Overlapping Cortical Substrate of Biomechanical Control and Subjective Agency

John P. Veillette[1†]
Alfred F. Chao[1]
Romain Nith[2]
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

A.F.C.: Software and Writing - review & editing. **H.C.N.:** Conceptualization, Funding acquisition, Resources, Supervision, and Writing - review & editing. **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, Resources, and Writing - review & editing. **R.N.:** Resources and Writing - review & editing.

Acknowledgements

This work was supported by NSF BCS 2024923 to H.C.N. and P.L., and J.P.V. was supported by NSF GRFP DGE 1746045. Data acquisition was completed with equipment funded by NIH S10OD018448 to the Magnetic Resonance Imaging Research Center at the University of Chicago, and analysis used resources provided by the University of Chicago's Research Computing Center.

Abstract

Every movement requires the nervous system to solve a complex biomechanical control problem, but this process is mostly veiled from one's conscious awareness. Simultaneously, we also have conscious experience of controlling our movements—our sense of agency (SoA). Whether SoA corresponds to those neural representations that implement actual neuromuscular control is an open question with ethical, medical, and legal implications. If SoA is the conscious experience of control, this predicts that SoA can be decoded from the same brain structures that implement the so-called "inverse kinematic" computations for planning movement. We correlated human fMRI measurements during hand movements with the internal representations of a deep neural network (DNN) performing the same hand control task in a biomechanical simulation–revealing detailed cortical encodings of sensorimotor states, idiosyncratic to each subject. We then manipulated SoA by usurping control of participants' muscles via electrical stimulation, and found that the same voxels which were best explained by modeled inverse kinematic representations—which, strikingly, were located in canonically visual areas—also predicted SoA. Importantly, model-brain correspondences and robust SoA decoding could both be achieved within single subjects, enabling relationships between motor representations and awareness to be studied at the level of the individual.

Introduction

Even simple movements require the nervous system to solve a complex control problem. The human hand alone has more than 20 degrees of freedom (i.e., directions joints can move) and is controlled by over 30 muscles. Complicating the problem, each joint can be moved by multiple muscles, each of which move multiple joints (Bullock et al., 2012). Theoretical accounts posit our brains contain a vast, interacting set of internal representations of the body's positional states and kinematics, but little of this machinery is accessible to awareness (Blakemore et al., 2002). The act of moving *feels* straightforward and is accompanied by a sense of agency (SoA)—a feeling of "I did that." Does this conscious experience of directing movement reflect the actual machinery of control, and if so, what types of motor representation influence awareness?

Most extant research focuses on *prediction errors* between expected and actual sensory outcomes as determinants of SoA (Frith, 1987; Synofzik et al., 2008). However, not all errors diminish SoA—some mismatches can even overwrite our original intention in memory (Lind et al., 2014). Since errors only account for *negative* experiences of agency, not why we feel SoA in the first place, such "comparator" models fall short of predicting which mismatches are consciously detected (Frith, 2012).

The need for a predictive account of SoA is becoming urgent. Advanced neuroprostheses leverage the technology behind today's large language models to "autocomplete" instructions decoded from the brain (Metzger et al., 2023; Tang et al., 2023). These interfaces may pose a threat to the autonomy of those they aim to help if deviations from user intentions are not accessible to awareness. Some countries have already passed legislation to this effect (Fernández & Fernández, 2022). However, without a foundational understanding of which motor representations affect consciousness, it is unclear how to design neural interfaces around these constraints.

Normative theories of motor control posit two sorts of internal models: *forward kinematic models* to predict future states of the body from current neuromuscular output, and *inverse kinematic models* or *control policies* to generate neuromuscular output that achieves a desired state (Wolpert & Ghahramani, 2000). By some accounts, paired forward and inverse models are specialized for specific *sensorimotor contexts* (e.g. moving with rested vs. fatigued muscles), and forward model prediction errors are used not just to update the models for the current context, but also to identify when contexts switch (Wolpert & Kawato, 1998). The recent COIN model (Heald et al., 2021), which posits neuromuscular output is averaged over all control policies (weighted by the probability of being in each context) but subjects only have awareness of one context at a time, quantitatively predicts the contribution of explicit (i.e. conscious) and implicit components of adaption following rotation of subjects' visuomotor mapping. Following this logic, SoA may persist following a prediction error if the error can be explained by another sensorimotor context for which one already has a control model, but SoA is lost when one lacks an inverse kinematic model for the current best-guess context.

This hypothesis makes a testable prediction that SoA can be *decoded* from the same cortical areas where representations of an inverse kinematic model are putatively *encoded* (in the context for which that model would ordinarily apply). In the present study, participants performed hand gestures during functional MRI. To localize inverse kinematic representations, we predicted

participants' voxelwise brain activity during the task from activations of a deep neural network (DNN) that performed the same task with a simulated biomechanical hand (Caggiano et al., 2022). Then, in another session, we usurped control of subjects' muscles using functional electrical stimulation; by slightly preempting subjects' endogenous movements, we elicit an erroneous SoA over roughly half of involuntary movements as validated in our previous work (Veillette et al., 2023). We test whether we can decode participants' SoA over individual muscle movements from those same voxels that were best predicted by inverse kinematic representations days prior.

Results

Rather than testing a large number of participants for a short amount of time as in typical fMRI studies, we focused on collecting sufficient amounts of data to establish robust model-brain correspondences in each of four participants (i.e., 3.5 hours of BOLD data, collected over 5 hours). By focusing on the individual rather than the group as the unit of analysis, this "dense sampling" approach is more sensitive to neural patterns that are robust within individuals but idiosyncratic across them (Naselaris et al., 2021; Poldrack, 2017), which is critical for modeling neural encodings of high-dimensional feature spaces (Cross et al., 2021; Tang et al., 2023).

Consequently, each subject was treated as an $n = 1$ experiment, and all statistics were performed within each subject. The first subject is labelled "sub-01" and other subjects are given arbitrary letter codes, to denote that they are replications of the first $n = 1$ experiment. For each analysis, we report a p-value for each subject (with their subject ID as a subscript), and we combine p-values across subjects meta-analytically using Fisher's method—this combined p-value (denoted p_{all}) corresponds to the "global" null hypothesis that there is no effect in any subject, accounting for multiple comparisons. A significant result tells us that an effect exists in the population but does not imply it is representative—a limitation shared by most dense sampling studies—but we can put a Bayesian lower bound on the population prevalence of our main findings using an approach recently described by Ince and colleagues to verify our results are still likely to apply to a sizeable portion of the population (Ince et al., 2021).

Encoding of sensorimotor states across cortex is widespread but idiosyncratic

Subjects completed ten 10-minute runs of a motor task in which they performed a hand gesture voluntarily, attempting to match a visually-presented hand gesture, which switched every 5 seconds. During this, we recorded their joint angles and velocities using an MRI-compatible motion tracking glove (60 Hz sampling rate). While target positions were presented visually, participants could not see their own hands outside of the scanner. The unit of analysis was not these 5-second trials, but prediction of the blood oxygen level dependent (BOLD) measurement at each voxel in each fMRI measurement (every 2 seconds).

To build our inverse kinematics-informed encoding model, we used a DNN that was trained to generate specified hand gestures, like our human participants, in a biomechanical simulation. Specifically, we used the pretrained model released with *MyoSuite*'s hand pose simulation environment; this model maps a current task state (i.e., joint angles, velocities, and distance from target hand position) to a set of muscle activations that aim to achieve the target position (Caggiano et al., 2022). Though it does not reach human-level performance—and

importantly, it never saw human data during training—the model approximates a solution to the inverse kinematic problem, achieving a joint angle error (summed across joints and simulation time) of 1.94 radians after 1,000 training iterations (compared to 3.29 rad after the first training iteration) using the natural policy gradient method (Kakade, 2001). To translate this approximation into a neural encoding model, we input each measurement of our human participants' task state at each time point into the DNN and extracted all activations from all artificial neurons; these activations were used as features to linearly predict the subjects' continuous BOLD activity using a separate ridge regression for each voxel (see Figure 1).

