RESEARCH REPORT

European Journal of Neuroscience   FENS   WILEY

Bayesian multilevel hidden Markov models identify stable state dynamics in longitudinal recordings from macaque primary motor cortex

Sebastien Kirchherr[1,2] 🔘 | Sebastian Mildiner Moraga[3] 🔘 | Gino Coudé[1,2,4] |
Marco Bimbi[1,2] 🔘 | Pier F. Ferrari[1,2] 🔘 | Emmeke Aarts[3] 🔘 |
James J. Bonaiuto[1,2] 🔘

[1]Institut des Sciences Cognitives Marc Jeannerod, CNRS UMR 5229, Bron, France

[2]Université Claude Bernard Lyon 1, Université de Lyon, France

[3]Department of Methodology and Statistics, Universiteit Utrecht, Utrecht, Netherlands

[4]Inovarion, Paris, France

Correspondence
Sebastien Kirchherr, Institut des Sciences Cognitives Marc Jeannerod, CNRS UMR 5229, Bron, France.
Email: sebastien.kirchherr@isc.cnrs.fr

Abstract

Neural populations, rather than single neurons, may be the fundamental unit of cortical computation. Analysing chronically recorded neural population activity is challenging not only because of the high dimensionality of activity but also because of changes in the signal that may or may not be due to neural plasticity. Hidden Markov models (HMMs) are a promising technique for analysing such data in terms of discrete latent states, but previous approaches have not considered the statistical properties of neural spiking data, have not been adaptable to longitudinal data, or have not modelled condition-specific differences. We present a multilevel Bayesian HMM addresses these shortcomings by incorporating multivariate Poisson log-normal emission probability distributions, multilevel parameter estimation and trial-specific condition covariates. We applied this framework to multi-unit neural spiking data recorded using chronically implanted multi-electrode arrays from macaque primary motor cortex during a cued reaching, grasping and placing task. We show that, in line with previous work, the model identifies latent neural population states which are tightly linked to behavioural events, despite the model being trained without any information about event timing. The association between these states and corresponding behaviour is consistent across multiple days of recording. Notably, this consistency is not observed in the case of a single-level HMM, which fails to generalise across distinct recording sessions. The utility and stability of this approach is demonstrated using a previously learned task, but this multilevel Bayesian HMM framework would be especially suited for future studies of long-term plasticity in neural populations.

Sebastien Kirchherr and Sebastian Mildiner Moraga shared first authors

Emmeke Aarts and James J Bonaiuto shared last authors

1 | INTRODUCTION

The accelerating scale of recorded neurophysiological data has been accompanied by a shift from the view that individual neurons are the basic unit of computation, to one that explains computational processes in terms of neural population dynamics (Saxena & Cunningham, 2019). Electrophysiology studies routinely record data from ever larger neural populations using chronically implanted multi-electrode arrays. These datasets are challenging to analyse because of their dimensionality, but recent evidence suggests that neural population activity in low dimensional subspaces, or manifolds, may be more important in neural computation than the firing of particular single neurons (Barack & Krakauer, 2021; Gallego et al., 2017, Gallego, 2018; Humphries, 2021; Saxena & Cunningham, 2019). Chronically implanted electrode arrays present an additional analysis hurdle because the recorded signals can change dramatically over weeks and months of recording because of electrode drift or degradation (Barrese et al., 2016; Welle et al., 2020), or task-induced plasticity (Dayan & Cohen, 2011). The ability to understand long-term learning mechanisms requires such longitudinal recording, but the analysis of the resulting data at the individual neuron level is hindered by the difficulty in tracking single neurons over such long periods.

Hidden Markov models (HMMs; Rabiner, 1989; Zucchini et al., 2017) are a particularly promising approach for analysing the activity of large neural populations (e.g. Gat et al., 1997; Kemere et al., 2008; Mazurek et al., 2018; Ponce-Alvarez et al., 2012; Radons et al., 1994; Sadacca et al., 2016; Seidemann et al., 1996; van Kempen et al., 2021; Warwick 2011). HMMs are probabilistic models which infer unobservable (hidden) 'states' from a temporal sequence of observable data. These models can therefore be used to infer discretised lower dimensional state spaces that represent high dimensional neural activity. By examining the impact of experimental manipulations on these states, it becomes possible to gain insights into the underlying computations performed by the associated neural activity, which is often challenging to achieve solely through analysing firing rate data from large neural populations (Marcos et al., 2019). Given a suitable observational model governing the emission of observable data from unobservable states, HMMs do not require trial averaging or smoothing, making them suitable for trial-by-trial analyses.

Movements unfold in a specific spatiotemporal sequence, making it difficult to disentangle the underlying cognitive and motor mechanisms of goal-directed movements in which planning and feedforward and feedback-based control are fundamentally intertwined. The primary motor cortex is the largest source of corticospinal projections, making it uniquely situated to integrate these mechanisms and translate them to muscle coordination of the arm and hand. Because HMMs can identify discrete changes in continuous multivariate time series (Cunningham & Yu, 2014), they are a promising technique for decomposing dynamic M1 population activity into discrete states corresponding to the underlying cognitive and motor mechanisms of goal-directed movement. Previous HMM-based approaches to analysing neural spiking activity have shown that such models can identify similar patterns of neural population activity in premotor and primary motor cortex between execution and observation of the same actions (Mazurek et al., 2018), accurately predict changes in movement velocity in the primary motor cortex (Kadmon Harpaz et al., 2019), and segment baseline, plan, and perimovement epochs from dorsal premotor cortex activity (Kemere et al., 2008). However, these approaches typically use statistical models with assumptions that are unmet by neural spiking activity and, in general, have not considered non-stationary patterns of activity over long-term recordings. Efforts have been made to develop spike sorting algorithms to track single neurons over long periods of time (McMahon et al., 2014; Muthmann et al., 2021; Zhu et al., 2020) and techniques for identifying continuous latent variables from neural population activity over weeks or months (Gallego et al., 2020; Pandarinath et al., 2018). However, to our knowledge no such techniques exist for HMMs, which we have chosen because they are better suited for identifying discrete segments over longer trial durations.

In the current study, we present a novel Bayesian multilevel HMM (e.g. de Haan-Rietdijk et al., 2017; Raffa & Dubin, 2015; Rueda et al., 2013; Shirley et al., 2016; Zhang & Berhane, 2014) adapted to longitudinal spiking data recorded from multi-electrode arrays with a multivariate Poisson log-normal emission distribution. In our model, we assume that the pattern of neural activity detected over time by the electrodes depends on a

WILEY

sequence of hidden neural states that cannot be observed directly. Because the activity is measured as spike counts for multiple electrodes, we adopt a multivariate Poisson distribution conditional on the hidden state. The innovative aspect of our model is that we use a multilevel framework to accommodate the fact that different trials may have different patterns of activity. Each individual trial is thus allowed to have a specific sequence of hidden states, along with a specific set of transition probabilities between states and emission parameters (i.e. trial-specific Poisson means). These parameters are centred around group-level means, which reflect the overall patterns of neural activity across all trials.

