofzik et al. (2008) (Figure 1) we almost immediately encounter a problem: what is our *desired state*? Is it the imagined movement of our own hand or is it the movement of the hand on the screen? Since the experiment is focused on the SA felt for the digital hand on screen I will assume the *desired state* to be a movement of the virtual hand. The *motor command* at work here is another problem. Does *imagined movement* count as movement in the sense that Synofzik et al. or Frith had in mind? One has to realize that (given what we assumed up until know) the *desired state* is a gesture of the *virtual* hand, while the *motor command* a) does relate to ones actual hands and b) is about up- and down-movement instead of, say, making the okay-sign. We thus end up with an *efference copy* (assuming that *imagined movement* does generate those) of the up- and down-movement of our actual hand while the *desired state* is not even about that particular hand.

Remembering Figure 1, with no clear match between the *desired state* and the *predicted state* there is no aspect of the FoA coming from the comparator C_2. The other aspect of the FoA, that according to this model would come from the comparator C_3, needs a match between the *predicted* sensory feedback and the *actual* sensory feedback. The prediction is made based on the aforementioned *efference copy*, but for the *actual sensory* feedback one is (again) left with two options. Either one could argue that the *(estimated) actual state* in the model is about the imagined movement or that it is about the movement of the virtual hand. In line with the earlier assumption I pick the latter, which leads to a mismatch on two grounds: first because any comparison here would require the possibility of multi-modal feedback (that is: comparing actual visual feedback with predicted feedback based on the *efference copy*). Secondly, for these two to *match* one would have to make the assumption that the (in this case illusory) relationship that imaginary hand movement causes a movement of the virtual hand has been learned by participants on a pre-reflective level. What can be seen as given, though, is the ability of a comparator to learn, thus yielding a possible explanation on the level of comparators why we can learn to use tools effectively and why we can cope with non-transparent situations (Beursken, 2012, Carruthers, 2012).

If one were to assume that the incarnation of the model seen in Synofzik et al. (2008) supports multi-modal feedback there is still much that would need additional assumptions. What I am trying to get across here is that the comparators themselves are black boxes in case of multi-modal input. Case in point: the *comparator model* can be used as contributory explanans, but cannot explain the findings to a satisfiable degree in addition to the downside that even the more fruitful incarnations of it do limb because of their *magic components*, parts of their workings are pretty much unaccounted for. The results of Wegner et al. (2004) cannot be explained by any version of the *comparator model*: the participants remained still, thus yielding no motor commands, there is no reason for any sort of agency to be present here according to the *comparator model*[32]. In conclusion the *comparator model* as seen in Synofzik et al. (2008) can explain the findings reported here (at least partially) while any agency found in the experiments of Wegner et al. (2004) cannot be explained.

[32]The only thing that the *comparator model* can explain is the fact that agency ratings where higher when participants were allowed subtle mimicry movements compared to other sessions with no movement or distracting movement.

In depth: Concerning different comparator models. As said before Gallagher (2000) brought up two incarnations of the model, one for thought and one for motor action. Since *imagined movement* pretty much covers both aspects we will first try and see how well the models cited by Gallagher support my findings.

When looking at the model for movement (cf. Figure 2, p.12) one – again, as with the incarnation in Synofzik et al. (2008) – has to realize that (given what we assumed up until now) the *intended state* is a gesture of the *virtual* hand, while the *motor command* a) does relate to ones actual hands and b) is about up- and down movement instead of a gesture. Looking at C_{Agency} we end up with an efference copy of up- and down movement of our actual hand while the intended state is about the virtual hand. Due to the unimodal nature of this model it simply cannot cope with the comparison of different modalities. The comparator for ownership fails as well: the *actual state* might be reached, meaning that the virtual hand moved as we wanted it to, but since this model only allows for motor reafference there is no feedback in this condition that could match the *intended state*. This particular incarnation of the *comparator model* fails to explain the results obtained here. The best we could get out of it would be some FoA for the imagined movement, but none for the movement of the virtual hand.

The variant of the model that works for thought (Figure 3, p.13) is problematic in the sense that is *only* about thought, if anything it can tell us something about the perceived FoA for the imagined movement.

The instance of the *comparator model* presented by Carruthers (2012) (Figure 4, p.14) fares slightly better than the movement related model seen in Gallagher (2000), but only under the assumption that the participants have learned to make the connection between the imagined movement and movement of the virtual hand. On a side note: whether the participants have learned to make this connection or not is of importance neither to the models of Gallagher nor to the one by Synofzik et al. for they cannot handle the comparison of inputs of a different modality anyway.