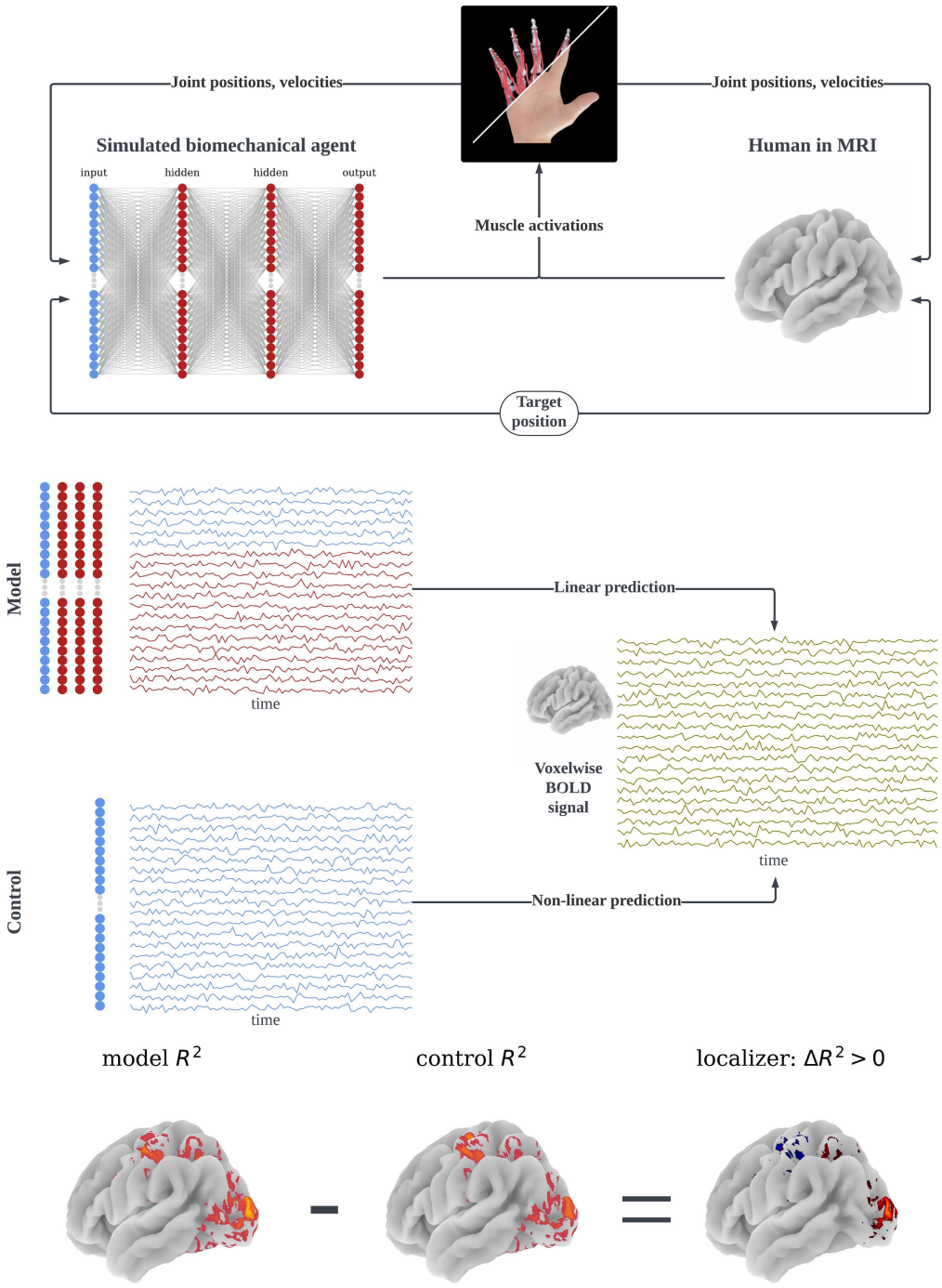

Figure 1: Approach to localizing inverse kinematic model representations in cortex. Subjects moved their hands in specified "target positions" while we recorded their hand movements in an fMRI scanner. We used the activations of a deep neural network (DNN) which performs the same hand pose task with a simulated biomechanical hand—i.e., approximates an "inverse kinematic model" or controller for the hand—as features to predict voxelwise brain activity over time. We compare the out-of-sample predictive accuracy of the inverse kinematic DNN model to a purely data-driven voxelwise encoding model, and the voxels that are better predicted by our biomechanics-informed model are interpreted as reflecting the representations of an inverse kinematic model.

The resulting encoding models describe each voxel's encoding properties as a 163-dimensional vector containing linear model weights for each neuron in the DNN, the activations of which are deterministic functions of the DNN's inputs. These voxelwise encoding models could predict substantial out-of-sample variance in voxel responses (block permutation test w/ R^2-max correction: FWER-corrected p = 0.0002 for all subjects, which is the lowest p-value our permutation test could obtain). To visualize the model parameters interpretably, we divided the DNN's inputs into three feature spaces—joint positions, joint velocities, and deviation from target positions—and computed Shapley values for each feature space, which quantify the relative contribution of each feature space to (per-voxel) predictions on the test set (Lundberg & Lee, 2017). As seen in Figure 2, the broad spatial extent of cortex predicted by the encoding model is relatively similar across subjects—showing recruitment of motor, somatosensory, and visual cortices—but the fine-grained encoding properties of individual voxels show little alignment across subjects despite generalizing across fMRI runs within a subject. All of this detail would be lost to group-averaging.

