Master's thesis on Brain and Cognition

Universitat Pompeu Fabra

Sense of agency for motor and mental actions

Edmundo Lopez Sola

Supervisor: Ruben Moreno Bote

Co-supervisor: Xerxes Arsiwalla

24th August 2020

Universitat Pompeu Fabra Barcelona

Abstract

Research in the past decade has converged on the idea that our experience of control over our actions and their consequences, i.e. the *sense of agency*, arises from the integration of multiple sources of information at different levels. The specific contribution of each of these cues remains unclear, and elucidating the mechanisms behind the construction of the sense of agency is of critical relevance for a better understanding of pathological disruptions of agency, as in schizophrenia, and for the attribution of moral and legal responsibility. In this study, we investigated the specific contribution of motor cues to the experience of agency, and we assessed whether a solid sense of agency can emerge from non-motor mental actions. The sense of agency was measured at the low perceptual level through the 'intentional binding' paradigm, a phenomenon correlated with self-agency; and at the high conceptual level through explicit reports of agency. We used a fake action-effect paradigm, where participants were led to think that their motor or mental actions could trigger a sound. In reality, there was no causal relationship between the participants' actions and the sensory events, but we successfully created an illusion of causality in a meaningful number of participants. We found that the sense of agency emerging at the conceptual level was of equal strength for non-motor mental actions than for motor actions. At the perceptual level, mental actions produced intentional binding in a similar manner than motor actions, with the exception of short action-effect delays. Furthermore, we found a consistent relationship between the explicit reports of agency and the intentional binding effect, which may begin to shed light on the mechanisms behind this phenomenon. The results of this study provide novel insight to the specific role of motor signals and intentional cues in the construction of the sense of agency, and open the door to future studies investigating pathological experiences of control over mental actions. The outcomes of the experiment are also relevant to the field of brain-computer interfaces, suggesting that a solid experience of agency may arise from the interaction with such interfaces.

Keywords— sense of agency, intentional binding, brain-computer interface, mental action, motor action

Contents

Abbreviations

BCI Brain-computer interface

CM Comparator model

EEG Electroencephalogram

FoA Feeling of agency

IB Intentional binding

JoA Judgement of agency

SoA Sense of agency

1 Introduction

"Sense of agency" (SoA) refers to the experience of controlling our actions and their consequences. Whenever we walk, type, or perform any voluntary action, we have an inner feeling of being the authors of those actions. How our mind constructs the sense of control over our actions has been the subject of a large body of research in the last twenty years, and is a relevant issue in cognitive science for many distinct reasons [1].

First, although we might take this experience of control for granted, the sense of agency can be dysfunctional under several pathological conditions. Patients with certain types of schizophrenia display what are called "delusions of control", a situation where patients report that some of their actions are not under their control (see [2] for an extensive review of abnormal experiences of agency in schizophrenia). Other pathologies, such as Tourette's syndrome or the alien hand syndrome present similar disruptions of the SoA [3, 4]. Elucidating the mechanisms underlying the construction of the SoA is crucial for a better understanding of these disorders.

Second, the sense of agency is a key factor in the attribution of responsibility, to ourselves and to others. Moral responsibility is a basic ingredient in the social apparatus of most cultures, and is one of the pillars of current legal systems. Illustrative examples of the relevance of the sense of agency from this point of view are the cases where a disrupted sense of agency (due to pathologies, during sleep-walking, or in the case of reflex actions) leads to diminished responsibility from the moral and legal perspective. Therefore, investigating the necessary conditions for individuals to feel control over their actions and their consequences might lead to relevant changes in our social attitudes and the legal system [5, 6].

Responsibility is tightly linked with the free will debate. Many researchers have avoided this philosophical discussion, mainly for the reason that whether or not there is free will, the experience of being in control *exists* without any doubt, and this experience can be scientifically studied. One could argue that, at the very least, the (perhaps mistaken) intuitive feeling of free will must be closely linked to the fact that we feel in control of our voluntary actions: who would believe in free will if no experience of agency was felt at all?[1]. A deeper scientific understanding of the experience of agency could shed light to the everlasting free will problem, or at least to the reasons why we feel that we are in control of our actions.

Different accounts have been put forward to explain how the sense of agency is constructed. One of the earliest theories, the "Theory of apparent mental causation", was proposed by D. Wegner [8], who suggested that our feeling of control over voluntary actions is based on the relationship between our intention to act and the action itself. Based on principles similar to Hume's theory of causality [9], the sense of agency would arise most strongly when our intention

[1] Indeed, some studies have shown a strong relationship between SoA and beliefs in free will [7]

is *prior* to the action, *consistent* with the action, and is the only potential cause of the action (is *exclusive*). Under Wegner's account, the contribution of the motor system is not specially relevant for the SoA.

Another category of theories does underline the role of motor signals in the construction of the SoA. These accounts are based on the comparator model (CM) for sensorimotor control, according to which our motor system is able to predict the outcome of a given motor action by sending an "efference copy" of the motor command through a predictive forward model. This results in a prediction of the sensory outcome of a given action, which can be compared with the actual sensory feedback [10, 11]. This predictive mechanisms are thought to be involved in a variety of adaptive functions, such as saccadic movements and the acquisition of motor skills [12]. Regarding the experience of agency, the CM theory suggests that actions or external events are self-attributed if the outcomes predicted by the forward motor models match the actual sensory feedback, and that no sense of agency arises in the case of a mismatch. In addition, a disrupted comparator would lead to abnormal experiences of control, as in the case of schizophrenia or other pathologies [10]. The existing literature supporting each of these theories is summarized in Appendix A.1.

This dichotomy between models based on postdictive inferential processes (Wegner's theory of apparent mental causation) and models relying on predictive mechanisms (the CM) has been progressively replaced by unifying accounts that consider both types of signals as crucial in the construction of the sense of agency [4, 13–15]. According to these views, the SoA arises from the integration of multiple cues of distinct origins, such as motor signals, sensory feedback, background beliefs and affective cues.

In fact, it has been recently argued that different cues could be integrated at different levels. Synofzik and colleagues [15] proposed distinguishing between a low-level, pre-reflective feeling of agency (FoA), and a high-order, explicit judgment of agency (JoA). The FoA would arise at the sensorimotor level from the interplay of predictive signals (motor predictions, efference copy) and postdictive inferences, and corresponds to the "raw" feeling of being in control of an action. The JoA refers to explicit attributions of authorship ("I did this") and emerges at the conceptual level, mainly based on the low-level FoA but also influenced by contextual cues and high-order beliefs (see Figure 1). Contextual information can also influence the FoA in the form of sensorimotor priors (dashed line in Figure 1), as shown by recent studies [16].

This distinction is particularly important when considering the paradigms developed in the past few years to measure the sense of agency. *Explicit measures* involve the participants' explicit feedback about their control over an action or its effect. These measures reflect the sense of agency at a conceptual level, i.e. the JoA. On the other side, *implicit measures* are based on metrics that are known to correlate with the experience of agency, but that do not require

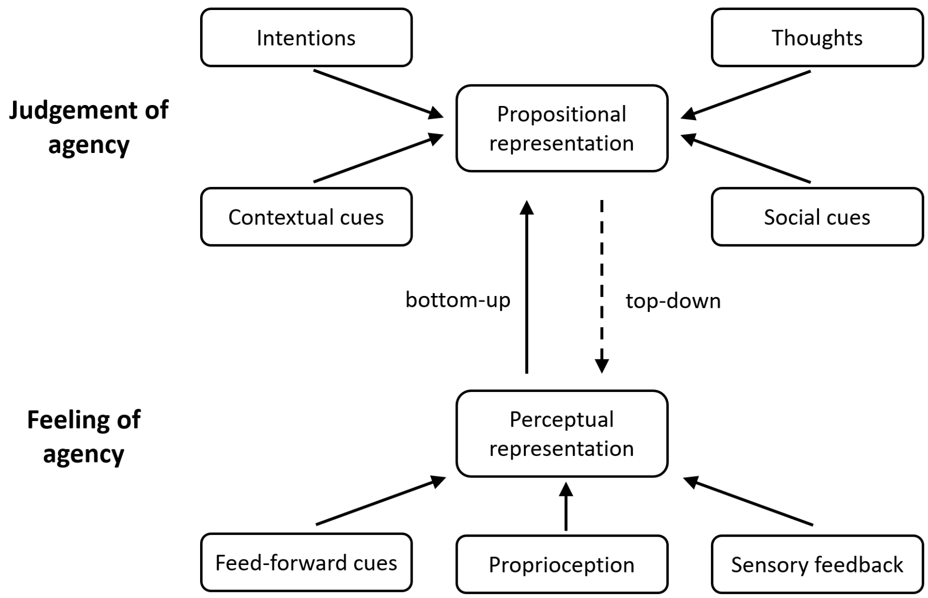


Figure 1: Two-level distinction of the sense of agency and multiple cue integration. Different cues are combined to construct the sense of agency: some are integrated at a low perceptual level (motor predictions, postdictive sensory cues), producing a feeling of agency; while others are combined at a high conceptual level (prior beliefs, intentions), giving rise to judgements of agency. Judgements of agency are largely based on the feeling of agency (bottom-up influence), but high-level cues such as prior beliefs or contextual information may also affect the lower level of agency (top-down effect). Figure adapted from [14].

participants to introspect about whether they caused an action or not. These non-conceptual measures are thought to target specifically the FoA. One of the most common implicit measures of sense of agency is the intentional binding (IB) paradigm. In a seminal study, Haggard and colleagues [17] found that time perception was altered upon execution of voluntary actions: the action, e.g. a key press, was perceived later in time, and its effect, e.g. a sound, was perceived as occurring earlier; so that both events were subjectively perceived to be closer in time (see Figure 2). Crucially, this effect was not found when the action was involuntary (a TMS pulse over the brain motor area was applied to contract the muscle that pressed the key). Waszak and colleagues have proposed a mechanistic account to explain why the intentional binding effect appears for voluntary actions [18]. According to their theory, the sensorimotor prediction of the effect of an action leads to the pre-activation of the perceptual representation of that outcome, which makes that this outcome reaches the threshold of awareness faster in time (thus, producing binding). See Appendix A.2 for a more detailed literature review of sense of agency measures.

There is a general consensus about the fact that the sense of agency depends on the integration of multiple cues at different levels, but the specific contribution of each of these cues to each level is still a matter of debate [4, 15]. As already mentioned, the relevance of motor signals has been emphasized in theories based on the CM, and some authors have suggested that without these

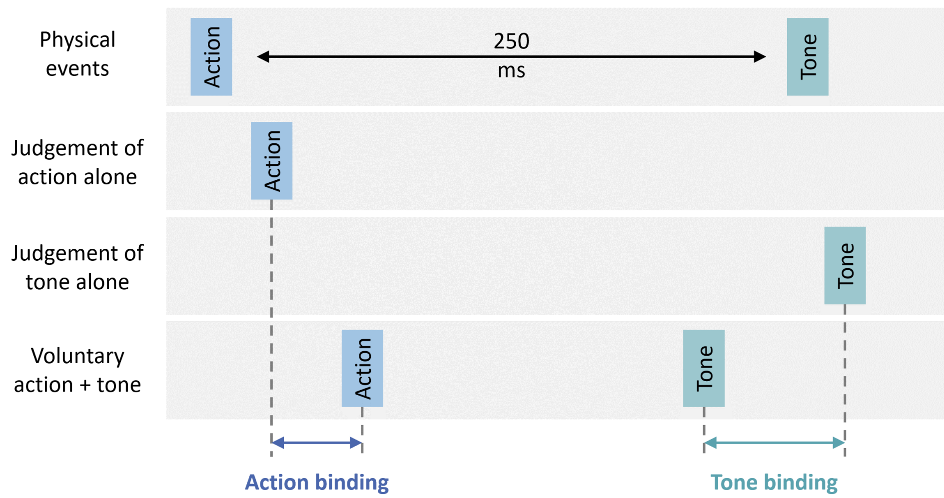


Figure 2: The intentional binding effect. Upon execution of voluntary actions with external sensory outcomes, the action and the outcome are perceived as being closer in time. There is binding over the action: a voluntary action is perceived later in time when followed by a tone than when performed alone or when the action is involuntary. There is also binding over the tone: the tone is perceived earlier in time when produced by a voluntary action than when not caused by the subject's action or when this action is involuntary. Figure adapted from [19].

motor predictions the SoA would be largely reduced [20–22]. However, to date most experimental paradigms do not allow drawing solid conclusions about whether the SoA could emerge or not in the absence of motor cues, because most of them have failed to use an appropriate contrast (non-motor) condition to investigate this issue. The problem with most paradigms comparing the SoA for motor and non-motor actions is that the non-motor contrast condition not only gets rid of motor signals, but also of the participants' intentionality to produce an action (see Appendix A.3 for detailed examples of such cases). Since intentionality is a crucial element in the construction of the SoA, in order to dissociate the concrete role of motor predictions from other mechanisms contributing to the SoA, one would need to design a paradigm where an *intentional* motor action is contrasted with an *intentional* action that does not involve motor signals.