Given the assumption that participants have made said connection we can proceed like this: with a goal state that is "make the virtual hand perform gesture X" the participant will issue motor commands for up- and down-movement of the related actual hand of hers – presuming that the link between imagined movement and movement of the virtual hand has been made. The *efference copy* is generated based on the up- and down-movement, the predicted sensory feedback (as Carruthers states) is based solely on this copy and thus limited to the realm of motor feedback. Since Carruthers allows for the *actual sensory feedback* to be in any modality, we can assume that the visual feedback (seeing the virtual hand move) does count. Still, to say that the comparison between feedback from different modalities does produce a match to some extend only works if we assume that the link between motor imagery and virtual hand movement has been established and is – in addition – learned to such an extent that it works on a pre-reflective level like the rest of the model.

If one were to try and, for the sake of comparison, predict the findings of Wegner et al. (2004) with these models all incarnations seen in Gallagher (2000) would fail due to their unimodal nature. The visual feedback of the arm of the *hand helper* moving could in no way produce a match for two reasons:

- participants were asked to keep their arms at rest in almost all condition, thus no intention to move would even be generated
- in the *distraction* condition, were participants had to perform unrelated movements upon hearing a cue, the only predicted FoA would be for participants' own arms (again, unimodal argument); the same goes for the *mimicry* condition, were participants were asked to execute the cued movements to a certain extent.

Trying to explain the results of Wegner et al. with the model seen in Carruthers (2012) the only experimental condition out of the several used by Wegner et al. that can be predicted is the *mimicry* setting: there is the motor intention to move ones hand according to the audio cue, based on which we get predicted sensory feedback. That feedback is then compared to actual sensory feedback (in any modality), thus allowing for a comparison between the visual feedback of the hands of the *hand helper* moving and the predicted motor feedback. However, this clashes with one conclusion of Wegner et al., that "the feeling of moving another person's hands, (. . .), was not dependent on participants' ability to move their own hands in the same way." (Wegner et al., 2004, p.845)

In conclusion the comparator model (no matter in which incarnation) can neither predict the results of van Acken nor the results of Wegner et al. (2004). When one extends the model so that it can be multi-modal it is only able to predict the results of van Acken under heavy assumptions regarding learning.

The *two-step account* by Synofzik et al. (2008) merges both the FoA and the JoA into one framework. The FoA, according to this model is evoked by *feed-forward cues*, *proprioception*, *sensory feedback* and the top-down influences from the high-level JoA. As elaborated there is little to no *proprioception* or *feed-forward cues* here to speak of, but there is visual *sensory feedback*, the strength of which I already mentioned before. As for the JoA *intentions, thoughts, social* and *contextual cues* as well as influences from the FoA need to be taken into account. Participants *intent* to move the virtual hand, so this is accounted for; they *think* about moving their own hand. If these thoughts do contribute or not is something that I believe to be linked to the question if the mapping is learned well enough. *Social cues* – assuming that these are meant to be cues provided by the social surroundings – are present due to the *absence* of any other *social agent* around. I assume that the knowledge from the briefing prior to the experiment counts as *contextual cue* here. Applied to the *Helping Hands* setting by Wegner et al. I can find no evidence for anything other than (visual) *sensory feedback* to explain a possibly present FoA. The JoA for the *Helping Hands* cannot be grounded in *intention*, as for *thought* and *contextual cues* both can be considered as primed due to the auditory cues via headphone and are thus accounted for. Looking at *social cues* the movement of the *hand helper* proves to be difficult: on the one hand they do move according to the audio cue, on the other hand they still are the hands of the *hand helper*. To let the so-called *mirror neurons* enter the fray would not help to ease this difficulty either since their contribution is pre-reflective while the JoA is not.

Looking at the *three-stage intentions* by Pacherie (2008) (Figure 6, p. 17) and their interpretation by Gallagher (2012) his five contributors come to mind. We would have to check for "basic efferent motor-control processes" and "pre-reflective perceptual monitoring of the effect of (an) action in the world" (Gallagher, 2012, p.29) to account for a FoA. The aforementioned problems with efference regarding hand-movement apply, but the second aspect of a FoA – seeing the virtual hand move – can be found. JoA can be composed by the formation of D-intentions, P-intentions and retrospective attribution. The D-intention here is the intention to make the virtual hand move, the P-intention is to let the actual hand undergo some imagined movement and in retrospect it is all to easy say "I did this" due to various reasons stated earlier.

The *Helping hands* does presumably have no FoA according to the *intentional cascade* as there are no motor-control processes present in the participants and the user undertakes no action[33], so none can be observed. This theoretical absence FoA contradicts the findings of Wegner et al. (2004), since their report told of ratings with an average above 1 on a scale from 1 to 7. As for a JoA neither a D-intention nor a P-intention can be spotted; the retrospective attribution seems to be absent as well. This means that for the *helping hands* setting not a single one of the contributors proposed by Gallagher (2012) can be found.