We trained the model on spiking neural data recorded from the primary motor cortex of the macaque monkey during a reaching, grasping, and placing task. Consistent with previous applications of HMMs to motor cortical activity (Kadmon Harpaz et al., 2019; Kemere et al., 2008; Mazurek et al., 2018) state transitions were tightly linked to task behavioural events, and state statistics such as the number of activations and lifetime were closely related to the duration of intervals between events. We show that the states identified by the HMM represent fast transitions between distinct spatiotemporal patterns of neural population activity which are generated during specific phases of the reaching, grasping, and placing task, and likely correspond to different motor control processes. Finally, we demonstrate that the multilevel HMM successfully identifies stable relationships between neural states and behaviour across multiple days of recording. Conversely, a single-level HMM is unable to generalise to neural activity recorded on different days, thereby constraining its capacity to capture such relationships.

2 | MATERIALS AND METHODS

2.1 | Multilevel Bayesian hidden Markov model

HMMs identify latent states from observable data by assuming that each state is only dependent on its previous state (the Markov assumption), and that only one state actively generates observations (emissions) at any given time point (i.e. states are mutually exclusive). The fitted model parameters include transition probabilities, which define the likelihood of transitioning from any given state to any other state, and emission probabilities, which govern the generation of observable data within each state. Classic frequentist HMMs applied to neural data have used a univariate multinomial distribution for emission probabilities (Bollimunta et al., 2012; Mazurek et al., 2018; Diomedi et al., 2021), encoding which of

N neurons is spiking in each time bin, and choosing a random neuron if multiple neurons spike in the same bin. Such an approach does not capture the multivariate nature of neural population activity measured with high density recordings, nor the statistical distribution of neural spiking, and breaks down as larger neural populations are recorded from. Previous HMM frameworks which have used multivariate and/or Poisson emission probabilities (Radons, 1994; Siedemann, Kemere et al., 2008; Warwick, 2011; Ponce-Alvarez et al., 2012) have been single level, meaning that they do not model the hierarchical nature of the data (e.g. data points come from a particular trial which was recorded in a particular session). They are therefore unable to model day-to-day variability in longitudinal neural recordings. We therefore used a multilevel Bayesian hidden Markov model framework (Shirley et al., 2012) to mitigate these shortcomings, modelling trial-specific random effects in transition and emission probabilities, multivariate Poisson log-normal emission probability distributions, and trial-specific condition covariates.

The adoption of the multilevel (i.e. hierarchical or mixed) framework allows estimation of parameters at the single trial and group level. It therefore yields group-level or global parameter estimates representing the expected mean values for the corresponding parameter over the full sample of trials, and additionally allows parameters for individual trials to slightly deviate from these values. This has a number of advantages over previous non-hierarchical approaches. First, all the available information in the full data set is used for the estimation of each trial-specific parameter, meaning that parameter estimation for trials with fewer observations leverages information from trials with more observations (Hox et al., 2018; Schoot & Miočević, 2020). As a result, the hierarchical structure of the model helps to regularise the estimation of trial-specific parameters, making them more robust to outliers (i.e. trials with extreme parameter values are pulled towards the group-level averages; Gelman & Pardoe, 2006). Second, the electrophysiological data inherently have a nested structure: measurements that occur as part of a same trial are generally more alike than measurements obtained from different trials. Allowing individual parameters to deviate from group-level estimates avoids the statistical caveats of ignoring the nested structure of the recordings, such as the underestimation of standard errors and overconfidence in statistical significance tests (Aarts et al., 2014; Hox et al., 2018; Schoot & Miočević, 2020). Lastly, the multilevel framework assigns individual parameters to each trial, resulting in a unique sequence of states that corresponds to specific patterns of neural activity. This approach is advantageous because it tailors the inference of hidden states to the specifics of

each trial, and ignoring the differences between trials by using a single-level model may lead to lower state decoding accuracy, as shown by McClintock (2021). Additionally, the multilevel framework ensures between-trial state consistency, which eliminates the need for post-hoc state matching between trials that would otherwise be necessary if separate single-level models were fitted to data from different recording sessions or days.

Using a Bayesian hidden Markov model (HMM) framework has advantages over traditional approaches that relied on maximum likelihood estimation and expectation–maximisation methods. Firstly, with the inclusion of trial-specific random effects, it enables the estimation of trial-specific parameters in transition and emission probabilities, which would not be computationally possible for both components with other estimation frameworks (Altman, 2007; de Haan-Rietdijk et al., 2017). Secondly, the Bayesian method provides additional features at no extra computational cost, such as local decoding of the states (i.e. the most likely state at each point in time for a trial) and credibility intervals on all model parameters, which avoids the need for additional steps to estimate parameter standard errors. Lastly, while non-informative priors are commonly used, the framework can also use (weakly) informative priors, which allows updating of the model with new data as it becomes available.

2.1.1 | Model specification of a basic HMM with Poisson emissions

A basic single-level Bayesian HMM with a Poisson emission distribution can be used to model the variability in the activation patterns of neurons (e.g. Kemere et al., 2008; Seidemann et al., 1996; van Kempen et al., 2021). Let $O_1, O_2, ..., O_T$ be a time series of spike counts (neural activations) of T measurements (bins) in a single trial ($N = 1$), measured by a single electrode ($K = 1$). The neural activity recorded by the K electrodes is assumed to be a result of a sequence of $S_1, S_2, ..., S_T$ with M possible latent neural states such that $S_t \in \{1, ..., M\}$ which cannot be measured directly. These neural states represent transient groups of neurons in the neural cortex that work together during the orchestration of a motor task and are assumed to follow a first-order Markov process. This means that the probability of switching from one neural state to the next only depends on the previous state visited.

Three sets of parameters describe the basic single-level HMM: the initial distribution of the states π with the initial probabilities $\pi_i = P(S_1 = i)$ of each state at the first-time measurement of the trial; the transition distribution A containing the probabilities $a_{ij} =$ $P(S_t = j | S_{t-1} = i)$ of switching from state i to state j; and the emission distribution B consisting on the parameters b_{ki} defining $P(O_{kt} = q | S_t = i)$, the probabilities of observing q spike counts in electrode k, given the neural state i being visited.

In this study, π is a column vector of dimensions $(1 \times M)$ representing the initial probabilities of the states $S_{t=1} \in \{1, ..., M\}$ at the beginning of a trial and is calculated from the transition distribution A after assuming a stationary distribution. A is a transition probability matrix of dimensions $(M \times M)$ with the transition probabilities a_{ij} that are modelled with a multinomial logistic regression using the experimental conditions of the trial (i.e. left, centre or right) as covariates in the linear predictor and that are assumed to be homogeneous over time. The observed spike counts, $O_{k1}, O_{k2}, ..., O_{kT}$, are assumed to follow a Poisson distribution conditional on the sequence of neural states, $S_{k1}, S_{k2}, ..., S_{kT}$, such that $Poisson(b_{ki}) \equiv P(O_{kt} = q | S_t = i)$ with mean Poisson parameters b_{ki} representing the mean expected spike counts for electrode k, given each state i. The Poisson emission distribution B consists of a matrix of dimensions $(KM \times KM)$ with K submatrices of dimensions $(M \times M)$ on its diagonal, each with the ith diagonal element representing the Poisson emission mean b_{ki} for the neural state i and electrode k. We assume that the multiple sequences (e.g. observations from multi-electrode probes) are conditionally independent given the neural states, so no correlations are assumed between them after controlling for the neural states.

