Illusion of agency over involuntary muscle movements reveals temporal dynamics of neural activity underlying self-agency judgments

John P. Veillette[1†], Pedro Lopes[2], Howard C. Nusbaum[1]

[1]Department of Psychology, University of Chicago
[2]Department of Computer Science, University of Chicago

[†]Correspondence should be addressed to John P. Veillette; E-mail: johnv@uchicago.edu

Author Contributions (CRediT Taxonomy)

J.P.V.: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing - original draft, and Writing - review & editing. **P.L.**: Conceptualization, Funding acquisition, Methodology, Resources, Software, and Writing - review & editing. **H.C.N.**: Conceptualization, Funding acquisition, Resources, Supervision, and Writing - review & editing.

Acknowledgements

P.L. and H.C.N. were supported by NSF NCS 2024923, and J.P.V. was supported by NSF GRFP DGE 1746045 and a Neubauer Family Distinguished Doctoral Fellowship.

Conflict of Interest Statement

The authors declare no competing interests.

Abstract

Sense of agency (SoA), the conscious sensation that "I" am causing my own bodily movements, is a fundamental feature of human experience and a core facet of self-awareness. While prior work has measured participants' SoA over action consequences (e.g., a tone resulting from a button press), the neural mechanisms underlying SoA over *action itself* remains poorly understood since control of *one's own muscles* is normally unambiguous. In this work, we preempt subjects' own movements during a reaction time task using electrical muscle stimulation (EMS). On each trial, subjects report whether their movements were self-caused or caused by EMS. Then, by adaptively adjusting the timing of EMS delivery using Bayesian Optimization, we caused subjects to erroneously attribute roughly half of such involuntary (EMS-induced) movements to their own volition. Comparing trials with illusory agency to those without a sense of agency, we identify multivariate patterns of EEG activity that predict trial-by-trial SoA over EMS-caused movements across subjects. We show these EEG-derived predictors transition from transient to sustained dynamics as predicted by dual-process theories of action monitoring. In an exploratory analysis, we additionally decode SoA from nonlinear fractal measures of trial-by-trial brain activity. Results suggest that while the complexity (i.e., fractal dimension) of activity throughout the cortex is robustly higher when subjects experience agency over a behavior, but electrodes over motor regions appear to be selectively less sensitive to local perturbation (i.e., suppressed long-range correlations) when this sense of agency is absent. This finding is consistent with an account of SoA as a mechanism that gates the motor system from irrelevant sensation.

Significance Statement

Our muscles are the primary means through which we affect the external world, and the sense of agency (SoA) over action through those muscles is fundamental to self-awareness. Disturbances in bodily SoA have been reported in some psychiatric and movement disorders; despite the clinical importance of understanding the biological basis of SoA, however, the mechanisms and functional role of SoA in the neuromuscular system are still poorly understood. Drawing on methods from human-computer interaction and adaptive experimentation, we optimize the timing of electrical muscle stimulation to control the perception of agency over electrically-actuated movements using human-in-the-loop Bayesian optimization. Using this paradigm while subjects undergo EEG recording, we show that early (sensorimotor) and late (post-consciousness) neural patterns, which show distinct temporal dynamics, both predict illusory SoA on a trial-by-trial basis. Moreover, fractal analysis of the EEG suggests plausible neural mechanisms for insulating the motor system from irrelevant sensory feedback.

1. Introduction

Voluntary movements tend to be accompanied by a subjective experience of volition. That is, when we move through the control of our muscles, we tend to feel that "I did that." This feeling of controlling one's own actions and, by extension, the consequences of those actions in the external world, is referred to as the sense of agency (SoA), which is considered one of the most basic building blocks of our experience of selfhood (Gallagher, 2000; Haggard, 2008) and a salient part of our conscious experience of daily life.

Thus, the sudden loss of agency is a highly salient event that can quickly reorganize the behavioral priorities of an agent (Haggard & Chambon, 2012). Consequently, malfunctions in SoA can be extremely distressing to patients, such as delusions of control in schizophrenia (Frith, 2012), alien hand syndrome (Panikkath et al., 2014), perceived control or non-control of a phantom limb (Ramachandran & Hirstein, 1998), and automatic "utilization behavior" in patients with frontal lobe damage (Lhermitte et al., 1986). While these pathological conditions are characterized by anomalies in SoA over one's body, including the direct control of one's muscles, research on SoA in the healthy nervous system has instead focused primarily on the sense of agency over external sensory events or the consequences of one's actions (Haggard, 2008, 2017).

SoA in direct neuromuscular control is challenging to study, since control over one's muscles is unambiguous in normal conditions—if your muscle moved in the absence of an external force, it must have been you that caused it to move. While subjects' muscles can be actuated with electrical or magnetic stimulation, stimulation-evoked movements are not usually accompanied by an experience of agency (Haggard et al., 2002). In general, muscle movements evoked by transcranial magnetic stimulation (TMS) applied to the motor cortex are not thought to be accompanied by corollary discharge from motor areas that would suppress the neural response to sensory feedback from movement (Chronicle & Glover, 2003), likely because corollary discharge seems to originate from upstream of M1 (Voss et al., 2006). Notably, this explanation leaves open the possibility that aligning actuated movements with subjects' movement initiations may result in a sense of agency.

Researchers have attempted aligning TMS-evoked movements with subjects' endogenous intention to move in order to induce a false sense of agency over the involuntary movement, but the implicit measure of agency used—the shrinking of the perceived time between action (involuntary button press) and outcome (a tone following the button press)—was not responsive to this manipulation (Haggard & Clark, 2003). This measure, however, specifically probed the implicit binding between an action and an action outcome (putatively reflecting a causal relationship between the muscle movement and its outcome), rather than subjects' judgements of agency pertaining to the muscle movement directly. Recently, a number of human-computer interaction studies have tested the latter possibility, and these studies have consistently reported that subjects, when asked, explicitly judge muscle movements induced by electrical muscle stimulation (EMS) as self-caused when muscle stimulation aligns with their intention to move (Kasahara et al., 2019, 2021; Tajima et al., 2022). In particular, this effect can be elicited by a paradigm in which EMS-actuated movements *preempt* subjects' endogenously produced movements in a simple cue-response reaction time task, even if it results in a "reaction" time that would have been faster than if the subject had reacted on their own (Kasahara et al., 2019).

This phenomenon—in which subjects perceive illusory agency over an involuntary muscle movement—advances our understanding of bodily agency, clearly rejecting the hypothesis that "such involuntary movements are never accompanied by a sense of agency" (Haggard, 2017). It

also presents an unprecedented opportunity to study SoA at the level of neuromuscular control in healthy populations; comparing neural responses to electrically-actuated movements between EMS trials perceived as self- and externally-caused can provide a window into how the sensation of movement is processed when it perceived as self-caused. It is understood that at some level a comparison between sensations and predictions about sensory feedback underlies both conscious SoA and the unconscious aspects of sensorimotor control that are critical for generating effective movement (Blakemore et al., 2002). However, the neural correlates of low-level sensorimotor monitoring – action-dependent modification of the early neural response to self- vs other-caused sensation – have recently been shown to occur in the absence of conscious SoA (Timm et al., 2016), raising questions about which neural processes differentiating self and other (at varying levels of abstraction) propagate into conscious awareness. It may be that sensorimotor mechanisms are more predictive for SoA over one's body, as is in question here, as opposed to SoA over an action outcome as has been investigated previously; that is, the "neural correlates of SoA" may differ according to which level of abstraction is most appropriate for the relevant agency judgement (Charalampaki et al., 2022).

Understanding how the neural mechanisms of SoA coordinate with low-level sensorimotor processes is critical to understanding how we incorporate sensory feedback during motor learning, since SoA serves the prerequisite function of determining *what is feedback* to begin with, providing a basis from which to bootstrap learning before a precise action-outcome mapping is known (Wen & Haggard, 2020). Indeed, early evidence suggests that motor learning is directly facilitated by SoA (Chambon et al., 2020; Kasahara et al., 2021; Wen et al., 2021), highlighting the functional role SoA may play in insulating the motor system from irrelevant (i.e. spurious) sensations that might otherwise be interpreted as sensorimotor feedback or as cues to the body's current state (Goupil et al., 2021). Working with, rather than against, this organizing mechanism seems to be critical in the therapeutic application of EMS to facilitating motor learning following stroke (Gandolla et al., 2016).

Thus, in the present work, we use a human-in-the-loop Bayesian optimization procedure to guide the timing of electrical stimulation to achieve a tight distribution of stimulation times around the latency at which subjects will experience roughly half of all EMS-caused movements as self-caused. This allows us to test whether the putative neural correlates of agency predict SoA judgments even in the absence of true volition. To this end, we use multivariate pattern analysis (MVPA) of subjects' brain electrical activity measured by EEG to identify patterns that differentiate electrically actuated (EMS) movements that are perceived as self- vs. other-caused, as well as the temporal dynamics of these patterns—hypothesizing that we would see transient patterns of activity predicting agency during the early sensory response, consistent with modification of the neural response by the corollary discharge that accompanies movement intent (Voss et al., 2006), followed by a switch to sustained representation as the perception of movement enters conscious awareness (see ***Discussion 4.2***). Finally, in an exploratory analysis, we apply nonlinear measures of signal complexity (i.e., fractal dimension, related to the amount of information a signal can encode) and of long-range correlation structure in the signal (i.e., Hurst exponent, which is related a system's global sensitivity to local perturbation). Within this fractal geometry framework, we propose a hypothesis in which the internal dynamics of cortex insulate the motor system from irrelevant feedback in the absence of SoA.

