

# Comparing Metacognitive Representations of Bodily and External Agency

 Angeliki Charalampaki,<sup>1,2,3</sup>  Heiko Maurer,<sup>4,5</sup> Lisa K. Maurer,<sup>4,5,6</sup> Hermann Müller,<sup>4,5,6</sup> and Elisa Filevich<sup>1,7</sup>

<sup>1</sup>Bernstein Center for Computational Neuroscience Berlin, Berlin 10115, Germany, <sup>2</sup>Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin 10115, Germany, <sup>3</sup>Department of Psychology, Faculty of Life Sciences, Humboldt-Universität zu Berlin, Berlin 10099, Germany, <sup>4</sup>Neuromotor Behavior Laboratory, Justus Liebig University, Giessen 35394, Germany, <sup>5</sup>Institute of Sport Science, Justus Liebig University, Giessen 35394, Germany, <sup>6</sup>Center for Mind, Brain and Behavior of the Universities Marburg, Giessen and Darmstadt 35032, Germany, and <sup>7</sup>Hector Institute for Education Sciences & Psychology, University of Tübingen, Tübingen 72072, Germany

## Abstract

We studied the role of movement and outcome information in forming metacognitive representations of agency. Human participants ( $N = 40$ ; 25 female, 15 male, 0 diverse) completed a goal-oriented task: a semivirtual version of a ball-throwing game. In two conditions, we manipulated either the visual representation of the throwing movement or its proximal outcome (the resulting ball trajectory). We measured participants' accuracy in a discrimination agency task, as well as confidence in their responses and tested for differences in the electrophysiological (EEG) signal using mass linear mixed-effect modeling. We found no mean differences between participants' metacognitive efficiency between conditions. However, through exploratory analyses, we found that metacognitive sensitivity did not correlate between the two conditions and that the EEG signal differed between the two conditions during the agency discrimination task. We cautiously interpret these results as suggesting that although both movement and outcome information contribute to participants' sense of agency, they may do so through distinct processes. These findings highlight the need for further research examining the potential neurophysiological differences corresponding to the conceptual distinction between bodily and external agency.

**Key words:** EEG; metacognition; motor control; movement; outcome; sense of agency

## Significance Statement

The sense of agency refers to the feeling of intentionally moving our body and, through that movement, affecting our surroundings. Empirical studies often explore this concept by focusing on either the bodily representations (how participants move their bodies) or the external representations (how this movement affects the environment) of voluntarily performed actions. In other words, these two types of representations are often assumed to be equivalent. This work challenges that assumption using a goal-oriented motor task, electroencephalographic data, and bias-free methods from metacognitive literature.

## Introduction

The subjective experience that we are the authors of our intentional actions is referred to as the sense of agency. This is an umbrella term that can refer to many, potentially distinct, cognitive processes. In fact, previous work has pointed at this breadth of the scope in the term and aimed to distinguish different operational definitions of agency (Pacherie, 2007;

Received April 24, 2025; revised Nov. 13, 2025; accepted Nov. 20, 2025.

The authors declare no competing financial interests.

Author contributions: A.C. and E.F. designed research; A.C. performed research; A.C., H.Ma., L.K.M., H.Mü., and E.F. contributed unpublished reagents/analytic tools; A.C. analyzed data; A.C., H.Ma., L.K.M., H.Mü., and E.F. wrote the paper.

We thank Pedro Espinosa Mireles de Villafranca and Paula Guiomar Alarcón de Antón for support with experiment setup and data collection.

This work was supported by a Deutscher Akademischer Austauschdienst (DAAD) scholarship awarded to A.C. and a Freigeist Fellowship to E.F. from the Volkswagen Foundation (grant number 91620) and by the Collaborative Research Center "Cardinal Mechanisms of Perception", funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; SFB/TRR 135 TP B6, grant number: 222641018).

*Continued on next page.*

Synofzik et al., 2008; Gallagher, 2012; Christensen and Grünbaum, 2017, 2018; Wen, 2019). In particular, there is an ongoing discussion on the similarities and differences between bodily and external operational definitions of agency (Kalckert and Ehrsson, 2012; Christensen and Grünbaum, 2017; Wen, 2019). The difference lies in whether the definition used in a study of agency over an action (“I threw the ball in the basket”) is strictly limited by the boundaries of our body (the movement: “I moved my arm to throw”) or the definition is broad enough to also include agency over the consequences of our actions that go beyond the boundaries of our body (proximal outcome: “The ball flight trajectory” or the distal outcome: “The ball landed in the basket”). However, in the literature, this dichotomy in the operational definition often goes unnoticed, despite its importance in designing experimental protocols and models that adequately address this concept. These two definitions are not only conceptually distinct but differ also in their neural correlates. A review of fMRI studies on agency found a distinct activation pattern based on whether the experimental protocol employed focused on either the bodily or external agency (Charalampaki et al., 2022), emphasizing the broader implications of this dichotomy and the necessity to study these two aspects of agency in relation to one another, to explore where they converge or diverge both at the behavioral and at the neuronal level. In what follows, we will first discuss the most common model used to explain the human sense of agency: the comparator model (Frith et al., 2000), which is often used interchangeably to describe both bodily and external sense of agency. There is evidence that both agency over the movement and agency over the outcome are reduced by similar types of experimental manipulations; but there are also conflicting results on their relative importance in our sense of agency, which challenges the notion that they can be used interchangeably. We will then motivate our study by outlining how borrowing methods from the metacognitive literature can inform this discussion. Finally, we will offer a brief literature-based overview of the EEG signatures of agency and describe how this motivated our exploratory analysis.

According to the comparator model, whenever we make an intentional action, a copy of the motor (efferent) commands issued to make the necessary movements is used to make predictions about the consequences of the movement. These predictions are then compared with the actual afferent sensory feedback (associated with our moving limbs). When these two match, we experience authorship of the movement; when they do not, we experience a loss of agency. While the comparator model was originally used to explain agency over our bodily movements (Frith et al., 2000), it is often extended to explain the agency we experience over the external consequences accompanying our actions (Haggard, 2017). That is, it is usually assumed that the comparator can also make comparisons between predictions and feedback for the sensory consequences that go beyond our bodies. However, this assumption has been challenged. In particular, because current experimental designs have either failed to or provided mixed results regarding the involvement of efference copies in predicting the sensory consequences of outcomes beyond the body (Christensen and Grünbaum, 2018; Dogge et al., 2019). Further, participants can experience agency even for actions carried out by another person (Wegner et al., 2004; Tieri et al., 2015). Meaning that agency can be experienced in the absence of an efference signal (but see Kilteni et al., 2021 for the importance of efference copies in the sensory attenuation). Therefore, it is important to question the suitability of the comparator model to explain the sense of agency over external consequences of movement.

On the one hand, one series of studies shows that subjective reports on participants' sense of agency are affected by manipulations of the visual representations of both a movement and its outcome. In these studies, participants reported experiencing a diminished sense of agency over an action when they fail to achieve a specific goal (van der Weiden et al., 2013), when the outcome was different than their predictions (Sato and Yasuda, 2005), or when the predicted outcome was delayed (Spengler et al., 2009; Farrer et al., 2013). Similarly, participants reported a lower sense of agency over a movement they made when the visual feedback of the movement they performed was manipulated spatially (van den Bos and Jeannerod, 2002; Farrer et al., 2003; Padrao et al., 2016) or temporally (Farrer et al., 2003; Krugwasser et al., 2019).

On the other hand, another line of work has aimed at directly comparing the relative precision of agency representations based on movement and outcome prediction violations, with conflicting results. While some studies found that participants were more sensitive to violations of outcome predictions as compared with violations of bodily predictions, other

The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or manuscript preparation.

Correspondence should be addressed to Angeliki Charalampaki at angelikichar@gmail.com.

Copyright © 2026 Charalampaki et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

studies have found the opposite. For example, [David et al. \(2016\)](#) introduced a sensorimotor mismatch by temporally delaying the visual feedback, either at the point of movement initiation or at the onset of the movement outcome. They found that participants' agency ratings decreased most sharply when the timing of the outcome was manipulated. Similarly, [Wen et al. \(2015\)](#) asked participants to click on a keyboard to control the position of a moving dot, aiming to make rapid movements to reach a visual target. The visual feedback was manipulated by introducing a temporal lag, and the level of control on the movement was manipulated by ignoring some of the participants' commands. The latter manipulation was meant to introduce a sensorimotor mismatch so as to affect whether participants reached their goal, sometimes improving their success rate. Participants' agency reports reflected whether they succeeded with their goal despite some of their commands being ignored, albeit only under some conditions, where the delay introduced was long. Finally, [Oishi et al. \(2018\)](#) extended these findings by showing that reaching a goal can retrospectively influence participants' agency: Participants moved a cursor toward a changing color target (on some trials, the cursor moved with a delay). However, independently of the delay introduced on the movement, reaching the intended outcome (successfully reaching the target while it was on a specific color, as opposed to it being of a different color) was sufficient for participants to report feeling stronger agency.

The studies summarized above point toward outcome prediction violations overriding movement prediction ones, suggesting in turn that the representation of outcome predictions might affect participants' agency more strongly than representations of movement predictions. But, as mentioned above, other studies directly contradict this conclusion, by showing that participants are more sensitive to violations of motor predictions as compared with violations of outcome predictions. For example, [Metcalf and Greene \(2007\)](#) asked participants to move a cursor toward downward-moving visual targets and, on some trials, manipulated the movement feedback by introducing "turbulence" on the cursor's movement. They also manipulated the outcome by introducing false outcome hits (viz., the target appeared as being "magically" touched). Without turbulence, participants' ratings of agency (or control) matched the proportion of targets they hit. In other words, in the absence of movement prediction violations, the more targets they hit, the higher their control rating. However, when there were either movement manipulations or outcome manipulations, participants' control ratings more closely followed the relation between their motor act and the visual feedback of it, irrespective of how many targets it appeared as they had touched. In a similar study, [Metcalf et al. \(2013\)](#) extended their findings by manipulating both the motor aspect of the action (turbulence) and the probability of a target being exploded upon touch. They showed that while both types of violations negatively affected participants' agency, violations of the movement predictions had a more notable effect on participants' ratings of control, and this was independent of participants' performance, measured as the number of targets participants hit.