	Congruency	$CM_{Synofzik}$	Two-step	Three-stage intentions
Wegner et al. (2004)	✗	✗	✗	✗
van Acken	✗	✗	✓	✓

Table 3
Explanatory power for different models – CM is used as abbreviation for comparator model, Two-step refers to the two-step account by Synofzik et al. (2008), three-stage intentions refers to the intentional cascade by Pacherie as handled by Gallagher (2012); ✓ = the model can explain the data, ✗ = the model cannot explain the data.

8 Conclusion

What comes out of this thesis is the confirmation that it is possible to measure SA in BCI context. Looking back at the research question postulated earlier I can say that the non-functional BCI setting generated a stronger SA than the *Helping Hands* condition of Wegner et al. (2004). While it seems relatively easy to evoke a fake SA an important question remains: would the BCI community want to fake it and under what circumstances could it be beneficial?

The hypothetical benefits of a higher SA in BCI users are manifold: to begin with a higher SA might show itself in a higher signal-to-noise ratio for signals from the brain such that instead of tuning the algorithmic signal detection side we would have a tool in our hands to tune the signal source. Furthermore it stands to reason that a high SA for a BCI

[33]Given that we assume that doing nothing does not count as an action. Even if it does, this still would only yield FoA for inaction, not for the movements of the *hand helper*.

evokes some sense of ownership (not necessarily in an embodied sense but in the sense of owning like one might own a pencil or owning your favourite shirt) and could thus trigger the *mere ownership effect* (Beggan, 1992). That effect states that the attitude towards an object gets more positive if one owns the object compared to the attitude towards the same object if one does not own it. As such the *mere ownership effect* could in turn lead to a more positive evaluation of the BCI by the user. A user that *likes* her BCI might be more forgiving towards errors. In addition positive affect (alongside glucose and resting) is known to replenish mental resources to an extent; training or using a BCI would feel less tiresome for users who view *their* BCI in a positive light.

The recent thesis by Beursken (2012) tried to shed some light on the role of a transparent versus an non-transparent mapping of a mental task to a resulting action and the relation of mapping to a perceived SA. In the experiment conducted by Beursken users had to use a non-functional BCI to control virtual hands on a screen much like the one presented in this thesis. They were told that they could operate it while everything they saw was pre-rendered. What participants did see where two virtual hands which reacted differently depending on the setting:

- in the transparent condition imagined movement of the participant's right hand seemingly caused the right virtual hand to move up and down into resting position, vice versa for the left hand
- in the non-transparent condition imagined movement of the participant's right hand seemingly caused both virtual hands to perform the "okay" sign; imagined movement of the left hand caused both virtual hands to perform the "thumbs up" sign

According to her findings a transparent mapping generates a higher agency rating than a non-transparent mapping. If earlier assumptions about SA (like the proposed link between SA ratings and signal-to-noise ratio) turn out to be true then the work of Beursken strongly suggests that interface designers should go for transparent mapping instead of non-transparent mapping.

In depth: On the results of Beursken. When comparing the agency ratings obtained by Beursken (2012) to the ones reported in this thesis one thing is noteworthy: the average ratings obtained in this thesis are all higher than the one reported by Beursken. The striking point about this is that the setup of this thesis counts as non-transparent. Combined with the findings of Beursken, suggesting that transparent mapping generates a higher SA than non-transparent mapping, this contradicts each other. However, this can be explained when looking at the differences between the two experiments. The main differences are:

- **number of virtual hands** – one hand used here versus two hands used by Beursken (2012)
- **availability of instructions** – the instructions where visible at all times after the initial reading by the participants here while they were taken away from them in the experiment by Beursken
- **setup** during setup participants saw real-time recording of the electrodes (via the so called *buffer view* of Brainstream); this was not the case for Beursken

Due to the number of differences we cannot – in all conscience – point to a single one of them and claim that this particular one carries the most explanatory load. Neither can we rule out interaction effects between the three. What can be given is an estimation of the effect of each difference viewed on its own.

The use of two virtual hands in a scenario that requires imagined movement of the two hands of the participants seems beneficial at first glance for it should feel more transparent and thus should yield a higher SA according to Beursken. If the so called non-transparent two-hands condition of Beursken is generally more or less transparent than the single-hand condition of this thesis is debatable.

As for the availability of the instructions: in cases of doubt participants could easily look them up in the experiment reported in this thesis, thus easing their memory load compared to the non-transparent condition of Beursken and making it arguably more transparent since it was *obvious* at all times which imagined movement would cause which reaction of the virtual hand.

The pre-task showing of the real-time recordings (including suggestions to generate artifacts by, say, raising eyebrows or gnawing teeth) arguably reinforced the participants belief that the

BCI (or at the very least: the EEG) would work. Harking back at Aarts et al. (2005), who found that priming effect-information enhanced the perceived JoA, one can argue that this should strengthen the experienced JoA.