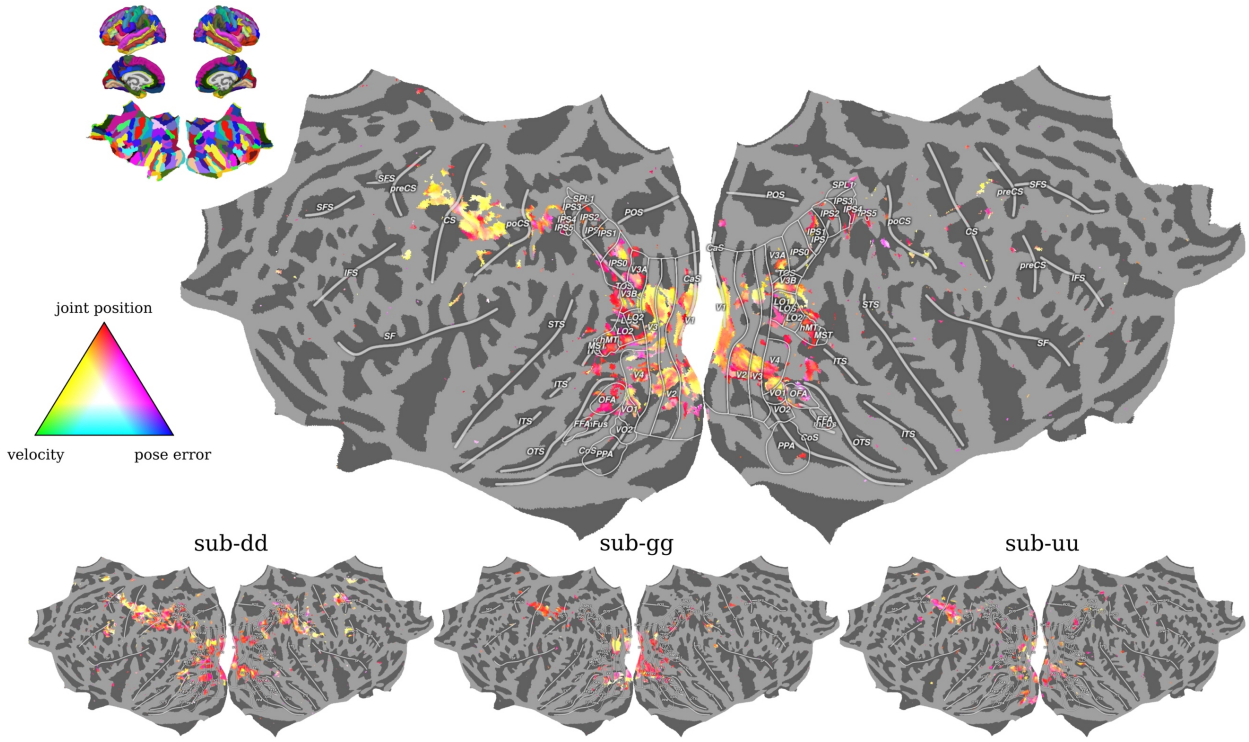

Figure 2: Voxelwise response properties while moving the hand. Relative importance of sensorimotor task state features during the motor control task were estimated using Shapley values. Each voxel that is significantly predicted by the DNN-based encoding model (FDR < 0.05) is assigned an RGB color based on these feature importances. For example, a voxel colored red is predicted by DNN units that respond exclusively to hand position, but a voxel colored pink behaves like DNN units that respond to combinations of hand position and deviance from target position—i.e., likely encode the target position.

Modeled inverse kinematic representations predict voxels in early visual cortex

While the inverse kinematic DNN-informed model outperforms chance prediction in much of cortex, this is insufficient evidence that the DNN's representations are uniquely good predictors

of voxel responses. Another explanation is that the inputs to the DNN—parametrizing the sensorimotor task state—or any nonlinear projection of those input features (of which the DNN activations are just one random choice) are sufficient to predict voxel activity. To this end, we fit a control encoding model, in which a separate kernel ridge regression (which can learn linear or nonlinear functions) is used to predict voxel activity from just the input layer to the neural network. This control model also explains substantial variance in all subjects (FWER-corrected $p = 0.0002$ for all subjects) but with no inverse kinematic prior.

To isolate voxels that are uniquely well-predicted by the inverse kinematic DNN's representations, we subtracted the variance explained (i.e., out-of-sample R^2) by the control model from that explained by the DNN-informed model in each voxel (see Figure 3). After correcting for multiple comparisons, we find that modeled inverse kinematic representations improve prediction in small patches of, strikingly, early visual cortex (V1, V2, and V3) in three out of four subjects (paired block permutation test w/ TFCE correction: FWER-corrected $p_{01} = 0.0006$, $p_{dd} = 0.0300$, $p_{gg} = 0.0494$, $p_{uu} = 0.946$; $p_{\text{all}} = 0.0005$), though inverse kinematic responsive voxels did not spatially align across subjects at the voxel level (see Figure 3). Notably, subjects could not see their hands while their heads were in the MRI scanner, so visual monitoring of their hand cannot explain the visual cortical location.

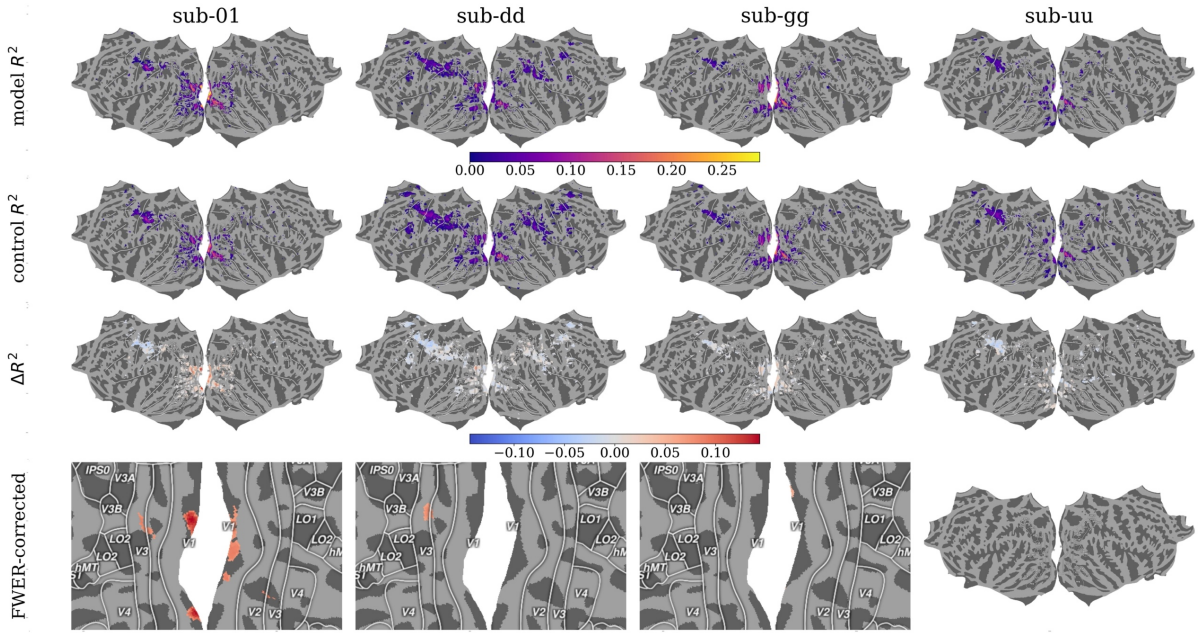

Figure 3: Inverse kinematic model representations predict voxel activity in visual cortex. Out-of-sample variance in voxelwise activity explained by the control model was subtracted from that explained by the DNN approximating an inverse kinematic model. Voxels significantly better explained by inverse kinematic model representations, after correcting for multiple comparisons (FWER < 0.05), are interpreted as encoding similar representations.

Our main hypothesis test entails decoding SoA from voxels selectively predicted by inverse kinematic representations. Since we were unable to identify any such voxels in sub-*uu*, this subject was dropped from subsequent analyses. It is not clear why our model did not perform as well in sub-*uu*; the control model also underperformed in this subject But the significant results in 3-out-of-4 subjects allows us to conclude that neural activity resembling our model's representations is

present in the early visual cortex of at least 31.0% of the population (MAP = 73.7%, 95% HDI: [29.4%, 97.3%]) with 95% posterior probability (Ince et al., 2021).

Inverse kinematic responsive voxels predict sense of agency over electrically-actuated muscle movements

In a second session, participants returned to complete ten 10-minutes runs of a cue-response reaction time task, pressing a button with their ring finger as quickly as possible following a visual prompt. After the first (baseline) block, we applied functional electrical stimulation (FES) to the *flexor digitorum profundus* muscle to produce a muscle movement which would cause an involuntarily press of the button around the time participants would respond on their own. After each trial, they were asked to discriminate whether they or the muscle stimulator caused the button press. Using a Bayesian adaptive procedure developed in our prior work (Veillette et al., 2023), we continuously adjusted the stimulation latency until subjects responded that they caused the button press ~50% of the time (see Figure 4a)—which, for most subjects, is robustly before they could have possibly begun to move (Kasahara et al., 2019; Tajima et al., 2022; Veillette et al., 2023). Trial-by-trial agency judgements in this task can be decoded from electroencephalogram (EEG) within the first 100 ms following the onset of muscle stimulation, indicating that variation in self-reports usually reflects sensorimotor processes rather than post hoc guessing (Veillette et al., 2023).