It is intuitive to think that in a situation where an intentional mental event (e.g., thinking about moving an object) would be followed by the desired outcome (the object moved), the match between intention and sensory feedback would lead to the emergence of a sense of agency. Interestingly, the use of Brain-Computer Interfaces (BCI) allows subjects to produce controlled outcomes in the outside world, such as moving a prosthetic arm or controlling a cursor in a computer screen, only by means of mental actions [23]. We refer to "mental actions" as particular thoughts that can trigger outcomes in the external world. Several studies have investigated the SoA arising from the interaction with BCIs [24–28](see Appendix A.4), most of them evidencing that a SoA may emerge from mental actions under appropriate conditions, at least at the conceptual level (JoA) when assessed with explicit reports. This is coherent with the aforementioned idea that voluntary non-motor actions should lead to a subjective experience of control.

At this stage, several questions remain unsolved: first, it is unclear if the SoA emerging from mental actions is comparable to the strong experience of agency originated by motor movements. Second, if mental actions lead to a SoA, as proven by the studies with BCIs, it is not known whether this SoA emerges only at a conceptual level (JoA) from the combination of thoughts, intentions and contextual cues, or if a low-level FoA may also appear at the non-conceptual level in the absence of motor signals. Regarding the latter issue, some authors have recently argued that predictive mechanisms based on counterfactuals and high-order cognitive sources (intentions, beliefs, etc.), different from predictive motor models, could also play a key role in several body-related action-outcomes, including agency attribution [12]. Thus, one could expect that a low-level FoA will emerge from pure mental actions.

In the present study, we investigated these issues by contrasting the SoA arising from motor actions with the one emerging from mental actions. To that purpose, we have used an action-effect paradigm where a given motor or mental action is followed by a sound. In the motor action, participants had to press a key. The mental action condition was based on the concept of BCIs. Critically, in this experiment a “fake” BCI was used: participants were led to believe that one specific thought could be detected by a computer software (see “Materials and Methods”) and subsequently produce a sound. Other studies have succeeded at creating an illusion of control using phony BCIs [25, 26]. The idea is that a causal action-effect link is not necessary for the emergence of the SoA, it is sufficient that the action and its alleged effect are close enough in time and that the subject *believes* there is a causality relationship between both events. Importantly, in order to ensure that both experimental conditions are adequately similar, in the motor action condition there is also no actual causal relationship between the motor action and the sound.

Based on previous studies [22, 29, 30], we manipulated the time interval between the action and the effect. Temporal contiguity between an action and its outcome is a solid authorship indicator, and shorter intervals typically lead to a stronger SoA [30, 31]. Additionally, in order to introduce a certain degree of ambiguity in the attribution of authorship, participants were told that the sound could be produced by their action or by the computer. Other studies have employed similar paradigms to explore the judgements of agency [32]. To clarify at which level the SoA for mental actions is constructed, both explicit measures (authorship reports) and implicit measures (intentional binding) will be used, each of them reflecting the high-level JoA and the low-level FoA respectively.

The main hypothesis of the study is that mental actions can produce a sense of agency, and that this sense of agency will be comparable to that of motor actions. In particular, we hypothesize that the JoA, measured at the conceptual level by explicit reports, will be produced in a similar manner by motor actions and by mental actions. We predict that for shorter action-effect intervals, participants will attribute the sensory effect more frequently to themselves than to the computer; and that this frequency will be comparable for motor and for mental actions. This prediction goes in line with the findings of the BCI studies mentioned above, and disputes the comparator theory of agency, according to which a lower SoA would be reported in the absence of motor cues.

We also hypothesize that a low-level feeling of agency will arise from mental actions, following the idea that the brain can engage in action-outcome predictions without the need of predictive motor mechanisms, but rather based on counterfactuals [12]; and that other mechanisms such as inferential postdictive processes may also contribute to the feeling of agency [15]. This is of special interest for the current research on agency, since, to our knowledge, this is the first study investigating low-level agency in the absence of motor cues. Our prediction is that mental actions will produce intentional binding to an extent similar to that of motor actions. It is worth noting that some studies have suggested that the IB effect only arises when motor cues are present [22], while others have argued that it is a general signature of intentionality, independent of motor commands [33, 34]. Others have even claimed that it reflects the mere experience of causality [35]. The present study will help clarify whether the IB effect is specifically linked to motor actions or not.

Finally, given that a higher sense of agency is expected for short action-effect delays, the prediction would be that the IB effect will be found for short, but not for long intervals, and that this will be the case both for motor and mental actions. However, some studies using the IB paradigm have not found different binding for different time intervals [16], and the relationship between IB and explicit measures of agency is currently unclear [36, 37]. This study will offer a novel insight to the relationship between implicit and explicit measures of agency, both with and without motor cues.

Critically, these predictions are valid only for participants that actually believe that their motor or mental actions are causally linked to the sound (in reality they are not). Therefore, it is crucial for the experiment that as many participants as possible feel that their actions can produce the alleged effect. In this sense, we expect more participants to have causal beliefs for the motor condition than for the mental condition, given the increased difficulty of creating causal beliefs for mental actions (see Section 2.2.2).

2 Materials and methods

It is worth mentioning that the experiment was initially thought to be run in a laboratory environment. However, due to the SARS-CoV-2 pandemic, the experiment was adapted to be performed online. More details will be given in the following paragraphs.

2.1 Participant selection

50 participants (23 male, 22 female, 5 unknown) were recruited through the online platform Prolific (`https://www.prolific.co/`). To ensure that the instructions of the experiment were correctly understood by the participants, only English native speakers were selected.

Seven participants were excluded from the study due to system incompatibilities and one participant was excluded due to poor performance in the tasks (more details are given in Appendix B.2).

2.2 Experimental design

2.2.1 Software tools

The study was designed with the open-source python-based software Psychopy3 [38]. In order to launch the experiment online, the experiment was translated into JavaScript and uploaded to the online platform Pavlovia (`https://pavlovia.org/`). The study URL from Pavlovia was provided to the participants for the execution of the experiment in their own devices. The full python and Javascript codes for the study are available at `https://gitlab.pavlovia.org/edmundolopez/agency_exp`.

2.2.2 "Fake" BCI: a facial motion detection software

For the experiment to be successful, it was crucial that as many participants as possible believed that their mental actions could trigger an external event through the phony BCI. Participants were told that they would participate in an experiment to evaluate the users' interaction with a facial motion detection software. They were told that this software could detect micro-gestures in the faces of the participants, which were recorded with their device webcam. Particularly, the software could recognize involuntary micro-expressions in their faces when thinking about a word related to a face gesture, in this case the word 'Smile'. During the trials, participants had to think about the word 'Smile' at a particular moment and they were told that this thought would produce a micro-gesture in their face that would be detected by the software which would eventually produce a sound. Thus, they would produce a sensory event with a mental action.

Critically, participants were instructed to avoid producing any voluntary face movement during the trial.[2]

2.3 Experimental procedure

Participants were first given a brief introduction to the alleged purpose of the experiment: assessing the users' interaction with a facial motion detection software. They were told that the main interest of the study was the sense of control and the sense of time.

The experiment consisted of three distinct blocks, one for each experimental condition: first, the SoA of the subjects was assessed through explicit and implicit measures in a motor action task; then, the same measures were used to evaluate the SoA under a mental action task; finally, in the last block we measured the baseline error made by participants when estimating the time of the sound, which was necessary to calculate the intentional binding in the two previous conditions.

2.3.1 Motor action condition

Participants were told that in the first block of the experiment their sense of control and time over motor actions would be measured, without using the facial detection software. The motor action condition was divided in two parts: in the first part, participants performed three test trials and twelve real trials, and their sense of agency was measured via explicit report after each trial. The second part had the same number of trials and after each trial the intentional binding was measured.

The design of the trials was inspired by the work of Libet and Haggard, who used for their studies a clock and a dot rotating inside the clock with a period of 2.56 seconds [17, 39]. In the study carried out by Haggard and colleagues, participants had to decide when to perform the action (a key press), and this action triggered a sound after a given interval. After each trial, participants had to report which was the dot position when they performed the action or when they heard the sound. In that manner, the authors could measure the intentional binding produced by voluntary actions. The present experiment used a modified version of this design. One main difference is that participants were told at which specific moment to perform the action. The reason for this was that the experimental design required the subjects' actions not to be causally linked with the sound; by instructing the participants to perform the action at a specific time, the sound could be programmed to occur after that time (thus, after the action), thereby creating an illusion of causality. In brief, we assumed that if participants pressed the key at a given time and the sound

[2]The initial idea was to use an EEG-montage in the laboratory and informing the participants that their EEG-signal when producing a given thought could be decoded to trigger an external event. However, the experiment had to be performed online due to the SARS-CoV-2 pandemic, and this alternative solution was adopted.

appeared right after their action, they would feel that they produced the sound, even if no actual causal relation existed between both events and even if they were "forced" to perform the action at an indicated time.

In each trial, participants were shown a clock and a dot rotating with a period of 2.55 seconds (the period was slightly modified with respect to Libet's version in order to correspond with an integer number of frames, see Appendix B.1). A red cross appeared after 500 ms in a given position of the clock, and remained on screen during 500 ms. Participants were instructed to press the 'space' key when the dot reached the position where the cross had appeared. Crucially, the time when the dot reached that position was always the same (2.55 seconds), so that the sound could be programmed to occur at varying intervals from the moment in which participants performed the action (if following the instructions properly). It is worth mentioning that the cross appeared at a different location each time, so that participants could not habituate to the clock position where the key press would be made. The trial structure is shown in Figure 3.