As a whole, this body of literature demonstrates that both bodily and external prediction violations, probed with manipulations of the feedback on either the movement or its outcome, can affect explicit reports. Importantly however, the different results also emphasize the importance—and challenges—of differentiating between different experimental manipulations in the study of agency. Beyond the conceptual differences driven by different experimental operationalizations of the sense of agency, there are two methodological aspects that limit the interpretability of the findings: All of the experiments summarized above measure participants' sense of agency using subjective reports, which are prone to biases and this could explain the apparent diversity of the findings regarding the relative importance of movement and outcome predictions in our subjective experience of agency. Also, some of the studies compared participants' sensitivity in detecting prediction violations of movement and outcome using manipulations that differ substantially between the two conditions, for example, manipulating the movement feedback with turbulence versus manipulating the outcome feedback by controlling the probability of an outcome occurring ([Metcalf et al., 2013](#)). This challenges the comparison of the precision of these two sources of information as we do not know how these two different levels of action description relate to one another. To deal with these complications, while studying agency with explicit reports, one can instead measure the precision of explicit representations of agency using tools borrowed from research in metacognition ([Fleming and Lau, 2014](#)). That is, one can measure metacognitive representations of the agency judgments that lie on one representational level above, allowing for a comparison of their relative precision. This is the approach we chose.

In addition to behavioral data, we chose to collect EEG data to investigate possible neuronal-level differences arising from violations of bodily and external prediction linked with the sense of agency during action execution. Currently there is limited literature regarding electrophysiological markers of agency. Existing studies have primarily examined changes in the EEG signal amplitudes. In other words, they have examined the positive and negative deflections of the EEG signal (event-related potentials, ERPs) either during ([Padrao et al., 2016](#); [Han et al., 2021](#); [Ren et al., 2023](#); [Gomez-Andres et al., 2024](#)) or following action execution ([Kühn et al., 2011](#); [Bednark and Franz, 2014](#); [Vastano et al., 2020](#)). However, these studies differ from one another significantly both in the experimental designs and in the ways agency was measured. Some involved Eriksen Flanker stimuli followed by hand movements without behavioral measures of agency ([Padrao et al., 2016](#); [Gomez-Andres et al., 2024](#)); others required pressing a button to prevent a sound, measuring agency implicitly through auditory sensory attenuation reflected in EEG ([Han et al., 2021](#)). Additional variations include using a go/no-go task with feedback on performance ([Ren et al., 2023](#)), button presses ([Sedarus et al., 2017](#)), cursor movements toward targets ([Bednark and Franz, 2014](#)) with explicit agency attribution judgments, or button presses followed by an auditory stimulus, where agency was measured with explicit self-other ratings ([Kühn et al., 2011](#)) or implicit measures of agency ([Vastano et al., 2020](#)). Given these methodological differences, it is unsurprising that ERP components linked to agency

differ substantially in their temporal and spatial characteristics. Negative deflections observed include the frontal event-related negativity occurring ~100 ms after erroneous actions (Padrao et al., 2016; Gomez-Andres et al., 2024) and N400 components over parietal areas in response to externally generated errors (Padrao et al., 2016). Positive deflections have been identified as well, such as the P3 component over frontocentral regions, which was modulated by self-other agency judgments (Kühn et al., 2011) or varied with levels of agency under ambiguous action–outcome coupling (Bednark and Franz, 2014), and a P600 component showing differences between self and externally generated errors (Gomez-Andres et al., 2024).

Other studies have focused on the EEG frequency content and power changes across frequency bands, but these results remain inconsistent. Some evidence points toward increased alpha and beta oscillation power correlating with increased agency (Zito et al., 2023), while others report decreased alpha power with increased agency (Kang et al., 2015; Wen et al., 2017). Additionally, reduced frontal alpha-band connectivity has been associated with increased action control (Kang et al., 2015). Beta oscillation phase coherence changes have shown modulation based on control level across widespread regions, whereas gamma oscillation phase coherence varied with agency level over central and left parietal areas (Kang et al., 2015). It has also been proposed that coupling of gamma-band oscillations between parietal and presupplementary motor areas relates to positive agency (Ritterband-Rosenbaum et al., 2014). Lastly, findings regarding theta oscillations and their role in agency perception remains similarly complex. Enhanced theta power over left frontocentral and parietal electrodes has been linked to positive, and over temporoparietal electrodes to negative agency (Jeunet et al., 2018), whereas decreased theta synchronization over centroparietal regions has been associated with increased agency (Zito et al., 2023). In the absence of widely accepted neural markers of sense of agency, we performed a series of exploratory analysis on EEG signals to identify differences associated with variations in the salience of agency experience.

In this study, participants made movements in a deterministic, goal-oriented motor task in which we could tightly manipulate both the visual representations of the movement and the outcome. The motor task was a semivirtual version of the “Skittles” game (Müller and Sternad, 2004; Fig. 1). In the Skittles game, in which a ball is attached to a pole with a rope, the aim is to grab and throw the ball in such a way to hit the target placed behind the pole. In this task, we could manipulate both the visual representations of the movement and the outcome. On each trial, participants completed a motor action (grab and throw the ball to hit the target) twice in succession; they selected which of the two actions they felt were more in control (only one matched their action) and then rated their confidence in that decision. We hypothesized that representations of movement and outcome differentially affect participants’ metacognitive representations of their sense of agency and that this would also be reflected as behavioral differences. We also explored the data and sought for any potential differences at the neural level based on the underlying electroencephalography (EEG) data. Specifically, we focused on the FCz electrode, as prior evidence suggests that centro-frontal regions are involved in online monitoring of action execution, outcome processing, and other agency-related processes, such as the attenuation of the sensory consequences associated with voluntary actions. Given the exploratory nature of this analysis and the precise timing of the experimental design, we examined time-locked differences in EEG signal amplitude.

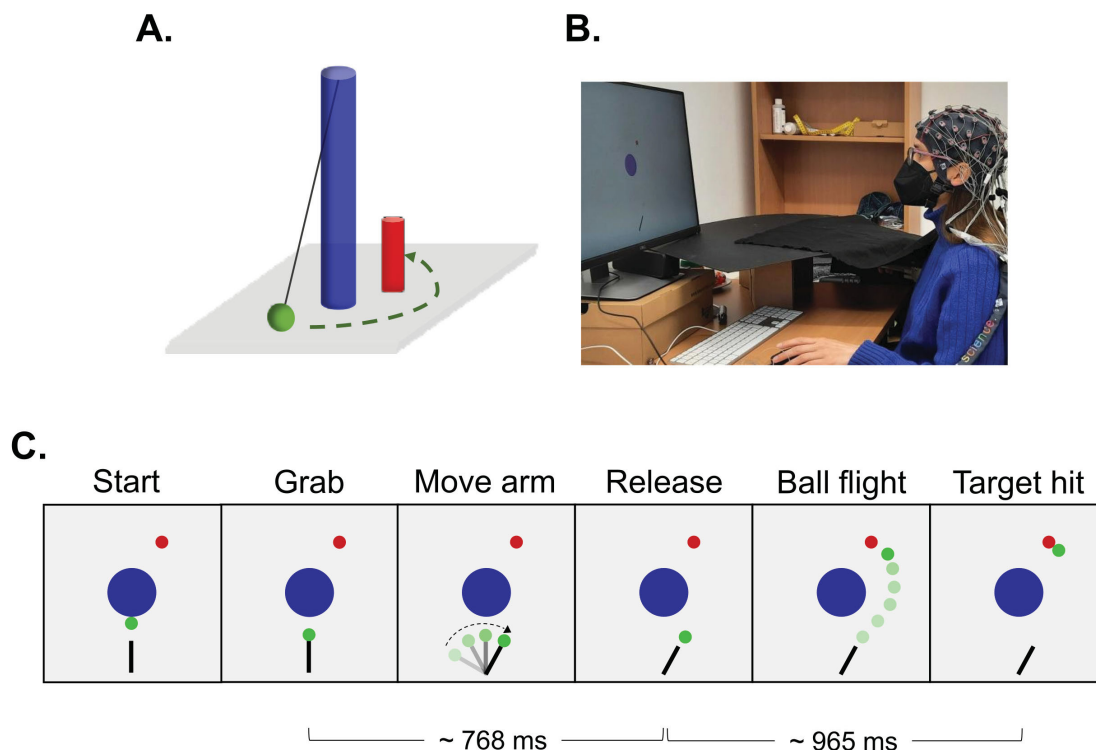
## Materials and Methods

We assessed whether violations of visual representations of a movement and its consequences on the environment equivalently affect participants’ sense of agency using behavioral and electroencephalography data (EEG). The study was preregistered (<https://osf.io/SJURK>), and we report any deviations from the preregistered plan. The ethics committee of the Institute of Psychology of the Humboldt-Universität zu Berlin approved the experimental procedures (2017-17), which were in line with the Declaration of Helsinki. Human subjects were recruited at a location which will be identified if the article is published.

### Participants

Our preregistered plan was to collect 40 clean datasets from participants who passed the inclusion criteria: no neurological or psychiatric history, normal or corrected-to-normal vision, right-handedness, and proficiency in English. The sample size was determined based on a previous study in which the Skittles task was embedded in a motor metacognitive experimental design (Arbuzova et al., 2021). To that extent, we collected data from 52 participants. We only included participants who correctly selected the interval in which they were in control (the online feedback matched their action), 60% or more of the total number of trials belonging to each condition. Therefore, we excluded five participants with very low discrimination accuracy (<60%) in at least one of the two conditions. In deviating from our preregistered plan, we further excluded the datasets from seven participants with especially variable discrimination performance throughout the experimental session. We deemed performance to be variable for these participants based on a visual inspection of the adaptively changed manipulation magnitude of the visual feedback (see the description of the staircasing procedure below) prior to analyzing the data. This resulted in a final sample of 40 participants (mean age = 26.60 years  $\pm$  5.60, 25 female, 15 male, 0 diverse, mean handedness score = 90.60  $\pm$  14.50; Oldfield, 1971). Due to technical issues, the EEG signal from two participants was not recorded, but we nevertheless used these participants’ data in the behavioral analysis. Participants signed a written informed consent and for their time, were compensated with €8/h or course credits.