2. Materials and Methods

2.1 Participants and Ethics Statement

25 University of Chicago undergraduate students participated in the study; however, two subjects were subsequently excluded for noncompliance with task instructions (i.e., one admitted to letting the electrical stimulator perform the task without attempting a volitional button press, and another pressed the button continually to speed through the task instead of when cued). Participants were recruited through the University of Chicago's human subject recruitment system, SONA Systems. All subjects gave written, informed consent before participating. All of the methods performed in the study were in accordance with relevant safety and ethics guidelines for human subject research and were approved by the Social and Behavioral Sciences Institutional Review Board at the University of Chicago (IRB19-1302). This study was not a clinical trial.

2.2 Experimental Design

Subjects completed three blocks of trials. In the first and final blocks (50 trials each), each trial was as follows: (a) a fixation cross was displayed on the screen; (b) after a random wait period (2-4 seconds, uniformly distributed), a cue to press the response button as quickly as possible was displayed on the screen ("Go!"); after which, (c) the subject pressed a button (on a *Cedrus RB-620* button box; California, United States) voluntarily and without any electrical stimulation; finally, (d) the trial ended with another fixation cross displayed on the screen.

In the middle (EMS) block (300 trials), (a), (b), (c) and (d) from the above remained the same, but this time, electrical muscle stimulation (EMS) was applied to the forearm in order to move the subjects' finger (see ***2.4*** for stimulation details), causing their finger to press the button involuntarily after the cue to move. A typical trial in the stimulation block is shown in Figure 1.

The latency of the EMS relative to the cue to move (b) was guided by a Bayesian Optimization (BO) procedure (see 2.3 below) to be as near as possible to the latency at which subjects will experience 50% of movements to be self-caused, which prior studies have shown is substantially before the subjects' average reaction time (Kasahara et al., 2019, 2021). (However, the Bayesian optimization procedure only considered latencies between 50 and 600 milliseconds, so stimulation did always occur in that interval.) As a result, trials in which the subject pressed the button before stimulation was delivered were rare (see ***Results 3.1***). If the subject pressed the button before stimulation was delivered, stimulation occurred immediately upon the button press so that the onset of the electrical stimulation, which causes a perceptible though painless tingling sensation on the skin, was always temporally confusable with that of the movement.

As depicted in Figure 1, after each button press in the stimulation block, the subject was asked to judge whether they caused the button press or if the electrical stimulation caused the button press. This response was fed into the BO algorithm to inform the timing of stimulation in subsequent trials (see below). The experiment interface was implemented using *PsychoPy* (Peirce, 2007).

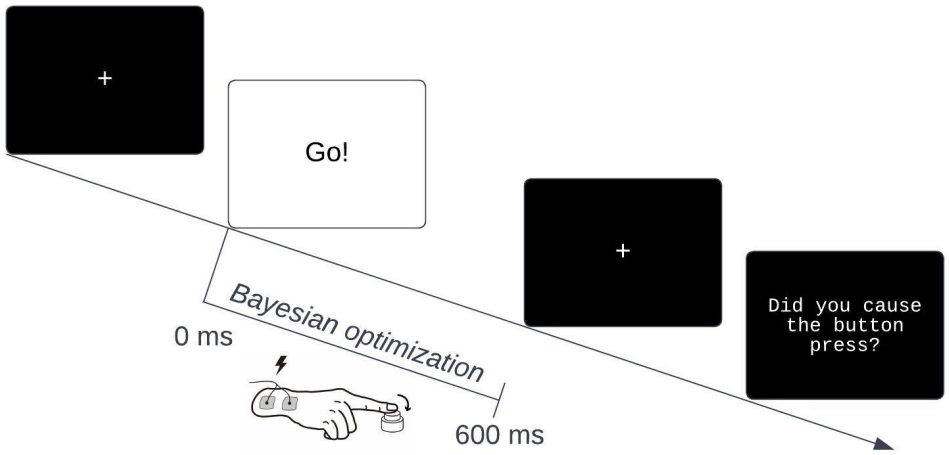

Figure 1: Graphical depiction of a typical trial in the stimulation block. Stimulation is hard-coded to occur before 600 milliseconds after the cue to move, but a Bayesian Optimization procedure biases it to occur at the latency the subject is estimated to perceive agency over the resulting muscle movement with 50% probability. In practice, this almost always occurs before the subject is able to move on their own.

2.3 Bayesian Optimization

We adaptively fit a Bayesian logistic regression, updated after each trial, to estimate the stimulation latency (relative to the movement cue) at which the subject would judge that they caused the muscle movement on that trial with 50% probability. We parametrized the logistic regression as in Formula (1) so that the parameter α would correspond to that 50% threshold.

$$\begin{aligned} \mathrm{logit}(p) &= \beta(x-\alpha) \\ y &\sim \mathrm{Bernoulli}(p) \end{aligned} \qquad (1)$$

Using the 50 reaction time trials from the first block of the experiment to compute each subject's mean reaction time μ_{RT} with standard deviation σ_{RT}, we began the stimulation block with informative priors on α and β to start our model in a good position to learn the true threshold quickly. For α, we used the log-normal distribution with mean $\mu_{\mathrm{RT}} - 40$ ms and standard deviation σ_{RT}, since the "preemptive gain" estimated in previous studies is roughly 40 milliseconds (Kasahara et al., 2021). For β, the slope parameter, we used the log-normal distribution with mean 0.017 and standard deviation 0.005, which generously encompasses the slopes in the subject-specific logistic regressions reported by a previous study (Kasahara et al., 2019).

On each trial, the stimulation latency is chosen by taking a draw from the current posterior distribution for α, truncated between 50 and 600 milliseconds after the movement cue. After each trial, the subject judges whether they caused the muscle movement, and their response is used to update the posterior, which is approximated via variational inference. (However, if the subject pressed the button before stimulation was delivered, and thus actually caused the movement, the trial was counted as being judged self-caused regardless of how the subject responded.) In this way, as the posterior becomes more precise (i.e., we are more certain about the true threshold), the stimulation times will hone in on that threshold. But while we are still uncertain, the posterior we are sampling from is wide, and we will explore a wider range of latencies. Thus, this algorithm navigates the exploration-exploitation tradeoff naturally; it is a modified version of Thompson

sampling, a BO approach which benefits from strong regret-minimization guarantees (Agrawal & Goyal, 2012).

This procedure was implemented in *Python* using the probabilistic programming language *Pyro* for variational inference (Bingham et al., 2019), and all associated code is available (see *Data and Code Availability*).

2.4 Electrical Muscle Stimulation

Before the experiment began, two EMS electrodes were applied to the skin above the *flexor digitorum profundus* muscle on the right forearm, which is an easily accessible muscle that moves the ring finger. (Contrary to Figure 1, subjects used their right ring finger to press the button during the experiment, in large part because this finger is practically quite easy to move with EMS.)

Stimulation was performed with a *RehaStim 1* device (*HASOMED* GmbH, Mageburg, Germany). On each trial, muscle actuation consisted of a single, biphasic pulse of constant current stimulation lasting 900 microseconds (400 μs positive, 400 μs negative, separated by 100 μs), since biphasic pulses result in less perceptible tingling of the skin than monophasic pulses.

Before beginning the experiment, we calibrated the stimulation amplitude to the minimum intensity required to reliably move the subject's finger. The calibration procedure was as follows: (a) The subject placed their ring finger on the button that would be used during the experiment and was instructed not to move their hand. (b) Starting at an intensity of 1 mA, we stimulated the subjects' arm 10 times. If fewer than 10 button presses were registered, then we iterated the intensity by 1 mA and repeated. (c) We stopped increasing the intensity upon achieving 10 consecutive actuated button presses, or if a conservative safety limit of 25 mA was reached. (In practice, stimulation intensities can safely go much higher, but might eventually become uncomfortable.) (d) If stopping in (c) was due to reaching the safety limit, the electrode positions were adjusted and the procedure was repeated until stimulation elicited button presses reliably.

2.6 Manipulation Check and Outlier Removal

We performed a brief analysis to demonstrate the effectiveness of our Bayesian optimization procedure at preempting subjects' movements.

First, we identified trials in which stimulation was ineffective at eliciting a muscle movement, which were excluded from subsequent behavioral analysis, as well as from EEG analysis. To identify such trials, we (a) removed trials in which subjects pressed the button prior to stimulation, since such movements were obviously caused by the subject, (b) removed trials in which the recorded response time was greater than 600 ms, outside the stimulation time window, then (c) fit a log-normal distribution to the lag between stimulation and the time of the subsequent button press and removed the trials in the most extreme 5% of that fit distribution. Briefly, (a) and (b) removed an average of 18.4 trials and (c) removed an average of 12.4 trials per subject. The result of outlier removal is visualized, for convenience, in Figure 3. After this step, measured "reaction" time in the stimulation block should be a nearly linear function of stimulation latency.

We then fit a Bayesian multilevel model to the trial-by-trial reaction times in each experiment block using the *Bambi* package (Capretto et al., 2022) as an interface to (Patil et al., 2010) using the default (weakly informative) priors, which are conservative in the sense that they should shrink estimates of the RT in each condition toward each other (Westfall, 2017). This is done in the interest of replicating the finding that EMS-induced movements can substantially preempt

subjects' natural movements in this paradigm, in which case the measured "reaction" time should be lowest in the stimulation block (Kasahara et al., 2019, 2021). We report the empirical percentage of stimulation trials that subjects perceived as self-caused, which should be close to 50% if our BO procedure was successful.

2.6 EEG Acquisition

EEG was recorded with 64 active Ag/AgCl electrodes (*actiCHamp*, Brain Products, Munich, Germany) placed according to the extended 10-20 system. At the time of recording, the electrodes were referenced to Cz and sampled at 10,000 Hz. Two of the 64 electrodes (which would have been AF7 and AF8 on the typical *actiCAP* layout) were dropped below the left and right eyes so that they could later be re-referenced to become EOG channels.