Notice that in the basic model, the same values of A and B are used for all the trials in the sample; as a result, inferring the most likely sequence of hidden states is performed with the Viterbi algorithm (Viterbi, 1967) based on the observations specific to each trial but using the same parameter values $\lambda = \{A, B\}$. Although the HMM can be estimated using direct likelihood maximisation, following a computationally efficient recursive implementation of the Expectation–Maximisation (EM) algorithm known as the Baum–Welch algorithm (Baum & Petrie, 1966), we perform an iterative Bayesian estimation based on Scott (2002) to keep the results consistent with the multilevel HMM (which cannot be easily estimated using traditional EM procedures). We refer the reader to the online supporting information for the full model specification.

2.1.2 | Model specification of the Bayesian multilevel HMM with Poisson emissions

The multilevel HMM model incorporates individual random effects in both transition and emission distributions,

extending the basic HMM model. This results in each trial having its own set of trial-specific parameters $\lambda_n = \{A_n, B_n\}$, denoted by sub-index n for each of the N trials, which are centred around group-level estimates $\lambda = \{A, B\}$. Now, $O_{nk1}, O_{nk2}, ..., O_{nkT}$ denote the number of spikes per bin recorded in the trial n of $1, ..., N$ trials, by electrode k of $1, ..., K$ electrodes, over T measurements. Thus, each trial counts with individualised transition probabilities a_{nij} and Poisson emission means b_{nki}. Same as for the single-level HMM, the transition probabilities a_{nij} are modelled with a multinomial logistic regression using the experimental conditions of the trial as covariates, and they are assumed homogeneous over the trial. The trial-specific transition intercepts α_{nij} in the multinomial logistic regression are allowed to deviate from the group-level intercepts $\bar{\alpha}_{ij}$ (intercepts for the group-level transition probabilities) following a multivariate normal distribution. A multivariate normal distribution is chosen as prior because it is a flexible and well-known distribution that has several desirable properties such as a single peak and probabilities evenly distributed on either side of the mean (Gelman et al., 2015). The means of the multivariate normal distribution represent the expected values for the trial-specific transitions given the information across the sample of trials (i.e. the group-level means), and the covariance, the expected degree of dispersion or uncertainty on the trial-specific transitions. The emission distribution of the spike counts O_{nkt} at time t and electrode k for trial n now also follows a Poisson distribution, although with trial-specific means b_{nki} that depend on the state i: $Poisson(b_{nki}) \equiv P(O_{nkt} = q | S_{nt} = i)$. The trial-specific Poisson means for each state and each electrode, b_{nki}, are lognormally distributed around the group-level emission log means $\bar{b}_{ki}$ (notice the absence of the subindex n). The lognormal distribution is adopted as prior to ensure that the Poisson means are positive and because it can accommodate the wide range of count values often observed in count data with its long tail (Aguero-Valverde, 2013; Congdon, 2019; Ma et al., 2008).

As mentioned in the previous section, incorporation of trial-specific parameters following a hierarchical approach grants additional estimation properties such as leveraging of information across trials and reliability to outliers. However, including individual deviations from the group level in both transition and emission parameters is challenging to estimate using traditional expectation–maximisation or maximum likelihood estimation (Altman, 2007; Scott, 2002). Because of this, an iterative Bayesian estimation framework is followed (for more details, see the supporting information). Finally, note that the state decoding (inferring the most likely sequence of hidden states for a trial) is performed using the Viterbi algorithm with trial-specific parameters

$\lambda_n = \{A_n, B_n\}$, as opposed to the decoding in the single-level HMM where the same values are used for all trials (i.e. $\lambda = \{A, B\}$).

2.2 | Reaching, grasping and placing task

Two adult female rhesus macaques (*Macaca mulatta*) trained on a reaching, grasping and placing task served as the subjects. The animal handling as well as surgical and experimental procedures complied with the European guideline (2010/63/UE) and authorised by the French Ministry for Higher Education and Research (project # 2016112713202878) in force on the care and use of laboratory animals, and were approved by the ethics committee CELYNE (comité d'éthique Lyonnais pour les neurosciences expérimentale, C2EA 42). After initial training, we performed a sterile surgery to implant six floating multielectrode arrays (FMA, Microprobes for Life Science, Gaithersburg, MD, USA) in the right (monkey 1) or left (monkey 2) cortical hemisphere. Each array comprised 32 platinum/iridium electrodes (impedance .5 MΩ at 1 kHz) with lengths ranging from 1 to 6 mm and with an inter-electrode spacing of 400 μm. One electrode array was implanted in the primary motor cortex (M1), two were implanted in the ventral premotor cortex (F5), one in the dorsal premotor cortex (F2), and two in the prefrontal cortex (45a and 46/12r), as estimated according to a previous magnetic resonance imaging scan. For the purposes of this study, we analysed data from the M1 array of each monkey.

Throughout the task, the monkey sat in front of a table containing a handle and a semicircular groove. A metallic cube, the target, was placed at 11 cm from the handle, in a slot within a metallic grasping platform which could be placed within the groove (Figure 1b). Contact with the handle, target object and the bottom of the groove was recorded by a circuit which detected changes in resistance. The task was programmed and controlled by EventIDE software (OkazoLab Ltd). Trials started when the monkey grasped the handle. After a variable delay period (500–1000 ms), an auditory go signal (900 Hz tone) instructed the monkey to release the handle, reach for the target object, grasp it, lift it and place it anywhere into the groove beside the grasping platform (see Videos S1 and S2 for different views of the same example trial). The monkey was then rewarded with a few drops of water. The data were recorded from 1 s before the handle grasp until the reward. The task was run in blocks of 10 trials per condition (with the grasping platform placed in the left, right, or centre of the groove). The left and right target positions were each 25° from the centre of the semicircular groove, relative to the handle.
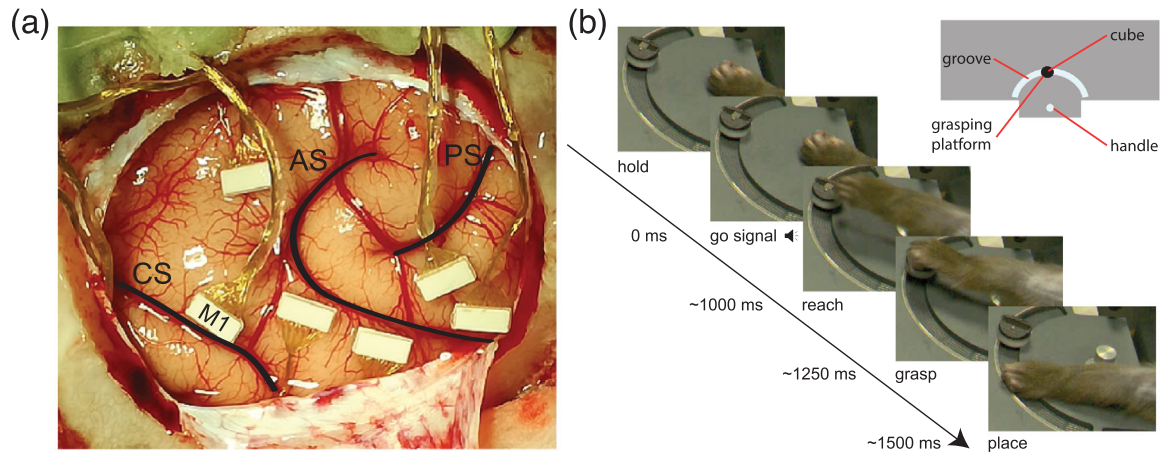
FIGURE 1 Experimental setup. (a) Localisation of the floating multi-electrode arrays (FMA) on the right cortical hemisphere of monkey 1. Six arrays were implanted in the frontal cortex, and here we analyse data from one which was located in the primary motor cortex (M1). Labels show the location of the arcuate sulcus (AS), central sulcus (CS) and principal sulcus (PS). (b) The monkey performed a cued reaching, grasping and placing task. In each trial, the monkey was required to grasp a handle until the go signal, then release the handle and reach and grasp a metallic cube with a precision grasp (using the index finger and thumb). Finally, after lifting the cube out of the slot, they had to place it anywhere into a groove in the table. The inset shows a schematic of the table from a top-down view with the handle, cube, grasping platform (containing the slot) and groove labelled.