**Figure 1.** **A**, The Skittles game. The aim of the game is to swing the ball (green) attached to the pole (blue) with a rope to hit the target (red). **B**, Experimental setup. Participants sat in front of the screen, placed their right hand on the custom-made lever (which was hidden below an occluder), and used a computer mouse with their left hand to register their responses. We recorded the EEG signals from 62 active electrodes placed on an elastic head cap. **C**, Virtual version of the Skittles game. For each action, the ball, lever, and target appeared simultaneously on the screen. Participants could then grab the ball to start their action. On average, for each action, participants released the ball 765 ms after grabbing it, and the ball traveled for a mean of 965 ms from the point of release to the target. From the moment participants released the ball, the ball disappeared from the screen for  $\sim 202$  ms ( $M = 202.36 \pm 0.21$ ) to avoid participants being able to identify the manipulated interval based on the jump in the position of the ball due to the manipulation.

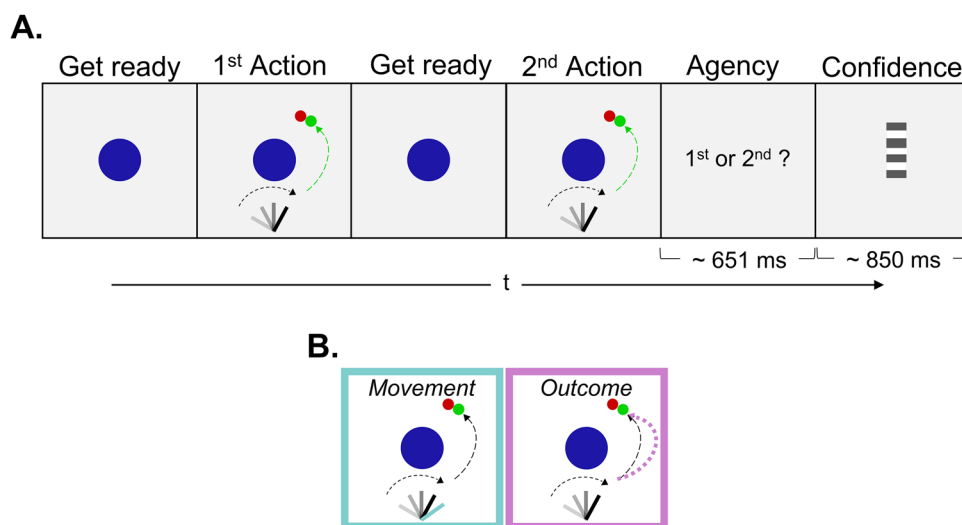
### Apparatus and stimuli

For the motor task, we adapted the virtual version of the “Skittles” game, first described by Müller and Sternad (2004). In the Skittles game, a ball is attached to a pole with a rope, and a target—the Skittle—is placed behind the pole (Fig. 1A). The goal is to hit the target by swinging the ball around the pole. In the virtual version of the Skittles game, participants had a top-down view of the scene: The central pole was positioned in the middle of the scene [blue circle, position  $(x,y) = (0,0)$  m, with a radius = 0.25 m]; the ball initially appeared hanging next to the pole (green circle, radius = 0.05 m); and the target was, when viewed from above, top-right of the pole [red circle, position  $(x,y) = (0.20,0.70)$  m, radius = 0.05]. The massless rope (spring constant  $k = 1$  N/m) was not visible on the scene but was taken into consideration when estimating the ball trajectory. Participants grabbed, moved their forearm, and released the ball using a custom-made lever. The lever rotated on the end nearest to the participants, so that they could swing their arm horizontally around the elbow point. The angle of their hand was recorded online with the aid of a goniometer (Novotechnik, RFC4800 Model 600, 12 bit resolution,  $0.1^\circ$  precision), and all the data from the lever were recorded with a sampling rate of 1 kHz using a LabJack T7 data acquisition device (LabJack). This allowed us to depict the virtual lever in the lower middle part of the scene [ $(x,y) = (0, -1.5)$  m] as a black rectangular (0.4 m length, 0.04 m width) that pivoted along with the physical lever. Participants grabbed the virtual ball by placing their index finger on the touch switch on the tip of the lever and released it by lifting their finger. The Skittles game is deterministic. That is, the ball trajectory can be determined at the point of ball release by the angle of the lever and the velocity of its tip (Müller and Sternad, 2004; Sternad et al., 2011). Therefore, we could display the ball trajectory in real time with a maximum possible flight duration of 2 s. Once participants lifted their finger, the ball disappeared for  $\sim 200$  ms ( $M = 202.36 \pm 0.21$ ) to avoid a jump in the ball position when the ball trajectory was manipulated (see below). Each motor task started with only the central pole displayed, and with a pseudorandomized delay (0.40–1.00 s), the lever, the ball, and the target appeared on the screen. We programmed the task in MATLAB (R2016b, MathWorks) and Psychtoolbox-3 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). Throughout the experiment, participants had no visual access to either their arm or the real lever (an opaque piece of cloth was held above their arm, without interfering with the movement), so they had to rely on the visual feedback they received on an LCD monitor (2,560  $\times$  1,440 pixels,  $0.61 \times 0.345$  m, refresh rate 60 Hz) located  $\sim 0.6$  m in front of them (Fig. 1B).

## Procedure

On each trial, participants performed the motor task two consecutive times (Fig. 2A). The online visual feedback they received matched their action in only one of the two intervals. In the other, the visual feedback was manipulated in one of two possible ways: either the position of the lever (Movement condition) or the trajectory of the ball (Outcome condition) did not match their own (Fig. 2B). Two previous studies used the semivirtual Skittles game and found that participants had accurate representations of both movement parameters—the velocity of the movement and the hand's angle position at the point of ball release—but also the ball flight trajectory and the distal outcome (Arbuzova et al., 2021; Charalampaki et al., 2023). We used a visuospatial manipulation to introduce violations of the predictions of the movement and proximal outcome for two reasons: first, because temporal delay could affect participants' ability to monitor their action rather than simply introducing a mismatch between the prediction and the feedback of their action (for an extended discussion on the topic, see Wen, 2019; but also, Osumi et al., 2019) and second, to prevent participants from recognizing the visually delayed feedback as a delayed version of their action (Farrer et al., 2008). In the Movement condition, we only manipulated the angle of the lever (see details below, Online staircases). Therefore, the virtual lever always moved synchronously with the physical one but had a different angle, and the ball (which appeared after 202 ms;  $M = 202.36 \pm 0.21$ ) flew according to the participant's actual throw (based on the angle of participants' real hand and velocity at the point of ball release). In the Outcome condition, the virtual lever precisely matched the participant's hand's movement. We manipulated only the ball trajectory, ensuring that it would always match the distal outcome of the participants' real throw. Because on each trial (see below) participants compared two actions, we aimed to match the goal of each one participants' action (Charalampaki et al., 2023). Therefore, we positioned the target slightly to the right of the pole to increase the chances of participants hitting it and emphasized in the instructions that they should always aim to hit it. At the end of the second action, participants discriminated in a two-interval forced choice task, in which of the two intervals they felt more in control (Fig. 2A). To do so, they could press the left or right computer mouse buttons with their left hand to report the first or second interval, respectively. Following the discrimination response, participants rated how confident they were that the interval they selected was the one in which they had complete control over the action seen on the screen, using a continuous virtual scale from "Very unsure" to "Very confident." Participants moved up and down a computer mouse with their left hand, which controlled the position of a bar that appeared on a pseudorandom position over the virtual scale on each trial and then pressed either button on the mouse to register their confidence. During this point of the trial, participants could alternatively press the spacebar on a computer keyboard to mark the trial for exclusion (e.g., they pushed the wrong button in the discrimination step).

To familiarize participants with the virtual version of the task, participants first completed eight trials of the motor task in which the visual feedback always matched their action. After that, they did 40 training trials, half of which were Movement and half Outcome trials presented in pseudorandomized interleaved order. In these training trials, the trial ended after the discrimination response. Finally, participants completed 16 trials in which they also rated their confidence in their decision to become familiar with the steps of the main experiment. The main experiment consisted of 504 trials, split into 9 blocks,



**Figure 2.** **A**, Experimental paradigm. In the main experiment, participants threw the ball twice in succession. The visual feedback accurately reflected their actions in only one of the two intervals. In the other interval, the feedback was manipulated based on the trial condition. At the end of the second action, participants selected which interval they felt more in control. Finally, participants rated how confident they were about their decision. **B**, Experimental conditions. We manipulated the online visual feedback participants received on the screen in two conditions. In the Movement condition, we manipulated only the position of the virtual lever, which moved synchronously with the physical one but deviated a few degrees to the left or right (the black bar depicts the real position, the cyan bar shows the manipulated feedback participants receive). In the Outcome condition, we manipulated only the ball trajectory. The ball flew in an alternative trajectory (depicted with pink) with the same distal outcome (hit/miss the target) as the participants' actual trajectory.

with breaks in between. Each block consisted of 56 pseudorandomized and interleaved Movement and Outcome condition trials.

The whole experiment, including EEG electrodes elastic cap preparation, took ~3 h.

### Online staircases

To keep participants' discrimination accuracy at ~71% (Levitt, 1971), we used a separate online staircase (2-down, 1-up) for each condition. Both staircases operated on the velocity manipulation ( $|\Delta v|$ ). That is, in both cases, on each trial we added or subtracted the staircased  $|\Delta v|$  to the actual lever velocity. We then displayed, on each time point, the lever position that corresponded to the manipulated velocity (on Movement trials) or the ball position given the manipulated release velocity (on Outcome trials). Specifically, on Movement trials we manipulated the angle of the lever by multiplying the  $|\Delta v|$  by the time difference  $\Delta t$  between two screen refresh frames in order to obtain the corresponding lever angle difference. On Outcome trials we manipulated the trajectory of the ball by changing the velocity of the real movement in  $|\Delta v|$ , in order to create a manipulated ball flight trajectory. We pseudorandomized a parameter to determine whether we added or subtracted  $|\Delta v|$ , in order to present a manipulated trajectory to the right or left of the real one, respectively. We deviated from this process if the alternative ball trajectory ended in a throw with a distal outcome (hit/miss the target) that did not match participants' throw. In these instances, we first multiplied the  $|\Delta v|$  with the opposite sign, and if this again led to an outcome incongruent from the real one, we used the nearest possible value that would. If none of these steps worked, we used the original staircase value and marked the trial for exclusion.

Importantly, on Movement trials, the ball flight trajectory in both intervals corresponded to the trajectory determined by the angular velocity of the participant's hand at the point of ball release. In other words, on Movement trials, the ball flight trajectory was displayed intact in both intervals. On Outcome trials, the virtual lever's angle in both intervals corresponded to the participant's hand position. In other words, on Outcome trials the virtual lever's angle was intact in both intervals.