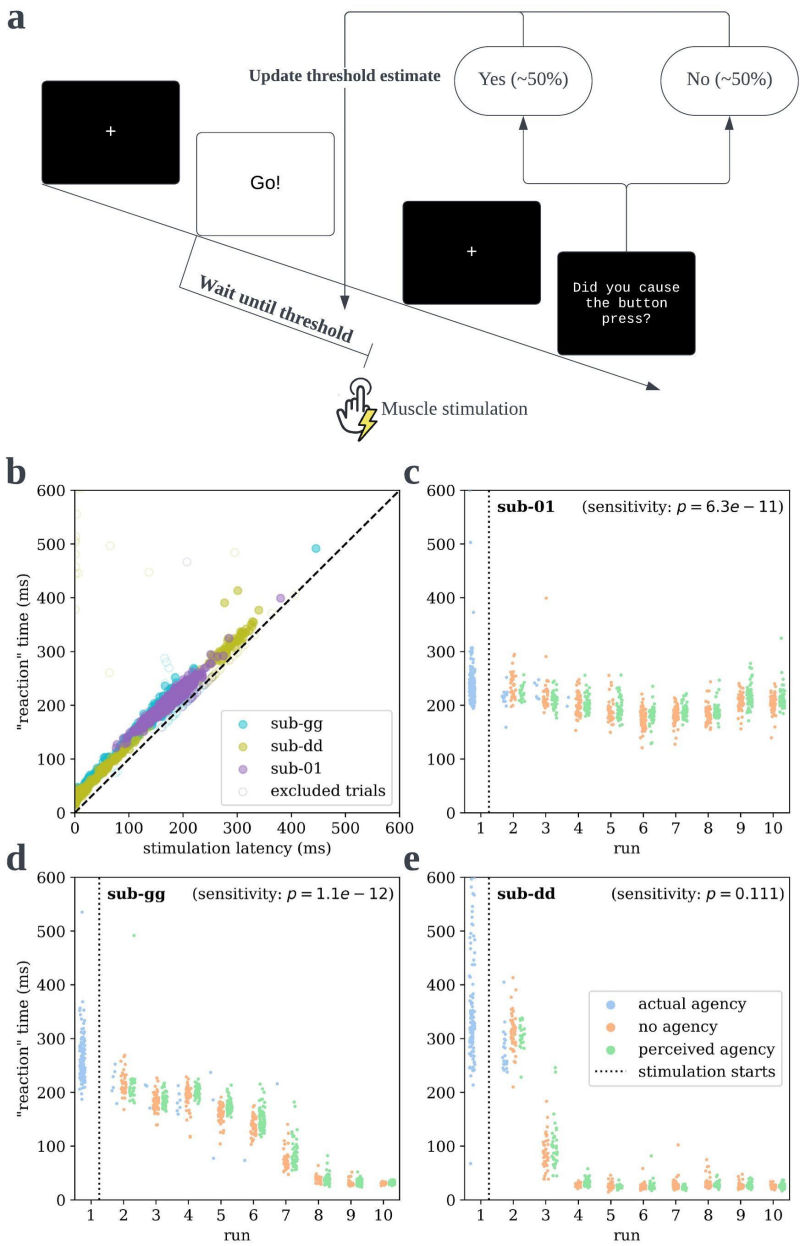

Figure 4: Manipulating sense of agency (SoA) by usurping control of subjects' muscles with electrical stimulation. (a) Subjects complete a cue-response reaction time task, but we used muscle stimulation to preempt their self-produced movements on most trials. Subjects were asked to discriminate, after each trial, whether they or the muscle stimulation caused their finger to press the button. Their response was used to update a threshold estimate on which the timing of stimulation was based, such that we could guarantee they believe they caused roughly 50% of movements. (b) As expected from our prior work, this 50% threshold is substantially earlier than subjects move autonomously, as illustrated by the fact that button presses ("response" times) are essentially a linear function of stimulation latency—seen here for all trials across all subjects. (c-e) This adaptive procedure tracks non-stationarity in subjects' threshold, such that subjects rarely move faster than the muscle stimulator (blue) after the first run with stimulation. Consequently, we obtain distributions of "no agency" and "perceived" but false agency trials with highly overlapping stimulation latencies, so subjects' experienced SoA can be dissociated from the task manipulation. Logistic regression *p*-values for the sensitivity of subjects' SoA responses to the latency of muscle stimulation are shown.

Our three remaining participants were, similarly, preempted by FES on the overwhelming majority of trials, as indicated by the fact that "response" times—i.e., the time of the button press—are a linear function of the stimulation latency (see Figure 4b). However, sub-*dd* was not clearly sensitive to the latency of muscle stimulation around their threshold (logistic regression w/ random effect for fMRI runs: p_{01} = 6.3e-11, p_{dd} = 0.1105, p_{gg}= 1.1e-12; p_{all} = 1.2e-20). This, combined with the fact that the Bayesian adaptive procedure could quickly push sub-*dd* to the earliest possible stimulation latency while maintaining a 50%-50% response rate (see Figure 4), suggests they were likely guessing at a high rate rather than relying strongly on movement latency as a cue. This behavioral insensitivity mirrored sub-*dd*'s neural decoding results, as out-of-sample classification accuracy for this subject exceeded chance levels but not classifier sensitivity.

To decode SoA from the brain, we used logistic lasso-PCR (Wager et al., 2011) to predict single-trial agency judgments from voxel activity. We quantified decoding performance with the leave-one-run-out cross-validated accuracy and with the cross-validated area under the receiver-operating characteristic (AUROC), a bias-free analog of predictive accuracy—in other words, a measure of the classifier's sensitivity. We applied the same decoder to (1) just the inverse kinematics selective voxels, called the "theory mask" as our hypotheses predicts strong decoding from these, (2) all voxels predicted by the control encoding model, called the "visuomotor mask" reflecting cortical areas involved in the control of the hand, and (3) all of the voxels in cortex, which serves as a rough upper bound on the potentially decodable information available in cortex. Model weights of all decoders, and their ROC curves, are visualized in Figure 5. SoA could be decoded from the theory mask with above-chance *accuracy* in all three subjects (accuracy_{01} = 0.561, accuracy_{dd} = 0.546, accuracy_{gg} = 0.599; permutation test: p_{01} = 0.0006, p_{dd} = 0.0266, p_{gg} = 0.0002; p_{all} = 6.7e-7). However, the theory mask predicted SoA with above chance *sensitivity* in only the two subjects whose agency judgments were themselves behaviorally sensitive to the latency of FES (AUROC_{01} = 0.600, AUROC_{dd} = 0.521, AUROC_{gg} = 0.638; permutation test: p_{01} = 0.0002, p_{dd} = 0.167, p_{gg}= 0.0002; p_{all} = 1.3e-6), and this result held even after controlling for stimulation latency (logistic regression w/ random effect for fMRI runs: p_{01} = 8.7e-6, p_{dd} = 0.3578, p_{gg}= 2.4e-8; p_{all} = 3.7e-7). These results allow us to conclude that neural activity in inverse kinematic coding areas in EVC predicts consciously perceived agency with above-chance accuracy in at least 44.5% of the sampled population (MAP = 100%, 95% HDI [44.5%, 100%]) and with above-chance AUROC in at least 21.0% of the population (MAP = 65.9%; 95% HDI: [18.8%, 95.4%]) with 95% posterior probability (Ince et al., 2021).

Decoding from the visuomotor mask significantly outperformed the theory mask in just one subject (AUROC_{01} = 0.696, AUROC_{dd} = 0.486, AUROC_{gg} = 0.603; one-tailed permutation test vs. theory: p_{01} = 0.0002, p_{dd} = 0.8738, p_{gg}= 0.8914; p_{all} = 0.0075), as did the whole-cortex decoder (AUROC_{01} = 0.707, AUROC_{dd} = 0.536, AUROC_{gg} = 0.647; one-tailed permutation test vs. theory: p_{01} = 0.0002, p_{dd} = 0.3037, p_{gg} = 0.3639; p_{all} = 0.0015). Cortex outperforms the visuomotor mask, interestingly, in the other two subjects (one-tailed permutation test vs. visuomotor: p_{01} = 0.2985, p_{dd} = 0.0238, p_{gg}= 0.0390; p_{all} = 0.0118).