Experiment events were marked in the EEG recording using TTL triggers, delivered using an *RTBox* (Li et al., 2010) in lieu of a built-in parallel port on the experiment control computer.

We additionally placed a photo-sensor (*Brain Products*, Munich, Germany) on the subjects' screen to record screen brightness at the same sampling rate as the EEG. The timestamps of the TTL triggers corresponding to the onset of the visual cue to move (see section 2.2) was corrected after the recording using the ground-truth event time measured using the photo sensor. The difference between the TTL trigger, which was sent in parallel with the start of the reaction time trial, and the actual time of the screen flip recorded with the photo sensor was used to correct reaction time measurements taken with the response box as well.

Before beginning the experiment, we tested the hardware latency between the EMS pulses and the TTL pulses marking them in the EEG recording; to this end, we used a custom cable that allowed us to record the output voltage of the stimulation device directly with the EEG amplifier. Over 500 repetitions, we measured a latency of mean 1.23 ms with a jitter of standard deviation 0.04 ms. This constant 1.23 ms delay in the event timestamps was corrected offline after the EEG recordings. All timestamps found in the open dataset associated with this article (see ***2.9 Data and Code Availability***) have already been corrected for convenience.

Additionally, the precise subject-specific positions of the 62 head electrodes were measured at the end of each recording using a *CapTrak* (*Brain Products*, Munich, Germany).

2.6 EEG Preprocessing

EEG was preprocessed in *Python* using *MNE-Python* package (Gramfort et al., 2014). First, the electrodes below the left and right eye, which were referenced to Cz at the time of recording, were re-referenced against forehead electrodes Fp2 and Fp1, respectively, to become bipolar EOG channels montaged to capture both horizontal and vertical eye movements.

Subsequently, the electrical artifact from the EMS needed to be removed. The EMS pulse was brief (less than a millisecond), so it could be removed by linearly interpolating over the interval starting 5 ms before and ending 10 ms after the event timestamp. However, since by the end of the stimulation block, stimulation was delivered at a narrow distribution of latencies after the cue to move, which occurs at the 60 Hz screen refresh rate of the subjects' monitor, stimulation did not occur at a random phase of the background 60 Hz line noise. Thus, we found data quality was improved by removing line noise prior to interpolating over the stimulation artifact. However, applying an FIR notch filter before removing the pulse-like (flat power spectrum) stimulus would result in distortion of the filtered signal. To this end, we fit a multi-taper estimation of the sinusoidal components at the line noise frequency and its first two harmonics and removed those

components (as implemented in MNE). Then, after performing linear interpolation, we applied an FIR notch filter at 60 Hz and its harmonics up to the upper edge of the bandpass filter that would be applied subsequently (see below) to thoroughly clean the data of line noise, and then resampled the data to 5,000 Hz to improve the speed of computation for subsequent preprocessing steps.

Next, we applied common preprocessing operations in adherence with the standardized *PREP* preprocessing pipeline for EEG data (Bigdely-Shamlo et al., 2015) using the implementation in the *PyPREP* package (Appelhoff et al., 2022). This pipeline robustly re-references the EEG signal to the average of all electrodes and interpolates electrodes it determines have poor signal quality; since data quality is sensitive to the order in which these operations are performed (i.e., bad channel identification depends on the reference, and the average reference should not include bad channels), the pipeline applies the two steps iteratively until convergence. Criteria for identifying bad channels includes extreme amplitudes, lack of correlation with other channels, lack of predictability by neighboring channels, and unusual high frequency noise; see the PREP paper for a full description (Bigdely-Shamlo et al., 2015). Precise, subject-specific electrode positions were used for the interpolation of bad channels, and a record of which channels were interpolated are available in subject-specific preprocessing reports (see 2.9).

Usually, line noise is also removed as part of the PREP pipeline, but since we removed line noise when interpolating over the stimulation artifact, we did not perform this operation again during PREP. (Note that removal of the stimulation artifact needed to be performed before PREP, since PREP include a detrending step that uses a highpass filter, which would result in signal distortion if the pulse-like stimulation artifact was still present.)

Next, we filtered the data to the frequency band used for analysis. We used a low cutoff of 1 Hz to remove low-frequency drift, and we used two separate high cutoffs, 30 Hz and 70 Hz, for the two types of analysis to follow preprocessing (see 2.7 and 2.8 below). Thus, the following preprocessing steps were applied twice, once for each copy of the data.

We then removed EOG contamination of the EEG signal. We decomposed to EEG data into 15 independent components (ICs) using the *FastICA* algorithm implemented in MNE (Hyvarinen, 1999). Then, we correlated each IC with the EOG channels, *z*-scored the correlation coefficients, and deemed an IC to contain eye artifact if the absolute value of its *z*-score exceeded 1.96. Those ICs were zeroed out to remove them from the original data. Plots of the scalp topographies of removed ICs for each subject can be found in their preprocessing reports (see 2.9).

Subsequently, we segmented the data into epochs starting 100 ms before the onset of stimulation and ending 500 ms after stimulation. We then estimated the peak-to-peak rejection threshold that would be used to identify trials containing unwanted artifacts using the *Autoreject* package (Jas et al., 2017), which estimates the optimal threshold as that which minimzes the 5-fold cross-validated root-mean-squared difference between the mean of the training folds and the median of the testing fold, a robust proxy metric for signal-to-noise. The resulting per-subject rejection thresholds are recorded in each subjects' preprocessing report (see 2.9).

Since the visual evoked response to the movement cue is unlikely to be over by the time of stimulation, we attempted to remove the visual evoked response from our epoched data to minimize confounds. To do so, we computed evoked responses to both the visual and electrical stimuli simultaneously using a regression-based overlap correction on the continuous (non-epoched) data, excluding second-long chunks of the data in which peak-to-peak amplitude exceeds the rejection threshold (Smith & Kutas, 2015); conceptually, this is very similar to the way

generalized linear models (GLMs) are used to deconvolve hemodynamic responses in fMRI. Then, the overlap-corrected visual evoked response was aligned with the epoched version of the data and subtracted out. Thus, the average visual response to the movement cue was removed from the stimulation-locked epochs. Subject-level evoked responses can be found in our open dataset and are visualized in the subject-specific preprocessing reports (see 2.9).

Finally, the rejection threshold was applied to the cleaned and overlap-corrected epochs, removing trials still contaminated by artifacts. The surviving epochs were down-sampled to twice their high-cutoff frequency for computational expediency and saved for further analysis. This epoched data is available in our open dataset, and subject-level trial yields are recorded in the accompanying quality check reports (see 2.9).

2.7 EEG Decoding: Multivariate Pattern Analysis

For MVPA, we used the copy of the epoched data filtered from 1-30 Hz, since a conservative high-cutoff smooths signals of interest, which may occur at slightly different times on a per-epoch or per-subject basis, such that they can be aligned across subjects.

To assess the temporal dynamics of patterns which differentiate trial perceived as self- and EMS-caused, we employ the temporal generalization method (King & Dehaene, 2014). In this approach, a linear classifier is trained on the pattern of voltages at each timepoint (using a training set), and then its classification performance is quantified at every other timepoint (using a test set). In our case, we quantify classification performance using the area under the receiver-operator curve (ROC-AUC), a non-parametric metric of class separation that is more robust than accuracy. This results in a $n_{\text{times}} \times n_{\text{times}}$ matrix M, where M_{ij} is the performance of the classifier trained at time i evaluated at time j. For our linear classifier, we use a logistic regression fit with generalized estimating equations (GEE) as implemented in the *statsmodels* package (Liang & Zeger, 1986; Seabold & Perktold, 2010). GEE is a semi-parametric fitting procedure for linear models that does not assume the data comes from a certain distribution, resulting in parameter estimates that are robust to heterogeneity; moreover, it allows the fitting of the marginal (i.e., population) parameters of a multilevel model without explicitly computing the subject-level model, which allows us to account for subject-level random effects without the computational burden of a multilevel model, which is in general too high for MVPA.

Transient patterns, which predict agency but only appear for a short time, are characterized by strong classification performance only on the diagonal of M; sustained patterns show strong decoding performance off the diagonal as well, indicating that the same neural process persists over time. This method has been argued to provide insight into the duration of and dynamic transitions between mental representations as a cognitive process unfolds in time (King & Dehaene, 2014).

We employ a 10-fold stratified cross-validation scheme where subjects are randomly split into non-overlapping trials across folds (i.e., stratification entails an attempt to ensure an equal proportion of agency and non-agency trials across folds). Thus, classifiers are only evaluated on trials from subjects that did not appear in their training set, ensuring identified patterns generalize across subjects. Specifically, we used this cross-validation scheme as implemented in the *scikit-learn* package (Pedregosa et al., 2011). This 10-fold cross-validation scheme was repeated 10 times, with 10 different random splits into folds so that results are robust against random effects of particular splits of the data. Then, ROC-AUC scores obtained across the 10-by-10 splits were compared to the chance performance (ROC-AUC = 0.5) using a one-sided, one-sample *t*-test at

every (train-time, test-time) pair M_{ij}. However, the independence assumption of the standard *t*-test are violated due to observations being repeated across training and test sets (Dietterich, 1998); to account for this, we use the variance correction proved by Nadeau & Bengio to contain false positive rates in this context when computing our *t*-statistics (Bouckaert & Frank, 2004; Nadeau & Bengio, 1999).