The starting value of both staircases was 1, the step size was 0.15 in the training session and 0.10 in the main experiment, and the minimum possible value of the staircase was 0.21 for the training and 0.01 in the main experiment. We used larger step sizes in the training staircases as a means to roughly approximate a good starting point for each participant's discrimination ability, which would then be more finely staircased during the experiment.

### Data analysis

#### Data exclusion

In keeping with our preregistered plan, we excluded from both the behavioral and EEG data those trials in which the discrimination reaction time was above 8 s (mean excluded trials  $\pm$  SD:  $1 \pm 2$ ), trials marked for exclusion by the participants (mean  $\pm$  SD:  $4 \pm 6$  trials), and those trials in which the intervals differed in terms of the distal outcome (mean  $\pm$  SD of trials participants missed hitting the target in either of the intervals:  $63 \pm 55$  and mean trials in which participants threw the ball toward the left of the pole in at least one interval:  $5 \pm 11$ ). Overall, we included a mean of  $405 \pm 70$  trials from each participant (Movement trials:  $198 \pm 38$  and Outcome trials:  $207 \pm 33$ ).

In addition to our preregistered exclusion criteria, we excluded the (behavioral and EEG) data from the first blocks from those participants for whom the staircases were especially variable at the start of the experiment. Specifically, we excluded the first block from eight participants, the first two blocks from four participants, and the first four blocks from two participants. Data exclusion took place prior to any further data analysis.

#### Behavioral data

**Metacognitive representations of agency.** This design allowed us to quantify how well people can inspect their agency judgments. That is, we could use an estimate of metacognitive efficiency (Mratio; Maniscalco and Lau, 2012), which reflects the relationship between participants' discrimination accuracy and confidence rating. Mratio aims to quantify the proportion of information available for confidence ratings, relative to that available in the discrimination decisions. Thus, as a unit-free measure, it allows for the comparison between conditions without being contaminated by potential biases in subjective reports. Participants with high Mratio typically rate with high-confidence trials in which they correctly identified in which of two intervals they had more control of the perceived action and rate with low confidence those trials in which they instead identified the interval in which they had less control of the perceived action. In contrast, confidence ratings from participants with low Mratio typically cannot differentiate between the trials in which they selected the correct interval from those they did not.

To estimate the Mratio, we used the maximum likelihood estimation method as implemented in the "metaSDT" R package (Craddock, 2021). For this analysis, we had to discretize the continuous confidence ratings, and we chose to do so using six equidistant bins. Before discretizing the ratings, we normalized them per participant (by subtracting their minimum confidence rating and dividing by the difference between their maximum and minimum confidence values, like Charalampaki et al., 2023), to account for possible differences in the range of confidence ratings used by each participant. Further, in line with the recommendations, to avoid having empty bins that would hinder the fitting procedure (Maniscalco and Lau, 2012), when fitting the models for each participant, we added  $1/(2 \times \text{number of bins}) = 1/12$  to each input vector containing the count of each confidence rating bin type.

We used the “BayesFactor” package (Morey and Rouder, 2018) to obtain the Bayes factors in favor of H1 ( $BF_{10}$  values) and the “rstatix” package (Kassambara, 2023) to estimate effect sizes and ran parametric correlations using the “ggstatsplot” package (Patil, 2021), all implemented in R (version 4.2.2, R Core Team, 2022). Prior to the statistical analyses, we log-transformed  $d'$  and scaled the confidence ratings (by mean-centering to the condition-specific mean, and normalizing by the condition-specific SD) because they were not normally distributed. We report mean values and standard deviations ( $M = \text{mean value} \pm \text{SD}$ ).

### *Electrophysiology recordings*

We recorded electrophysiology (EEG) data using 62 active electrodes fitted on an elastic cap following the 10–20 system g.tec position montage (g.tec medical engineering). We placed two reference electrodes (one on each earlobe) and three electrodes around the right eye (above, below, and outer canthus) to record horizontal and vertical eye movements. We used AFz as the ground electrode and the electrode on the right earlobe as the online reference. The signal was digitized at a sampling rate of 1,200 Hz, online bandpass filtered between 0.5 and 200 Hz, and notch filtered between 48 and 52 Hz to remove line interference. We monitored the electrodes' impedances throughout the experiment to keep them below 10 K $\Omega$ .

### *EEG analysis*

**Preprocessing pipeline.** We preprocessed the EEG signal using the EEGLAB toolbox (Delorme and Makeig, 2004) and custom scripts implemented in MATLAB (R2020a, MathWorks). In line with our preregistered plan, we rereferenced the EEG signal to the average signal of the earlobe electrodes, filtered the signal using separate finite high-pass (1 Hz passband edge), and used low-pass (100 Hz passband edge) Hamming windowed-sinc finite impulse response (FIR) filters, notch filtered using a filter with passband edges at 49 and 50 Hz, downsampled to 600 Hz, and run Independent Component Analysis (ICA; *runica*). We first flagged the components, using built-in EEGLAB functions, and then visually inspected them to reject those involving eye movements and muscle contraction (focusing on those with a probability of over 0.90 for belonging to the eye or muscle component category).

**Epoch extraction.** After downsampling the EEG data to 250 Hz, we segmented them into epochs time locked to two events of interest: discrimination agency response and confidence response.

#### *Report cues*

##### (A) Discrimination agency response

For the discrimination response, we segmented the data into epochs time locked from  $-150$  to  $500$  ms relative to when participants received the cue to register their choice of interval in which they experienced being more in control of (i.e., the appearance on the screen of the agency question). We baseline corrected these epochs using a time window of  $-150$  to  $0$  ms. We excluded the epochs time locked to the discrimination response cue from one participant because they were noisy.

##### (B) Confidence response

For the confidence response, we used epochs that started  $-150$  ms prior and ended  $500$  ms after participants received the cue to report their confidence in the preceding discrimination response. We baseline corrected for the time window  $150$ – $0$  ms relative to the confidence response cue.

After epoching the data to each time-locking event, we first automatically rejected noisy epochs using the function *pop\_autorej* with a threshold of  $70 \mu\text{V}$  for extremely large fluctuations (Mean number of rejected epochs time locked at: Cue for discrimination:  $M = 102.87 \pm 41.36$ ; Cue for confidence:  $M = 107.11 \pm 42.10$ ). We then visually inspected the output to manually reject any remaining noisy epochs (Discrimination cue epochs:  $M = 151.66 \pm 55.39$ ; Confidence cue epochs:  $M = 164.45 \pm 56.10$ ). Finally, we excluded a small set of epochs from those trials in which, due to a technical error, at least one of the triggers fired more than once, introducing uncertainty in the timing (Rejected epochs time locked at: Discrimination cue:  $M = 1.50 \pm 1.46$ ; no epochs were removed time locked to the confidence cue). Therefore, for the final analyses we used the following number of epochs at each time-locking event: Discrimination cue epochs:  $M = 234.32 \pm 70.53$ , min = 79; Confidence cue epochs:  $M = 224.97 \pm 65.46$ , min = 76. We chose to focus our analysis on the FCz electrode, because ERP studies suggest that neuronal correlates of agency, be that online monitoring of movement execution, outcome feedback, and both implicit and explicit agency measures can be identified in this area (Kühn et al., 2011; Bednark and Franz, 2014; Padrao et al., 2016; Vastano et al., 2020; Han et al., 2021; Ren et al., 2023; Gomez-Andres et al., 2024) and scalp topography.

### **Statistical analyses**

To compare the amplitude of the EEG signal for the epochs created, we fitted linear mixed-effect models using the “afex” package (Singmann et al., 2023). To account for trial-level variability and maintain the within-subject structure of



the data, we included all trials for each participant, without prior averaging. We then used cluster-based permutation tests to control for family-wise error rate (Maris and Oostenveld, 2007). For the latter, we permuted the factors' labels (used on each model) on each trial and fitted linear mixed-effect models on each time point 5,000 times. For this purpose, we used sum-to-zero coding (*contr.sum*) to define the contrasts among the factors and their interactions. This approach centers each factor, allowing its effects to be interpreted relative to the grand mean rather than a reference level. For each main effect and interaction, we got clusters of neighboring time points within the epoch when there was a significant difference between the conditions ( $p < 0.05$ ). Then, we estimated the sum of the  $F$  statistics for each cluster and saved the largest value. By repeating this process, we obtained the distribution of the sum statistics related to finding an effect by chance. We then tested whether the significant effects referred to time points that were both lower than our alpha (0.05) and fell above the 95th percentile of the time points of the sum-statistics distribution. These were exploratory analyses, in which we examined long epochs. Since the exact timing of the monitoring of the action and the experience of diminished agency is unknown, we prefer not to interpret the timing of the clusters but instead suggest that these exploratory analyses may reveal potential differences in the EEG signal.

For illustration purposes only, to explore how the main effects of these factors and their interactions across the scalp, we fit the data from the remaining 61 channels using the same linear mixed-effect models and then plotted the topographic maps of the mean beta estimates across the time points included in each of the significant clusters using the *eegUtils* package (Craddock, 2022) in R.