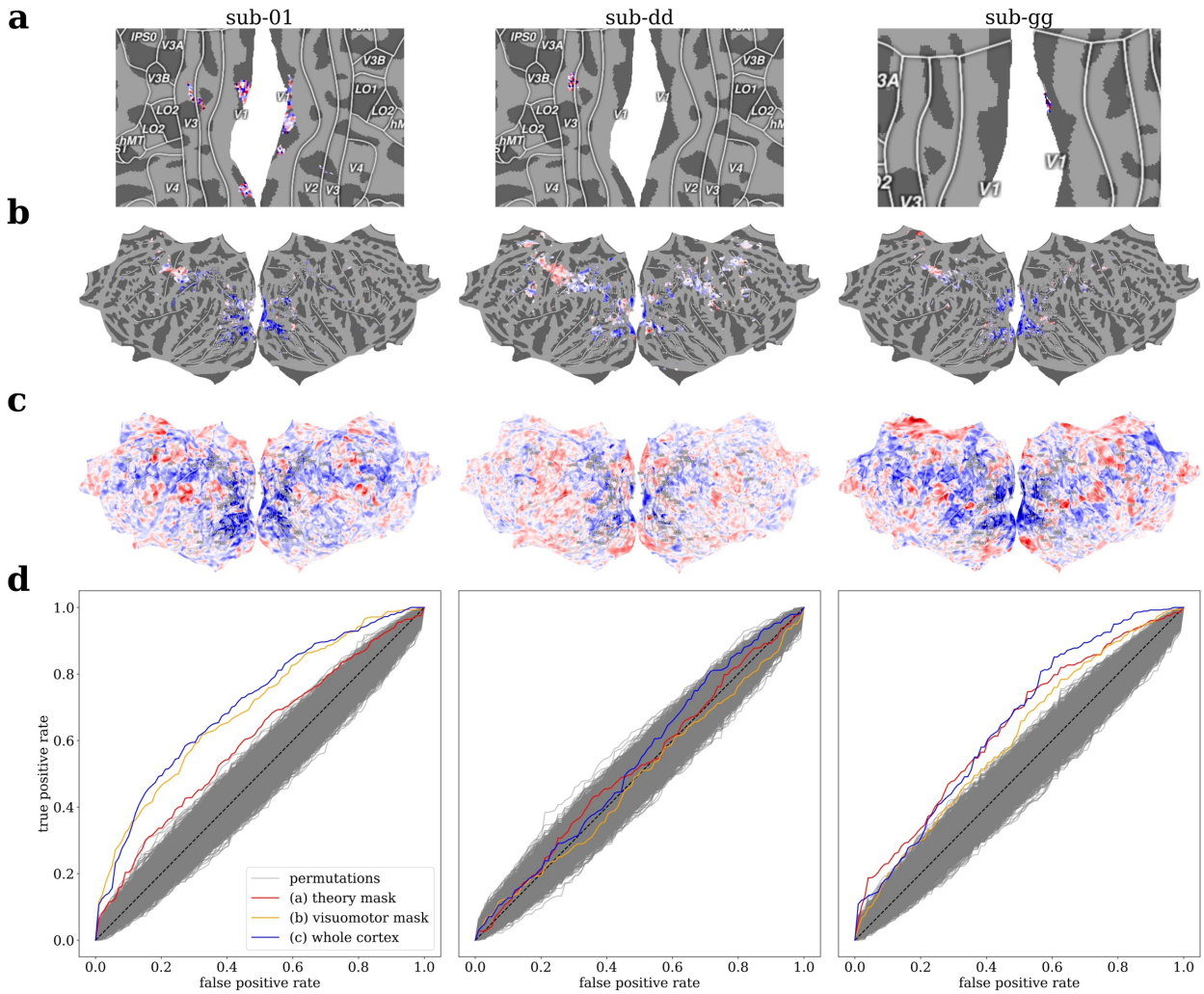

Figure 5: Decoding sense of agency (SoA) from brain activity during muscle stimulation. Linear model weights for classifying between stimulation-evoked muscle movements that subjects erroneously reported were self-caused from those they did not, scaled to be between -1 and 1, are shown for (a) the "theory mask" containing only voxels that were best predicted by inverse kinematic model representations during the motor task days earlier, (b) a "visuomotor mask" containing all voxels that were predicted above-chance by the control encoding model, representing all voxels putatively involved in hand visuomotor control, and (c) all of cortex. (d) Receiver-operator characteristic curves illustrating the cross-validated performance of the three decoding models, compared to a null distribution generated by shuffling the test-set labels. The theory-driven models exceed chance sensitivity in the same two subjects whose SoA judgments were behaviorally sensitive to the latency of muscle stimulation, and they exceed chance accuracy in all three subjects.

Discussion

The fact that our inverse kinematic encoding model successfully selected a set of voxels from which sense of agency (SoA) could be reliably decoded days later is consistent with our hypothesis that SoA is a conscious indicator that the brain can formulate a control policy for the current sensorimotor context. More conservatively, our results broadly support the idea that our internal representations for implementing motor control—not just related prediction errors (Haggard, 2017)—are involved in generating SoA.

However, the superior performance of visuomotor and whole-cortex decoders in sub-01 suggests that, at least in some subjects, our theory does not capture the full set of motor representations which may percolate into consciousness as SoA. This actually illustrates a strength of our analytic approach; rather than simply rejecting the null hypothesis and concluding our theory-based decoding is above chance, our design yields several useful diagnostic indicators to guide further research. The visuomotor mask consists of those voxels that are predictable in a data-driven manner in the motor task. Such voxels were not predicted by our theory as operationalized but *could* have been predicted by a different theory or different computational specification of control model. Similarly, it is unlikely sub-*uu*'s brain does not perform inverse kinematic computations but our specific model did not capture their idiosyncratic representations; though prevalence calculations suggest those subjects in which our model was successful were far from outliers (Ince et al., 2021). Investigating differences in the expression of inverse kinematic representations and their relationship with SoA on a subject-wise basis is uniquely enabled by our individuating analytic approach and dense sampling data, which we have made publicly available to support such follow-up work. Contrastingly, when the whole-cortex decoder outperforms the visuomotor decoder, it suggests there are voxels containing SoA-predictive information which could not, in principle, have been predicted from the motor task data—consistent with the contemporary view of SoA as arising from a combination of predictive, postdictive, and contextual cues (Synofzik et al., 2008). Future work, then, could explore alternative task paradigms for fitting encoding models, perhaps using a more ethologically valid motor task (e.g., object manipulation) or different neuroimaging modality.

That voxels best explained by our inverse kinematic model were exclusively found in early visual cortex (EVC) is surprising. In retrospect, there was reason to suspect this may be the case. We know encoding of movements is not restricted to canonically motor areas but ubiquitous throughout (especially sensory) cortex (Musall et al., 2019; Stringer et al., 2019). Hand movement type can be decoded from human EVC prior to movement initiation (Monaco et al., 2020), hand-selective visual regions appear to reflect potential grasp movements when viewing 3D tools (Knights et al., 2021), and reach direction can be decoded from EVC even in congenitally blind subjects (Bola et al., 2023).