The orientation of the slot in the grasping platform depended on the condition. In the centre condition, the slot was oriented at 0° with respect to the animal, and in the left and right conditions, the slot was oriented at 45° to left and right, respectively. The monkey completed 2–4 blocks of 10 trials per day by condition (monkey 1: left: $M = 17.1$, $SD = 5.4$, centre: $M = 16.7$, $SD = 5.1$, right: $M = 16.7$, $SD = 7.2$; monkey 2: left: $M = 16.0$, $SD = 4.1$, centre: $M = 13.1$, $SD = 4.4$, right: $M = 15.2$, $SD = 3.9$), for a total of 25–70 trials per day (monkey 1: $M = 50.5$, $SD = 13.3$; monkey 2: $M = 44.3$, $SD = 11.9$). Any trial in which the handle was released before the go signal or the grasping movement was not properly executed within 3 s was aborted and not rewarded. Each monkey used the hand contralateral to the implanted array to perform the task (monkey 1: left hand, monkey 2: right hand).

2.3 | Neural recordings

The wideband neural signal (bandpass filtered at .1 to 7500 kHz) was recorded at 30 kS/s, and amplified and digitised (16-bit; .192-μV resolution) with an Intan Tech-based (Intan Technologies, Los Angeles, CA, USA) open-source acquisition system (Open Ephys; Siegle et al. 2017). This system uses a 256-channel Intan RHD2000 series acquisition board and 32-channel head stages (RHD2132). Spike detection was performed offline using Trisdesclous (Garcia & Pouzat, 2015). The common reference was removed to reduce ambient noise. Spikes were then detected from each electrode using a threshold of

2 times the median absolute deviation (MAD) and ana-lysed as multi-unit activity (MUA) in 10 ms bins. All elec-trodes in which at least one well-isolated spike waveform was detected were selected for the following analyses. We thus used a sample of 21 electrodes out of 32 for monkey 1, and 25 out of 32 electrodes for monkey 2. Custom made detection panels were used to record the moments when the monkey's hand released the handle, the hand contacted the target object, and when the object was placed in the groove. An Omniplex 16-channel recording system (Plexon, Dallas, TX, USA) was used to simulta-neously record these behavioural events. Trials were dis-carded if the response time (time between the go signal and handle release) was less than 100 or greater than 1500 ms, the reach duration (time between handle release and object contact) was less than 100 or greater than 1000 ms, or the placing duration (time between object contact and placing the object in the groove) was less than 100 or greater than 1200 ms, leaving 19–68 trials per day for monkey 1 ($M = 43.9$, $SD = 15.46$, $N = 439$; left: $M = 14.8$, $SD = 5.74$; centre: $M = 14.4$, $SD = 5.15$; right: $M = 14.7$, $SD = 7.73$), and 23–49 per day for mon-key 2 ($M = 38.3$, $SD = 9.87$, $N = 383$; left: $M = 14.2$, $SD = 3.91$; centre: $M = 10.8$, $SD = 3.55$; right: $M = 13.3$, $SD = 3.37$).

2.4 | Model training

Model training was performed using R (R Core Team, 2021) and the developer version of the package

mHMMbayes (Aarts, 2019). In *mHMMbayes*, models are fitted using a hybrid Metropolis within Gibbs Markov Chain Monte Carlo (MCMC) algorithm, which expands on classic HMM implementations by using Bayesian estimation as outlined in Scott (2002). Fitting the Bayesian multilevel HMM requires specifying a set of starting values and a set of hyper-parameters for the hyper-priors of the group-level transition and emission parameters. The MCMC chains for the group-level parameters were initialised using random values centred on the final corresponding parameter estimates (i.e. group-level transition and emission means) obtained with a conventional expectation–maximisation HMM trained on spike count data aggregated over trials as starting values (depmixS4; Visser & Speekenbrink, 2010). In fitting the model, agnostic non-informative hyper-priors were specified for all group-level parameters. An ergodic topology was assumed for the transitions between states (i.e. transitions to and from all states were possible), so no restrictions were imposed over the transition parameters. The Bayesian multilevel HMM was trained with binned (10 ms bin width) spike counts from each electrode used with trial-specific intercepts and covariates encoding each trial's condition. The model was fit with 4000 iterations, with the first 2000 being discarded to dissipate the effect of starting conditions (burn-in). Convergence of all sample-level parameters was checked by verifying that the multivariate potential scale reduction factor, $\hat{R}$, was lower than a threshold of 1.05 (Brooks & Gelman, 1998) for two additional chains with randomised starting values. The sequence of most likely states given the neural data (i.e. state decoding), along with each state's forward probabilities, was determined for each trial with the Viterbi algorithm (Forney, 1973; Viterbi, 1967) based on trial-specific parameters.

To assess whether the model presented here offered a better fit to the spike count data than traditional methods, we also fitted a basic (i.e. single-level) Bayesian HMM following the same training procedure.

Model selection was performed by comparing relative model fit using AIC (Akaike, 1974) and BIC (Schwarz, 1978). The ability of the selected model to reproduce the original multielectrode data was determined using Bayesian posterior predictive checks (PPCs; for more details see, Gelman & Shalizi, 2013; McElreath, 2020). For the PPCs, the fitted model was used to simulate 500 new data sets, after which we assessed the extent to which a set of summary statistics (mean, standard deviation, maximum counts and proportion of zeros) on all electrodes (21 for monkey 1 and 25 for monkey 2) at the aggregated sample level in the simulated data recovered the values empirically observed in the electrophysiological data.