### Code accessibility

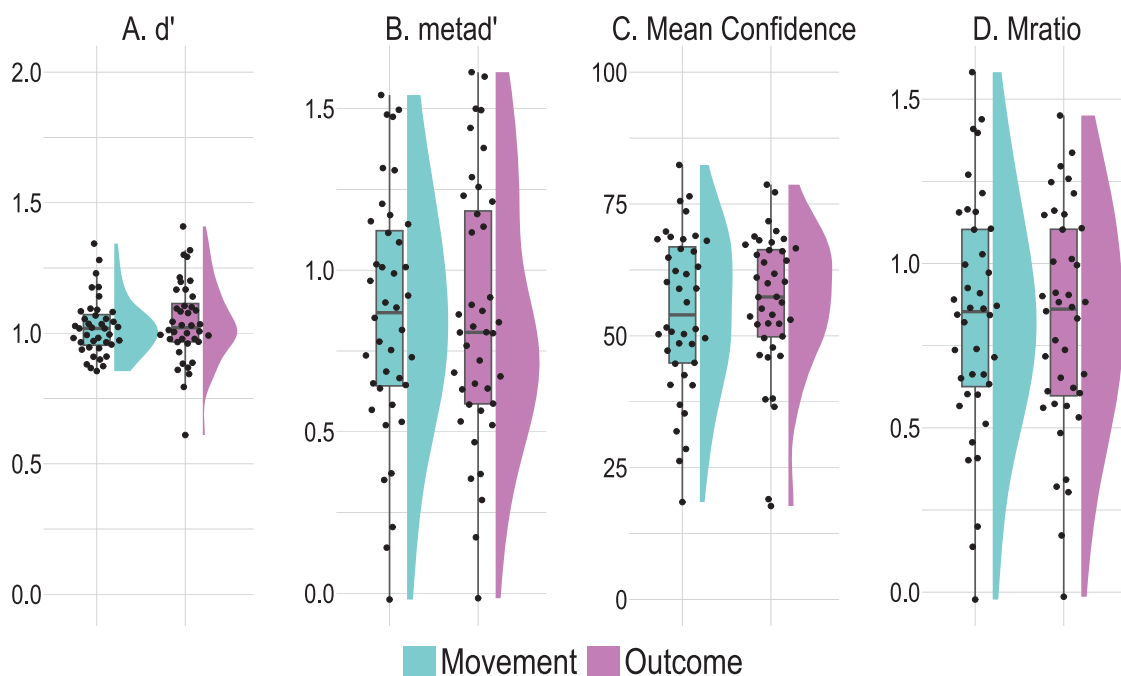
This study was preregistered (<https://osf.io/SJURK>), and any deviations from the preregistered plan are explicitly reported. The preprocessed data, as well as R and MATLAB scripts for the analyses, are freely available at (<https://osf.io/sm7rg>).

## Results

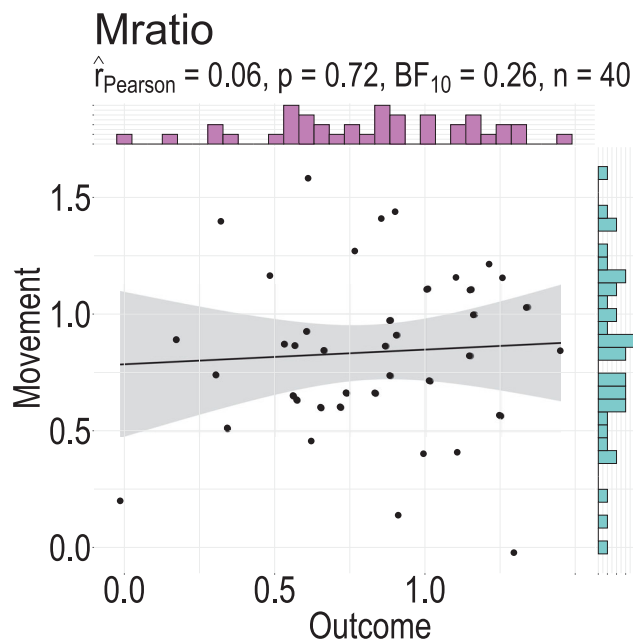
### Behavioral

First, we compared participants' discrimination performance—measured with  $d'$ —and found that there was no significant difference between the two conditions ( $d'_{\text{movement}} = 1.02 \pm 0.11$ ;  $d'_{\text{outcome}} = 1.04 \pm 0.15$ ;  $t_{(39)} = -0.42$ ,  $p = 0.677$ , Cohen's  $d = -0.07$ ,  $BF_{10} = 0.19$ ; Fig. 3A). That is, the manipulation did not significantly affect the discriminability of the intervals in terms of control. This result confirms that the staircase procedure employed to keep participants' accuracy stable in the discrimination task worked, allowing us to validly compare participants' Mratio values between the two conditions. Further, we found that mean confidence ratings did not differ between the two conditions (Mean confidence<sub>movement</sub> =  $54.45 \pm 15.03$ ; Mean confidence<sub>outcome</sub> =  $56.31 \pm 13.51$ ;  $t_{(39)} = -1.71$ ,  $p = 0.095$ , Cohen's  $d = -0.27$ ,  $BF_{10} = 0.64$ ; Fig. 3B). That is, we get weak evidence that participants' confidence was comparable when they assess the accuracy of their representations of agency, both when the manipulation affected the visual feedback of their movement or the proximal outcome. Additionally, we found that participants' metacognitive sensitivity, measured with  $\text{metad}'$ , was similar between the two conditions ( $\text{metad}'_{\text{movement}} = 0.86 \pm 0.38$ ;  $\text{metad}'_{\text{outcome}} = 0.86 \pm 0.40$ ;  $t_{(39)} = 0.04$ ,  $p = 0.968$ , Cohen's  $d = 0.006$ ,  $BF_{10} = 0.17$ ; Fig. 3C). Crucially, contrary to our preregistered hypothesis, we found that metacognitive efficiency, measured with Mratio, did not differ between the two conditions (Mratio<sub>movement</sub> =  $0.84 \pm 0.36$ ; Mratio<sub>outcome</sub> =  $0.82 \pm 0.34$ ;  $t_{(39)} = 0.25$ ,  $p = 0.806$ , Cohen's  $d = 0.04$ ,  $BF_{10} = 0.18$ ; Fig. 3D). This shows that metacognitive representations of agency are equally sensitive to violations of predictions of the representations of the movement and outcome. These results held after removing two participants with negative Mratio. We found no significant difference in participants' discrimination performance ( $d'$ ) between the two conditions ( $d'_{\text{movement}} = 1.03 \pm 0.02$ ;  $d'_{\text{outcome}} = 1.04 \pm 0.03$ ;  $t_{(37)} = -0.23$ ,  $p = 0.82$ , Cohen's  $d = -0.04$ ,  $BF_{10} = 0.18$ ). Further, there was no difference in participants' confidence rating (Mean confidence<sub>movement</sub> =  $54.67 \pm 2.40$ ; Mean confidence<sub>outcome</sub> =  $56.21 \pm 2.23$ ;  $t_{(37)} = -1.47$ ,  $p = 0.151$ , Cohen's  $d = -0.24$ ,  $BF_{10} = 0.47$ ) or metacognitive sensitivity ( $\text{metad}'_{\text{movement}} = 0.90 \pm 0.05$ ;  $\text{metad}'_{\text{outcome}} = 0.87 \pm 0.06$ ;  $t_{(37)} = 0.40$ ,  $p = 0.694$ , Cohen's  $d = 0.06$ ,  $BF_{10} = 0.19$ ). Lastly, participants' Mratio did not differ between the two conditions (Mratio<sub>movement</sub> =  $0.88 \pm 0.05$ ; Mratio<sub>outcome</sub> =  $0.83 \pm 0.05$ ;  $t_{(37)} = 0.70$ ,  $p = 0.497$ , Cohen's  $d = 0.11$ ,  $BF_{10} = 0.22$ ).

In an exploratory analysis, we then examined the relationship between metacognitive representations of agency in the two conditions. We wanted to test whether participants' sensitivity to movement and outcome prediction violations were not only comparable but also correlated. We found that the correlation was not significant ( $t_{(38)} = 0.36$ , Pearson's  $r = 0.06$ ,  $p = 0.72$ ,  $CI = [-0.26, 0.36]$ ,  $BF_{10} = 0.26$ ,  $n = 40$ ; Fig. 4). Excluding the two participants with negative Mratio from the correlation analysis had no effect on the results ( $t_{(36)} = 0.24$ , Pearson's  $r = 0.04$ ,  $p = 0.81$ ,  $CI = [-0.28, 0.35]$ ,  $BF_{10} = 0.25$ ,  $n = 38$ ). We interpret the absence of differences between the Mratios in the two conditions as evidence that visual manipulations of the movement and the outcome might similarly affect metacognitive representations of agency judgments at the population level, and the lack of correlation may point toward individual level variability. In other words, some participants are more sensitive to movement feedback manipulations compared with outcome feedback manipulations, while for others the opposite is true. Importantly, this individual level variability is not attributable to the staircase procedures employed to keep participants' performance comparable across conditions. Stimulus variability was not associated with confidence ratings and did not account for the Mratio effects (see also Extended Data Fig. 4-1).



**Figure 3.** **A**,  $d'$ : Discrimination performance. **B**,  $metad'$ : Metacognitive sensitivity. **C**, Mean Confidence ratings and **D**, Mratio: Metacognitive efficiency. The violin plots depict the smoothed distributions of the data per condition, and the boxplots depict the interquartile range. The black circles represent the estimates per participant.



**Figure 4.** Correlation between the Mratios of the two conditions. No significant correlation was found between the Mratios of the Movement and Outcome condition. The black line represents the regression line, and the shaded area illustrates the 95% confidence interval. The black circles represent the Mratio for each participant and the histograms the distribution of the data for each condition (Extended Data Fig. 4-1 indicates that this pattern cannot be directly explained by differences in the magnitude of manipulation in each condition or by the confidence participants reported as a result).

### Electrophysiology

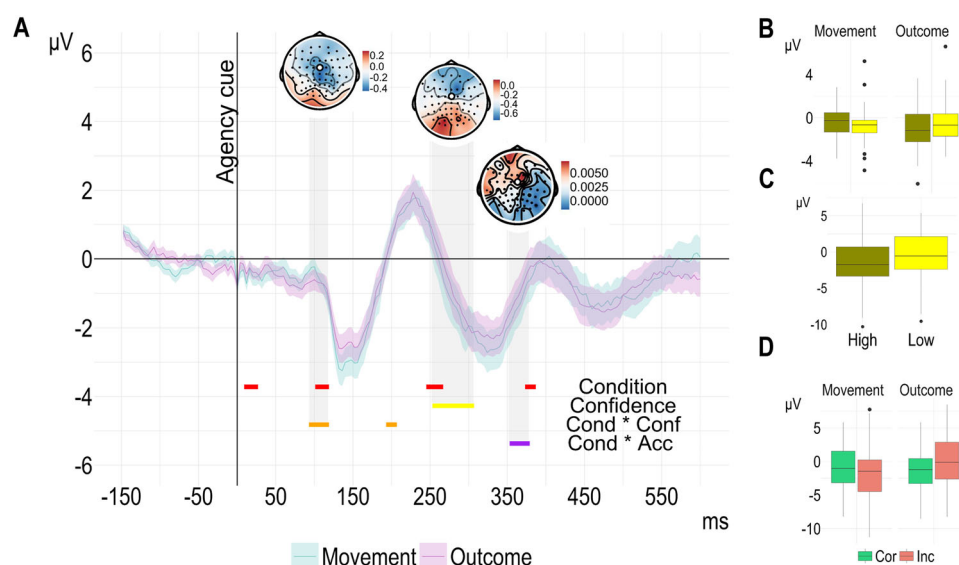
We then looked at the EEG signal to explore whether the different representation violations due to the movement or proximal outcome manipulations are linked with distinct brain signals.