While there is consensus that some inverse kinematic computation likely occurs in the cerebellum (Wolpert et al., 1998)—not the focus of the present study as cerebellum is usually argued not to influence awareness (Koch, 2018)—little neuroimaging work has studied cortical inverse models. Some researchers argue such computations are performed directly in motor cortex (Schweighofer et al., 1998), but experimental evidence suggests this occurs upstream of premotor cortex in visuomotor pathways (Ghasia et al., 2008; Xivry & Ethier, 2008). There is control-theoretic justification for inverse kinematic computation to occur near primary sensory areas; fast "suboptimal" control policies often benchmark better than theoretically optimal but slower policies in practice, as the benefits of more rapidly updating (imperfect) actions given feedback seem to outstrip that of step-by-step optimality (Howell et al., 2022). However, we suspect that further inverse kinematic computation may occur nearer to motor cortex but on a faster timescale, and fMRI seems too slow a neuroimaging modality to capture very fast signals. Indeed, a very recent rodent study took a similar approach to our own—correlating activations of a DNN trained in a biomechanical simulation—and found evidence of inverse kinematic computation in motor cortex,

where they had implanted electrodes, but they did not record activity from any sensory or other upstream cortical areas (Aldarondo et al., 2024).

Modeling cortical control of the musculature with such granularity has only recently become feasible. While motor control is a well-developed field with numerous empirically successful computational models, such models are usually specified at the level of macroscopic movements or net force output instead of that of musculoskeletal kinematics—and for good reason. Embodied control is too high-dimensional a problem to specify model parameters by hand, and biomechanical simulation has been too slow to learn parameters with reinforcement learning. Just recently, it became possible to port existing biomechanical models into robotic-grade physics simulators (Wang et al., 2022). Now, a growing library of simulation environments features not just healthy musculoskeletal systems but clinically informative cases incorporating sarcopenia, tendon transfer surgeries, and assistive exoskeletons (Caggiano et al., 2022). To capitalize on such computational advances, it is critical to model neural activity at the level of individual subjects rather than directing all resources to group studies (Naselaris et al., 2021). Even healthy brains show variation in functional organization (as in Figure 2), but the promise of neuroscience is to improve the lives of those whose neural responses may look *least* like the average brain—with neurological, motor, and musculosketal pathologies. The present work demonstrates that biomechanically detailed control models can be used to predict human brain activity, affording the opportunity to probe specific relationships between motor system computations, behavior, and subjective experience—all at the level of the individual.

Methods

1. Participants and Ethics Statement

Four participants (3 female, 1 male), between the ages of 24 and 28 years, participated in the study. All subjects gave their written, informed consent before participating, and all procedures were approved by the Social and Behavioral Sciences Institutional Review Board at the University of Chicago (IRB23-1323).

2. Experimental Design

2.1. Session 1: Motor Task

Upon arriving at the MRI facility, subjects put on an MRI-compatible Data Glove (5DT, Inc.) which was used to measure their joint angles throughout the subsequent recording. The Data Glove only outputs raw sensor values for each joint instead of joint angles directly, though these uncalibrated sensor values linearly vary with to the true joint angles. To calibrate glove, we recorded participants moving their gloved hand outside of the scanner with a Leap Motion Controller (Ultraleap), from which we estimated ground truth joint angles using the *RoSeMotion* software (Fonk et al., 2021). We used this data to estimate the linear mapping from the glove sensors to true joint angles (including the distal interphalangeal joints for which the glove does not have direct sensor coverage) on a per-subject basis.

Subjects were instructed to replicate hand gestures that were presented to them as pictures (Avotec Silent Vision SV-6011 LCD projection system), which switched every five seconds while in the MRI scanner. As stimuli, we used the same eight isometric and isotonic hand configurations used in the *Ninapro* database (Jarque-Bou et al., 2020). During the task, we exhausted every possible transition between the gestures at least twice during each of the ten, 10-minute runs. Overall, this session lasted 2.5 hours.

2.2. Session 2: Agency Task

Upon subjects' arrival, we applied two functional electrical stimulation (FES) electrodes to their forearm over the *flexor digitorum profundus* muscle, which were connected to a RehaStim 1 constant current stimulator (HASOMED) through a waveguide. Before the experiment, we ran a calibration procedure in which we raised the intensity of the FES stimulation until it could cause subjects' ring finger to press a button (Celeritas optical response pad) ten times in a row to ensure we could adequately move their muscles with FES. Each instance of stimulation consisted of 3 consecutive, 400 microsecond (200 pos, 200 neg) biphasic pulses.

The task procedure was the same as used in our prior research (Veillette et al., 2023), but with more trials spread across ten blocks. The first 10-minute run was a typical cue-response reaction time task, in which subjects were asked to press the response pad with their ring finger as quickly as possible after they see a cue to move. In the remaining 9 runs, subjects were instructed to still attempt to complete the reaction time task on their own, but if they were not fast enough, the muscle stimulator would move their finger to press the button for them. If subjects succeeded in moving before FES, it would trigger stimulation immediately so the muscle movement and FES were always temporally confusable. After each trial, subjects were asked to discriminate whether they or the muscle stimulator pressed the button. (If they were unsure, as it can be surprisingly difficult to discern, they were told to provide their best guess.)

We used subjects' response times from the first run to set a prior on the parameters of a logistic function describing the probability that they would report causing the button press, based on the observation from previous work that FES-caused movements can occur up to 40-80 ms prior to self-caused movements before subjects notice the have been preempted more than half the time (Kasahara et al., 2019). After each trial, we used their agency judgments to update this model, and we draw the next stimulation latency from the posterior distribution of threshold at which they would report agency or non-agency with equal probability. To account for nonstationary between runs, the uncertainty in the posterior is reset at the beginning of each run, though the posterior mean is retained.

As we know from our prior research (Veillette et al., 2023), this Bayesian adaptive procedure produces distributions of agentic and non-agentic trials with very similar distributions of stimulation latencies—so that participants' subjective experience of agency can be dissociated from the stimulation parameters. In practice, this 50%-50% threshold is sufficiently early that subjects are rarely able to respond before the stimulator (see Figure 4c-e), which can be verified by checking that measured "response" times are a linear function of FES latency (see Figure 4b), and those trials that do not follow the line can be removed following the criteria in our prior work (Veillette et al., 2023). In the present study, this yielded 877, 853, and 926 FES-caused trials across

the nine stimulation blocks for subjects 01, *dd*, and *gg*, respectively, that we used for our decoding analysis. Only FES-caused trials were used for decoding, as other, unobserved aspects of the muscle movement may differ between FES-caused and self-caused movements.

Additionally, we can rule out the possibility that self-reported agency judgments at this threshold latency are just random, as we have found that single trial judgements can be decoded (cross-validated across subjects) from EEG within the first 100 ms after the onset of muscle stimulation and remain decodable for at least another 400 ms, indicating that the agency judgments in this task usually have an origin in sensorimotor processing (Veillette et al., 2023). Of course, even if the average subjects respond nonrandomly, some subjects might still have very high guess rates and be effectively random. This outlier case can be identified if subjects' responses are insensitive to FES latency (see analysis).

An unanalyzed 10-minute eyes-open resting state scan was collected during this session after the fourth run, which was there merely to serve as a brief break for the subject. The experiment session lasted 2.5 hours in total.

3. Statistical Analysis

3.1. Voxelwise Encoding Models

To approximate an inverse kinematic model, we used the baseline model for *MyoSuite*'s *MyoHandPoseRandom-v0* simulation environment, which was trained using the natural policy gradient method to control a biomechanically realistic hand (Caggiano et al., 2022). This deep neural network (DNN) approximates an inverse kinematic model, as takes a current state as input and outputs a set of muscle activations that will move its biomechanically realistic hand closer to the target position.

At each timepoint during the motor task, we input the human participants' joint angles, velocities, and angular distance from the current target hand position—as measured by the motion tracking glove—into the DNN, and we extracted the full set of activations from the model's artificial "neurons." We used these activations, including the input layer, as features from which to predict brain activity in our voxelwise encoding model.

Voxelwise encoding models were fit using the *Himalaya* package (Dupré la Tour et al., 2022). Features from the DNN were filtered to the same rate as the MRI data, and then duplicated with four temporal delays (2, 4, 6, and 8 seconds) to account for the lag between neural activity and the hemodynamic response. A separate linear ridge regression was fit for each voxel, resulting in 652 weights (4 times 163 DNN units) for each voxel, were averaged across delays to produce a 163-dimensional vector describing each voxel's response properties. The regularization parameter for each ridge regression was chosen by grid search to maximize the leave-one-run-out cross-validated R^2 within the training set.