As previously mentioned, one of the advantages of our multilevel HMM is that trial-specific parameters are estimated and used on the Viterbi algorithm for state decoding. However, this poses a challenge when attempting to implement a cross-validation approach, because no trial-specific parameters are immediately available for out-of-sample trials. State decoding would have had to be performed using group-level estimates, jeopardising the main advantage of using a multilevel model. As a result, we made the decision not to use cross-validation to assess state decoding accuracy, and instead perform a small Monte Carlo simulation to obtain an indication of the decoding precision of the models where the ground truth (i.e. the actual sequence of states) was known. The multilevel HMMs trained on the empirical data of the two monkeys were used to simulate 100 new data sets for each monkey with the same number of trials and observations per trial as the respective empirical data. The simulated data sets, for which the true sequence of states was known, were then used to fit a multilevel HMM and a basic HMM, and the precision of state decoding using the Viterbi algorithm (with trial-specific parameters in the multilevel HMM) was evaluated in terms of accuracy, balanced accuracy, F1 score and Cohen's kappa.

2.5 | Neural and behavioural data analysis

To compare state onset and offset times to behavioural event timings, we identified episodes lasting at least 100 ms in each trial where a particular state was identified as the most likely using the Viterbi algorithm. If a state was activated multiple times in a trial, we used the first activation. The partial Spearman's correlation coefficient was used to relate state onset and offset times to each behavioural event time (movement onset, object contact and placing), accounting for correlations between other event timings and the onset or offset of all other states. Statistical significance was tested using permutation tests (10,000 permutations) in which the event timings, but not the state onset and offset times, were shuffled in each iteration. The stability of these correlations over time was evaluated using permutation tests (10,000 permutations) in which the recording session of each trial was shuffled.

Response time (the time between the go and movement onset events), reach duration (time between the movement onset and object contact events) and placing duration (time between the object contact and placing events) were computed for each trial and compared between conditions for each monkey using linear mixed models with condition (left, centre, right) as a fixed effect

and day-specific offsets as a random effect. These kinematic variables were also compared between monkeys using the same models with the addition of subject and its interaction with condition as fixed effects. The number of state activations per trial was given by the number of episodes where a state was identified as the most likely for at least 100 ms. To compute state lifetime, the first activation was used if a state was activated multiple times in a trial, as above. State activations and lifetime were compared across states and conditions using generalised linear mixed models (with Poisson distribution and log link function for activations, and normal distribution and identity link function for lifetime) with condition, state, and their interaction as fixed effects, and trial nested within day offsets as random effects. The relationship between response time, reach duration, and placing duration and the number of activations and lifetime of each state was evaluated using partial Spearman correlation coefficients, accounting for all other behavioural durations and state activations or lifetimes.

Multi-unit firing rate activity around state transitions was compared by aligning firing rates to the onset and offset of the maximum duration activation of each state around a time window lasting from 100 ms prior to the onset to half of the mean state lifetime after onset, and from half the mean state lifetime prior to offset to 100 ms after offset. The firing rate from each electrode was then baseline corrected using the mean firing rate in the 100 ms prior to the state onset. For each electrode, we compared the mean, baseline-corrected firing rate in the 100 ms prior to state onset to that in the first half of the state lifetime using linear mixed models with state, time period (baseline or first half of lifetime), and their interactions as fixed effects, and trial nested within day offsets as random effects. In the same way, we compared the mean, baseline-corrected firing rate in the first half of the state lifetime to that in the second half.

Temporal and spatial shuffling was performed by permuting the data, either by time point or by electrode, 100 times. The resulting permuted data were run through the HMM to obtain forward probabilities and sequences of the most likely state. For each iteration, the partial Spearman coefficient between state onsets and offsets and the behavioural events (accounting for correlations between other event timings and the onset or offset of all other states, as above) were evaluated, and the resulting coefficients were used as the null distribution for comparison with the corresponding unshuffled coefficients.

All mixed models were run in R (R Core Team, 2021) using *lme4* (v1.1.26; Bates et al., 2018). Factor significance was determined using type II Wald X^2 tests using the *car* package (v3.0.10; Fox et al., 2019), and pairwise Tukey-corrected follow-up tests were performed using estimated marginal means using the *emmeans* package (v1.5.3; Lenth et al., 2020) with Kenward–Roger approximated degrees of freedom. Code for all analyses is freely available (https://github.com/danclab/motor_mea_mbhmm).

3 | RESULTS

Multiple multilevel Bayesian HMMs with three to eight states were fit to single-trial multi-unit spiking activity (Figure 2a) recorded from the primary motor cortex of two macaque monkeys during a reaching, grasping and placing task. Model comparison (AIC and BIC) was used to select the optimal model, yielding a model with six states for monkey 1 and five states for monkey 2 (Figure S1). All group-level parameters in the optimal models met the convergence criterion.

3.1 | Simulations

We started by evaluating the goodness of fit of the optimal models selected for the two monkeys with posterior predictive checks. The results of the posterior predictive checks (PPCs) show that the models were able to recover the overall patterns of electrical activity aggregated over trials for the 21 and 25 electrodes of the first (Figure S2) and second monkeys (Figure S3), respectively, which indicates a good fit to the empirical data. In addition, for some of the electrodes, the PPCs reveal a small overestimation of the mean aggregated counts, coupled with a small underestimation of the proportion of bins with zero spike counts (e.g. monkey 1 electrodes 6, 10 and 21; Figure S2). These minor deviations from good fit appear to be more pronounced for the data of the second monkey than for the first monkey.