### Report cues

We then sought for differences in the EEG signal between conditions at the point participants were prompted to (A) report their discrimination decision and (B) their confidence rating. For readability, from now on, we will call them “Discrimination cue” and “Confidence cue” epochs, respectively. The aim of these analyses was to test for changes in the brain signals resulting from the different types of feedback participants received while performing the Skittles game. Specifically, we aimed to determine whether the brain signals recorded while participants were prompted to register their agency discrimination and confidence reports respectively differed significantly depending on whether the manipulation was introduced during the arm movement (Movement trials) or during the outcome phase (Outcome trials) of the action they had just completed. To compare for differences among the time points, we tested the following linear mixed-effect model  $\text{EEG amplitude} \sim \text{condition} * \text{confidence} * \text{accuracy in discrimination response} + (1 | \text{participant})$  applied to each time point of the Discrimination cue and Confidence cue epochs. For both, we used a time window starting at 0 s and finishing at 500 ms from when participants were prompted. These time windows were chosen based on the average time participants needed to respond (discrimination response:  $651.36 \pm 148.63$  ms; confidence response:  $844.10 \pm 238.37$  ms), to focus our analysis on the underlying signal linked with the discrimination and confidence task and avoid motor artifacts and other processes contaminate the neural signals. We expected that if the type of feedback plays a significant role in how participants experience their sense of agency during the action, then the neural signature would differ between conditions (Movement vs Outcome) during Discrimination cue and Confidence cue epochs. Such a result would align with the absence of correlation between the Mratios of the two conditions. It would also provide preliminary evidence that bodily and external agency are not simply facets of single, unified experience, and emphasize the need to study how each component contributes to an overall agency.

#### (A) Discrimination agency response

When we explored differences in the epochs time locked to the discrimination cue, we found four significant clusters linked to the effect of condition (onset  $\sim 12$ , 104, 248 and 372 ms); one cluster linked to confidence (starting  $\sim 256$  ms); one cluster linked to the interaction between condition and discrimination accuracy (onset  $\sim 356$  ms); and one cluster based on the interaction between condition and confidence (onset  $\sim 96$  ms). As shown in Figure 5, by averaging the signal of all the time points included in the confidence cluster, we found that high confidence was linked with smaller EEG amplitude compared with low confidence. However, when averaging the signal of all the time points included in the cluster associated with the interaction between condition and confidence, we noticed that the amplitude of the EEG signal was greater for high-confidence Movement trials compared with low-confidence Movement trials and smaller for high-confidence



**Figure 5.** EEG amplitude for the two conditions (FCz electrode). **A**, Time series: the signal was time locked to the onset of the discrimination cue (when participants provide their agency judgment). For illustration purposes, in this figure we averaged the epoched signal per participant and condition and plotted the mean of the data from the 37 participants included in this analysis. The ribbons around the mean represent the 95% confidence interval. The significant time point clusters from the cluster-based permutation analysis [EEG amplitude difference  $\sim \text{Condition} * \text{Accuracy} * \text{Confidence} + (1 | \text{participant})$ ] are color-coded to indicate the effect they represent, marked on the bottom-right. The topographical maps show the mean beta estimates across the time points included in the clusters of the EEG amplitude marked within the gray area for each of the 62 channels recorded. The boxplots show the mean EEG amplitude across the time points included in the clusters (areas marked with gray) linked with the interaction between condition and confidence (**B**), the main effect of confidence (**C**), the interaction between condition, and accuracy (**D**). To simplify the visualization, we median split the continuous confidence ratings for each participant and used the resulting binary factor of confidence (“High” vs “Low”) only for this figure.

Outcome trials compared with low-confidence Outcome trials (onset ~96 ms). This result suggests that the confidence level is linked with the amplitude of the EEG signal already when participants are prompted to provide their discrimination response and varies by the type of feedback manipulation introduced during the action. Similarly, by averaging the signal of all the time points included in the cluster associated with the interaction between condition and discrimination accuracy (onset ~356 ms), we see that the amplitude of the EEG signal was greater for correct Movement trials compared with incorrect Movement trials and greater for incorrect Outcome trials compared with correct Outcome trials. This result suggests that the neural signal differs between conditions during the Discrimination cue epoch and implies the presence of different underlying mechanisms engaged based on the type of agency representation being manipulated.

### (B) Confidence response

The permutation analysis on the epoched data time locked to Confidence cue revealed five significant clusters. The first was linked to the effect of condition (onset ~24 ms), the second to the effect of discrimination accuracy (onset ~36 ms), the third and fourth to the effect of confidence (onset ~136 ms; ~216 ms), and the fifth cluster indicated an interaction between confidence and discrimination accuracy (onset ~40 ms). The time points included in the cluster indicating an interaction between confidence and discrimination accuracy overlap with those associated with the main effect of discrimination accuracy. Therefore, we focus here on the interaction effect. By averaging the signal across all time points within the cluster associated with the two-way interaction between discrimination accuracy and confidence, we found that the amplitude of the EEG signal was greater when participants correctly identified the interval in which they were in control compared with when their decision was incorrect. This difference was more pronounced for the low-confidence trials compared with high-confidence trials. Overall, though, low-confidence trials were associated with smaller EEG amplitude compared with high-confidence trials, as evidenced by averaging the signal of all the time points included in the second confidence cluster (onset ~216 ms). Finally, by averaging the signal over the time points included in the condition cluster, we found that EEG signal amplitude was greater in Outcome trials than that in Movement trials. Taken together, the results from the response cues suggest, first, that violations of movement representation prediction are associated with distinct neural processes compared with violations of outcome representations when participants assess their agency through explicit agency reports and then rate their confidence in their decisions and second, that both response accuracy and confidence levels can predict the neural signal during these assessments.

## Discussion

We tested the role of movement and outcome information in metacognitive representations of agency. That is, we measured the clarity of the subjective experience of agency when either the visual feedback of a movement (Movement condition) or its proximal outcome (Outcome condition) is manipulated. Participants performed a goal-oriented action two consecutive times. Then, they discriminated on which of the two actions they felt more in control and rated their confidence in that decision. Our study showed that Mratios of the Movement and Outcome conditions did not differ but they also did not correlate. We also found evidence for distinct EEG signals when we prompted participants to register their agency (discrimination) and confidence reports. Here, we will argue that, overall, the results of the behavioral and EEG data are in line with the idea (Kalckert and Ehrsson, 2012; Christensen and Grünbaum, 2017; Wen, 2019) that bodily and external agency should not be collapsed into a single construct. In what follows, we discuss each of these pieces of evidence separately and then bring them together to discuss how they might inform more theoretical accounts of agency. We will also, however, highlight the limitations of this study and encourage further work that might provide stronger evidence for or against a neurophysiological distinction of the two conceptually distinct kinds of agency.

The absence of differences in Mratio between the two kinds of manipulations suggests that metacognitive representations of agency are equally sensitive to movement and outcome prediction violations. This result goes against our preregistered hypothesis that the mean Mratio would differ between the two conditions. Instead, we found in fact evidence for the null hypothesis of no differences between the conditions. Hence, this result suggests that the salience, or precision, of experiencing a violation of predictions following an action is not consistently different, for the sampled population, depending on whether those predictions are bodily or external. Because we measured the precision of metacognitive representations, and not, like previous literature, merely mean agency ratings, a direct comparison between these results and previous studies must be taken with caution. These two representational levels—agency and metacognitive judgments about it—have been argued to be distinct, both behaviorally (Charalampaki et al., 2024) and computationally (Constant et al., 2022). Nevertheless, these results do complement the field. Previous studies examining the relative importance of movement and outcome prediction violations in participants' sense of agency have yielded conflicting findings. Some studies have argued that participants' sense of agency is more sensitive to outcome violations than to movement violations (Wen et al., 2015; David et al., 2016; Oishi et al., 2018). Other studies have provided evidence for the opposite, namely, that participants' sense of agency is more sensitive to movement violations relative to outcome violations (Metcalf and Greene, 2007; Metcalf et al., 2013).

In addition to using metacognitive judgments of agency, our study differs from previous work in two important ways: (1) the use of qualitatively comparable feedback manipulations in the two conditions and (2) an experimental design in which



the distal goal was reached on almost every trial for all participants. Regarding the first point, in much of previous work, the manipulations compared are typically qualitatively different from one another. For instance, in movement manipulations, visual feedback is usually continuous and presented for a sustained period of time; whereas in outcome manipulations, visual feedback is usually discrete, presented for a shorter duration, and typically consists of a sound or color change at the end of the action (Wen et al., 2015; David et al., 2016; Oishi et al., 2018) or the outcome congruency (Metcalf and Greene, 2007; Metcalfe et al., 2013). This comparison is problematic because previous work has shown that the duration of the feedback can overshadow or conceal the true effect of movement and outcome prediction violations. Schmitter et al. (2021) found that continuous visual feedback differed from discrete feedback not only in participants' accuracy in detecting a delay but also in the BOLD activity patterns. The authors interpreted these results as indicating that distinct predictive mechanisms underlie the processing of the two different types of manipulation (Schmitter et al., 2021). In our study, we manipulated the visual feedback of both the movement and the proximal outcome continuously. Put simply, we manipulated either the visual representation of participants' hand movement or the ball flight trajectory, both of which have a long duration relative to the short-lived outcome representations employed in previous studies. Therefore, we argue that the Skittles task allowed us to compare the two types of action representations in a tightly controlled manner, minimizing response and comparison biases. This does not undermine the findings from previous studies but rather highlights the limitations of comparing simple subjective reports between experimental conditions that are known to participants.