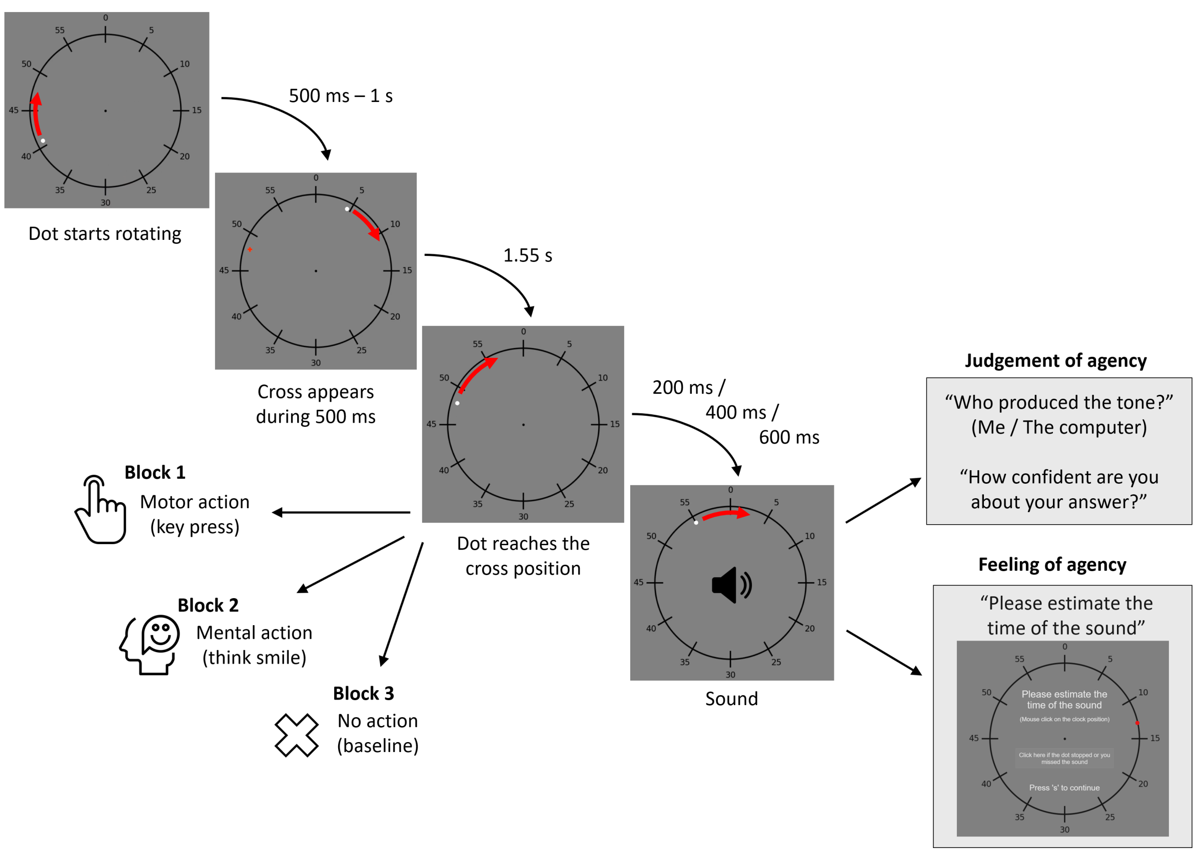


Figure 3: Trial structure. The stimulus presented was identical in all trials of the three blocks of the experiment, with the only difference that in each block participants were instructed to perform different actions at the cued time (the moment when the dot reached the position of the cross).

During the trials, the sound always appeared, independently of whether the 'space' key was pressed or not, and at which moment. The sound appeared at three different intervals from the cued time: 200, 400 and 600 ms. Each interval appeared in total four times, and the order of appearance of each interval was randomized. Similar time intervals have been used in previous work that manipulated the action-effect delay [22, 36]. Before the trials, participants were told that their key press would trigger a sound, but that in some of the trials the sound would be produced by the computer instead; introducing a certain level of ambiguity in the authorship of the sound, as done in other studies [32, 33].

In the first part of this block, the JoA was assessed with explicit measures. Participants were asked after each trial the following question: "Who produced the tone?". They could choose between "Me" and "The computer". After this response, they had to respond in a Likert 1-7 scale to the question "How confident are you about your answer?". The scale had in its extremes the labels "Not at all" and "Very much". In the second part of the motor action condition, the FoA was evaluated. Participants had to indicate after each trial the clock position corresponding to the instant where they heard the sound. This procedure is analogous to the one used by multiple studies to measure the intentional binding effect over sensory outcomes [16, 17].

In our experiment only the tone binding was measured, due to the fact that the action binding cannot be measured in the mental action condition – there is no manner in which the actual time of the thought could be obtained. In any case, it is expected that the execution of voluntary actions (mental or motor) leads to intentional binding for the sound, as suggested by other studies [16, 40]. In fact, it has been proposed that action binding and tone binding are driven by distinct mechanisms, and that they shall be measured separately [40, 41]

Additional instructions were added throughout the experiment to ensure that participants performed the task correctly and to increase the illusion of a causal relationship between their motor action and the outcome. A message appeared when participants did not press the 'space' key or pressed it too early or too late in two consecutive trials, reminding them to perform the key press at the correct time; so that they did not figure out that the sound was unrelated to their action. Furthermore, contrarily to the real trials, in the test trials the sound was not independent of the key press, which increased the feeling of causality (see Appendix B.3 for more details).

2.3.2 Mental action condition

Participants then completed the second block of the experiment, where their SoA for mental actions was assessed. Before the real trials, several instructions about the (fake) facial motion detection software were given to the participants. Again, the purpose of these instructions was to increase the chance that participants believed that the software could detect their thoughts and subsequently produce a sound, thereby creating an illusion of causality between their mental actions and the sensory outcome.

First, participants were told that the software had to be calibrated. They were shown their webcam feed and were instructed to position their face inside a square that was presented over their webcam feed (Figure 4, left). Then, they were asked to produce a real smile for the software to recognize their face gestures. They repeated this process three times. After that, they were told to think about the word 'Smile' without making any voluntary movement to calibrate the micro-gesture recognition (Figure 4, right). Finally, they were told that the software was successfully calibrated, and that it was now able to detect their facial micro-gestures when thinking about the word 'Smile'.

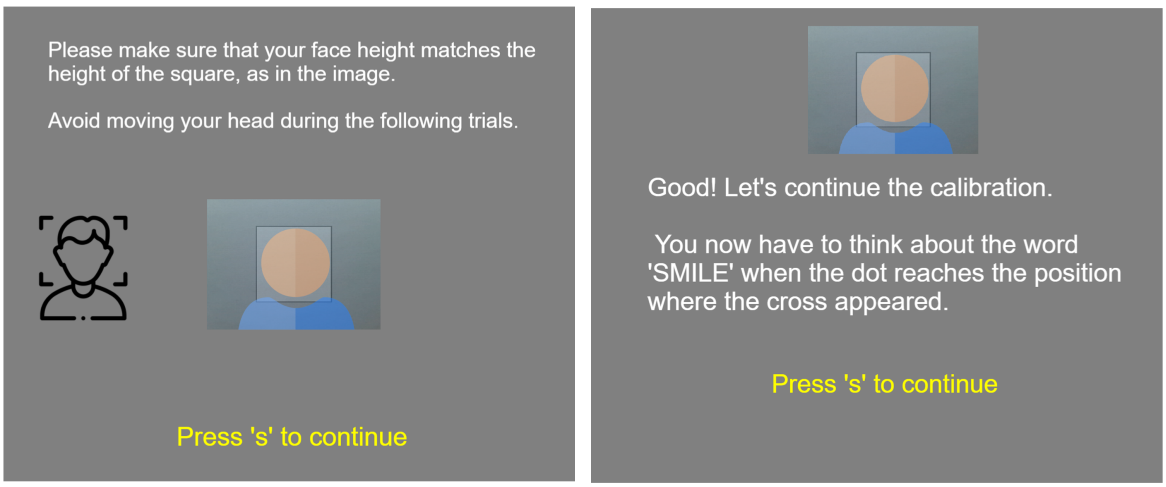


Figure 4: Example of the instructions presented to the participants before the trials of the mental action condition. Detailed instructions were given in order to increase the credibility of the mental action paradigm: participants were instructed to position their face in a precise location at the beginning of the block and before each trial (left); and they underwent a fake calibration process of the facial motion detection software (right).

Participants then completed two parts, each of them consisting of one test trial and twelve real trials. As in the motor action condition, in the first part the JoA was measured by explicit report; in the second part, the FoA was measured by the intentional binding effect. Before each single trial, participants were told to position their face again for the software to operate properly. During each trial, participants viewed the clock and the rotating dot, a cross appeared in a given position after 500 ms, and they were instructed to think about the word 'Smile' when the dot reached the position where the cross had appeared. Again, the sound appeared at varying intervals from the cued time (200, 400, 600 ms, four times each in a randomized order). If participants performed the task correctly, they should perceive that their thought was followed by the sound (with more or less delay).

Given that the sound appeared independently of the action of the participants, it was possible that a participant that did not perform the task correctly (he would not think about any particular word, or he would do it too early or too late), would realize that there was no connection between their thought and the sound. To ensure that participants performed the task correctly, a message appeared every four trials warning them to correctly perform the task, and mentioning that otherwise the software would de-calibrate and their data would not be used.

2.3.3 Baseline condition

This last block consisted of three test trials and twelve real trials. In each trial, participants viewed the clock and the rotating dot, and a cross appearing after one second, but they were instructed to not perform any action. They were told that their webcam was not recording and that their key press would not have any effect. The sound appeared at different intervals from the moment when the dot reached the position where the cross had appeared (200, 400 and 600 ms, four times each in a randomized order).

After each trial, participants were asked to indicate in the clock the position where they heard the sound, so that the baseline error in the estimation of the sound time could be measured. This metric is required to compute the intentional binding effect: the estimated time of the sound under voluntary actions must be compared with the estimated time of the sound in a baseline non-active condition [17, 37].

2.3.4 Post-experimental questions and debriefing

Once the subjects completed all three experimental conditions, they were asked two questions regarding their experience of control over the different tasks: "In the first block of the experiment (press key - sound), do you think that in some trials the sound was produced by your key press?"; "In the second block of the experiment (think 'SMILE' - sound), do you think that in some trials the sound was produced by your thought?"

The possible answers where "Yes" and "No". These questions are crucial for the study because they provide information about whether participants believed or not they could eventually produce the sound with their motor or mental actions. If a participant replied 'Yes' to the first or third question, this means that he thought that at least in one trial he produced the sound; therefore, the participant believed the "story" about the facial motion detection software, or about his/her key press producing the sound. It is critical to separate participants who had causal beliefs from those who didn't because a consistent sense of agency could be measured only if the subjects believed that their motor or mental actions could trigger the sound at all.

At the end of the experiment, participants were informed about the real goal of the experiment. They were told that there was no real facial motion detection software and no causal relation between their key press and the sound.

3 Results

Analysis of the data was carried out in Python using the *statsmodels* Python module for statistical models [42] and the *SciPy* Python ecosystem [43].

3.1 Post-experimental questionnaire

Out of 42 participants, 32 (76%) reported that they thought that in some trials their key press produced the sound, and 15 participants (36%) reported that they thought that in some trials their mental action (thinking 'Smile') produced the sound. All participants that believed in a causal relationship between their thought and the sound also believed in a causal relationship between their key press and the sound. Therefore, we used the data of 32 participants for the motor action condition, and of 15 participants for the mental action condition.

Only the data of those participants was used because it was expected that if a given subject didn't believe that his/her actions could produce the tone, the sense of agency would not arise. Analyzing the data of subjects who didn't have causal beliefs could be misleading because some subjects could have thought that there was no causal connection from the beginning, while others could have realized about that during the experiment. A large number of participants (50) were initially selected so that the number of participants for which causal beliefs were created was high enough.

3.2 Behavioral performance

Participants' performance in the motor action condition was evaluated by measuring their key press time and comparing it with the cued time (the instant when the rotating dot reached the position where the cross had appeared). The mean error in the key press across participants was 96 ms (STD = 0.122 s), which means that in general participants pressed the space bar after the cued time but not much later, and that this pattern was consistent across participants.