It is noteworthy, that some BCI debates also touch on ethical and legal grounds. To highlight some there is the ethical problem that repeated usage of some BCIs might change a BCI user's *identity* (Nijboer, Clausen, Allison, & Haselager, 2011) as well as legal problems around the notion of voluntary acts and the voluntary or involuntary aspects of BCI-powered devices (Haselager, 2012).

While the illusory SA evoked with the virtual hand casts no notable ethical shadow there are other settings imaginable where the illusion to have caused *anything* might be ethically intolerable. Especially so if recent papers on BCI-powered devices list "military operations using unmanned aerial vehicles for reconnaissance" as possible applications (Solovey, Schermerhorn, Scheutz, andd Sergio Fantini, & Jacob, 2012, p.9). But as soon as the military deviates from reconnaissance such a UAV can be armed and thus harm intentionally or they can – like their unarmed counterparts – simply be involved in accidents. As soon as this becomes reality the notion "if thoughts could kill..." needs to be rewritten because then thoughts might literally kill.

Even if many of these scenarios might not be realizable in the next years or even decades there is no reason not to tackle them. Another point worth discussing is that the sometimes high error rate of BCI applications pared with a strong perceived SA might be undesirable. Going back to the UACV (unmanned aerial combat vehicle) example the following hypothetical scenario can be conjured up: imagine controlling a fully armed UACV. Due to an error of the BCI the UACV fires and claims an unintended victim. While this might be traumatizing enough on its own, imagine being the operator, feeling a strong SA towards the UACV and thus feeling guilty for (not) having shot someone unintentionally.

If future research indicates, that faking the SA in more complex settings is also possible in the BCI world, then not only the paranoid among us should think about a demand for open source code for BCI applications. With that I mean that the source code – and thus the inner workings – of the applications should be open for anybody to review, thus assuring that the application does internally what it externally might only claim to be doing. This should at the very least be demanded in clinical and research backgrounds so that at least the technical experts can always verify that potential users are not just being tricked[34]. Other (not mutually exclusive) solutions for this problem might be a system of certificates, licenses etc., although I personally would not advice it.

The issue whether or not the BCI community would *want* for their applications to have a strong SA cannot be resolved until both the risks and benefits have been clearly assessed.

9 Future Research

The interaction between the three principles of *priority*, *consistency* and *exclusivity* has – as far as I am aware – not been modelled computationally and formally yet, save for some pioneering work by Hindriks, Wiggers, Jonker, and Haselager (2011) with Bayesian networks. It would be nice to see if their model can also predict the findings reported here. As for other theories surrounding SA there seem to be no other computational models around.

For the BCI & neuroscientist community there remains the unanswered questions if from a greater Sense of Agency there comes greater power for your EEG signal or not.

Developing a standard agency questionnaire and checklist for BCI could be a promising step for the community and would also contribute to the development of theories how what type of SA is generated. Such a list could, for example, consist of

- a checklist of authorship indicators
- questions tapping into *intentional binding* (Obhi & Hall, 2011)
- questions about the *proprioceptive drift* (Tsakiris et al., 2007)
- questions on the SO regarding the device

[34] I realize, that my demand for open source brings with it the idea that, aside from just viewing code, anybody is also free to rewrite code. So unless we work out a reliable trust system there is still the danger of someone you trust giving you some rewritten application, but it should be a much smaller danger compared to closed source applications.

- questions on perceived co-actors
- questions on the reflective SA

and could probably be of great benefit. Far be it from me to have read all corresponding literature regarding agency questionnaires, but if I had to suggest some starting points I would like to point towards the questionnaires by Wegner et al. (2004) and Perez-Marcos et al. (2009) as well as Beursken (2012)[35]for inspiration.

Another issue that I would like to raise is, that the SA need not be seen as something that can only *be* or *not be*; I suggest viewing it as a two-component sense. Most models only point towards different contributors for SA and check for presence or absence while I do believe that some models could carry a greater explanatory load if they were to differentiate between strengthening and weakening factors. With such expansions models – that up until now only predict *a* SA for both the experiment reported in this thesis and for the experiments by Wegner et al. (2004) – could differentiate between different strengths of SA.

It is true that other authors have already suggested that the SA is no on/off sense and is something that is gradual, but I suggest to differentiate between positive and negative contributors to get a more refined picture. To elaborate: there is a vast number of possible contributors to the SA to be found throughout literature. Synofzik et al. (2008) names a few and Wegner et al. (2004) offers several *authorship indicators*. Their exact contribution and interaction remains to be researched. The model that I envision sees agency not as a binary rating differentiating between zero and full agency but as the combined result of positive contributions on one side and negative contributions on the other, much like how putting weights on the scales of a balance scale tips its pointer. Since many contributors already – at least partially – cover environmental cues such a model could not also predict the presence or absence of agency but ultimately the target, that is: differentiate between self-agency and the recognition that someone or something else is the agent.