To account for the possibility that the task features alone drove prediction, rather than inverse kinematic features, a separate control encoding model was fit similarly but only had access to the input features of the neural network as predictors. Instead of linear ridge regression, this

encoding model used kernel ridge regression such that it could also learn nonlinear mappings between task features and voxel activity. This ensures that, when our DNN-based encoding model outperforms the control, it is because the inverse kinematic features are *particularly* good features for predicting brain activity (i.e., better than a data-driven nonlinear mapping)—not because any arbitrary nonlinear transformation of the input space improves performance.

Encoding models were fit to the first seven MRI runs, and then tested (cross-validated) on three hold-out runs which used a separate MR field map measurement such that fMRI preprocessing for these runs was totally independent. Models' out-of-sample R^2 were compared to chance using a block permutation test (5,000 permutations) in which continuous blocks of 10 TRs are kept together on each permutation so that the autocorrelation structure of the data is preserved in the null distribution (Huth et al., 2016; LeBel et al., 2023). In visualizations (depicted in Figures 2-3) and in the "visuomotor mask" used for decoding, we apply a false-discovery rate (FDR) correction and keep voxels where FDR < 0.05, as in other voxelwise encoding studies (Huth et al., 2016; LeBel et al., 2023; Tang et al., 2023). In-text, we report the lowest familywise error rate corrected p-value across the brain as a "global" p-value for each subject, corrected for multiple comparisons using an R^2-max procedure (Nichols & Holmes, 2002). When comparing the DNN-based and control models, we use a paired block permutation test, in which blocks of model predictions are shuffled between models rather than in time, and we control the familywise error rate using threshold-free cluster enhancement (Smith & Nichols, 2009).

To visualize DNN-based encoding models, we subdivided the inputs of the DNN into three features spaces, and we computed Shapley values using the *DeepLIFT* method with the *shap* package (Lundberg & Lee, 2017; Shrikumar et al., 2017). Shapley values describe how much each feature or feature space contributes to the predictions of a model or, roughly, how much the model's predictions (in our case, of a voxel) would change if features were not included. For visualization (depicted in Figure 2), we show each feature space's test-set Shapley values divided by the sum of the Shapley values for all feature spaces, so each voxel's color represents the relative contribution of each feature space for explaining that voxel.

3.2. Decoding Models

For decoding, voxel activity for each trial was estimated by computing the beta series for the FES stimulation events using the "least squares all" method (Rissman et al., 2004). We used logistic regression with lasso (L1) regularization to predict participants' single trial agency judgments (from the agency task) from the full-rank set of principle components of the voxel activity—called logistic lasso-PCR. Lasso-PCR is commonly used for whole-brain decoding models, as the principle components transformation usefully deals with the high spatial autocorrelation of fMRI measurements, and thus the method scales quite well with the number of voxels included as predictors—but the PCA transformation can be inverted to easily project model weights back into interpretable voxel space (Wager et al., 2011).

We fit decoding models using three nested feature sets: (1) our "theory mask," consisting of those voxels the inverse kinematic DNN model predicted better than the control, (2) a "visuomotor model" consisting of all the voxels predicted above-chance by the control model, and (3) all the voxels in cortex. Models performance was quantified as the leave-one-run-out cross-

validated area under the receiver operator characteristic curve (AUROC), which is interpreted similarly to accuracy (0.5 is chance, 1 is perfect, 0 is worst) but is not dependent on a threshold criterion and thus invariant to "how much" agency a subject must feel before claiming they caused a movement—and is usefully robust to class imbalances.

Models are compared to chance (i.e., at an AUROC of 0.5) using a one-tailed permutation test. Nested models are compared to each other, using a one-tailed paired permutation test. (A one-tailed test is used when comparing models with nested feature sets, since it is assumed that all the information captured by the smaller set of models is also present in the larger set, and if the larger set underperforms the only explanation is theoretically uninteresting overfitting with the larger feature set.)

To account for the possibility that, since agency judgments are sensitive to stimulation latency—that is, after all, how we experimentally manipulate agency—we might just be decoding agency, we ran a logistic regression (with random effects for each run) predicting agency judgments from (a) the stimulation latency and (b) the out-of-sample prediction of the "theory mask" model for each trial. To ensure robustness against violations of normality assumptions (e.g., as a biproduct of cross-validation), we fit this logistic regression using generalized estimating equations (GEE) instead of maximum likelihood (Liang & Zeger, 1986). Since this model statistically controls for the stimulation latency, we can interpret a nonzero coefficient assigned to the brain-based prediction as evidence that the decoder captures variation in sense of agency that is *not* explained by the stimulation latency. Additionally, the coefficient for the stimulation latency quantifies the subjects' sensitivity to the timing of their muscle movement, and thus serves as a manipulation check.

4. fMRI Acquisition and Preprocessing

MRI data were collected with a 3T Philips Achieva at the Magnetic Resonance Imaging Research Center at the University of Chicago. Functional scans were collected using gradient echo EPI with TR = 2.00 s, TE = 0.028 s, flip angle = 77 deg, in-plane acceleration (SENSE) factor = 2, voxel size 3.13x3.13x3.0 mm, matrix size = (64, 62) with 32, FOV = 200 mm. FOV covered cortex in its entirety in all subjects. Pepolar field maps were collected between every 2-4 functional scans to be used for susceptibility distortion correction. High-resolution (1x1x1 mm) anatomical scans were collected during Session 1 on the same 3T scanner with a T1-weighted MP-RAGE sequence.

Data was first preprocessed with fMRIPrep 23.2.0, which was used to perform susceptibility distortion correction, slice time correction, brain extraction, co-registration of functional and anatomical scans, computation of confound (e.g., motion) time series, cortical surface reconstruction, and projection of BOLD data onto the Freesurfer *fsaverage* template surface, in which subsequent analyses were performed (Esteban et al., 2019; Fischl, 2012). CompCor confounds generated by fMRIPrep were regressed out of the raw BOLD time series data (Behzadi et al., 2007) using the *nilearn* package, either prior to fitting voxelwise encoding models for the motor control task or as part of the estimation of beta series (see Methods 3.2) for the agency task.

5. Data and Code Availability

Anonymized raw data and preprocessed derivatives, including detailed data quality reports and the fitted encoding/decoding models, have been made publicly available on *OpenNeuro* (https://doi.org/10.18112/openneuro.ds005239.v1.0.1). The experiment code used for data collection for the hand tracking task (https://doi.org/10.5281/zenodo.12610625) and the muscle stimulation task (https://doi.org/10.5281/zenodo.12610710) are archived on *Zenodo,* as well as all analysis code (https://zenodo.org/doi/10.5281/zenodo.12610621).