The 'space' key was not pressed in 13 trials (out of the 763 total trials in the motor action condition), those trials were removed. Trials with outlier key press times (± 3 STD of the group mean) were also removed (11 trials, less than 2% of the total).

In the mental action task, behavioral performance could not be measured: we had to expect participants to perform the task correctly (thinking about the word 'Smile' when the dot reached the position of the cross), and we could only encourage them to do so by reminding them the task instructions and emphasizing the role of the webcam and the facial motion detection software.

3.3 Explicit reports – Judgement of agency

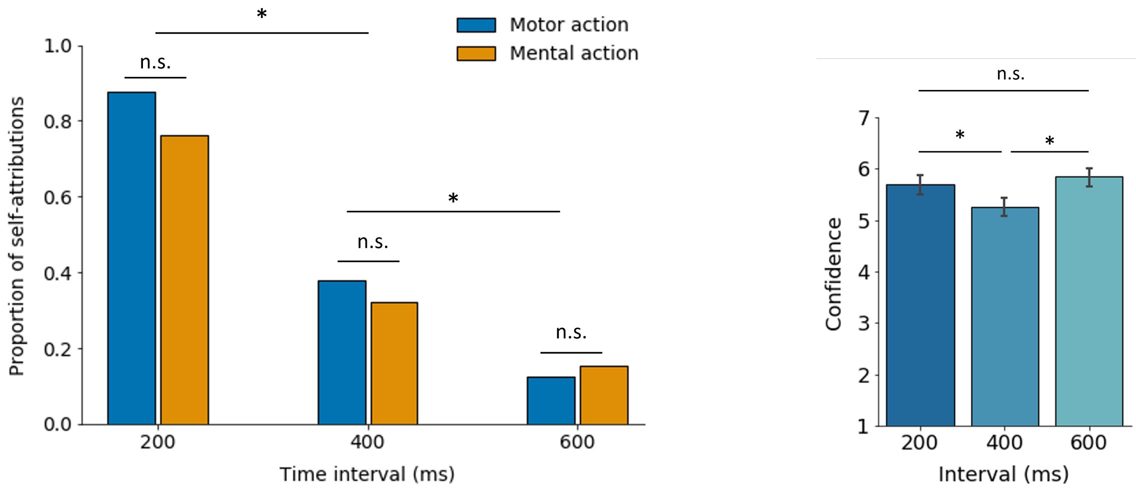
In order to verify the hypothesis that the JoA would arise similarly for motor actions than for mental actions, we analyzed if the number of self-attributions (responding 'Me' to the question "Who produced the tone?") compared to the number of attributions to the computer was modulated by the condition (motor action or mental action) and by the action-effect delay.

We used the data of all trials where an explicit attribution of agency was reported. In 10 trials no report was given, and in 2 trials participants reported not having heard the sound. In total, data from 543 trials was available (around 4 trials per participant in each condition and interval).

We ran a binomial logistic regression with authorship attribution ('Me' / 'The computer') as dependent variable, and condition (motor / mental action) and time interval (200 / 400 / 600 ms) as main predictors. We found a significant main effect of time interval ($z = -10.026$, $p < 0.0001$), no effect of condition ($z = -1.632$, $p = 0.103$) and no interaction between condition and time interval ($z = 1.573$, $p = 0.116$). More details about the binomial logistic regression are given in Appendix C.1. Post-hoc Chi-Square tests revealed that the number of agency self-attributions was significantly larger for shorter intervals than for medium intervals ($\chi^2 = 83.7$, dof = 1, $p < 0.0001$), and was also significantly larger for medium intervals than for long intervals ($\chi^2 = 24.2$, dof = 1, $p < 0.0001$). All pairwise comparisons were Bonferroni-corrected for multiple comparisons. To further assess whether there was a difference in the authorship attribution for the motor and mental conditions, we ran separate Chi-Square tests for each time interval. No significant difference was found for any of the time intervals (200 ms, $\chi^2 = 2.89$, dof = 1, $p = 0.089$; 400 ms, $\chi^2 = 0.34$, dof = 1, $p = 0.558$; 600 ms, $\chi^2 = 0.1$, dof = 1, $p = 0.752$). Figure 5a shows the proportion of agency self-attributions for the different time intervals and for the two experimental conditions. The statistical significance in the comparison between groups is shown in the figure only for illustrative purposes (the statistical tests have not been carried out on the proportion of self-attributions).

Summarizing, participants thought that their action had produced the sound much more frequently when the interval between their action and the sound was short than when it was long. These results are in line with the prediction that the JoA is modulated by the time interval between the action and the sound. Crucially, there was no significant difference in the amount of self-attributions for motor and for mental actions, indicating that the sense of agency at the conceptual level emerges in a similar manner in the presence or absence of motor commands.

The confidence reported by participants for their authorship attributions was also analyzed. The mean confidence rating was 5.6 (STD = 1.3) in a 1–7 Likert scale ranging from 'Not at all' to 'Very much'. A between-participants two-way ANOVA on confidence ratings with time interval and action type as factors revealed a significant effect of time interval ($F(2, 532) = 10.35$,



(a) (b)

Figure 5: Results of the explicit measures of agency. (a) Proportion of self-attributions for each condition, groups divided by interval. A significant effect of time interval was found, shorter intervals leading to a higher number of self-attributions. No effect of condition and no interaction were found. $^{*}p < 0.0001$; 'n.s.' = non-significant ($p > 0.017$); Chi-Square tests, Bonferroni corrected. (b) Mean confidence rating for the attribution of agency divided by time interval, reported in a 1–7 Likert-scale. As expected, the confidence for medium intervals is significantly lower than for short and long intervals. Error bars represent 95% confidence intervals. $^{*}p < 0.05$; 'n.s.' = non-significant ($p > 0.05$); Tukey HSD test on pairwise comparisons.

$p < 0.0001$), no significant effect of action type ($F(1, 532) = 3.65$, $p = 0.056$), and no interaction ($F(2, 532) = 1.23$, $p = 0.292$).[3] A Tukey HSD test showed that the confidence rating for medium intervals (400 ms) was significantly smaller than for short intervals (200 ms, $p < 0.05$) and also than for long intervals (600 ms, $p < 0.05$). No significant difference was found between the confidence for short and long intervals ($p = 0.51$). This is an expected result because participants are more likely to think that they caused the sound for short intervals, and that the computer did so for long intervals, but they are less confident in attributing the authorship of the sound to themselves or to the computer when an intermediate interval is used. Figure 5b summarizes the confidence ratings for the different time intervals.

Of note, all the analysis carried out in this section are between participants. The reason for that is that we had a different number of subjects for the action and the motor condition (see Section 3.1), thus, we couldn't compare both conditions for all participants.

[3] It is a matter of discussion whether the results of Likert scale ratings can be treated as continuous variables [44]. We have done so in the present study we have done so, given that this assumption is widely used, specially for 1–7 scales.

3.4 Implicit measurements – Intentional binding

We analyzed the participants' time estimations under three conditions: motor action, mental action and baseline (no action). In 4 trials, participants reported not to have heard the sound and in 20 trials no time estimation was given. We removed the outlier time estimations, which correspond to estimations beyond ±2.5 STD of the group mean time estimation (in average less than 1 trial per subject), as done in other studies [40]. The total number of trials was 884, approximately 3 per participant per interval per condition.

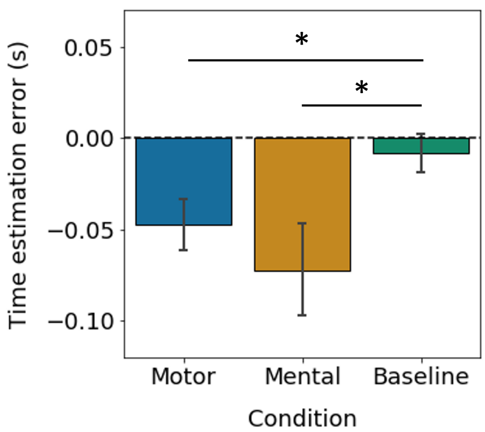
To evaluate whether intentional binding was found in the motor action and mental action conditions, we compared the error in sound time estimations under these conditions with the error in the baseline condition. The time estimation error is calculated by subtracting the actual time of the sound from its estimated time in a given trial. For each subject, we calculated the mean baseline error and the mean error under the motor or the mental action condition. Figure 6a shows the average estimation error for all three experimental conditions. Negative errors indicate that the sound was perceived earlier than it actually occurred, and an error significantly more negative than the baseline error reflects intentional binding.

Paired samples t-tests revealed that the sound was significantly anticipated in the motor action condition compared to the baseline condition ($t(30) = -3.55, p < 0.05$)[4], evidencing that intentional binding was present under motor actions. Critically, we also found that there was significant binding in the mental action condition ($t(14) = -2.54, p < 0.05$), which confirms the prediction that a low-level FoA may arise from non-motor actions. Independent samples t-test showed that there was no significant difference between the estimation error in the motor action and the mental action condition ($t(44) = 0.68, p = 0.497$), suggesting that intentional binding for the sound was produced in the same manner for both types of actions.[5].

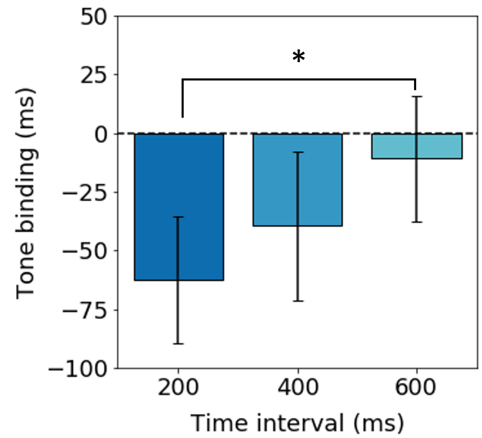
The previous results correspond to the time estimations for intervals collapsed. Although collapsing the data for all intervals is a common way to study intentional binding [16, 22], in this study we also wanted to assess if the time interval played a role in the intentional binding effect (as it did for the higher-level SoA). First, we conducted a repeated measures one-way ANOVA on time estimation errors in the baseline condition with time interval as factor. As expected, no significant effect of time interval was found ($F(2, 62) = 1.5,\ p = 0.231$), suggesting that in the baseline condition participants did similar estimation errors for the three time intervals. There-

[4]For these two tests, different data sets were used (participants who had causal beliefs over their motor actions and participants who had causal beliefs over their mental actions); thus, no correction for multiple comparisons was applied. More details are given in Appendix C.2.

[5]We treated the samples as independent because the number of subjects in the motor action condition and the mental action condition was different. We further confirmed that the difference in the estimation error between the two conditions was not significant with a paired sample t-test using only subjects for which data was available in both conditions ($t(13) = 0.49, p = 0.636$)



(a)



(b)

Figure 6: Intentional binding over mental and motor actions. (a) Mean error in the estimation of the sound time (with respect to the actual time) for each experimental condition. Errors more negative than the baseline error indicate intentional binding. Significant binding was found for both the motor and mental action condition, indicating the presence of low-level agency. Bars represent 95% confidence intervals. $^*p < 0.05$, paired sample t-tests. (b) Average binding scores for the three time delays used in the experiment, calculated as the mean estimation error of each participant in the action conditions (mental or motor) minus the mean baseline error of that participant. Shorter delays lead to higher binding, in consonance with the explicit measures of agency. Bars represent 95% confidence intervals. $^*p < 0.017$, paired sample t-test, Bonferroni corrected.

fore, we averaged the baseline error for each participant over the three intervals and subtracted it from their estimation errors in the other two conditions, thus obtaining a direct measure of binding.