Since the "shape" of classification performances in the matrix M is of interest, then cluster-level inference (e.g., on the size of contiguous subsets of M_{ij}), rather than inference on the individual M_{ij}'s, is desirable in principle. However, the cluster-level tests used in neuroimaging do not control the false-positive rate for claims about the extent of effects, such as how far they extend off the diagonal, which are critical to the interpretation of the time-by-time generalization results (Sassenhagen & Draschkow, 2019). Instead, we use All-Resolutions Inference (Rosenblatt et al., 2018), as implemented in our *MNE-ARI* package (Veillette, 2022), to compute simultaneous lower bounds on the true positive proportion in each cluster across all $n_{\text{times}} \times n_{\text{times}}$ possible clustering thresholds. This approach to dealing with the multiple comparisons problem adequately conveys uncertainty about the localization of true effects within clusters, while controlling the false-positive rate in the strong sense. For instance, if the proportion of true positives in a cluster is low, then it is unclear which M_{ij} in the cluster are the true positives; conversely, if the proportion is high (e.g., >95%), then the localization is quite certain.

2.8 EEG Decoding: Fractal Analysis (Exploratory)

To compute fractal measures of the EEG data, we use the copy of the epoched data filtered between 1 Hz and 70 Hz, since fractal metrics are fundamentally about how the behavior of a time series varies across timescales, and filtering out high frequencies removes more granular timescales. Nonetheless, we still filter at 70 Hz because it is a topic of active debate whether high gamma band (>70 Hz) can be reliably detected using scalp EEG (Darvas et al., 2010; Fries et al., 2008). To improve spatial specificity of this analysis (since the time domain is no longer of interest given the properties of these fractal measures), we computed the current source density transformation of our scalp-recorded voltages to reduce the effect of volume conduction before computing the following fractal metrics. Both simulations and empirical results indicate that this method is effective at reducing the radius of cortical tissue to which each electrode's signal is sensitive to a few centimeters (Kayser & Tenke, 2015).

The first metric of interest is the Higuchi fractal dimension, which is an efficiently computable approximation of the fractal dimension (Higuchi, 1988). Generally speaking, this quantity is a measure of the complexity or roughness of a time series (or of its underlying dynamical attractor), which is directly related to the entropy of a signal (i.e., an upper bound on the amount of information it can encode) but can be computed more reliably from small time series. The fractal dimension of both the background EEG and the EEG response to perturbation is highly predictive of states of consciousness (Kesić & Spasić, 2016; Ruiz de Miras et al., 2019), consistent with information theoretic accounts of conscious awareness (Oizumi et al., 2014). Moreover, some evidence suggests that neural activity corresponding to internally generated percepts (e.g., mind wandering) is of a higher fractal dimension than that resulting from sensory stimulation, but it is unclear whether this difference is accounted for by the different modalities (i.e., mind wandering vs. external sensation) or to the self-attribution of the percept, which our analysis should clarify.

The second metric of interest is the Hurst exponent, which is a measure of long-range temporal dependencies in a time series (Qian & Rasheed, 2004). While, in a self-affine (i.e., strongly self-similar) time series, the fractal dimension D and the Hurst exponent H are related by the deterministic relationship $D + H = 2$, these quantities can vary independently in time series that depart from strict self-similarity (Gneiting & Schlather, 2004), and these metrics have been reported to diverge in EEG data (Martis et al., 2015). As such, fractal dimension is a local property of the signal, whereas Hurst is a global property which quantifies how sensitive the larger time series is to local changes (Gneiting & Schlather, 2004). In the cognitive neuroscience literature, this quantity has been interpreted as corresponding to how responsive the underlying neural system is to external input (Churchill et al., 2016; Kardan et al., 2020; Zhuang et al., 2022), and it is notably suppressed in those suffering from psychiatric conditions associated with an impaired sense of agency (Sokunbi et al., 2014; Stier et al., 2021).

It is important to note, as a caveat, that the Hurst exponent cannot be reliably estimated from time-series as small as our single-trials. While Hurst is in principle a scale-invariant metric, in practice, time series as long as 10,000 time points are required to produce reliable estimates with conventional estimates (Eke et al., 2000, 2002; Oliver & Ballester, 1998), and while our time series were in principle sampled at 10,000 Hz, filtering reduces the effective sampling rate to twice the upper passband edge in practice. Despite the unreliability of the Hurst estimates at the trial level, however, *within*-subject variation in trial-by-trial Hurst exponents estimated from EEG are sensitive to cognitive functions (Kardan et al., 2020). In other words, while our individual trial-by-trial Hurst estimates may not be reliable indices of the mathematically-defined Hurst exponent, their variation may still be meaningful.

We estimate fractal dimension using Higuchi's algorithm (Higuchi, 1988) and the Hurst exponent using rescaled range analysis (Qian & Rasheed, 2004), using the implementations provided in the MNE-Features package (Schiratti et al., 2018). Then, for each electrode and subject, we compute the ROC-AUC between each metric and subjects' self-reported agency. Since these metrics are determined a priori and not from the data, no cross-validation scheme is necessary; a one-sample, two-sided *t*-test is used compare subject-level decoding performance to chance (ROC-AUC = 0.5).

2.9 Data and Code Availability

Code for running the experiment can be found on *GitHub* (github.com/apex-lab/agency-experiment) and in a permanent archive on *Zenodo* (doi.org/10.5281/zenodo.7894011). Similarly, all data analysis code, including EEG preprocessing code, can be found at github.com/apex-lab/agency-analysis and https://doi.org/10.5281/zenodo.7894007. All data, including both raw data, preprocessed derivatives, and post-preprocessing quality check reports for each subject, can be found on *OpenNeuro* (doi.org/10.18112/openneuro.ds004561.v1.0.0).

3. Results

3.1 Efficacy of Bayesian Optimization Procedure

The Bayesian Optimization (BO) procedure resulted in trial-by-trial stimulation latencies honing in on some threshold estimate throughout the stimulation block. A representative time course is shown in Figure 2.

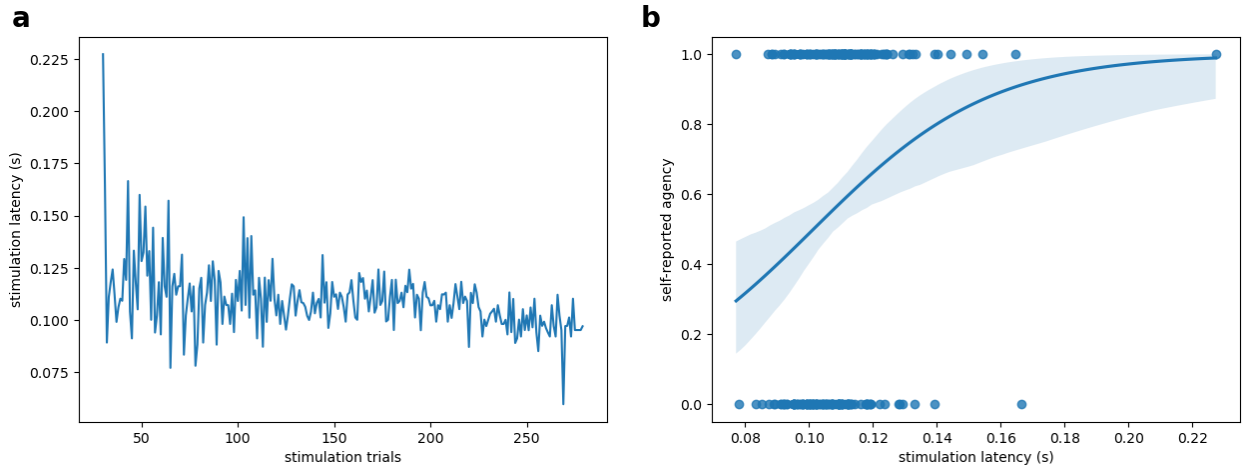

Figure 2: Trial-by-trial stimulation latency over the course of the stimulation block for a representative subject. (a) Stimulation latency hones in on a stable value over time. (b) A logistic regression computed after the experiment and shows that stimulation times are close to the retro-actively estimated 50% threshold, even though that threshold was not known in advance. The subject featured here is "sub-07" in the associated dataset.

After removing trials in which stimulation failed to produce a muscle movement (and therefore the "reaction" time was not a function of stimulation latency), our multilevel model of the recorded reaction times estimated a 99.9% posterior probability that button presses occurred earlier in the stimulation block than in either of the other blocks. In particular, we estimate that "reaction" times resulting from EMS-actuated movements were between 17.5 ms and 65.0 ms faster than true reaction times in the first (pre) block with 95% probability, and between 13.8 ms and 43.9 ms faster than those in the final (post) block. A nominal speedup between the pre and post blocks was observed with 90.6% probability (95% HDI: [-6.7 ms, 32.8 ms]), suggesting that subjects may have improved their reaction times by the end of the task, but not enough to account for the much lower reaction times in the stimulation block. Posterior distributions for the (group) mean response times in each condition are shown in Figure 3.