Next, we explored the expected state decoding accuracy with a Monte Carlo simulation based on the modelling results for the optimal models on the two monkeys. The mean percent balanced accuracy averaged over trials and simulation repetitions of 91.6% (CI95 [91.3–92.0%]) for monkey 1 and 87.5% (CI95 [86.3–88.7%]) for monkey 2 (see Table S1 for extended performance metrics). The mean balanced accuracy varied across states, ranging from 73.0% for state 2 to 96.6% for state 1 in monkey 1, and from 60.8% for state 2 to 98.6% for state 3 in monkey 2 (Table S2). Given the good fit to the data (Figures S2 and S3), these results indicate that a high level of decoding accuracy can be expected for the empirical data.
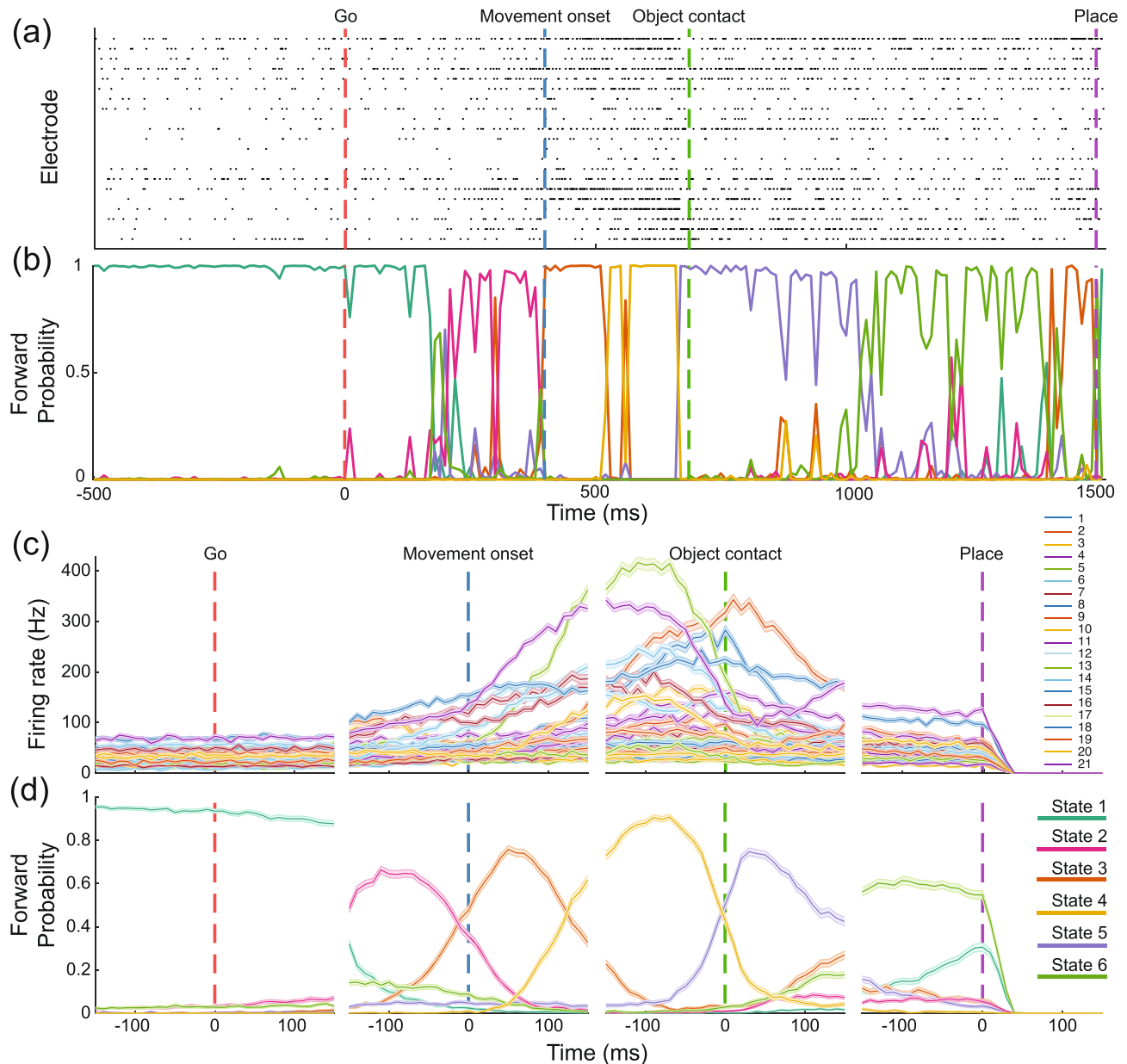
FIGURE 2 The HMM identifies states that correspond to different phases of the action. (a) Raster plot of a single example trial representing the spiking activity from 21 electrodes from the M1 array of monkey 1. (b) Forward probabilities for each state given by the model for the same trial shown in Figure 2a. (c) Mean multi-unit firing rate of 21 electrodes from the M1 array of monkey 1 over 10 days (439 trials) aligned to behavioural events (go: go signal tone, hand movement onset: beginning of the reaching movement, object contact: grasping of the object, place: placing the object). Solid lines represent the mean firing rate of each electrode and the shaded areas indicate the standard error. (d) Mean forward probabilities (solid lines, shaded areas represent standard error) generated by the model for each state, averaged over the same trials and aligned to the same behavioural events as in Figure 2c.

3.2 | Distinct M1 population activity states occur during different phases of the reaching, grasping and placing task

For each state, the model generates a forward probability, representing the probability at each point in time that the neural population is in that state given the MUA (Figure 2b). States were labelled in order of their trial-averaged order of occurrence. Trials had different durations due to variability in response time and movement kinematics, so in order to determine the task phase specificity of each state, we aligned both multi-unit firing rates and each state's forward probability with each behavioural event (go signal, hand movement onset, object contact and placing), and averaged over trials. The event-aligned multi-unit mean firing rate exhibited slow

ramping dynamics prior to the movement onset, and peaking just before object contact (monkey 1: Figure 2c; monkey 2: Figure S4A). In contrast, the mean forward probabilities of each HMM state indicate that discrete state transitions tended to occur around the same time as the movement onset, object contact and object placing behavioural events (Figure 2d). At the start of each trial, state 1 predominated, but sometime following the go signal, there was a transition to state 2, followed by a transition to state 3 after the hand movement onset. During the initial reach, there was a transition to state 4, and then to state 5 after the hand contacted the object. Finally, during the final reaching movement to place the object in the groove, there was a transition to state 6. This sequence of state transitions was similar for the five states identified from monkey 2 (Figure S4B). Importantly, behavioural event timings were not provided to the HMM during training, it was simply fit to the multi-unit spiking activity recorded during each trial. This suggests

that the identified states represent actual states of the neural population corresponding to different processes involved in motor control, rather than trivially reflecting the temporal structure of the task.