A repeated measures one-way ANOVA on sound binding scores (calculated as the mean time estimation error in the action conditions minus the mean time estimation error in the baseline condition for each participant) with time interval as factor revealed a significant main effect of time interval ($F(2, 62) = 3.87$, $p < 0.05$). Post-hoc pairwise paired t-tests showed that the binding effect was significantly larger for short intervals than for long intervals ($t(29) = -3.31$, $p < 0.017$). The significance level was Bonferroni-corrected to 0.017 for multiple comparisons. This confirms the prediction that a stronger binding would appear for short intervals, in line with the results of the explicit measures of SoA summarized in Section 3.3. The binding effect was also larger for short than for medium intervals, and for medium than for long intervals, not significantly though ($t(29) = -1.34$, $p = 0.19$; $t(29) = -1.94$, $p = 0.06$ respectively). Figure 6b displays the binding measured for the action conditions (motor and mental together) for all three time intervals.

Finally, we wanted to evaluate if, for each time interval, there was a different binding effect for the motor and for the mental actions.[6] This allowed us to assess the relationship between implicit and explicit measures of the sense of agency: we could see if the pattern of results observed in the explicit reports (high agency for short intervals, low agency for long intervals, and no difference between motor and mental actions) is also found for intentional binding.

For each time interval, we calculated every subject's mean estimation error in each of the three experimental conditions (motor action, mental action, baseline). Figure 7 shows the average estimation error for all three experimental conditions in each time interval. For each interval, we conducted pairwise t-tests to compare the estimation errors in each action condition (motor and mental) with the baseline error:

- For short intervals (200 ms), the sound was significantly anticipated in the motor action with respect to the baseline ($t(30) = -5.66$, $p < 0.0001$), but there was no significant difference between the mental action and the baseline ($t(14) = -0.71$, $p = 0.487$). Therefore, for short action-effect delays motor actions lead to intentional binding over their effects, but mental actions do not seem to produce binding. However, an independent samples t-test suggested that the difference in time estimation errors between the motor and mental condition is not significant ($t(45) = -1.21$, $p = 0.228$); while paired sample t-test on the subject data-set that had causal beliefs in both the motor and the action condition (14 subjects) revealed a marginally significant difference in time estimation errors in the two conditions ($t(13) = -2.27$, $p = 0.04$). Giving the contradictory nature of these results, they must be interpreted with caution, and they will be further analyzed in the "Discussion" section.
- For medium intervals (400 ms), we found that the sound was significantly anticipated with respect to the baseline for both the motor action condition ($t(30) = -3.32$, $p < 0.001$), and the mental action condition ($t(14) = -4.44$, $p < 0.0001$). This suggests that for medium intervals binding over the tone was produced by both motor and mental actions, as predicted. Furthermore, no significant difference was found between the time estimation errors in both conditions, as shown by an independent samples t-test ($t(45) = 1.42$, $p = 0.157$) and by a paired t-test on the reduced 14 subject data-set (see above, $t(13) = 1.03$, $p = 0.323$).
- For long intervals (600 ms), there was no significant difference in time estimation errors between the motor action and baseline condition ($t(28) = 0.23$, $p = 0.820$), nor between the mental action and baseline condition ($t(13) = -1.59$, $p = 0.135$). Thus, for long intervals, no binding

[6] This analysis could have been done with a repeated-measures factorial ANOVA with time interval and condition as factors, thus including the interaction between them. However, since the data was unbalanced (more subjects believed the motor condition than the mental condition), such a repeated-measures analysis was not possible. A between participants factorial ANOVA was thought not to be representative of the hypothesis we wanted to test, given than for each participant we could find different estimation errors depending on individual differences. Therefore, we decided to conduct the analysis in the way that has been presented so far.

over the tone appeared, in line with the fact that participants attributed the authorship of the sound to the computer in most of the long interval trials. A significant difference was found between the time estimation errors in the action and the motor conditions, as revealed by an independent samples t-test ($t(45) = 2.80$, $p < 0.01$) and by a paired t-test on the 14 subject data-set ($t(12) = 2.21$, $p = 0.048$). This difference suggests that participants made higher estimation errors in the mental action condition, but this is not of special relevance for the present analysis given that in none of the conditions we found significant binding.

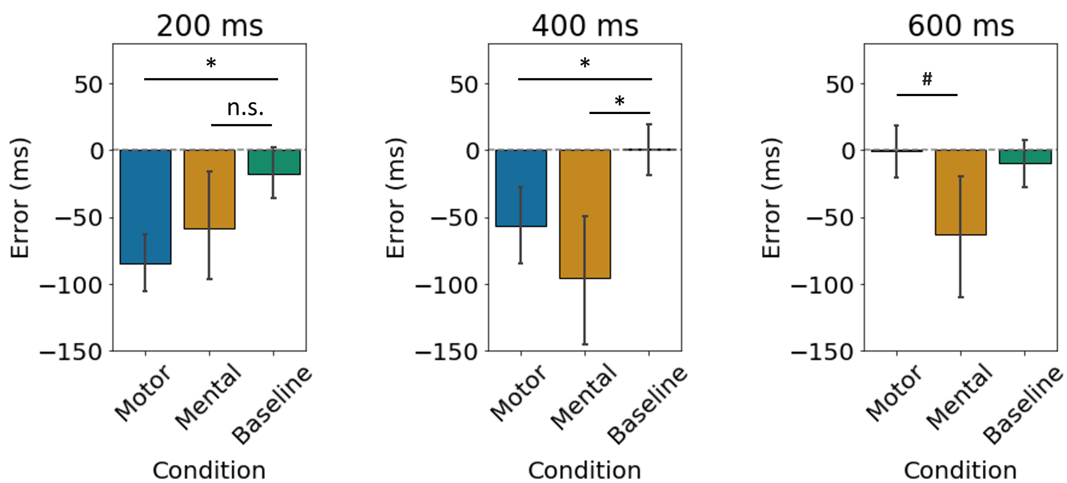


Figure 7: Comparison of the sound time estimation error for the three experimental conditions, data divided by time interval. For short intervals, significant binding (difference with baseline error) was found only for motor actions. For medium intervals, the binding over the sound was significant for both motor and mental actions, in line with the JoA results. For long intervals, the binding was not significant for any of these two conditions, also in consonance with the explicit reports. Bars represent 95% confidence intervals. $^{*}p < 0.01$, 'n.s.' = non-significant, paired samples t-test. $^{\#}p < 0.01$, independent samples t-test.

4 Discussion

In the present study, we have investigated the sense of agency arising from motor actions and mental actions in an action-effect paradigm, using implicit and explicit measures of agency.

One relevant outcome of this study is that we have successfully created a sense of agency in the participants by using an action-effect paradigm where only the causal beliefs prior to the tasks were manipulated. Indeed, the stimulus was identical in the three experimental conditions: the dot rotated and after a given interval the sound appeared, independently of the participants' action. However, before each condition, participants were led to believe that either a motor action (a key press) or a mental action (thinking about a particular word) could trigger the sound; or that without performing any action the sound would appear (baseline condition).

The experimental design was based on the assumption that a sense of agency would be created if the prior causal beliefs were sufficiently established. The results of the experiment confirm this assumption: more than two thirds of the total number of participants claimed to be in control of the sound in the motor action block, and more than one third of the subjects thought that their mental action could trigger the sound in the mental action block. The fact that a sense of agency emerged for those participants purely from the manipulation of prior causal beliefs is a critical success of this experiment.

4.1 High- and low-level agency

We assessed the high-level JoA by measuring the number of times that a participant would attribute the sound to his/her own action or to the computer, and comparing it for motor and mental actions. We validated the finding that shorter intervals lead to higher agency under motor actions, as confirmed by the higher number of self-attributions for short action-effect delays, and the progressive reduction of self-attributions for medium, and then subsequently for large delays. Interestingly, this same pattern was found under mental actions, and no significant difference was found in the number of self-attributions for mental and motor actions, confirming our hypothesis. This result offers a novel insight to the possibility that a strong SoA arises under non-motor actions, at least at a conceptual level (judgement of agency). Although it was already evidenced by studies involving BCIs that the SoA might emerge from mental actions [24, 28], the present results suggest not only that mental actions produce JoA, but that they do so to a similar extent than motor actions; and that the SoA disappears progressively under longer action-effect time delays in a similar manner than for motor actions. Within the framework of the cue integration theory [4, 15], these results indicate that in the absence of motor signals, other cues such as prior beliefs or non-motor predictions may contribute to the construction of judgements of agency, compensating for the lack of motor cues. This contradicts the view that motor commands are necessary for a robust sense of agency to emerge, as proposed by accounts based on the CM [20].

We also evaluated the low-level agency by measuring the intentional binding over the sound produced by motor and mental actions. We replicated the finding that voluntary motor actions produce binding over the effect of those actions: the sound was perceived as occurring earlier than it actually occurred. Critically, we also found that mental actions produced binding over the sound, and that there was no significant difference in the magnitude of the intentional binding effect for motor and for mental actions, confirming our predictions. This is a major finding of the present study, from which two separate conclusions can be extracted. First, that the intentional binding effect is not be exclusively linked to motor actions as some authors suggest [18, 41], but is rather a signature of low-level agency for intentional actions in general, as proposed by others [33][7]. Second, that mental actions give rise to a sense of agency not only at a conceptual level, but also at a low, perceptual level. This is a novel finding in the research on agency, since, to our knowledge, the current study is the first one to investigate the FoA under non-motor actions.

We hypothesize that the construction of the FoA from mental actions relies on general predictive mechanisms, such as counterfactuals, combined with inferential postdictive mechanisms not related to motor cues [15]. In fact, it could be that the FoA for both motor and mental actions is in part based on cognitive predictions independent of motor-based forward models, as suggested by Dogge and colleagues [12]. Based on the pre-activation account of intentional binding proposed by Waszak and colleagues [18], we hypothesize that these general cognitive predictions could pre-activate the perceptual representation of the sensory effect, leading to intentional binding (the original proposal by Waszak and colleagues suggested that these predictions were motor predictions).

We also investigated whether a consistent relationship between explicit and implicit measures of agency could be found: if this was the case, the intentional binding effect should be substantially larger for shorter intervals, given that in most short-delay trials participants attributed the authorship of the sound to themselves. Indeed, in the motor action condition we found significant binding for short and medium intervals but no binding for large intervals. This confirms that there is a coherent relationship between the explicit reports of the participants and the implicit measurement of the sense of agency through the intentional binding paradigm. This result is of special relevance because other studies investigating the relationship between explicit and implicit measures of agency have not found a fully consistent relationship between both measures, concluding that they might be driven by dissociable mechanisms operating over distinct timescales [36]. Here instead, we have found that the intentional binding over the sound is a good predictor of the explicit agency reported by the participants (see Appendix C.3 for an additional analysis of the relationship between implicit and explicit measures). This provides further evidence that the low-level feeling of agency is the major basis for the construction of a higher level judgement of agency, as hypothesized by Synofzik and colleages (see Figure 1) [15].