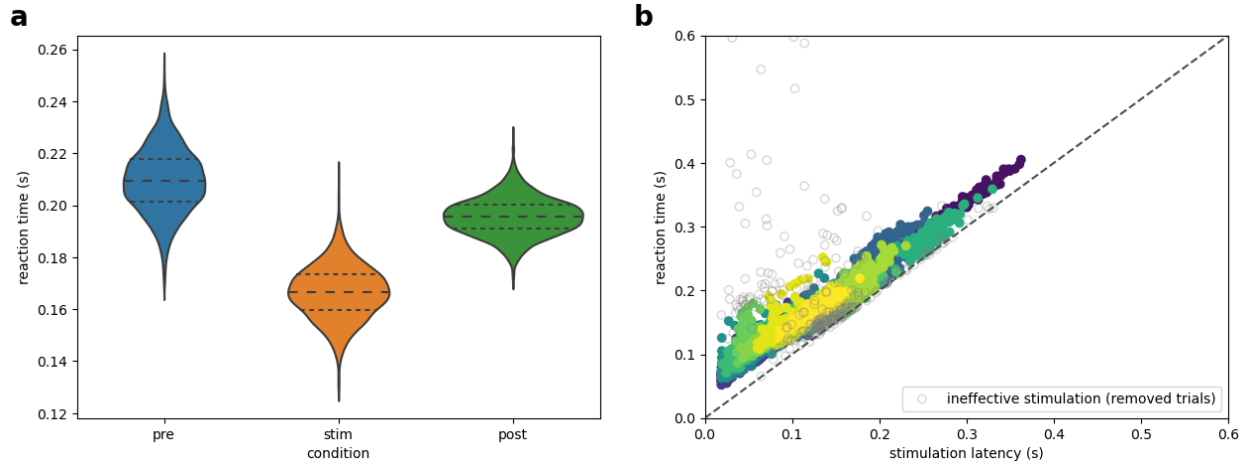

Figure 3: Electrical muscle stimulation consistently preempted subjects' volitional movements. (a) Posterior distributions of the mean reaction times in each condition show that EMS-induced muscle movements occur earlier than subjects' natural muscle movements. (b) After outlier removal, measured "reaction" times (shown for all trials

and subjects) are a nearly linear function of the stimulation latency, indicating that movements in the remaining trials are, in fact, EMS-actuated.

While EMS-caused movements were substantially faster than subject-caused movements, and the time of button press was clearly a function of stimulation latency (see Figure 3), subjects still reported that they caused roughly half of the movements. Overall, after outlier removal (see Methods and Figure 3), 51.98% of all trials across all subjects were judged as self-caused. On average, subjects reported that they caused 50.99% (SD: 14%) of movements. In other words, the Bayesian Optimization procedure was effective at controlling the proportion of trials in which movements were experienced as self-caused, generating a roughly 50-50 split of agency vs. non-agency trials.

3.2 Multivariate Pattern Analysis

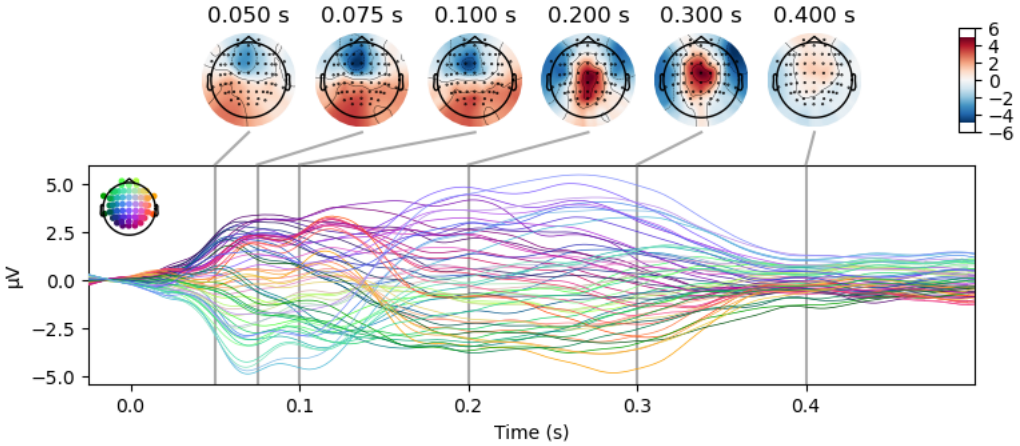

Figure 4: The grand-average evoked EEG response to muscle stimulation.

Our MVPA procedure showed above-chance decoding performance *across subjects* (more often such decoding is performed within subjects), reaching up to ROC-AUC = 0.587. While we report the true-positive proportion within clusters across all clustering threshold (see Figure 5), we will focus primarily on the clusters in which the true positive proportion exceeds 95%, since these clusters are where we are sufficiently certain about the localization of the effect (Veillette, 2022).

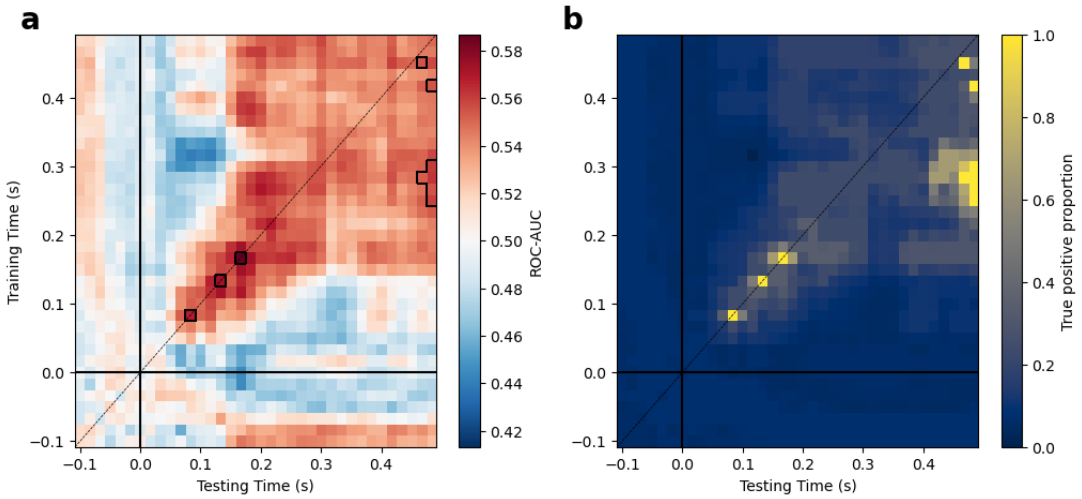

Figure 5: Temporal generalization of neural patterns predicting sense of agency. (a) Classification performance (ROC-AUC) for decoding subjects' judgement of agency for individual muscle movements, cross-validated across subjects. Results are shown for all (train-time, test-time) pairs to visualize the temporal dynamics of patterns that predict SoA. (b) Lower bounds on the true-positive proportion (TPP) within clusters, computed across all clustering thresholds. The value represented at each (train-time, test-time) pair is the highest TPP of any cluster in which that pair is included, thus larger values reflect greater certainty in the localization of effects.

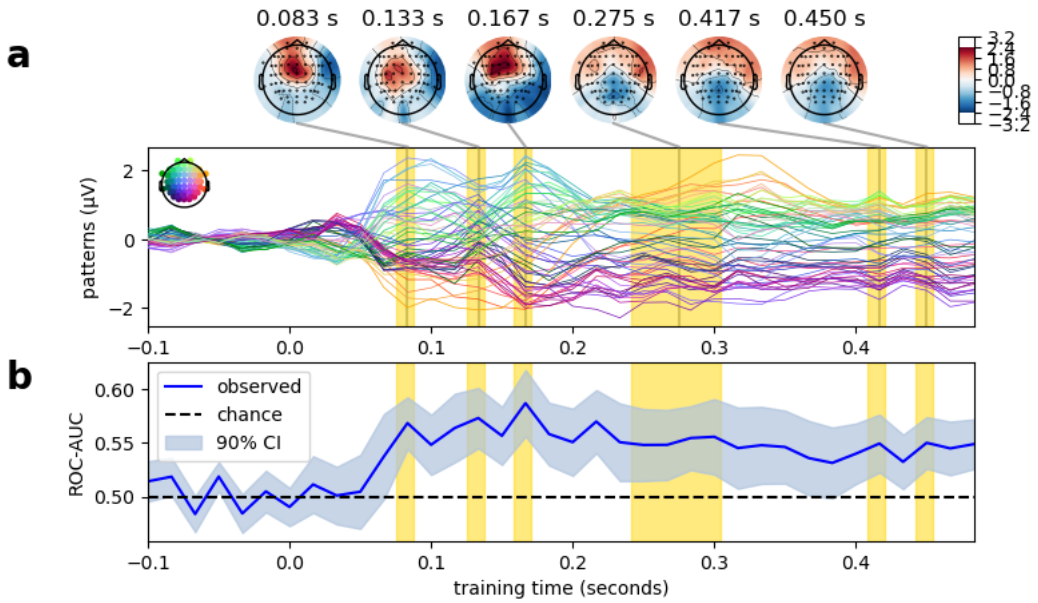

Figure 6: Voltage patterns that predict sense of agency. (a) The EEG topographies that the linear classifiers trained at each timepoint select for, reconstructed by inverting the trained classifier parameters using Haufe's trick (Haufe et al., 2014). (b) The decoding performance when testing at each train time (identical to the values on the diagonal in Fig. 5a). Training times are highlighted in yellow if included in a cluster with TPP > 0.95 at any test time (Fig. 5b).

The earliest such cluster occurs 83 ms after the onset of muscle stimulation ($p < 4.5 \times 10^{-6}$). When comparing the patterns our decoding model selects for (see Figure 6) to the average evoked response (see Figure 4), one notes that the polarity of the pattern that predicts SoA is opposite the average response, indicating that the classifier would predict a self-agency judgment as the result

when the sensory response is suppressed—a finding consistent with sensory attenuation (Voss et al., 2006). However, classifiers trained at earlier times do not generalize to predict SoA at later times (see Figure 5). Later in the epoch, however, the temporal dynamics of the predictive patterns change to reflect a single, sustained neural signature that predicts SoA starting by at least 250 ms after stimulation and persisting at least until the end of the epoch ($p < 0.003$).