[35]The questionnaire used there was in terms based on the one used in this thesis (which in terms was based on the one of Wegner et al. (2004)), but also contains some noteworthy additions.

Acknowledgements

I wish to thank the following people for their contribution: Jeroen Derks for the creation of the virtual hands, Niklas Weber for several discussions about philosophy & LaTeX and reviewing earlier versions of this text, Philipp Jakubeit and Fabian Urbaniak for reviewing earlier versions of this text, Robin Orthey for general feedback, those among the participants who participated for free, roughly half the BCI crowd among the Artificial Intelligence department for giving me some pointers here and there (with a special mention of Linsey Roijendijk) and last but not least both Rutger Vlek (who really knows where his towels are) and Evine Beursken for their assistance and/or supervision during the actual experiments.

References

Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and Cognition*, *14*, 439–458. doi: 10.1016/j.concog.2004.11.001.

Aarts, H., Wegner, D. M., & Dijksterhuis, A. (2006). On the feeling of doing: Dsyphoria and the implicit modulation of authorship ascription. *Behaviour Research and Therapy*, *44*, 1621–1627. doi: 10.1016/j.brat.2005.12.004.

Andersen, H. (2006). Two causal mistakes in Wegner's Illusion of Conscious Will. In *Philosophy of Science Assoc. 20th Biennial meeting (Vancouver).* Retrieved from http://philsci-archive.pitt.edu/3008/

Bayne, T., & Pacherie, E. (2007). Narrators and comparators: The architecture of agentive-self. *Synthese*, *159*, 475–491. 10.1007/s11229-007-9239-9.

Beggan, J. (1992). On the social nature of nonsocial perception: The mere ownership effect. *Journal of Personality and Social Psychology*, *62*, 229–237. doi: 10.1037/0022-3514.62.2.229.

Beursken, E. (2012). *Transparency in BCI: The effect of the mapping between an imagined movement and the resulting action on a user's sense of agency.* Bachelor thesis, available at http://www.ru.nl/ai/bachelor/voltooide-bachelor/, Radboud University Nijmegen, deparment of Artificial Intelligence.

Blankertz, B., Dornhege, G., Krauledat, M., Schroeder, M., Williamson, J., Murray-Smith, R., & Müller, K.-R. (2006). The Berlin Brain-Computer Interface presents the novel mental typewriter Hex-o-Spell. In *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course 2006.* Retrieved from http://eprints.pascal-network.org/archive/00002255/

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. Retrieved from http://psychtoolbox.org/PTB-3/citations.html

Carruthers, G. (2010). A problem for Wegner and colleagues' model of the sense of agency. *Phenomenology and the Cognitive Sciences*, *9*, 341–357. doi: 10.1007/s11097-010-9150-6.

Carruthers, G. (2012). The case of the comparator model as an explanation of the sense of agency and its breakdowns. *Consciousness and Cognition*, *21*, 30–45. doi: 10.1016/j.concog.2010.08.005.

Frith, C. (2012). Explaining delusions of control: The comparator model 20 years on. *Consciousness and Cognition*, *21*, 52–54. doi: 10.1016/j.concog.2011.06.010.

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for Cognitive Neuroscience. *Trends in Cognitive Sciences*, *4*, 14–21. doi: 10.1016/S1364-6613(99)01417-5.

Gallagher, S. (2006). Does consciousness cause behaviour? An investigation of the nature of volition. In W. Banks, S. Pockett, & S. Gallagher (Eds.), (p. 109-124). Cambridge, MA: MIT Press. Retrieved from http://pegasus.cc.ucf.edu/ gallaghr/gallonline.html

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

Gallagher, S. (2011). Strong interaction and the sense of agency. *Humana-Mente: Journal of Philosophical Studies*, *15*, 55–76. Retrieved from http://pegasus.cc.ucf.edu/ gallaghr/gallonline.html

Gallagher, S. (2012). Multiple aspects in the sense of agency. *New Ideas in Psychology*, *30*, 15–31. doi: 10.1016/j.newideapsych.2010.03.003.

Grübler, G. (2011). Beyond the responsibility gap. Discussion note on responsibility and liability in the use of brain-computer interfaces. *AI & Society*, *26*(4), 377–382. doi: 10.1007/s00146-011-0321-y.

Haselager, P. (2012). Did I do that? Brain-Computer Interfaces and the sense of agency. *In Prep.*

Hindriks, K., Wiggers, P., Jonker, C., & Haselager, P. (2011). *Towards a computational model of the self-attribution of agency* (Vol. 6703).