This consistent relationship between implicit and explicit measures was also found for mental actions, in particular for medium and long intervals (binding for medium delays and no binding for long delays). However, the binding over the sound was not significant for short intervals, which was not expected because participants had reported a high amount of self-attributions for short delays in the mental action condition. This result can be interpreted in two alternative ways: either no low-level feeling of agency emerged for non-motor actions under short delays, or

[7] It must be noted that the results are also consistent with the hypothesis put forward by Buehner [35] that the IB effect could reflect pure causality and not intentionality. In any case, both accounts agree in the fact that motor cues are not necessary for the IB effect to appear

a feeling of agency emerged but was not captured by the intentional binding effect. We support the latter option, for the reasons that follow. First, the high-level judgements of agency for 200 ms intervals were higher than for medium intervals. It is widely accepted that the judgements of agency, even if influenced by other factors, are largely based on the low-level feeling of agency. If no FoA arose for 200 ms, but did so for 400 ms, one would expect higher JoA for medium than for short intervals; but this was not the case. Second, one possible explanation for the absence of sound binding for mental actions and short intervals is that the outcomes of mental actions might be predicted slower than the outcomes of motor actions. Indeed, in our life-long experience interacting with the external environment generally requires bodily movements (BCI are an exception to this). Thus, it is coherent that the predictive mechanisms for the sensory outcomes of motor movements operate at fast timescales [45], while the predictions of sensory events triggered by mental actions, which are less common, may be less efficient. In the framework of the pre-activation account of IB proposed by Waszak and colleagues [18], this could mean that the predictive mechanisms for mental actions were not fast enough to pre-activate the perceptual representation before the stimulus arrived, and therefore no intentional binding was found. In the absence of an efficient sensory prediction, the FoA might have been constructed through other mechanisms, such as postdictive inferential processes based on the external sensory input and on cognitive priors, as proposed by the cue integration theory [15]. Nevertheless, the results of the current study do not allow us to draw a solid conclusion with respect to this matter, and it shall be further investigated (by using other implicit measures) whether a feeling of agency may arise for mental actions and short action-effect intervals.

In summary, we have found that the sense of agency emerging at the conceptual level was as strong for non-motor mental actions than for motor actions, and followed a similar pattern for different time intervals. At the perceptual level, the sense of agency for mental actions produced intentional binding also in a similar manner than for motor actions, with the exception of short action-effect delays, a matter which will need further investigation. Furthermore, we have found a consistent relationship between the explicit reports and the intentional binding effect, suggesting that this phenomenon might be driven by general predictive mechanisms not limited to motor predictions.

4.2 Limitations

This study presents several limitations. First, one could argue that the alleged non-motor actions involve motor signals to a certain extent, because participants are told that the software detects facial micro-expressions. Even if participants are told not to produce any voluntary gesture, it is possible that some motor commands are involved in the execution of the action, for example by forcing a 'static' facial expression to avoid producing voluntary gestures. Our prediction is that the results would have been very similar had this experiment been carried out with a BCI that didn't involve any kind of motor signal; we argue that any possible contribution of motor cues to the mental action condition is negligible. Another limitation is that many participants did not believe that they were in control in the mental action condition. We think that using a more credible paradigm like an EEG-based BCI (as initially planned) would largely increase the number of participants with causal beliefs.

Finally, in usual SoA experiments, participants get to choose when to perform an action, which action to perform or whether to act or not. In our everyday life, the conjunction of these three components (what to do, when to do it and whether to do it or not) gives rise to a strong feeling of intentionality in our actions. The relevance and neuroscientific evidence of the 'what', 'when' and 'whether' components of intentional action have been reviewed elsewhere [46]. However, in the present study participants were told which action to perform and when to perform it (and if they followed the instructions they couldn't choose whether to perform it or not). Our design was limited by the fact that there was no real causal relationship between the mental actions and the effect, so we had to instruct the subjects to execute a particular action at a particular time. In any case, we found that participants reported a SoA under our experimental paradigm. This could be explained by the fact that even if the instructions required the participants to perform an action at a given time, they could always choose not to follow the instructions, which could make that the 'what', 'when' and 'whether' components were still partially present. We hypothesize that under a real BCI paradigm, where participants could choose when to act or which action to perform, the sense of agency would be stronger than the one measured in this study. This line of research remains an interesting possibility for the future.

4.3 Further implications

The findings of this study have important implications for the field of brain-computer interfaces. Previous studies have already evidenced that subjects performing BCI-mediated actions may feel in control of those actions [24, 28]. The results presented here suggest that this sense of agency does not only arise at a pure conceptual level, where subjects would attribute to themselves an external event produced by their mental actions based on introspection and contextual information. Instead, it is likely that a sense of control emerges for BCI-actions at a basic non-conceptual level, similarly to the SoA for bodily movements. This implies that the interaction with BCIs may produce a solid sense of control, making them promising technologies for the future in a wide range of applications. In particular, the robust feeling of agency evidenced in this study argues in favor of the use of BCIs for the control of external devices, such as prosthetic limbs [47], and for other applications demanding a strong sense of control, such as communication through brain interfaces [48]. Importantly, the high SoA measured for non-motor mental actions indicates that external devices could be controlled in a robust manner not only by BCIs based on motor commands (typically through motor imagery), but also by non-motor-based BCIs, such as steady state visually evoked potentials (SSVEP) BCIs. Nevertheless, it is possible that sensorimotor BCIs provide more robust solutions for applications demanding precise and fast control, as suggested by the absence of intentional binding for mental actions and short delays in this experiment, and as proposed by Nierula and colleagues [28].

Besides, several authors have raised the problem of moral responsibility over BCI-mediated actions: users might be uncertain about their agency in a BCI context, making difficult the attribution of moral (thus, legal) responsibility [49, 50]. Moreover, in most legal systems this attribution of responsibility is based on the fact that only "voluntary acts" are punishable, an "act or action" being defined as a "bodily movement" (USA Model Penal Code, Section 2.01). Reflexes and unconscious movements are qualified as involuntary and, thus, are not punishable. However, BCI-mediated actions would not be proper actions under the definition of "act", even if those are voluntary. The current findings support the idea that BCI-mediated actions shall be considered voluntary actions, given that subjects have a sense of agency for mental actions to a similar level, and with a similar certainty than for motor actions. Thus, we argue that legal systems shall be adapted to include wider definitions of agency, notably with respect to mentally-triggered external events. The manner in which the notion of agency might be extended to mental actions in the framework of current legal systems is still a matter of debate and will be of critical relevance for the future of BCI applications [27, 49, 50].

Finally, we suggest that BCI-based experiments such as the one used in this study could help understand pathological disruptions of the sense of agency, in particular the phenomenon of thought insertion in schizophrenia [51]. Voss and colleagues explored the mechanisms behind the abnormal experiences of agency exhibited by schizophrenic patients through an IB paradigm [52]. Similarly, we propose that investigating the sense of control over mentally-triggered external events in patients with thought insertion symptomatology could help elucidating the mechanisms disrupted under this pathology. It has been suggested that the authorship of thoughts is not a feeling arising at a low level, but rather at a high conceptual level [51]. Thus, one would expect finding a similar intentional binding (low level) for mentally-triggered events in patients suffering from thought insertion and in healthy subjects; but schizophrenic patients would probably display abnormal attributions of authorship (high level) for the outcomes of these mental actions. This line of research remains to be explored in future experiments.

Statement of contribution

Ruben Moreno Bote and Xerxes Arsiwalla collaborated in the formulation of the conceptual ideas and research goals of this study. The development of the methodology and software was carried out by the author of the thesis, as well as the investigation process, data collection and statistical analyses. The Center for Brain and Cognition (UPF) assisted in the development of the JavaScript code to display the users' webcam during the experiment. The final manuscript was revised by Xerxes Arsiwalla.

References

[1] Moore, J. W. What Is the Sense of Agency and Why Does it Matter? *Frontiers in Psychology* **7** (2016). URL `https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01272/full`. Publisher: Frontiers.

[2] Synofzik, M. & Voss, M. Disturbances of the Sense of Agency in Schizophrenia. In Michela, B. (ed.) *Neuropsychology of the Sense of Agency: From Consciousness to Action*, 145–155 (Springer Milan, Milano, 2010). URL `https://doi.org/10.1007/978-88-470-1587-6_8`.

[3] Delorme, C. *et al.* Illusion of agency in patients with Gilles de la Tourette Syndrome. *Cortex* **77**, 132–140 (2016). URL `http://www.sciencedirect.com/science/article/pii/S0010945216300053`.

[4] Moore, J. W. & Fletcher, P. C. Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition* **21**, 59–68 (2012). URL `http://www.sciencedirect.com/science/article/pii/S1053810011002005`.

[5] Frith, C. D. Action, agency and responsibility. *Neuropsychologia* **55**, 137–142 (2014). URL `http://www.sciencedirect.com/science/article/pii/S0028393213002911`.

[6] Haggard, P. Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience* **9**, 934–946 (2008). URL `https://www.nature.com/articles/nrn2497`. Number: 12 Publisher: Nature Publishing Group.

[7] Aarts, H. & van den Bos, K. On the foundations of beliefs in free will: intentional binding and unconscious priming in self-agency. *Psychological Science* **22**, 532–537 (2011).

[8] Wegner, D. M. & Wheatley, T. Apparent mental causation. Sources of the experience of will. *The American Psychologist* **54**, 480–492 (1999).

[9] Hume, D. *A Treatise of Human Nature: A Critical Edition* (Oxford University Press, 2007).

[10] Frith, C. D., Blakemore, S.-J. & Wolpert, D. M. Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews* **31**, 357–363 (2000). URL http://www.sciencedirect.com/science/article/pii/S0165017399000521.

[11] Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. An internal model for sensorimotor integration. *Science* **269**, 1880–1882 (1995). URL https://science.sciencemag.org/content/269/5232/1880. Publisher: American Association for the Advancement of Science Section: Reports.

[12] Dogge, M., Custers, R. & Aarts, H. Moving Forward: On the Limits of Motor-Based Forward Models. *Trends in Cognitive Sciences* **23** (2019).

[13] Lau, H. C., Rogers, R. D. & Passingham, R. E. Manipulating the experienced onset of intention after action execution. *Journal of Cognitive Neuroscience* **19**, 81–90 (2007).

[14] Synofzik, M., Vosgerau, G. & Newen, A. Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition* **17**, 219–239 (2008). URL http://www.sciencedirect.com/science/article/pii/S1053810007000268.

[15] Synofzik, M., Vosgerau, G. & Voss, M. The experience of agency: An interplay between prediction and postdiction. *Frontiers in Psychology* **4** (2013). Place: Switzerland Publisher: Frontiers Media S.A.

[16] Desantis, A., Roussel, C. & Waszak, F. On the influence of causal beliefs on the feeling of agency. *Consciousness and Cognition* **20**, 1211–1220 (2011).

[17] Haggard, P., Clark, S. & Kalogeras, J. Voluntary action and conscious awareness. *Nature Neuroscience* **5**, 382–385 (2002). URL https://www.nature.com/articles/nn827. Number: 4 Publisher: Nature Publishing Group.

[18] Waszak, F., Cardoso-Leite, P. & Hughes, G. Action effect anticipation: Neurophysiological basis and functional consequences. *Neuroscience & Biobehavioral Reviews* **36**, 943–959 (2012). URL http://www.sciencedirect.com/science/article/pii/S0149763411002028.

[19] Haggard, P. Sense of agency in the human brain. *Nature Reviews Neuroscience* **18**, 196–207 (2017). URL https://www.nature.com/articles/nrn.2017.14. Number: 4 Publisher: Nature Publishing Group.

[20] Frith, C. The self in action: Lessons from delusions of control. *Consciousness and Cognition* **14**, 752–770 (2005). URL http://www.sciencedirect.com/science/article/pii/S1053810005000607.

[21] Sato, A. Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition* **110**, 74–83 (2009).

[22] Moore, J. W., Wegner, D. M. & Haggard, P. Modulating the sense of agency with external cues. *Consciousness and Cognition* **18**, 1056–1064 (2009).

[23] Wolpaw, J. R. *et al.* Brain-computer interface technology: a review of the first international meeting. *IEEE transactions on rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society* **8**, 164–173 (2000).

[24] Evans, N., Gale, S., Schurger, A. & Blanke, O. Visual Feedback Dominates the Sense of Agency for Brain-Machine Actions. *PLOS ONE* **10**, e0130019 (2015). URL `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0130019`. Publisher: Public Library of Science.