3.3 Fractal Analysis (Exploratory)

Notably, trial-by-trial fractal dimension predicted SoA at almost every electrode (see Figure 7), reaching an ROC-AUC of 0.614 at electrode C1 ($p < 0.027$), even after the current source density transformation of the EEG signal was applied to attenuate the effects of volume conduction (see Methods). This suggest that the (local) complexity of the brain activity is increased uniformly throughout cortex following muscle movement when that movement is perceived as self-caused (as compared to when it is not perceived as self-caused).

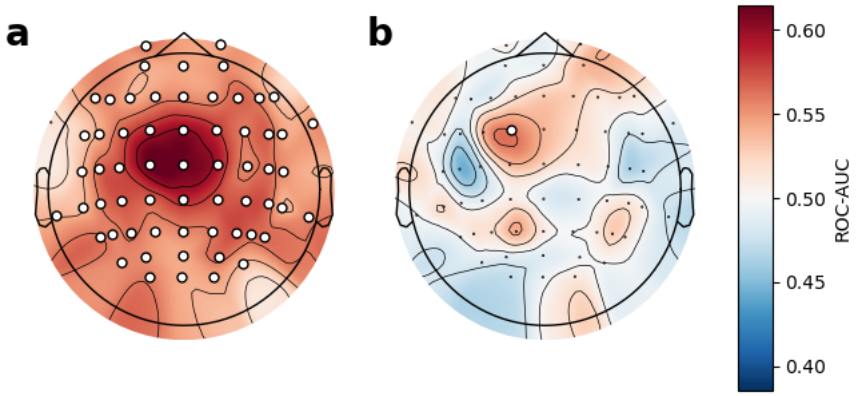

Figure 7: Classification performance for predicting trial-by-trial sense of agency from single-electrode fractal metrics. (a) Classification performance for Higuchi fractal dimension. (b) Classification performance for the Hurst exponent. Electrodes included in clusters in which the true positive proportion exceeds 95% are marked with white.

On the other hand, the Hurst exponent only predicted SoA at a single electrode at position FC1 (ROC-AUC = 0.559, p = 0.0006), located above cortical regions involved in motor control and planning, contralateral to the arm in which stimulation occurred (though we did not vary the arm used for stimulation, so we would caution against interpreting this as a strictly contralateral effect, though it is suggestive). This finding suggests a much more selective modulation of long-range temporal dependencies, such that the activity of specific frontocentral cortical regions becomes globally less to local perturbations in the absence of SoA—or, conversely, frontocentral areas are more sensitive to inputs in the presence of SoA. However, it should be noted that this analysis using fractal measures was exploratory, and consequently it should be interpreted as motivating hypotheses about the dynamical characteristics of neural activity that may give rise to the sense of agency, rather than as providing confirmatory evidence about such hypotheses.

4. Discussion

The decoding results featured in our main analysis clarify several theoretical issues about the relationship between sense of agency (SoA) and sensorimotor processing. These implications are discussed in Sections 4.1 and 4.2 below. Moreover, our exploratory decoding results, based on fractal measures of the EEG time-series, suggest some novel neural mechanisms by which the internal dynamics of cortex module their own sensitivity to external input as a function of SoA.

These potential mechanisms are discussed as hypotheses for future investigation in Sections 4.3 and 4.4 below.

4.1. Conclusion: Unconscious sensorimotor processes inform agency judgements.

Computational accounts of motor control suggest that internal models, which predict the sensory consequences of movements, are used to guide action selection (Wolpert et al., 1995, 2001). In such accounts, the prediction error—the difference between actual and expected sensory feedback—is used to update internal models, resulting in motor learning. These results are consistent with the hypothesis that early sensory responses are altered (usually suppressed) when they align with movement intentions, consistent with an inhibition of thalamo-recipient cortical populations via corollary discharge from motor planning regions (Voss et al., 2006); in this view, amplified sensory responses reflect a deviation between predicted and actual sensory feedback, which would otherwise be suppressed (Friston, 2005), culminating in a negative agency judgement downstream. This general framework is referred to as the "comparator model."

Early accounts of SoA suggested that this same low-level sensorimotor error signal determines the sense of agency, since volitional movement (but not uncontrolled external sensations) should be accompanied by an internal prediction (Blakemore et al., 2002; Frith, 2012). However, if all unpredicted sensations were deemed to be externally caused, then it would be illogical to use those same prediction errors to update internal action-outcome models. As a result, researchers have noted that (1) not all sensory prediction errors results in a non-agency judgement (Blakemore et al., 2002), and (2) sensations can be deemed self-caused even when their content is unpredictable (Wen & Haggard, 2020). Consequently, recent theoretical perspectives have suggested that the mechanisms highlighted in the comparator model alone is not sufficient for explaining agency judgments (Frith, 2012; Haggard, 2017; Synofzik et al., 2008).

In this vein, while the modification of early neural responses to sensation were once considered an implicit measure of SoA, the once-assumed relationship between low-level sensory processing and SoA has recently been called into question (Dewey & Knoblich, 2014; Voss et al., 2006). In particular, multiple studies have demonstrated that not all brain signals related to sensorimotor prediction error inform subsequent agency judgements (Kühn et al., 2011; Ohata et al., 2020; Timm et al., 2016), though precisely which sensorimotor processes *do* feed into judgements of agency remains a topic of investigation. Multiple EEG studies have failed to find a relationship between early sensory event related potentials and subsequent agency judgements (Kühn et al., 2011; Timm et al., 2016), and Voss and colleagues show that comparator-model-like suppression of early responses to self-caused sensation occur in both the presence and absence of SoA (Voss et al., 2006). These findings have been interpreted as meaning that temporally early (<200 ms), "exogeneous" neural responses (i.e., those that are a direct consequence sensory input and reflect unconscious sensorimotor processing) do not inform agency judgements, but later "endogenous" neural responses (e.g., P3 component) associated with conscious attention do (Kühn et al., 2011).

Decades of research tell us that these early (<200 ms) sensory responses reflect preconscious, rather than conscious, processing of the sensory stimulus (Dehaene & Changeux, 2011; Libet et al., 1967; Sergent et al., 2005). It is notable, then, that we are able to decode agency judgements from the EEG a mere 83 ms following the onset of muscle stimulation, directly falsifying the claim that early, unconscious sensorimotor processes do not predict SoA (Kühn et al., 2011). In fact, we demonstrate the opposite of Voss and colleagues' finding; while suppression of the early EEG response to self-caused sensation may occur in the presence or absence of SoA, modification of

early neural responses to sensation occur with SoA even when that sensation was not actually self-caused (as we exclusively analyzed trials in which movements were caused by EMS).

Thus, our findings re-implicate unconscious sensorimotor processes in the formation of conscious agency judgements. It may be that this relationship holds preferentially for agency judgements about the control of the musculature, as in the present study, while previous EEG studies have primarily investigated SoA over downstream consequences of motor actions (e.g., a tone resulting from a button press) (Kühn et al., 2011; Timm et al., 2016). It is rare that control over one's muscles would be usurped, so it is sensible that sensory prediction errors directly concerning muscle movements might be weighted more heavily in agency judgments than, for instance, external auditory input—if, for instance, such errors are weighted by the precision of the original prediction as in a Bayesian observer model (Legaspi & Toyoizumi, 2019).

4.2 Conclusion: Multiple neural processes, distinguished by their temporal dynamics, underlie sense of agency.

While, as argued above, our data support the claim that unconscious sensorimotor processes contribute to agency judgements, they do not suggest that this is the only type of process that contributes to SoA. In more recent theoretical accounts of SoA, sensorimotor prediction error is combined with prior (or prospective) information in the formation of agency judgments (Haggard, 2017; Legaspi & Toyoizumi, 2019; Synofzik et al., 2008). In particular, some models posit that this integration occurs in a multistep process (Ohata et al., 2020; Synofzik et al., 2008). This view dovetails with dual-process models of movement initiation, which suggest that action monitoring occurs on at least two distinct timescales through a non-conscious and a conscious pathway, respectively (Del Cul et al., 2009). Previous studies have estimated that the latter, conscious pathway that represents high-level intentions begins to dominate EEG decoding results following 200 ms post stimulus (Charles et al., 2014).

Importantly, we found that the earlier (<160 ms) patterns of EEG activity that predicted agency judgements were transient; while it remained possible to decode SoA from ~83 ms post-stimulation onward (see Figure 6b), the patterns that predicted SoA evolved over time (see Figure 5a). In other words, the putative neural representation underlying the observed decoding performance was continually changing, which has been argued to reflect a chain of sequential transformations between neural/mental representations of the sensory stimulus (King & Dehaene, 2014).

By at least 250 ms (but perhaps as early as 167 ms, see Figure 5) following the onset of electrical stimulation, however, we observed that the temporal dynamics of SoA predicting patterns undergo a distinct transition from a sequence of transient neural representations to a single, sustained pattern that differentiates SoA and non-SoA trials for the remainder of the epoch. The timing of this shift is quite consistent with that predicted by dual-process models of action processing (Charles et al., 2014; Del Cul et al., 2009), and the sustained nature of the predictive voltage patterns is consistent with the temporal signature of high-level novelty/error detection that has been argued to require conscious awareness (Dehaene & King, 2016) and previously proposed to inform agency judgements (Kühn et al., 2011).

The temporal generalization method used here allowed us to differentiate between these two distinct neural processes (evident in Figure 5) even though the separation is not self-evident in the time-by-time decoding accuracies (Figure 6) or in a typical evoked response analysis (King &

Dehaene, 2014). Thus, our decoding results are consistent with an account in which multiple, distinct neurocognitive processes contribute to judgments of agency (Synofzik et al., 2008). In particular, our findings suggest that unconscious neural processing of sensory input occurs over a sequential chain of steps, followed by a sustained process that differentiates self-caused from externally-caused sensation and may represent high-level, conscious action monitoring (Del Cul et al., 2009; Kühn et al., 2011).