Although it appears that state transitions occur in specific relation to behavioural events, this could have been a result of trial averaging. We therefore examined the temporal relationship between state transitions and behavioural events at the single trial level. In each trial, increases and decreases in state forward probabilities closely matched the timing of the movement onset, object contact and placing events (monkey 1: Figure 3; monkey 2: Figure S4C). In each trial, there was a transition from state 1 to state 2 occurring before the hand movement onset, from states 2 to 3 closely locked to the hand movement onset, from 3 to 4 between the hand movement onset and object contact, from 4 to 5 locked to object contact, and from 5 to 6 between object contact and the placing event. We next used the Viterbi algorithm
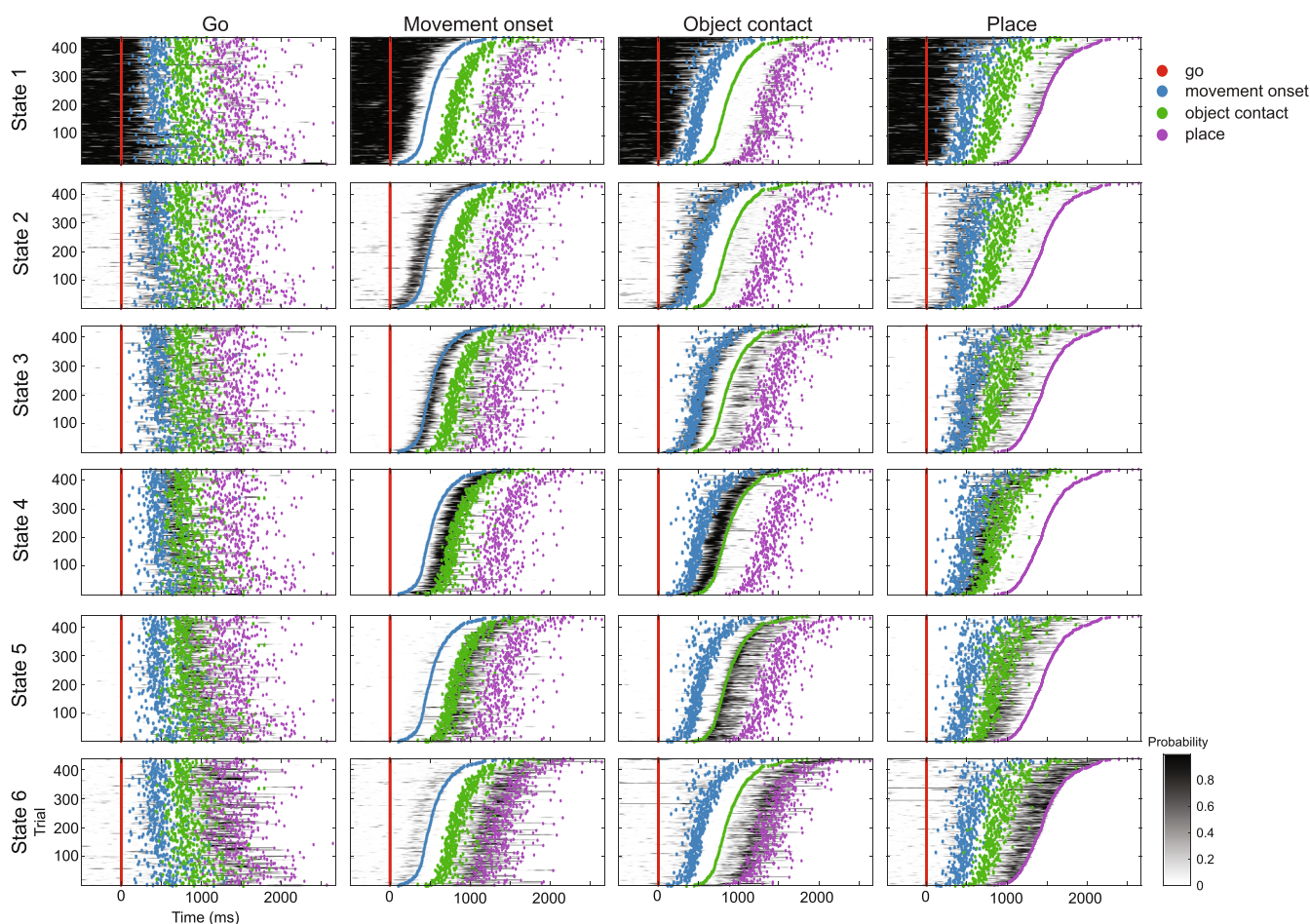

FIGURE 3 States are tightly linked to behavioural events at the single trial level. Each row shows the forward probabilities from one of the model states for each trial, over all days of recording, for monkey 1. Each column shows the trial forward probabilities sorted by one of the behavioural events (from left to right: go signal, hand movement onset, object contact and placing). When trials are sorted by the appropriate event, the alignment between state forward probabilities and event timing can be clearly seen over all trials.

-WILEY

(Forney [1973](#); Viterbi, [1967](#)) to estimate the most likely state at each point in time, for every trial, from the forward probabilities. We used the sequence of most likely states to determine the onset and offset times of each state in every trial. We then ran a permutation test on the partial Spearman's correlation coefficient (accounting for correlations between event timings and the onsets and offsets of other states) between the onset and offset time of each state and the timing of each behavioural event. This revealed that for monkey 1, the onset and offset of state 2 were most correlated with the movement onset time (Table [S3](#)). Only the onset of state 3, the offset of state 4 and the onset of state 5 were correlated with the object contact event. The offset of state 5 and the onset of state 6 were the transitions most correlated with the placing event. For monkey 2, the onset and offset of state 2 were correlated with movement onset time, the onset of state 3 and the onset and offset of state 4 were correlated with object contact, and the onset and offset of state 4 and the onset of state 5 were correlated with the placing event (Table [S3](#)). We then evaluated these correlations over the duration of the recordings, in groups of 2 days to ensure a large enough number of trials for each correlation. In both monkeys, these correlations were sustained over 10 days of recording (all $p < .02$; Figure [4](#)). A permutation test shuffling the trials by recording day confirmed that there were no consistent changes in these relationships over sessions (all $p > .10$; except for monkey 1, days 1–2: $p < .014$; monkey 2, days 5–6 and 9–10: $p < .044$). The onset and offset of various states were therefore tightly linked to the movement onset, object contact and placing events at the single trial level, with earlier occurring states tending to be related to movement onset, states in the middle of the trial related to object contact,

and states occurring later related to the placing event. Moreover, this linkage between states and behavioural events was stable over the entire period of recording.

3.3 | M1 population state statistics are related to behavioural state durations

We next compared response time (the time between the go signal and the hand movement onset events), reach duration (the time between the hand movement onset and object contact events), placing duration (the time between the object contact and placing events) and state statistics between task conditions. States were characterised by their number of activations per trial and their lifetime (the time that state remains active, taken for each trial as the maximum duration over all activations in that trial).

There was no difference in response time between conditions (monkey 1: $X^2(2) = .66$, $p = .717$; monkey 2: $X^2(2) = 4.37$, $p = .113$), but conditions varied in terms of reach duration (monkey 1: $X^2(2) = 78.68$, $p < .001$; monkey 2: $X^2(2) = 9.09$, $p = .011$; Figure [S5](#)). For monkey 1, reach durations to the left target location ($M = 408$, $SE = 8.40$ ms) were longer than to the right ($M = 343$, $SE = 8.58$ ms, $t[430] = 5.83$, $p < .001$) and centre ($M = 312$, $SE = 8.48$ ms, $t[432] = -8.67$, $p < .001$), and longer to the right compared to the centre target location ($t[436] = -2.79$, $p = .015$). For monkey 2, reach durations to the left target ($M = 230$, $SE = 13.2$ ms) were shorter than to the right ($M = 264$, $SE = 13.3$, $t[372] = -3.01$, $p = .008$). Placing duration was only different for monkey 1 ($X^2(2) = 30.18$, $p < .001$) and longer in the left ($M = 641$, $SE = 15.7$ ms) compared to the right
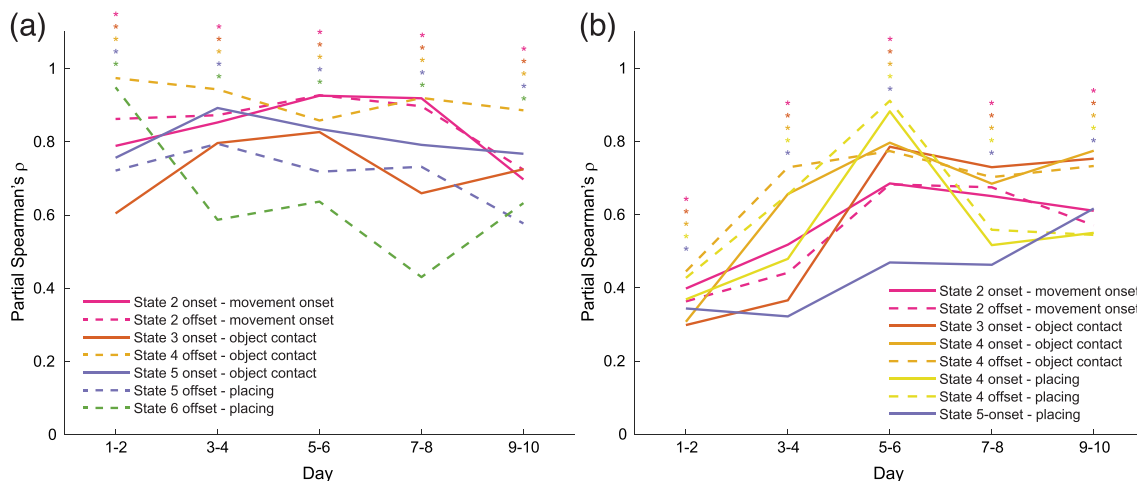