[25] Vlek, R., van Acken, J.-P., Beursken, E., Roijendijk, L. & Haselager, P. BCI and a User's Judgment of Agency. In Grübler, G. & Hildt, E. (eds.) *Brain-Computer-Interfaces in their ethical, social and cultural contexts*, The International Library of Ethics, Law and Technology, 193–202 (Springer Netherlands, Dordrecht, 2014). URL `https://doi.org/10.1007/978-94-017-8996-7_16`.

[26] Lynn, M. T., Berger, C. C., Riddle, T. A. & Morsella, E. Mind control? Creating illusory intentions through a phony brain–computer interface. *Consciousness and Cognition* **19**, 1007–1012 (2010). URL `http://www.sciencedirect.com/science/article/pii/S1053810010001078`.

[27] Limerick, H., Coyle, D. & Moore, J. W. The experience of agency in human-computer interactions: a review. *Frontiers in Human Neuroscience* **8** (2014). URL `https://www.frontiersin.org/articles/10.3389/fnhum.2014.00643/full`. Publisher: Frontiers.

[28] Nierula, B. *et al.* Agency and responsibility over virtual movements controlled through different paradigms of brain-computer interface. *The Journal of Physiology* (2019).

[29] Engbert, K., Wohlschläger, A., Thomas, R. & Haggard, P. Agency, subjective time, and other minds. *Journal of Experimental Psychology: Human Perception and Performance* **33**, 1261–1268 (2007). Place: US Publisher: American Psychological Association.

[30] Sato, A. & Yasuda, A. Illusion of sense of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition* **94**, 241–255 (2005).

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

[32] Aarts, H., Custers, R. & Wegner, D. M. On the inference of personal authorship: enhancing experienced agency by priming effect information. *Consciousness and Cognition* **14**, 439–458 (2005).

[33] Desantis, A., Hughes, G. & Waszak, F. Intentional Binding Is Driven by the Mere Presence of an Action and Not by Motor Prediction. *PLOS ONE* **7**, e29557 (2012). URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029557. Publisher: Public Library of Science.

[34] Synofzik, M., Vosgerau, G. & Lindner, A. Me or not me – An optimal integration of agency cues? *Consciousness and Cognition* **18**, 1065–1068 (2009). URL http://www.sciencedirect.com/science/article/pii/S1053810009001020.

[35] Buehner, M. J. Understanding the Past, Predicting the Future: Causation, Not Intentional Action, Is the Root of Temporal Binding. *Psychological Science* **23**, 1490–1497 (2012). URL https://doi.org/10.1177/0956797612444612. Publisher: SAGE Publications Inc.

[36] Ebert, J. P. & Wegner, D. M. Time warp: Authorship shapes the perceived timing of actions and events. *Consciousness and Cognition* **19**, 481–489 (2010). URL http://www.sciencedirect.com/science/article/pii/S1053810009001548.

[37] Moore, J. W. & Obhi, S. S. Intentional binding and the sense of agency: a review. *Consciousness and Cognition* **21**, 546–561 (2012).

[38] Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* **51**, 195–203 (2019). URL https://doi.org/10.3758/s13428-018-01193-y.

[39] Libet, B., Gleason, C. A., Wright, E. W. & Pearl, D. K. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) The unconscious initiation of a freely voluntary act. *Brain* **106**, 623–642 (1983). URL https://academic.oup.com/brain/article/106/3/623/271932. Publisher: Oxford Academic.

[40] Wolpe, N., Haggard, P., Siebner, H. R. & Rowe, J. B. Cue integration and the perception of action in intentional binding. *Experimental Brain Research* **229**, 467–474 (2013). URL https://doi.org/10.1007/s00221-013-3419-2.

[41] Wolpe, N. & Rowe, J. B. Beyond the "urge to move": objective measures for the study of agency in the post-Libet era. *Frontiers in Human Neuroscience* **8** (2014). URL https://www.frontiersin.org/articles/10.3389/fnhum.2014.00450/full. Publisher: Frontiers.

[42] Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference* **2010** (2010).

[43] Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261–272 (2020). URL https://www.nature.com/articles/s41592-019-0686-2. Number: 3 Publisher: Nature Publishing Group.

[44] Carifio, J. & Perla, R. J. Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences* **3**, 106–116 (2007). URL https://thescipub.com/abstract/jssp.2007.106.116.

[45] Seth, A. K., Suzuki, K. & Critchley, H. D. An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers in Psychology* **2** (2012). URL https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00395/full#B141. Publisher: Frontiers.

[46] Brass, M. & Haggard, P. The What, When, Whether Model of Intentional Action:. *The Neuroscientist* (2008). URL https://journals.sagepub.com/doi/10.1177/1073858408317417. Publisher: SAGE PublicationsSage CA: Los Angeles, CA.

[47] Clausen, J. *et al.* Help, hope, and hype: Ethical dimensions of neuroprosthetics. *Science* **356**, 1338–1339 (2017). URL https://science.sciencemag.org/content/356/6345/1338. Publisher: American Association for the Advancement of Science Section: Policy Forum.

[48] Grau, C. *et al.* Conscious Brain-to-Brain Communication in Humans Using Non-Invasive Technologies. *PLOS ONE* **9**, e105225 (2014). URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0105225. Publisher: Public Library of Science.

[49] Steinert, S., Bublitz, C., Jox, R. & Friedrich, O. Doing Things with Thoughts: Brain-Computer Interfaces and Disembodied Agency. *Philosophy & Technology* **32**, 457–482 (2019). URL https://doi.org/10.1007/s13347-018-0308-4.

[50] Haselager, P. Did I Do That? Brain–Computer Interfacing and the Sense of Agency. *Minds and Machines* **23**, 405–418 (2013). URL https://doi.org/10.1007/s11023-012-9298-7.

[51] Vosgerau, G. & Voss, M. Authorship and Control over Thoughts. *Mind & Language* **29**, 534–565 (2014). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mila.12065. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mila.12065.

[52] Voss, M. *et al.* Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain* **133**, 3104–3112 (2010). URL https://academic.oup.com/brain/article/133/10/3104/320647. Publisher: Oxford Academic.

[53] Wegner, D. M. *The Illusion of Conscious Will* (A Bradford Book, Cambridge, Mass., 2003), edición: 1 edn.

[54] Wegner, D. M., Sparrow, B. & Winerman, L. Vicarious agency: experiencing control over the movements of others. *Journal of Personality and Social Psychology* **86**, 838–848 (2004).

[55] Wegner, D. M. & Sparrow, B. Authorship processing. In *The cognitive neurosciences, 3rd ed*, 1201–1209 (Boston Review, Cambridge, MA, US, 2004).

[56] Blakemore, S.-J., Wolpert, D. M. & Frith, C. D. Abnormalities in the awareness of action. *Trends in Cognitive Sciences* **6**, 237–242 (2002). URL `http://www.sciencedirect.com/science/article/pii/S1364661302019071`.

[57] Blakemore, S. J., Wolpert, D. & Frith, C. Why can't you tickle yourself? *Neuroreport* **11**, R11–16 (2000).

[58] Gentsch, A., Schütz-Bosbach, S., Endrass, T. & Kathmann, N. Dysfunctional forward model mechanisms and aberrant sense of agency in obsessive-compulsive disorder. *Biological Psychiatry* **71**, 652–659 (2012).

[59] Gallagher, S. The Natural Philosophy of Agency. *Philosophy Compass* **2**, 347–357 (2007).

[60] Sidarus, N., Vuorre, M., Metcalfe, J. & Haggard, P. Investigating the Prospective Sense of Agency: Effects of Processing Fluency, Stimulus Ambiguity, and Response Conflict. *Frontiers in Psychology* **8** (2017). URL `https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00545/full`. Publisher: Frontiers.

[61] Moore, J. & Haggard, P. Awareness of action: Inference and prediction. *Consciousness and Cognition* **17**, 136–144 (2008). URL `http://www.sciencedirect.com/science/article/pii/S1053810007000049`.

[62] Hughes, G., Desantis, A. & Waszak, F. Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin* **139**, 133–151 (2013). Place: US Publisher: American Psychological Association.

A Extended Literature

A.1 Sense of agency theories

D. Wegner's "Theory of apparent mental causation", first published together with a series of experiments supporting the theory [8], and later on extensively detailed in his influential book [53], suggested that our feeling of control over voluntary actions arises from the relationship between our intentions and our actions. A strong SoA would appear based on the principles of priority, consistency and exclusivity (see main text). Since they propose that the SoA appears following inferential processes after the sensory event, these theories are usually referred to as *postdictive* or *inferential* accounts. Several experiments have provided support to Wegner's theory, by showing that the sense of agency increased in conditions where conscious intentions are consistent with action-effects compared to conditions where they are inconsistent [54]. Other studies demonstrated that when other potential causes for an action are given (violating the exclusivity principle), participants may attribute authorship to other agents, even if they were the authors of the action [16, 55].

Agency theories based on the comparator model for sensorimotor prediction have been proposed by Frith [10] and Blakemore [56]. These accounts, first conceived to explain pathological disruptions of agency such as schizophrenia or anosognosia, propose that the experience of control for an action emerges from the motor predictions related to the bodily movement of the action. They are referred to as *predictive* accounts. For example, the CM can explain the attenuation of the sensory consequences of self-generated actions ("Why can we tickle ourselves?" [57]). Predictive theories of agency have been supported by several studies, where predictive motor mechanisms have shown to play a key role in the attribution of agency [30] and in the faulty sense of agency under some pathologies [52, 58].

Cue integration theories propose that both predictive and postdictive mechanisms play an important role in the emergence of the SoA, thus reconciling both theories. Sources of information such as motor signals, sensory feedback, background beliefs or intentions are all integrated to construct the experience of agency. Some authors have proposed that this cue integration could take place in a way similar to the Bayesian integration of sensory cues: the reliability of each agency cue would determine its relative contribution to the resulting sense of agency [2, 4, 22]. This view has received support from many studies in the last few years. The results of Moore and colleagues [22] suggest that both internal motor cues and external cues (in the form of primes) are integrated to construct the sense of agency, but also that internal cues issued from motor signals receive higher weighting because they are more reliable sources of information.

The two-level distinction of the sense of agency (low-level feeling of agency, high-level judgement of agency) was proposed by Synofzik and colleagues [14, 15], and similar distinctions (first-order phenomenal agency versus high-order cognitive agency) had been also put forward by other authors [59]. Different cues would be integrated at different levels (see Figure 1): perceptual cues and internal motor commands typically influence the low-level agency, while beliefs, intentions and environmental cues are more influential at the conceptual level. It is worth mentioning that high-order information can also influence the low-level feeling of agency, as shown by Desantis and colleagues [16]: in their study, participants had to press a key which produced a sound. Their feeling of agency was measured with the intentional binding paradigm. The authors found that the FoA of the subjects was reduced when they thought that another person could have triggered the sound, even if it was always them who produced it, suggesting that high-level causal beliefs also influence the low-level FoA.

A.2 Sense of agency measures

Explicit measures refer to explicit reports of agency, which usually require participants to introspect about their sense of control over a given action or sensory event. For example, subjects may be asked to attribute one action or its effect to himself or to someone else under ambiguous situations of control ("Did you do that?") [32], or reporting the amount of control felt over a given stimulus ("How much control did you feel") [21, 60]. In the present study, the explicit measures are based on the "self / other" distinction ("Who produced the tone?", possible answers: "Me", "The computer"). We argue that this two-option question reflects better the experience of agency than a question about the degree of agency: either participants felt they produced the sensory event or not, and the fact that they can be more or less confident about their judgement might be collected in a separate question (as done in the present experiment).