4.3. Hypothesis: Brain activity is more complex following an action perceived as self-caused.

Further, the ability of fractal measures to predict SoA on a trial-by-trial basis in our exploratory analysis motivates compelling hypotheses concerning the relationship between the sense of agency and the internal dynamics of cortex. In particular, we see a robust increase in the complexity (fractal dimension) of the EEG across seemingly the whole of cortex (see Figure 7a) in the wake of a movement perceived as self-caused. In the literature, increasing fractal dimension has been associated with consciousness awareness (Kesić & Spasić, 2016), motivated by a correspondence with information theoretic complexity measures that have become a focus of consciousness research (Oizumi et al., 2014; Ruiz de Miras et al., 2019). Moreover, fractal dimension has been associated with internally-generated qualia (Ibáñez-Molina & Iglesias-Parro, 2014), but not with the sense of agency per se. While fractal dimension is suppressed in certain psychiatric conditions that affect action awareness (Sokunbi et al., 2014; Stier et al., 2021), this is the first demonstration to our knowledge that the complexity of brain signals covaries with this aspect of moment-to-moment self-consciousness. Given that, in dynamical systems theory, the fractal dimension of a time series reflects the dimension of the attractor that governs the underlying data generating process, one could speculate that the nervous system sorts self-caused from externally-caused stimuli by confining their neural responses to distinct manifolds (Kohn et al., 2020).

4.4. Hypothesis: Intrinsic dynamics of frontocentral cortical regions reflect agency-dependent gating of motor system to sensory input.

Lastly, we observe that the Hurst exponent, a measure of long-range temporal dependency within a time series, predicts the trial-by-trial experience of agency, but only in one frontocentral electrode. In the cognitive neuroscience literature, the Hurst exponent has been interpreted as corresponding to how responsive the underlying neural system is to external input, as it reflects how sensitive the full time series is to local perturbation (Churchill et al., 2016; Kardan et al., 2020; Zhuang et al., 2022). Thus, a tempting explanation of this finding is that specific frontocentral cortical regions, speculatively those involved in action initiation and control, are more sensitive to input from sensory regions when sensations are perceived as self-caused. Indeed, the internal dynamics of cortex are known to govern the cortical response to sensory input (MacLean et al., 2005). Thus, while sensory prediction errors may propagate throughout cortex regardless of SoA (Adams et al., 2013; Friston, 2005; Timm et al., 2016), the internal dynamics of motor areas may modulate receptivity to such errors as a function of whether those error are deemed self-caused—i.e., those prediction errors that would be informative to updating an internal model (Wolpert et al., 1995, 2001). In such a view, SoA may play a functional role in governing information flow in cortical networks to facilitate motor learning. Given the unreliability of Hurst exponents estimated from short time series, however, this explanation remains speculative. More direct tests of this hypothesis could be a fruitful avenue for future research, providing a possible mechanism for the functional role of SoA during motor learning (Kasahara et al., 2021; Wen & Haggard, 2020).

4.5. Outlook

While SoA has become a topic of increased attention in recent decades, most research in the area has focused on the experience of agency over downstream consequences of one's actions as they affect the external world rather than the more basal experience of directing one's own muscles (Haggard, 2008, 2017). We introduce the use of human-in-the-loop Bayesian optimization, in combination with electrical muscle stimulation, to experimentally manipulate the subjective experience of controlling the musculature. As we showcase here, this approach enables novel behavioral and neuroimaging investigations into the substrate of embodied self-awareness. Our results provide confirmatory evidence for the predictive relationship between low-level sensorimotor processes and SoA, which had recently been called into question (Dewey & Knoblich, 2014; Timm et al., 2016). This relationship can be understood through the lens of dual-process theories of action selection and monitoring (Del Cul et al., 2009), which are supported by our observation of at least two neural processes, with distinct temporal dynamics, that predict the SoA over individual muscle movements. Lastly, our exploratory analysis suggests that nonlinear properties of cortical dynamics reflect the sense of agency, motivating new hypotheses about the functional role of SoA as it relates to the trafficking of sensory information in motor networks.

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Agrawal, S., & Goyal, N. (2012). Analysis of Thompson Sampling for the Multi-armed Bandit Problem. *Proceedings of the 25th Annual Conference on Learning Theory*, 39.1-39.26. https://proceedings.mlr.press/v23/agrawal12.html

Appelhoff, S., Hurst, A., Lawrence, A., Li, A., Mantilla, R., Yorguin, J., O'Reilly, C., Xiang, L., & Dancker, J. (2022). *PyPREP: A Python implementation of the preprocessing pipeline (PREP) for EEG data*. Zenodo. https://doi.org/10.5281/zenodo.6363575

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*. https://www.frontiersin.org/articles/10.3389/fninf.2015.00016

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P., Horsfall, P., & Goodman, N. D. (2019). Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*, *20*(28), 1–6.

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, *6*(6), 237–242. https://doi.org/10.1016/S1364-6613(02)01907-1

Bouckaert, R. R., & Frank, E. (2004). Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. In H. Dai, R. Srikant, & C. Zhang (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 3–12). Springer. https://doi.org/10.1007/978-3-540-24775-3_3

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2022). Bambi: A Simple Interface for Fitting Bayesian Linear Models in Python. *Journal of Statistical Software*, *103*, 1–29. https://doi.org/10.18637/jss.v103.i15

Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., & Palminteri, S. (2020). Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*, *4*(10), Article 10. https://doi.org/10.1038/s41562-020-0919-5

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. *Consciousness and Cognition*, *101*, 103307. https://doi.org/10.1016/j.concog.2022.103307

Charles, L., King, J.-R., & Dehaene, S. (2014). Decoding the Dynamics of Action, Intention, and Error Detection for Conscious and Subliminal Stimuli. *Journal of Neuroscience*, *34*(4), 1158–1170. https://doi.org/10.1523/JNEUROSCI.2465-13.2014

Chronicle, E. P., & Glover, J. (2003). A Ticklish Question: Does Magnetic Stimulation of the Primary Motor Cortex Give Rise to an 'Efference copy'? *Cortex*, *39*(1), 105–110. https://doi.org/10.1016/S0010-9452(08)70078-9

Churchill, N. W., Spring, R., Grady, C., Cimprich, B., Askren, M. K., Reuter-Lorenz, P. A., Jung, M. S., Peltier, S., Strother, S. C., & Berman, M. G. (2016). The suppression of scale-free fMRI brain dynamics across three different sources of effort: Aging, task novelty and task difficulty. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep30895

Darvas, F., Scherer, R., Ojemann, J. G., Rao, R. P., Miller, K. J., & Sorensen, L. B. (2010). High gamma mapping using EEG. *NeuroImage*, *49*(1), 930–938. https://doi.org/10.1016/j.neuroimage.2009.08.041

Dehaene, S., & Changeux, J.-P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

Dehaene, S., & King, J.-R. (2016). Decoding the Dynamics of Conscious Perception: The Temporal Generalization Method. In G. Buzsáki & Y. Christen (Eds.), *Micro-, Meso- and Macro-Dynamics of the Brain*. Springer. http://www.ncbi.nlm.nih.gov/books/NBK435757/

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

Dewey, J. A., & Knoblich, G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing? *PLoS ONE*, *9*. https://doi.org/10.1371/journal.pone.0110118

Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

Eke, A., Hermán, P., Bassingthwaighte, J. B., Raymond, G. M., Percival, D. B., Cannon, M., Balla, I., & Ikrényi, C. (2000). Physiological time series: Distinguishing fractal noises from motions. *Pflugers Archiv European Journal of Physiology*, *439*(4), 403–415. Scopus. https://doi.org/10.1007/s004240050957

Eke, A., Herman, P., Kocsis, L., & Kozak, L. R. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement*, *23*(1), R1–R38. Scopus. https://doi.org/10.1088/0967-3334/23/1/201

Fries, P., Scheeringa, R., & Oostenveld, R. (2008). Finding Gamma. *Neuron*, *58*(3), 303–305. https://doi.org/10.1016/j.neuron.2008.04.020

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Frith, C. (2012). Explaining delusions of control: The comparator model 20years on. *Consciousness and Cognition*, *21*(1), 52–54. https://doi.org/10.1016/j.concog.2011.06.010

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Gandolla, M., Ward, N. S., Molteni, F., Guanziroli, E., Ferrigno, G., & Pedrocchi, A. (2016). The Neural Correlates of Long-Term Carryover following Functional Electrical Stimulation for Stroke. *Neural Plasticity*, *2016*, e4192718. https://doi.org/10.1155/2016/4192718

Gneiting, T., & Schlather, M. (2004). Stochastic models which separate fractal dimension and Hurst effect. *SIAM Review*, *46*(2), 269–282. https://doi.org/10.1137/S0036144501394387

Goupil, L., Johansson, P., Hall, L., & Aucouturier, J.-J. (2021). Vocal signals only impact speakers' own emotions when they are self-attributed. *Consciousness and Cognition*, *88*, 103072. https://doi.org/10.1016/j.concog.2020.103072

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Haggard, P. (2008). Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*(12), Article 12. https://doi.org/10.1038/nrn2497

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

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology*, *22*(10), R390–R392. https://doi.org/10.1016/j.cub.2012.02.040

Haggard, P., & Clark, S. (2003). Intentional action: Conscious experience and neural prediction. *Consciousness and Cognition*, *12*(4), 695–707. https://doi.org/10.1016/S1053-8100(03)00052-7

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

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, *31*(2), 277–283. https://doi.org/10.1016/0167-2789(88)90081-4

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10*(3), 626–634. https://doi.org/10.1109/72.761722