Kolb, B., & Whishaw, I. Q. (2006). *An introduction to brain and behaviour* (2nd ed.). Worth Publishers.

Lynn, M. T., Berger, C. C., Riddle, T. A., & Morsella, E. (2010). Mind control? Creating illusory intentions through a phony brain-computer interface. *Consciousness and Cognition*, *19*, 1007–1012. doi: 10.1016/j.concog.2010.05.007.

Neuper, C., Müller-Putz, G. R., Scherer, R., & Pfurtscheller, G. (2006). Motor imagery and EEG-based control of spelling devices and neuroprostheses. *Progress in Brain Research*, *159*, 393-409. doi: 10.1016/S0079-6123(06)59025-9.

Nijboer, F., Clausen, J., Allison, B. Z., & Haselager, P. (2011). The Asilomar survey: Stakeholder's opinions on ethical issues related to brain-computer interfacing. *Neuroethics*, 1–38. doi: 10.1007/s12152-011-9132-6.

Obhi, S. S., & Hall, P. (2011). Sense of agency in joint action: Influence of human and computer co-actors. *Experimental Brain Research*, *211*, 663–670. doi: 10.1007/s00221-011-2662-7.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. doi: 10.1155/2011/156869.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*, 179–217. doi: 10.1016/j.cognition.2007.09.003.

Perez-Marcos, D., Slater, M., & Sanchez-Vives, M. V. (2009). Inducing a virtual hand ownership illusion through a brain-computer interface. *NeuroReport*, *20*, 589–594. doi: 10.1097/WNR.0b013e32832a0a2a.

Raymond, E. S. (1996). *The New Hacker's Dictionary* (3rd ed.). MIT Press.

Roijendijk, L. (2009). *Variability and nonstationarity in Brain Computer Interfaces*. Masters thesis, available at http://www.ru.nl/ai/master/completed-master/, Radboud University Nijmegen, deparment of Artificial Intelligence.

Solovey, E. T., Schermerhorn, P., Scheutz, M., andd Sergio Fantini, A. S., & Jacob, R. J. K. (2012). Brainput: Enhancing interactive systems with streaming fNIRS brain input. In *Proceedings of the acm chi 2012 human factors in computing systems conference.* Retrieved from http://people.csail.mit.edu/erinsol/

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator-model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*, 219–239. doi: 10.1016/j.concog.2007.03.010.

Tsakiris, M., Schütz-Bosbach, S., & Gallagher, S. (2007). On agency and body ownership: Phenomenological and neurocognitive reflections. *Consciousness and Cognition*, *16*, 645–660. doi: 10.1016/j.concog.2007.05.012.

Uithol, S., Burnston, D., & Haselager, P. (2012). Why intentions might not be found in the brain.

van Duijn, M., & Ben, S. (2005). On the alleged illusion of conscious will. *Philosophical Psychology*, *18*, 699-714. doi: 10.1080/09515080500355210.

van Gerven, M., Farguhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., … Desain, P. (2009). The brain-computer interface cycle. *Journal of Neural Engineering*, *6*, doi: 10.1088/1741-2560/6/4/041001.

von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip. (Wechselwirkung zwischen Zentralnervensystem und Peripherie.). *Die Naturwissenschaften*, *37*, 464–476. doi:

10.1007/BF00622503.

Wegner, D. M. (2003). The mind's best trick: How we experience conscious will. *Trends in Cognitive Sciences*, *7*, 65–69. doi: 10.1016/S1364-6613(03)00002-0.

Wegner, D. M., Fuller, V. A., & Sparrow, B. (2003). Clever hands: Uncontrolled intelligence in facilitated communication. *Journal of Personality and Social Psychology*, *85*, 5–19. doi: 10.1037/0022-3514.85.1.5.

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology*, *86*, 838–848. doi: 10.1037/0022-3514.86.6.838.

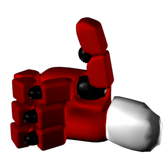
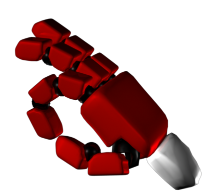
Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54*, 480–492. doi: 10.1037/0003-066X.54.7.480.