Regarding the distal goal, Kumar and Srinivasan (2014) showed how goal-dependent the relative role of movement and distal outcome information is: In a multi-agent task, they found that when participants successfully achieved their goal, they rated agency independently of their motor performance. However, when participants missed their goal, they relied more on the sensorimotor information (Did what I did match what I saw?) rather than the distal outcome congruency to rate their agency. Further, van der Weiden et al. (2010) showed that there are interindividual differences in the way participants represent their behavior (movement vs distal outcome) and that these representations can be prompted by the use of subliminal primes (prior to the action) or specific instructions ("make a specific movement" vs "reach a specific goal"). Together, these studies suggest, first, that reaching a goal can affect agency ratings following outcome manipulations differently than ratings following movement manipulations and, second, that different participants can differ in these relative effects. Importantly, we designed the virtual Skittles scene in such a way as to ensure participants would be successful in meeting their goal of hitting the target in the majority of the trials. On rare occasions, when participants missed the target in either of the intervals, we excluded the trials from further analysis. Hence, we argue that goal-based expectations were not a confound in our design.

Further, the second finding of our study was that participants' *Mratio* following movement manipulations did not correlate with *Mratio* following proximal outcome manipulations. Simply put, those participants in our sample who were most metacognitively sensitive to movement violations were not always those most sensitive to proximal outcome violations. This result suggests that a simple comparison between mean judgments of agency following different kinds of manipulations might not be sufficient to reveal the dissociations between them: Two different manipulations might affect the experience of agency to the same extent and nevertheless be processed via distinct mechanisms or contribute with different weights to the overall experience of agency (partial independent mechanisms). Because our preregistered hypotheses focused on the mean differences between conditions, we determined our sample size based on power analyses to detect mean differences rather than correlations; hence, we remain cautious with this interpretation.

Finally, when examining the EEG signal recorded at FCz during the time of the discrimination and confidence decisions, we found it to reflect the accuracy and confidence of the decision along with the type of violation prediction. At this point of the trial, participants are discriminating between two action epochs; thus, what we measure here is the underlying signal of the agency salience comparison between those two preceding action epochs. The association of the amplitude difference with the type of violation prediction suggests that the brain can dissociate the two types of violation based on the salience of the experience they affect. Together, the effects of condition at the time of the discrimination and confidence decision suggest that the salience of the agency experience might vary between kinds of manipulation. Further work should test this potential dissociation by extending our analysis to other cortical regions linked to the sense of agency and employing advanced methods (e.g., connectivity analysis, multivariate decoding) to better characterize the underlying signal of each.

### **Bodily and external representations and the underlying model of agency**

To reiterate, our study showed that *Mratios* of the Movement and Outcome conditions did not differ but they also did not correlate. We also found evidence for distinct EEG signals when we prompted participants to register their agency (discrimination) and confidence reports. These results emphasize the need to not conflate bodily and external agency into a single unified experience. The comparator model (Frith et al., 2000) is often used to explain the experience of (or reduction in) both the bodily and external sense of agency. There exists evidence to support the involvement of a comparator mechanism in the bodily sense of agency (but see Christensen and Grünbaum, 2018). However, the involvement of forward models in the comparison between predicted and perceived sensory consequences that go beyond the body has been challenged (Christensen and Grünbaum, 2018; Dogge et al., 2019). While a previous study has shown that participants can predict the distal outcome of their actions in the environment before even receiving explicit feedback about it (Maurer et al., 2022), it cannot be automatically assumed that a copy of motor commands is responsible for the predictions

about sensory consequences that go beyond the body (Christensen and Grünbaum, 2018; Dogge et al., 2019). The comparator could explain the role of motor incongruencies and, by extension, bodily representation violations in a diminished sense of agency. If there was evidence that the comparator model could make predictions based on the forward models for sensory information external to the body, it would mean that the comparator could explain the role of outcome incongruencies and, by extension, violations of bodily and external predictions should result in both correlated metacognitive agency representations and indistinguishable underlying signal when participants are prompted to report their discrimination agency decision. However, based on our results, it seems less likely that a single comparator mechanism is responsible for processing violations of both representations. While a formal model comparison is beyond the scope of this work, we speculate that a more inclusive model of agency that extends the core ideas put forward by the comparator model—like the Bayesian cue integration theory (Moore and Fletcher, 2012; Synofzik et al., 2013)—might better align with our results. The Bayesian cue integration theory, for example, posits that a sense of agency results from a weighted combination of different cues and puts forward a computational mechanism to integrate these cues, some of which result from a comparison similar to that supported by the simplest form of the comparator model, and other cues that may not. The weighting assigned to each of the cues can be determined by priors (prior knowledge and expectations). Because the model is, first, agnostic about how exactly these sources of information arise and, second, because it suggests that these cues are integrated, the cue integration model may be more suited to account for our results, specifically the differences found due to the salience of the agency experience. However, this claim remains speculative at this point and requires further validation. For instance, to conclusively argue for the distinct contributions of movement and outcome prediction violations, future studies might operationalize both types of violations within the same trial. Specifically, introduce violations of bodily representations in one of the intervals and violations of the external representations in the other, while varying the magnitude of manipulation for each one used within the trial. Then, as we did here, test for differences in participants' metacognitive representations of agency, behaviorally and at the neuronal level. This design would allow one to test if the absence of correlation we found in our study and the distinct brain signal is the result of a weighted contribution of different sources of information while forming representations of agency and run computational models to identify the way the brain assigns different weights to each agency representation.

## Limitations

In this study, we used *Mratio* to compare the precision of two potentially distinct agency representations: bodily and external. This metric allows for a direct comparison of the two otherwise not-comparable representations, but it may be influenced by first-order discrimination performance. To ensure this was not the case in our study, we implemented separate staircase procedures. This approach enabled us to examine how two types of prediction violations (movement related and proximal outcome related) influence metacognitive representations of agency, beyond mere differences in discrimination performance (but see Guggenmos, 2021; Rausch et al., 2023). However, as Rahnev and Fleming (2019) have noted, using staircases carries the risk of inflating estimates of metacognitive efficiency. Beyond this, we observed high variability in *Mratio*, which may be driven by factors such as stimulus strength and variability. Although we accounted for both the magnitude of the manipulation and staircase variability (used here as a proxy for stimulus variability), in our analyses of confidence ratings and *Mratio*, this remains an inherent limitation of our design. More broadly, when comparing the influence of bodily versus external information on the formation of agency representations, we are necessarily comparing two components of action that differ substantially in both space and time. Until more refined methods for assessing metacognitive abilities are available, we argue that our approach provided a suitable framework for addressing these questions.

Finally, we investigated the role of bodily predictions' violations by manipulating the visual representation of participants' hands, which appeared on the screen as a simple rectangle. Previous research has shown that participants can experience ownership over a variety of visual representations of their body parts (Kondo et al., 2018). Thus, we consider our results relevant to understanding the sense of bodily agency. Nevertheless, future work should aim to replicate and extend our findings using more realistic visual representations of bodily movements.

## Future directions

We used a metacognitive task to study the subjective experience of agency devoid of response biases. Another option would be to study agency with implicit measures. For example, temporal binding consists of biases in the perceived timing of the movement and the outcome of voluntarily performed actions (Haggard et al., 2002): movement initiation is perceived as happening later, and outcome as happening earlier, than they really did. Alternatively, sensory attenuation, where the sensory consequences of voluntarily performed actions are perceived as less intense (Blakemore et al., 2000) compared with those of passive movements, is also sometimes used to quantify the sense of agency. However, these measures do not directly address the subjective component of agency; rather, they act as a proxy. In particular, there is an ongoing discussion on the link between implicit and explicit measures of agency, the details of which are outside the scope of this manuscript. Briefly, the discussion revolves around whether these two kinds of measures target the same phenomenal experience. While some studies argue that they do (Imaizumi and Tanno, 2019), others argue against it (Dewey and Knoblich, 2014; Schwarz et al., 2019). Therefore, it would be interesting to compare the role of bodily and external representations on agency with implicit agency measures. Not only to inform the discussion of implicit versus explicit measures, but also to compare our results using a metacognitive measure and those from studies using implicit measures. Given the

nature of our design, where participants perform an action and receive visual feedback that is spatially incongruent with their action (either at the level of the movement or the level of the proximal outcome) adapting it directly into an implicit agency task may be challenging. However, an alternative motor task can be developed to clearly distinguish between the bodily and external aspects of agency allowing adjustments suitable for both metacognitive and implicit agency designs.

## Conclusions

We found that the sensory information of both the movement and the outcome is important for experiencing being the author of an action we intentionally perform. However, we should not treat these two sources of information as interchangeable as this overlooks a possible distinct contribution to an overall sense of agency as suggested by our results. Further research is needed to understand how movement and outcome representations inform our sense of agency.