Ibáñez-Molina, A. J., & Iglesias-Parro, S. (2014). Fractal characterization of internally and externally generated conscious experiences. *Brain and Cognition*, *87*, 69–75. https://doi.org/10.1016/j.bandc.2014.03.002

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

Kardan, O., Adam, K. C. S., Mance, I., Churchill, N. W., Vogel, E. K., & Berman, M. G. (2020). Distinguishing cognitive effort and working memory load using scale-invariance and alpha suppression in EEG. *NeuroImage*, *211*, 116622. https://doi.org/10.1016/j.neuroimage.2020.116622

Kasahara, S., Nishida, J., & Lopes, P. (2019). Preemptive action: Accelerating human reaction using electrical muscle stimulation without compromising agency. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. https://dl.acm.org/doi/abs/10.1145/3290605.3300873

Kasahara, S., Takada, K., Nishida, J., Shibata, K., Shimojo, S., & Lopes, P. (2021). Preserving Agency During Electrical Muscle Stimulation Training Speeds up Reaction Time Directly After Removing EMS. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3411764.3445147

Kayser, J., & Tenke, C. E. (2015). On the benefits of using surface Laplacian (current source density) methodology in electrophysiology. *International Journal of Psychophysiology*, *97*(3), 171–173. https://doi.org/10.1016/j.ijpsycho.2015.06.001

Kesić, S., & Spasić, S. Z. (2016). Application of Higuchi's fractal dimension from basic to clinical neurophysiology: A review. *Computer Methods and Programs in Biomedicine*, *133*, 55–70. https://doi.org/10.1016/j.cmpb.2016.05.014

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Kohn, A., Jasper, A. I., Semedo, J. D., Gokcen, E., Machens, C. K., & Yu, B. M. (2020). Principles of Corticocortical Communication: Proposed Schemes and Design Considerations. *Trends in Neurosciences*, *43*(9), 725–737. https://doi.org/10.1016/j.tins.2020.07.001

Kühn, S., Nenchev, I., Haggard, P., Brass, M., Gallinat, J., & Voss, M. (2011). Whodunnit? Electrophysiological Correlates of Agency Judgements. *PLOS ONE*, *6*(12), e28657. https://doi.org/10.1371/journal.pone.0028657

Legaspi, R., & Toyoizumi, T. (2019). A Bayesian psychophysics model of sense of agency. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-12170-0

Lhermitte, F., Pillon, B., & Serdaru, M. (1986). Human autonomy and the frontal lobes. Part I: Imitation and utilization behavior: A neuropsychological study of 75 patients. *Annals of Neurology*, *19*(4), 326–334. https://doi.org/10.1002/ana.410190404

Li, X., Liang, Z., Kleiner, M., & Lu, Z.-L. (2010). RTbox: A device for highly accurate response time measurements. *Behavior Research Methods*, *42*(1), 212–225. https://doi.org/10.3758/BRM.42.1.212

Liang, K.-Y., & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Libet, B., Alberts, W. W., Wright, E. W., & Feinstein, B. (1967). Responses of Human Somatosensory Cortex to Stimuli below Threshold for Conscious Sensation. *Science*, *158*(3808), 1597–1600. https://doi.org/10.1126/science.158.3808.1597

MacLean, J. N., Watson, B. O., Aaron, G. B., & Yuste, R. (2005). Internal Dynamics Determine the Cortical Response to Thalamic Stimulation. *Neuron*, *48*(5), 811–823. https://doi.org/10.1016/j.neuron.2005.09.035

Martis, R. J., Tan, J. H., Chua, C. K., Loon, T. C., Yeo, S. W. J., & Tong, L. (2015). Epileptic eeg classification using nonlinear parameters on different frequency bands. *Journal of Mechanics in Medicine and Biology*, *15*(03), 1550040. https://doi.org/10.1142/S0219519415500402

Nadeau, C., & Bengio, Y. (1999). Inference for the Generalization Error. *Advances in Neural Information Processing Systems*, *12*. https://proceedings.neurips.cc/paper/1999/hash/7d12b66d3df6af8d429c1a357d8b9e1a-Abstract.html

Ohata, R., Asai, T., Kadota, H., Shigemasu, H., Ogawa, K., & Imamizu, H. (2020). Sense of Agency Beyond Sensorimotor Process: Decoding Self-Other Action Attribution in the Human Brain. *Cerebral Cortex*, *30*(7), 4076–4091. https://doi.org/10.1093/cercor/bhaa028

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLOS Computational Biology*, *10*(5), e1003588. https://doi.org/10.1371/journal.pcbi.1003588

Oliver, R., & Ballester, J. L. (1998). Is there memory in solar activity? *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *58*(5), 5650–5654. Scopus. https://doi.org/10.1103/PhysRevE.58.5650

Panikkath, R., Panikkath, D., Mojumder, D., & Nugent, K. (2014). The alien hand syndrome. *Proceedings (Baylor University. Medical Center)*, *27*(3), 219–220.

Patil, A., Huard, D., & Fonnesbeck, C. J. (2010). PyMC: Bayesian Stochastic Modelling in Python. *Journal of Statistical Software*, *35*(4), 1–81.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Qian, B., & Rasheed, K. (2004, November). Hurst exponent and financial market predictability. *Proceedings of the IASTED International Conference*. IASTED conference on Financial Engineering and Applications, Cambridge, MA.

Ramachandran, V. S., & Hirstein, W. (1998). The perception of phantom limbs. The D. O. Hebb lecture. *Brain*, *121*(9), 1603–1630. https://doi.org/10.1093/brain/121.9.1603

Rosenblatt, J. D., Finos, L., Weeda, W. D., Solari, A., & Goeman, J. J. (2018). All-Resolutions Inference for brain imaging. *NeuroImage*, *181*, 786–796. https://doi.org/10.1016/j.neuroimage.2018.07.060

Ruiz de Miras, J., Soler, F., Iglesias-Parro, S., Ibáñez-Molina, A. J., Casali, A. G., Laureys, S., Massimini, M., Esteban, F. J., Navas, J., & Langa, J. A. (2019). Fractal dimension analysis of states of consciousness and unconsciousness using transcranial magnetic stimulation. *Computer Methods and Programs in Biomedicine*, *175*, 129–137. https://doi.org/10.1016/j.cmpb.2019.04.017

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335

Schiratti, J.-B., Le Douget, J.-E., Le Van Quyen, M., Essid, S., & Gramfort, A. (2018). An Ensemble Learning Approach to Detect Epileptic Seizures from Long Intracranial EEG Recordings. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 856–860. https://doi.org/10.1109/ICASSP.2018.8461489

Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and Statistical Modeling with Python*. 92–96. https://doi.org/10.25080/Majora-92bf1922-011

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), Article 10. https://doi.org/10.1038/nn1549

Smith, N. J., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology*, *52*(2), 169–181. https://doi.org/10.1111/psyp.12320

Sokunbi, M. O., Gradin, V. B., Waiter, G. D., Cameron, G. G., Ahearn, T. S., Murray, A. D., Steele, D. J., & Staff, R. T. (2014). Nonlinear Complexity Analysis of Brain fMRI Signals in Schizophrenia. *PLOS ONE*, *9*(5), e95146. https://doi.org/10.1371/journal.pone.0095146

Stier, A., Cardenas-Iniguez, C., Kardan, O., Moore, T., Meyer, F., Rosenberg, M., Kaczkurkin, A., Lahey, B., & Berman, M. (2021). *A Scale-Free Gradient of Cognitive Resource Disruptions in Childhood Psychopathology*. https://doi.org/10.1101/2021.08.24.457554

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Tajima, D., Nishida, J., Lopes, P., & Kasahara, S. (2022). Whose Touch is This?: Understanding the Agency Trade-Off Between User-Driven Touch vs. Computer-Driven Touch. *ACM Transactions on Computer-Human Interaction*, *29*(3), 24:1-24:27. https://doi.org/10.1145/3489608

Timm, J., Schönwiesner, M., Schröger, E., & SanMiguel, I. (2016). Sensory suppression of brain responses to self-generated sounds is observed with and without the perception of agency. *Cortex*, *80*, 5–20. https://doi.org/10.1016/j.cortex.2016.03.018

Veillette, J. (2022). *MNE-ARI: All-resolutions Inference for M/EEG in Python* (0.1.2). Zenodo. https://doi.org/10.5281/zenodo.5813807

Voss, M., Ingram, J. N., Haggard, P., & Wolpert, D. M. (2006). Sensorimotor attenuation by central motor command signals in the absence of movement. *Nature Neuroscience*, *9*(1), Article 1. https://doi.org/10.1038/nn1592

Wen, W., & Haggard, P. (2020). Prediction error and regularity detection underlie two dissociable mechanisms for computing the sense of agency. *Cognition*, *195*, 104074. https://doi.org/10.1016/j.cognition.2019.104074

Wen, W., Ishii, H., Ohata, R., Yamashita, A., Asama, H., & Imamizu, H. (2021). Perception and control: Individual difference in the sense of agency is associated with learnability in sensorimotor adaptation. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-99969-4

Westfall, J. (2017). Statistical details of the default priors in the Bambi library. In *ArXiv e-prints*. https://doi.org/10.48550/arXiv.1702.01201

Wolpert, D. M., Ghahramani, Z., & Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, *5*(11), 487–494. https://doi.org/10.1016/S1364-6613(00)01773-3

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Zhuang, C., Meidenbauer, K. L., Kardan, O., Stier, A. J., Choe, K. W., Cardenas-Iniguez, C., Huppert, T. J., & Berman, M. G. (2022). Scale invariance in fNIRS as a measurement of cognitive load. *Cortex*, *154*, 62–76. https://doi.org/10.1016/j.cortex.2022.05.009