Implicit measures do not require the introspection of participants about their experience of agency, but are based on explicit reports about other perceptual information that is known to correlate with the sense of agency. One example is the intentional binding effect, first evidenced by Haggard and colleagues [17] (see Figure 2). A large number of studies have supported the relationship between the sense of agency and the IB effect [29, 37, 61]; including studies with schizophrenic patients where an abnormal agentic experience has been associated to differences in the IB effect [52]. Another common implicit measure is the sensory attenuation paradigm: the sensory effects self-generated actions are usually perceived with less intensity than when the actions are not self-generated [57]. By asking participants about the intensity of a sensory event (loudness of a sound, pressure on a body part) produced or not by his own action, one can evaluate the sense of agency over those events [21]. A review of sensory attenuation and intentional binding experiments can be found in [62].

A.3 Motor vs. non-motor paradigms

We have argued that previous studies investigating the contribution of motor cues to the sense of agency have not used appropriate non-motor contrast conditions in the experimental paradigms. Let's consider the two following examples:

- In one experiment, Sato and colleagues [21] assessed the SoA of participants with implicit and explicit measures under two conditions: in the first one, participants pressed a key which triggered a sound; in the second one, they viewed the hands of a confederate in front of them (to facilitate the illusion that the hands belonged to the participant), and they observed how one of the hands pressed the key which eventually triggered a sound. The SoA measured in the second condition (absence of motor cues) was substantially lower than for the first condition.
- In another study, Moore and colleagues [22] compared a condition where participants voluntarily pressed a key to trigger a sound with a condition where a motor attached to the finger of the participant made the finger press the key, which then triggered a sound. Again, the measured SoA measured was lower for the non-motor condition.

In both cases, not only motor signals are absent in the non-motor contrast condition, but also the participants' intentionality to produce an action is absent. One cannot draw from these results any solid conclusion regarding the specific contribution of motor signals, given that intentionality is a key element in the construction of the SoA.

A.4 BCI and the sense of agency

Recently, the sense of agency arising from BCI-mediated actions has been the subject of several studies. Nierula and colleagues [28] assessed the SoA (through explicit reports) emerging from moving an embodied virtual reality body with two different BCI paradigms – motor imagery and SSVEP. They found that a SoA emerged in both cases from the control of movements with the BCI, but that the SoA was stronger for the motor imagery BCI paradigm. These findings suggest that the presence of a motor plan (even in the absence of proprioceptive feedback and overt bodily movement) facilitates the attribution of authorship; but that also in the complete absence of motor cues the SoA may arise from pure mental actions.

In another experiment, Evans and colleagues [24] used an EEG-based BCI where participants could control a cursor through motor imagery. They found that an SoA emerged for BCI-mediated actions, and that the mechanisms behind this SoA were similar to those proposed for bodily movements: discrepancies between the mental action and the sensory feedback led to a decrease in the SoA. Their results also suggest that in the absence of proprioception and overt bodily movement, visual feedback is the main source of information for the construction of the SoA.

B Methods – additional details

B.1 Dot rotation period

The rotation of the dot was coded in the following manner: after each frame refresh, the dot changed its coordinates to the next position. In Psychopy, programming the timing of stimuli based on the frame rate is the best approach if a precise timing is required[8]. In our case, a precise timing was necessary, because for a smooth rotation of the dot, the dot had to be displaced every 50 ms approximately. Since we programmed the dot to be displaced every frame, we could achieve a displacement every 16.6 ms for a typical screen refresh rate of 60 Hz, which was sufficiently precise.

We selected a period of 2.55 seconds instead of the 2.56 seconds used by Libet and Haggard [17, 39] to obtain an integer number of frames for each complete rotation: for typical refresh rates of 40 Hz or 60 Hz, the total number of frames per rotation would be an integer number (102 frames and 153 frames, respectively).

B.2 Excluded participants

Because the dot rotation was linked to the frame refresh rate of the participants' display (see above), the code retrieved automatically this value at the beginning of the experiment. However, in some cases the retrieved frame rate did not correspond to the actual frame rate, for reasons that are still unknown. This led to a slower or faster dot rotation, which could be seen in the participants' data: the time when they pressed the 'space' key was much earlier or much later (for example, at 4 seconds in average) than the expected time (the group average key press time was 2.65 seconds). Upon posterior enquiry, these participants confirmed that their frame rate was different to the one retrieved by the code. Therefore, the results for these participants could not be compared with the rest. Seven participants were excluded for this reason.

Another participant was excluded due to poor performance: in the first block, the 'space' key was consistently pressed out of time; in both the first and second block the same explicit reports were always given for each trial, with always the same confidence level; and the error in the time estimation was beyond 3 STD of the mean group error.

[8] see `https://www.psychopy.org/builder/startStop.html`

B.3 Test trials

In the motor action condition, three test trials were performed before the actual trials. In the test trials, the sound was causally linked to the key press, as opposed to the real trials. If participants did not press any key, or if participants pressed the key too early or too late with respect to the cued time (more than 250 ms before or more than 450 ms after the correct time), no sound appeared. The objective was that if participants performed the task incorrectly in the test trials, they would not notice that there was no actual causal relationship between their action and the sound. If the key press was performed within the “correct” interval, the sound appeared 400 ms after the key press in the two first test trials and 600 ms after the key press in the third one.

In the mental action condition, only one trial was performed before the actual trials. Here, the test trial was identical to the real trials, and the sound appeared 200 ms after the cued time of the mental action.

In the baseline condition, three test trials preceded the actual trials. Participants were instructed to not perform any action, as in the actual baseline trials. The sound appeared 400 ms after the cued time in the two first test trials and after 600 ms in the third one.

C Further statistical analyses

C.1 Explicit reports

We ran a binomial logistic regression with authorship attribution ('Me' / 'The computer') as predicted variable and condition (motor / mental action) and time interval (200 / 400 / 600 ms) as main predictors, to evaluate the contribution of those factors to the judgements of agency (see Section 3.3). The detailed results of the generalized linear model (GLM) are shown in Table 1.

Generalized Linear Model Regression Results						
Dep. Variable:	Attribution of agency ['Me', 'The computer']	No. Observations:	543			
		Df Residuals:	539			
Model:	GLM	Df Model:	3			
Model Family:	Binomial	Log-Likelihood:	-270.04			
Link Function:	logit	Deviance:	540.07			
No. Iterations:	5	Pearson chi2:	558.			
	coef	std err	z	P>$\|z\|$	[95.0% Conf. Int.]	
Intercept	28.9398	2.899	9.984	0.000	23.258	34.621
Condition	-7.4632	4.574	-1.632	0.103	-16.427	1.501
Interval	-9.9012	0.988	-10.026	0.000	-11.837	-7.966
Condition * Interval	2.4582	1.563	1.573	0.116	-0.605	5.521

Table 1: Summary of the generalized linear model results. Agency attributions were used as dependent variable, with condition and interval as main predictors, including the interaction between both.

The results of the GLM show that the attributions of agency were significantly modulated by the action-effect time interval; but no significant effect of condition was found, and the interaction between condition and interval did not have a significant effect either. We ran an additional GLM model without the interaction to evaluate if the model improved, but the Log-Likelihood of this model did not increase with respect of the one displayed in Table 1.

C.2 Intentional binding measures

To assess whether mental and motor actions produced intentional binding, we compared the subjects' mean estimation error in the different experimental conditions with paired samples t-tests and independent samples t-tests (see Section 3.4). Before the tests, we verified the assumptions of normality and homoscedasticity for the contrasted groups. Shapiro-wilk tests revealed no significant departure from normality for none of the groups, and Levene's test for equality of variances confirmed the homoscedasticity assumption for all pairwise comparisons. The test statistics for these comparisons are summarized in Tables 2 and 3.

Condition	Number of subjects	Shapiro-Wilk test	
		Statistic	p-value
Motor action	31	0.97	0.507
Baseline error	31	0.96	0.797
Mental action	15	0.96	0.407
Baseline	15	0.88	0.056

Table 2: Statistics of the Shapiro-Wilk test of normality for the mean estimation errors in each experimental condition.

Groups tested	Type of test	Levene's test		t-test		
		F	p-value	df	t-value	p-value
Motor action vs. Baseline	Paired samples	2.75	0.102	30	-3.55	0.001
Mental action vs. Baseline	Paired samples	2.59	0.119	14	-2.54	0.024
Motor action vs. Mental action	Independent samples	0.73	0.398	44	0.68	0.497

Table 3: Results of Levene's test for equality of variances and of the t-tests carried out to assess the difference in time estimation errors between the different experimental conditions.

C.3 Explicit vs. implicit measures

We have seen that the results of the explicit and implicit measures used in the experiment are coherent with each other: for short intervals, participants reported more self-attributions, and an intentional binding was found (only for motor actions though); for long intervals, participants attributed most trials to the computer and no intentional binding was found in any of the conditions.

To further explore the relationship between these implicit and explicit measures, we studied whether the intentional binding scores were a good predictor of the attributions of authorship. Since intentional binding and authorship attributions were measured in distinct trials, we divided the data in the following way: for each participant, we calculated his mean intentional binding score (mean sound time estimation error for the mental and motor actions minus his mean baseline error) for each time interval. Then, for each participant and each time interval, we assigned a 1 or 0 depending on whether the participant had attributed more trials for this time interval to himself or to the computer. Thus, for each participant, we had three average binding scores (one per interval) and three values of "probable authorship attribution" (one per interval). We wanted to study if the intentional binding scores for the actions could predict the attributions of authorship.

We conducted a binomial logistic regression with Authorship attribution (1: "Me", 0: "The computer") as dependent variable and average binding score as main predictor. We found a significant main effect of average binding score ($z = -2.378$, $p < 0.05$), indicating that intentional binding scores are significant predictors of authorship attributions. Table 4 summarizes the results of the generalized linear model regression, and Figure 8 shows the analyzed data, fitted with a logistic regression.

Generalized Linear Model Regression Results

Dep. Variable:	Probable attribution [1:'Me', 0:'The computer']	No. Observations:	93
		Df Residuals:	91
Model:	GLM	Df Model:	1
Model Family:	Binomial	Log-Likelihood:	-60.459
Link Function:	logit	Deviance:	120.92
No. Iterations:	4	Pearson chi2:	93.9

	coef	std err	z	P>$\|z\|$	[95.0% Conf. Int.]	
Intercept	0.5600	0.247	2.270	0.023	0.077	1.043
Binding score	6.5787	2.767	2.378	0.017	1.156	12.001

Table 4: Results of the generalized linear model regression conducted to evaluate whether binding scores were good predictors of the participants' explicit agency attributions.

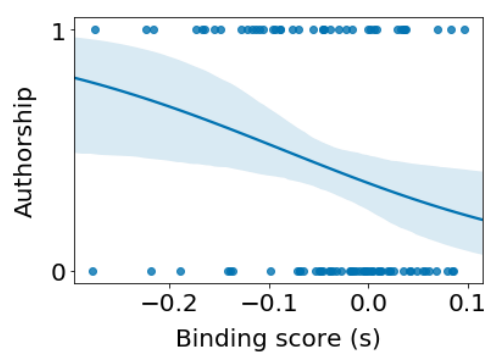


Figure 8: Scatter plot describing the data analyzed in the generalized linear model. Each data point corresponds to one participant and one time interval. The x-coordinate of each point represents the average intentional binding score for this participant at this interval, and the y-coordinate represents the most probable authorship attribution for the same participant at that interval (1: "Me", 0: "The computer"). The binomial logistic regression fitting the data is also shown, the shaded area represents the 95% confidence interval.