## References

- Arbuzova P, Peters C, Röd L, Koß C, Maurer H, Maurer LK, Müller H, Verrel J, Filevich E (2021) Measuring metacognition of direct and indirect parameters of voluntary movement. *J Exp Psychol Gen* 150:2208–2229.
- Bednark JG, Franz EA (2014) Agency attribution: event-related potentials and outcome monitoring. *Exp Brain Res* 232:1117–1126.
- Blakemore S-J, Wolpert D, Frith C (2000) Why can't you tickle yourself? *Neuroreport* 11:R11.
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
- Charalampaki A, Ninija Karabanov A, Ritterband-Rosenbaum A, Bo Nielsen J, Roman Siebner H, Schram Christensen M (2022) Sense of agency as synecdoche: multiple neurobiological mechanisms may underlie the phenomenon summarized as sense of agency. *Conscious Cogn* 101:103307.
- Charalampaki A, Peters C, Maurer H, Maurer LK, Müller H, Verrel J, Filevich E (2023) Motor outcomes congruent with intentions may sharpen metacognitive representations. *Cognition* 235:105388.
- Charalampaki A, Ciston AB, Filevich E (2024) Contributions of tactile information to the sense of agency and its metacognitive representations. *J Exp Psychol Gen* 153:2427–2440.
- Christensen MS, Grünbaum T (2017) Sense of moving: moving closer to the movement. In: *Sensation of movement* (Grünbaum T, Christensen MS, eds), pp 64–84. Abingdon, UK: Routledge.
- Christensen MS, Grünbaum T (2018) Sense of agency for movements. *Conscious Cogn* 65:27–47.
- Constant M, Salomon R, Filevich E (2022) Judgments of agency are affected by sensory noise without recruiting metacognitive processing. *Elife* 11:e72356.
- Craddock M (2021) metaSDT: Calculate type 1 and type 2 signal detection measures (version 0.6.0).
- Craddock M (2022) eegUtils: Utilities for electroencephalographic (EEG) analysis. Available at: <https://github.com/craddm/eegUtils>, <https://craddm.github.io/eegUtils>
- David N, Skoruppa S, Gulberti A, Schultz J, Engel AK (2016) The sense of agency is more sensitive to manipulations of outcome than movement-related feedback irrespective of sensory modality. *PLoS One* 11:e0161156.
- Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134:9–21.
- Dewey JA, Knoblich G (2014) Do implicit and explicit measures of the sense of agency measure the same thing? *PLoS One* 9:e110118.
- Dogge M, Custers R, Aarts H (2019) Moving forward: on the limits of motor-based forward models. *Trends Cogn Sci* 23:743–753.
- Farrer C, Franck N, Georgieff N, Frith CD, Decety J, Jeannerod M (2003) Modulating the experience of agency: a positron emission tomography study. *Neuroimage* 18:324–333.
- Farrer C, Bouchereau M, Jeannerod M, Franck N (2008) Effect of distorted visual feedback on the sense of agency. *Behav Neurol* 19:53–57.
- Farrer C, Valentin G, Hupé JM (2013) The time windows of the sense of agency. *Conscious Cogn* 22:1431–1441.
- Fleming SM, Lau HC (2014) How to measure metacognition. *Front Hum Neurosci* 8:443.
- Frith CD, Blakemore S-J, Wolpert DM (2000) Abnormalities in the awareness and control of action. *Philos Trans R Soc Lond B Biol Sci* 355:1771–1788.
- Gallagher S (2012) Multiple aspects in the sense of agency. *New Ideas Psychol* 30:15–31.
- Gomez-Andres A, Cerda-Company X, Cucurell D, Cunillera T, Rodríguez-Fornells A (2024) Decoding agency attribution using single trial error-related brain potentials. *Psychophysiology* 61:e14434.
- Guggenmos M (2021) Measuring metacognitive performance: type 1 performance dependence and test-retest reliability. *Neurosci Conscious* 2021:niab040.
- Haggard P (2017) Sense of agency in the human brain. *Nat Rev Neurosci* 18:196–207.
- Haggard P, Clark S, Kalogeras J (2002) Voluntary action and conscious awareness. *Nat Neurosci* 5:382–385.
- Han N, Jack BN, Hughes G, Elijah RB, Whitford TJ (2021) Sensory attenuation in the absence of movement: differentiating motor action from sense of agency. *Cortex* 141:436–448.
- Imaizumi S, Tanno Y (2019) Intentional binding coincides with explicit sense of agency. *Conscious Cogn* 67:1–15.
- Jeunet C, Albert L, Argelaguet F, Lécuyer A (2018) “Do you feel in control?” Towards novel approaches to characterise, manipulate and measure the sense of agency in virtual environments. *IEEE Trans Vis Comput Graph* 24:1486–1495.
- Kalckert A, Ehrsson HH (2012) Moving a rubber hand that feels like your own: a dissociation of ownership and agency. *Front Hum Neurosci* 6:40.
- Kang SY, Im C-H, Shim M, Nahab FB, Park J, Kim D-W, Kakareka J, Miletta N, Hallett M (2015) Brain networks responsible for sense of agency: an EEG study. *PLoS One* 10:e0135261.
- Kassambara A (2023) \_rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.7.2. Available at: <https://CRAN.R-project.org/package=rstatix>
- Kilteni K, Engeler P, Boberg I, Maurel X, Ehrsson HH (2021) No evidence for somatosensory attenuation during action observation of self-touch. *Eur J Neurosci* 54:6422–6444.
- Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in Psychtoolbox-3. *Perception* 36:1.
- Kondo R, Sugimoto M, Minamizawa K, Hoshi T, Inami M, Kitazaki M (2018) Illusory body ownership of an invisible body interpolated between virtual hands and feet via visual-motor synchronicity. *Sci Rep* 8:7541.
- Krugwasser AR, Harel EV, Salomon R (2019) The boundaries of the self: the sense of agency across different sensorimotor aspects. *J Vis* 19:14.
- Kühn S, Nenchev I, Haggard P, Brass M, Gallinat J, Voss M (2011) Whodunnit? Electrophysiological correlates of agency judgments. *PLoS One* 6:e28657.
- Kumar D, Srinivasan N (2014) Naturalizing sense of agency with a hierarchical event-control approach. *PLoS One* 9:e92431.

- Levitt HC (1971) Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49:467–477.
- Maniscalco B, Lau H (2012) A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious Cogn* 21:422–430.
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190.
- Maurer LK, Maurer H, Hegele M, Müller H (2022) Can Stephen Curry really know?—Conscious access to outcome prediction of motor actions. *PLoS One* 17:e0250047.
- Metcalfe J, Greene MJ (2007) Metacognition of agency. *J Exp Psychol Gen* 136:184–199.
- Metcalfe J, Eich TS, Miele DB (2013) Metacognition of agency: proximal action and distal outcome. *Exp Brain Res* 229:485–496.
- Moore JW, Fletcher PC (2012) Sense of agency in health and disease: a review of cue integration approaches. *Conscious Cogn* 21:59–68.
- Morey RD, Rouder JN (2018) BayesFactor: Computation of bayes factors for common designs. In R package version 0.9.12-4.2. Available at: <https://CRAN.R-project.org/package=BayesFactor>
- Müller H, Sternad D (2004) Decomposition of variability in the execution of goal-oriented tasks: three components of skill improvement. *J Exp Psychol Hum Percept Perform* 30:212–233.
- Oishi H, Tanaka K, Watanabe K (2018) Feedback of action outcome retrospectively influences sense of agency in a continuous action task. *PLoS One* 13:e0202690.
- Oldfield RC (1971) Edinburgh handedness inventory. *J Abnorm Psychol* 9:97–113.
- Osumi M, Nobusako S, Zama T, Yokotani N, Shimada S, Maeda T, Morioka S (2019) The relationship and difference between delay detection ability and judgment of sense of agency. *PLoS One* 14:e0219222.
- Pacherie E (2007) The sense of control and the sense of agency. *Psyche* 13:1–30.
- Padrao G, Gonzalez-Franco M, Sanchez-Vives MV, Slater M, Rodriguez-Fornells A (2016) Violating body movement semantics: neural signatures of self-generated and external-generated errors. *Neuroimage* 124:147–156.
- Patil I (2021) Visualizations with statistical details: the ‘ggstatsplot’ approach. *J Open Source Softw* 6:3167.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.
- Rahnev D, Fleming SM (2019) How experimental procedures influence estimates of metacognitive ability. *Neurosci Conscious* 2019:niz009.
- Rausch M, Hellmann S, Zehetleitner M (2023) Measures of metacognitive efficiency across cognitive models of decision confidence. *Psychol Methods*; Advance online publication.
- R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>
- Ren Q, Kaiser J, Gentsch A, Schütz-Bosbach S (2023) Prepared to stop: how sense of agency in a preceding trial modulates inhibitory control in the current trial. *Cereb Cortex* 33:8565–8580.
- Ritterband-Rosenbaum A, Nielsen JB, Christensen MS (2014) Sense of agency is related to gamma band coupling in an inferior parietal-preSMA circuitry. *Front Hum Neurosci* 8:510.
- Sato A, Yasuda A (2005) Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition* 94:241–255.
- Schmitter CV, Steinsträter O, Kircher T, van Kemenade BM, Straube B (2021) Commonalities and differences in predictive neural processing of discrete vs continuous action feedback. *Neuroimage* 229:117745.
- Schwarz KA, Weller L, Klaffehn AL, Pfister R (2019) The effects of action choice on temporal binding, agency ratings, and their correlation. *Conscious Cogn* 75:102807.
- Sidarus N, Vuorre M, Haggard P (2017) How action selection influences the sense of agency: an ERP study. *Neuroimage* 150:1–13.
- Singmann H, Bolker B, Westfall J, Aust F, Ben-Shachar M (2023) \_afex: Analysis of factorial experiments. R package version 1.3-0. Available at: <https://CRAN.R-project.org/package=afex>
- Spengler S, von Cramon DY, Brass M (2009) Was it me or was it you? How the sense of agency originates from ideomotor learning revealed by fMRI. *Neuroimage* 46:290–298.
- Sternad D, Abe MO, Hu X, Müller H (2011) Neuromotor noise, error tolerance and velocity-dependent costs in skilled performance. *PLoS Comput Biol* 7:e1002159.
- Synofzik M, Vosgerau G, Newen A (2008) I move, therefore I am: a new theoretical framework to investigate agency and ownership. *Conscious Cogn* 17:411–424.
- Synofzik M, Vosgerau G, Voss M (2013) The experience of agency: an interplay between prediction and postdiction. *Front Psychol* 4:127.
- Tieri G, Tidoni E, Pavone EF, Aglioti SM (2015) Mere observation of body discontinuity affects perceived ownership and vicarious agency over a virtual hand. *Exp Brain Res* 233:1247–1259.
- van den Bos E, Jeannerod M (2002) Sense of body and sense of action both contribute to self-recognition. *Cognition* 85:177–187.
- van der Weiden A, Aarts H, Ruys KI (2010) Reflecting on the action or its outcome: behavior representation level modulates high level outcome priming effects on self-agency experiences. *Conscious Cogn* 19:21–32.
- van der Weiden A, Ruys KI, Aarts H (2013) A matter of matching: how goals and primes affect self-agency experiences. *J Exp Psychol Gen* 142:954–966.
- Vastano R, Ambrosini E, Ulloa JL, Brass M (2020) Action selection conflict and intentional binding: an ERP study. *Cortex* 126:182–199.
- Wegner DM, Sparrow B, Winerman L (2004) Vicarious agency: experiencing control over the movements of others. *J Pers Soc Psychol* 86:838–848.
- Wen W, Yamashita A, Asama H (2015) The sense of agency during continuous action: performance is more important than action-feedback association. *PLoS One* 10:e0125226.
- Wen W, Yamashita A, Asama H (2017) Measurement of the perception of control during continuous movement using electroencephalography. *Front Hum Neurosci* 11:392.
- Wen W (2019) Does delay in feedback diminish sense of agency? A review. *Conscious Cogn* 73:102759.
- Zito GA, de Sousa Ribeiro R, Kamal E, Ledergerber D, Imbach L, Polania R (2023) Self-modulation of the sense of agency via neuro-feedback enhances sensory-guided behavioral control. *Cereb Cortex* 33:11447–11455.