References

Aldarondo, D., Merel, J., Marshall, J. D., Hasenclever, L., Klibaite, U., Gellis, A., Tassa, Y., Wayne, G., Botvinick, M., & Ölveczky, B. P. (2024). A virtual rodent predicts the structure of neural activity across behaviors. *Nature*, 1—3. https://doi.org/10.1038/s41586-024-07633-4

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

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

Bola, Ł., Vetter, P., Wenger, M., & Amedi, A. (2023). Decoding Reach Direction in Early "Visual" Cortex of Congenitally Blind Individuals. *Journal of Neuroscience*, *43*(46), 7868–7878. https://doi.org/10.1523/JNEUROSCI.0376-23.2023

Bullock, I. M., Borràs, J., & Dollar, A. M. (2012). Assessing assumptions in kinematic hand models: A review. *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 139–146. https://doi.org/10.1109/BioRob.2012.6290879

Caggiano, V., Wang, H., Durandau, G., Sartori, M., & Kumar, V. (2022). MyoSuite: A Contact-rich Simulation Suite for Musculoskeletal Motor Control. *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*, 492–507. https://proceedings.mlr.press/v168/caggiano22a.html

Cross, L., Cockburn, J., Yue, Y., & O'Doherty, J. P. (2021). Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments. *Neuron*, *109*(4), 724-738.e7. https://doi.org/10.1016/j.neuron.2020.11.021

Dupré la Tour, T., Eickenberg, M., Nunez-Elizalde, A. O., & Gallant, J. L. (2022). Feature-space selection with banded ridge regression. *NeuroImage*, *264*, 119728. https://doi.org/10.1016/j.neuroimage.2022.119728

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fernández, E. S., & Fernández, H. V. (2022). Neuro-Rights and Ethical Ecosystem: The Chilean Legislation Attempt. In P. López-Silva & L. Valera (Eds.), *Protecting the Mind: Challenges in Law, Neuroprotection, and Neurorights* (pp. 129–137). Springer International Publishing. https://doi.org/10.1007/978-3-030-94032-4_11

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Fonk, R., Schneeweiss, S., Simon, U., & Engelhardt, L. (2021). Hand Motion Capture from a 3D Leap Motion Controller for a Musculoskeletal Dynamic Simulation. *Sensors*, *21*(4), Article 4. https://doi.org/10.3390/s21041199

Frith, C. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine*, *17*(3), 631–648. https://doi.org/10.1017/S0033291700025873

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

Ghasia, F. F., Meng, H., & Angelaki, D. E. (2008). Neural Correlates of Forward and Inverse Models for Eye Movements: Evidence from Three-Dimensional Kinematics. *Journal of Neuroscience*, *28*(19), 5082–5087. https://doi.org/10.1523/JNEUROSCI.0513-08.2008

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

Heald, J. B., Lengyel, M., & Wolpert, D. M. (2021). Contextual inference underlies the learning of sensorimotor repertoires. *Nature*, *600*(7889), 489–493. https://doi.org/10.1038/s41586-021-04129-3

Howell, T., Gileadi, N., Tunyasuvunakool, S., Zakka, K., Erez, T., & Tassa, Y. (2022). *Predictive Sampling: Real-time Behaviour Synthesis with MuJoCo* (arXiv:2212.00541). arXiv. https://doi.org/10.48550/arXiv.2212.00541

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of population prevalence. *eLife*, *10*, e62461. https://doi.org/10.7554/eLife.62461

Jarque-Bou, N. J., Atzori, M., & Müller, H. (2020). A large calibrated database of hand movements and grasps kinematics. *Scientific Data*, *7*(1), 12. https://doi.org/10.1038/s41597-019-0349-2

Kakade, S. M. (2001). A Natural Policy Gradient. *Advances in Neural Information Processing Systems*, *14*. https://proceedings.neurips.cc/paper_files/paper/2001/hash/4b86abe48d358ecf194c56c69108433e-Abstract.html

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

Knights, E., Mansfield, C., Tonin, D., Saada, J., Smith, F. W., & Rossit, S. (2021). Hand-Selective Visual Regions Represent How to Grasp 3D Tools: Brain Decoding during Real Actions. *Journal of Neuroscience*, *41*(24), 5263–5273. https://doi.org/10.1523/JNEUROSCI.0083-21.2021

Koch, C. (2018). What Is Consciousness? *Nature*, *557*(7704), S8–S12. https://doi.org/10.1038/d41586-018-05097-x

Laurens, J. (2022). The statistical power of three monkeys. *BioRxiv*, 2022.05. 10.491373.

LeBel, A., Wagner, L., Jain, S., Adhikari-Desai, A., Gupta, B., Morgenthal, A., Tang, J., Xu, L., & Huth, A. G. (2023). A natural language fMRI dataset for voxelwise encoding models. *Scientific Data*, *10*(1), 555. https://doi.org/10.1038/s41597-023-02437-z

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

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' Acceptance of Real-Time Speech Exchange Indicates That We Use Auditory Feedback to Specify the

Meaning of What We Say. *Psychological Science*, *25*(6), 1198–1205. https://doi.org/10.1177/0956797614529797

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *30*. https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Metzger, S. L., Littlejohn, K. T., Silva, A. B., Moses, D. A., Seaton, M. P., Wang, R., Dougherty, M. E., Liu, J. R., Wu, P., Berger, M. A., Zhuravleva, I., Tu-Chan, A., Ganguly, K., Anumanchipalli, G. K., & Chang, E. F. (2023). A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, *620*(7976), 1037–1046. https://doi.org/10.1038/s41586-023-06443-4

Monaco, S., Malfatti, G., Culham, J. C., Cattaneo, L., & Turella, L. (2020). Decoding motor imagery and action planning in the early visual cortex: Overlapping but distinct neural mechanisms. *NeuroImage*, *218*, 116981. https://doi.org/10.1016/j.neuroimage.2020.116981

Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S., & Churchland, A. K. (2019). Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, *22*(10), 1677–1686. https://doi.org/10.1038/s41593-019-0502-4

Naselaris, T., Allen, E., & Kay, K. (2021). Extensive sampling for complete models of individual brains. *Current Opinion in Behavioral Sciences*, *40*, 45–51. https://doi.org/10.1016/j.cobeha.2020.12.008

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25. https://doi.org/10.1002/hbm.1058

Poldrack, R. A. (2017). Precision neuroscience: Dense sampling of individual brains. *Neuron*, *95*(4), 727–729.

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, *23*(2), 752–763. https://doi.org/10.1016/j.neuroimage.2004.06.035

Schweighofer, N., Arbib, M. A., & Kawato, M. (1998). Role of the cerebellum in reaching movements in humans. I. Distributed inverse dynamics control. *European Journal of Neuroscience*, *10*(1), 86–94. https://doi.org/10.1046/j.1460-9568.1998.00006.x

Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning Important Features Through Propagating Activation Differences. *Proceedings of the 34th International Conference on Machine Learning*, 3145–3153. https://proceedings.mlr.press/v70/shrikumar17a.html

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., & Harris, K. D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, *364*(6437), eaav7893. https://doi.org/10.1126/science.aav7893

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

Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*, *26*(5), 858–866. https://doi.org/10.1038/s41593-023-01304-9

Veillette, J. P., Lopes, P., & Nusbaum, H. C. (2023). Temporal Dynamics of Brain Activity Predicting Sense of Agency over Muscle Movements. *Journal of Neuroscience*, *43*(46), 7842–7852. https://doi.org/10.1523/JNEUROSCI.1116-23.2023

Wager, T. D., Atlas, L. Y., Leotti, L. A., & Rilling, J. K. (2011). Predicting Individual Differences in Placebo Analgesia: Contributions of Brain Activity during Anticipation and Pain Experience. *Journal of Neuroscience*, *31*(2), 439–452. https://doi.org/10.1523/JNEUROSCI.3420-10.2011

Wang, H., Caggiano, V., Durandau, G., Sartori, M., & Kumar, V. (2022). MyoSim: Fast and physiologically realistic MuJoCo models for musculoskeletal and exoskeletal studies. *2022 International Conference on Robotics and Automation (ICRA)*, 8104–8111. https://doi.org/10.1109/ICRA46639.2022.9811684

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, *3*(11), Article 11. https://doi.org/10.1038/81497

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7), 1317–1329. https://doi.org/10.1016/S0893-6080(98)00066-5

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347. https://doi.org/10.1016/S1364-6613(98)01221-2

Xivry, J.-J. O. de, & Ethier, V. (2008). Neural Correlates of Internal Models. *Journal of Neuroscience*, *28*(32), 7931–7932. https://doi.org/10.1523/JNEUROSCI.2426-08.2008