Appendix A

*

Introductory text displayed to participants - Dutch

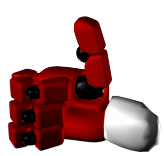
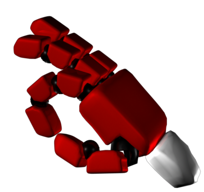
Hallo en welkom! Je gaat zo meteen deelnemen aan een experiment met betrekking tot "brain-computer interfaces", afgekort BCI. Tijdens het experiment vragen we om "voorgestelde bewegingen" te maken met je linker of je rechter hand. Je hoeft je handen hiervoor niet echt te bewegen (!), maar je alleen zo levendig mogelijk voor te stellen hoe je je hand in een regelmatig tempo omhoog en omlaag beweegt, en hoe het voelt als hij zou bewegen. Terwijl je je dit voorstelt kunnen je handen rustig voor je op tafel liggen. Terwijl je deze "voorstelde bewegingen" uitvoert meten we de activiteit van je hersenen met EEG sensoren, en uit de meting detecteren welke handbeweging je je inbeeldt. Op het beeldscherm voor je zie je straks een virtuele hand die reageert op de voorgestelde bewegingen die uit het EEG worden gedetecteerd. Als je je voorstelt dat je je linker hand beweegt maakt de hand op het scherm het gebaar "duim omhoog". Als je je voorstelt dat je je rechter hand beweegt maakt hij het gebaar "okay". Tijdens het experiment krijg je instructies te horen die je vragen om met de hand op het beeldscherm het gebaar "duim omhoog" of "okay" te maken. Om de virtuele hand zo goed mogelijk te besturen vragen we je om direct na het horen van deze instructie te beginnen met het voorstellen van een rechter- of linkerhand beweging en pas te stoppen nadat de hand op het beeldscherm weer helemaal tot rust is gekomen.
Hier zie je een overzicht van wat net beschreven is:

Instructie:	"Duim omhoog"	of	"Okay"
Jouw taak:	voorgesteld bewegen linker hand		voorgesteld bewegen rechter hand
Virtuele hand:			

Introductory text - English translation

Hello and welcome! You are about to participate in an experiment related to "brain-computer interfacing", a.k.a BCI. During the experiment we ask you to perform "imagined movement" with your right or left hand. You do not have to actually move your hands(!), actually you just have to imagine as vividly as possible how your hand moves up and down. While imagining that your hands can be rested on the table. While you are busy "imagining movement" we will measure your brain activity with EEG sensors and detect which movement you were imagining based on the measurements. On the screen before you you will see a virtual hand, reacting to your imagined movement. If you imagine moving your left hand the hand on screen will perform the "thumbs up" gesture. If you imagine moving your right hand the hand on screen will perform the "okay" gesture. During the experiment you will hear instructions asking you to perform the "thumbs up" or "okay" gesture using the hand on screen. To operate the virtual hand as good as possible we ask you to begin imagining a right or left hand movement as soon as you hear an instruction and continue doing so until the hand on screen has reached its resting position again.

Here you see an overview of the things just described:

Instruction:	"Thumbs up"	or	"Okay"
Your task:	imagined movement left hand		imagined movement right hand
Virtual hand:			

Appendix B

*

Example hand images

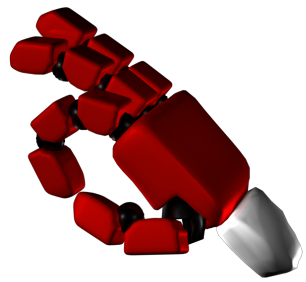


Figure B1. Hand displaying the okay sign

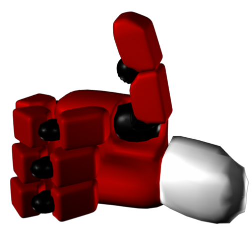


Figure B2. Hand displaying the thumbs-up sign

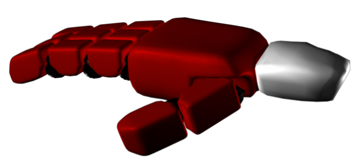


Figure B3. Resting hand as seen between movements

Appendix C

*

Questionnaire - Dutch

- Hoeveel controle heb je ervaren over de bewegingen van de hand?
- In hoeverre had je het gevoel dat je de hand bewust kon bewegen?
- In hoeverre zag de hand eruit alsof het de jouwe was?
- In hoeverre voelde de hand alsof het de jouwe was?
- Stoorde of irriteerde de hand je?
- Had je het gevoel dat de controle die je had toenam gedurende de tijd?
- Heb je een constant niveau van controle ervaren?
- Had je het gevoel dat je vermogen om zinvolle hersenensignalen te genereren toenam gedurende de tijd?
- Had je het gevoel dat de interpretaties van het EEG gedurende de tijd beter werden?

Questionnaire - English translation

- How much control did you feel that you had over the hands movement?
- To what degree did you feel you were consciously willing the hand to move?
- To what degree did the hand look like it belonged to you?
- To what degree did the hand feel like it belonged to you?
- Did the hand bother or annoy you?
- Did you feel an increase over time of the control you had?
- Did you feel a constant level of control?
- Did you feel as if your skill in generating meaningful brain signals increased over time?
- Did you feel as if the EEG interpretation improved over time?

Questionnaire from Wegner et al. (2004)

- To what degree did you feel you could anticipate the movements of the arms?
- How much control did you feel that you had over the arms' movements?
- To what degree did you feel you were consciously willing the arms to move?
- To what degree did the arms look like it belonged to you?
- To what degree did the arms feel like it belonged to you?
- Did the arms bother or annoy you?
- Did it feel as though you consciously caused the arms' movements?[36]

On the design of the questionnaire

Looking at the data by Wegner et al. (2004) we find that (next to the anticipation rating) the most prominent rating seems to be what they called their *vicarious agency* rating. This was computed by taking the means of two ratings[37] Wegner et al. reported to correlate with one another.

Since I used the same two questions – granted, we asked about hands, not arms – and since Wegner et al. did not report the individual findings for these questions I had to do the same and also created a new variable called *agency* in Table 2, p.35. Since the *vicarious agency* rating is the only SA related rating that Wegner et al. did report I chose to use and report my *agency* variable despite the fact that I could not reproduce the correlation reported in Wegner et al. (2004).

Comparing the questionnaire used here and the questionnaire presented in Wegner et al. (2004) readers may notice two things: first that the question "To what degree did you feel you could anticipate the movements of the arms?" is not present here. The reason for this is that I saw it unfit for a task that allegedly had participants in control. Additionally the question "Did it feel as though you consciously caused the arms' movements?", which was used in but one of the three experiments by Wegner et al. (2004) is missing here for

[36] Only present in the third experimental session of Wegner et al. (2004), cf. ibid for details on that

[37] "How much control did you feel that you had over the arms' movements?" and "To what degree did you feel you were consciously willing the arms to move?"

the simple reason that it had escaped my notice.

The reason that only questions related to Wegner et al. (2004) were in use instead of also incorporating questions from, say, Perez-Marcos et al. (2009) is that I became aware of the work of Perez-Marcos et al. only after the experiments had been run.

On the EEG analysis

The analysis was done using the Matlab program by Mathworks and the Fieldtrip toolbox (Oostenveld et al., 2011). After preprocessing steps[38] âĂŞ involving manual removal of artefacts and noisy channels âĂŞ both data sets went through the following processes per person: apply a baseline correction, remove the linear trend from the data and apply a common average reference (CAR). Next, the power in the frequencies of interest was calculated using Fast Fourier transformations with discrete prolate spheroidal sequences (dpss) focusing on 14 Hz with a spectral smoothing of 7 Hz, thus encompassing a range of 7-21 Hz. The time of interest was chosen to be the range from 2.5 seconds prior to the movement of the on-screen hand until the movement onset. In the case of van Acken, where the cue was given 5.5 seconds prior to hand movement in the early condition and 2.5 seconds prior in the normal condition, the 2.5 second interval was chosen to have segments of equal length for both conditions. Averages per subject and over subjects were calculated in each channel. Furthermore we calculated the normalized alpha modulation per subject and condition with the following formula: $P_M = (P_L - P_R) / (P_L + P_R)$, with P_L being the average power over trials where participants had to perform left-hand imagined movement and P_R being the average power of trials with left-hand imagined movement. We visualized our results with topographical plots.

Grand averages and further statistical testing. To answer the first research question regarding the EEG data we computed the average powerspectrum over all subjects; this was done for the left-hand and right-hand subsets (using the Fieldtrip function ft_freqgranaverage). Both results and their normalization were plotted. They were tested for statistically sig-

[38] conducted by L. Roijendijk

nificant differences with a dependant sample t-test via a Monte Carlo estimate based on clustering the limited 8-17 Hz band (cf. ft_freqstatistics; precise configurations included in the next Appendix segment).

Next the grand averages of the subsets of left-hand movement of condition 1 data and grand averages of right-hand movement of condition 1 data were normalized, the same was done, mutatis mutandis, for condition 2 data. Those were then run through the same statistical test mentioned before, yielding insight into the second research question (differences in lateralization when comparing the timing conditions).

The configuration settings for the statistical test done for the EEG analysis

This is a list of parameter settings used for all three statistical sets:

cfg.channel = 'EEG';

cfg.latency = 'all';

cfg.frequency = [7 21];

cfg.method = 'montecarlo';

cfg.statistic = 'depsamplesT';

cfg.correctm = 'cluster';

cfg.clusteralpha = 0.05;

cfg.clusterstatistic = 'maxsum';

cfg.minnbchan = 2;

cfg.tail = 0;

cfg.clustertail = 0;

cfg.alpha = 0.05;

cfg.numrandomization = 500;

cfg_neighb.method = 'distance';

cfg_neighb.layout = 'biosemi64.lay';

cfg_neighb *is then to be fed into the* ft_prepare_neighbours *function, yielding the content for* cfg.neighbours.