FIGURE 4 The link between states and behavioural events is consistent over days. (a) Partial correlation coefficients between state transitions and behavioural events over the duration of recording from monkey 1, in groups of 2 days. Asterisks indicate statistically significant correlations. (b) As in Figure 4a for monkey 2.

($M = 582$, $SE = 15.8$ ms, $t[436] = 3.41$, $p = .002$) and centre ($M = 548$, $SE = 15.8$ ms, $t[430] = -5.43$, $p < .001$) conditions, but there was no difference between placing duration for the centre and right conditions ([433] $= -1.96$, $p = .123$). Monkey 1 made more rapid responses and reaches than monkey 2 across conditions (response time: $X^2(1) = 59.91$, $p < .001$; reach duration: $X^2(1) = 279.18$, $p < .001$), but made slower placing movements than monkey 2 for the left target ($t(810) = 4.13$, $p < .001$) and faster placing movement for the centre target ($t[814] = -2.38$, $p = .018$).

The number of state activations per trial and their lifetime varied by state and condition in monkey 1 (activations: $X^2(10) = 95.94$, $p < .001$; Figure 5; lifetime: $X^2(10) = 66.00$, $p < .001$; Figure 6). State 1 had a shorter lifetime in the centre ($M = 533.5$, $SE = 11.4$ ms) compared to the left ($M = 635.5$, $SE = 11.2$ ms, $t[2616] = -6.39$, $p < .001$) and right ($M = 626.9$, $SE = 11.2$ ms, $t[2616] = -5.84$, $p < .001$) conditions. State 3 was activated more times the left condition ($M = .67$, $SE = .06$) than in the centre ($M = .36$, $SE = .07$, $z = -3.37$, $p = .002$) and right ($M = .39$, $SE = .07$, $z = 3.12$, $p = .005$) conditions. State
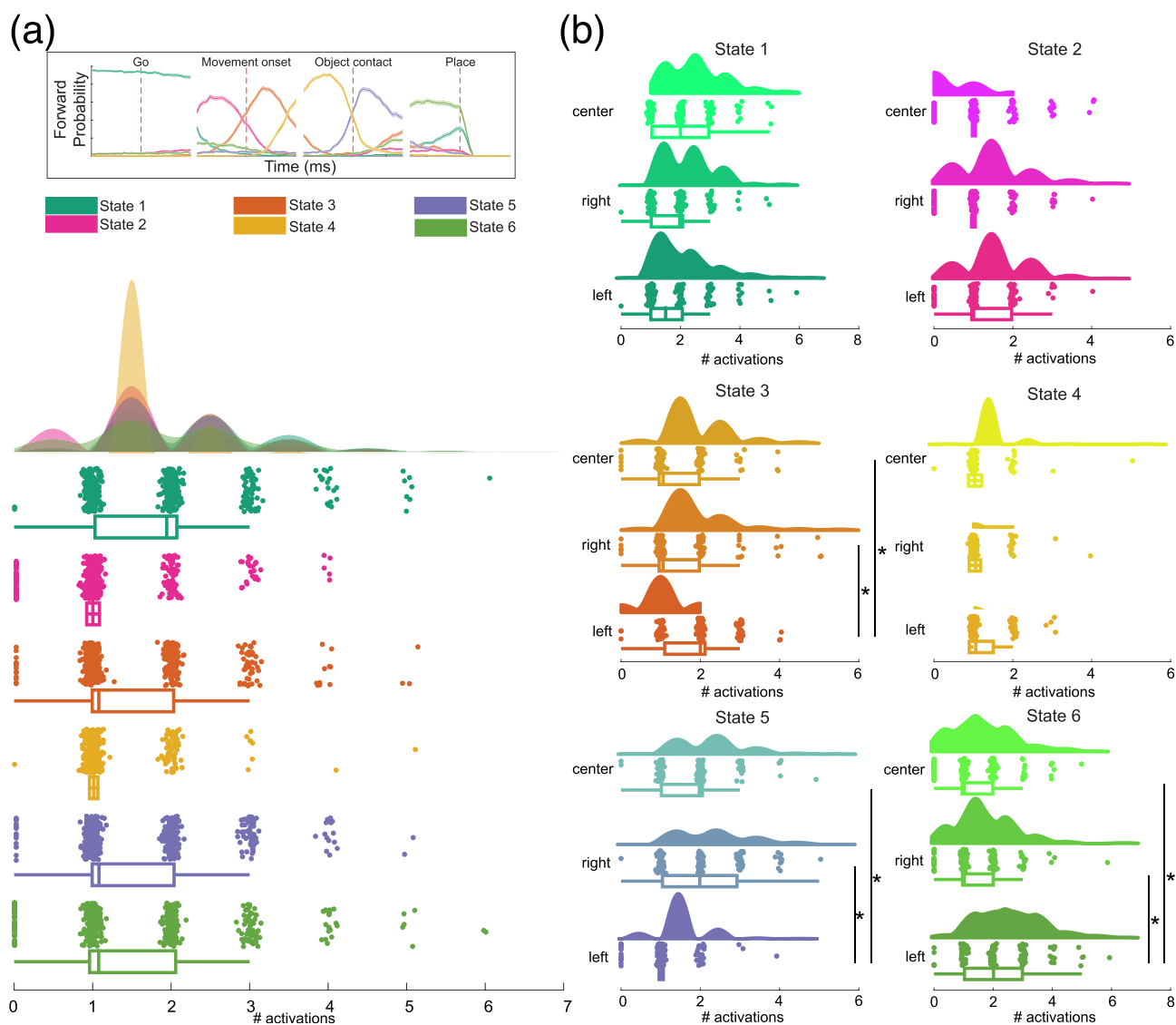

FIGURE 5 Number of state activations per trial compared between conditions. (a) The number of activations per trial for each state identified by the model. Shaded areas indicate the distribution density, scatter plots show the number of activations for individual trials (ordered by trial in the y dimension, and jittered in the x dimension), and box plots show descriptive statistics: median (in the box), interquartile interval (the box) and outliers (above whiskers). The inset shows the event-aligned mean state forward probabilities from Figure 2. (b) Number of activations per trial for each state compared by conditions: centre, right, left. Asterisks indicate significant pairwise comparisons.
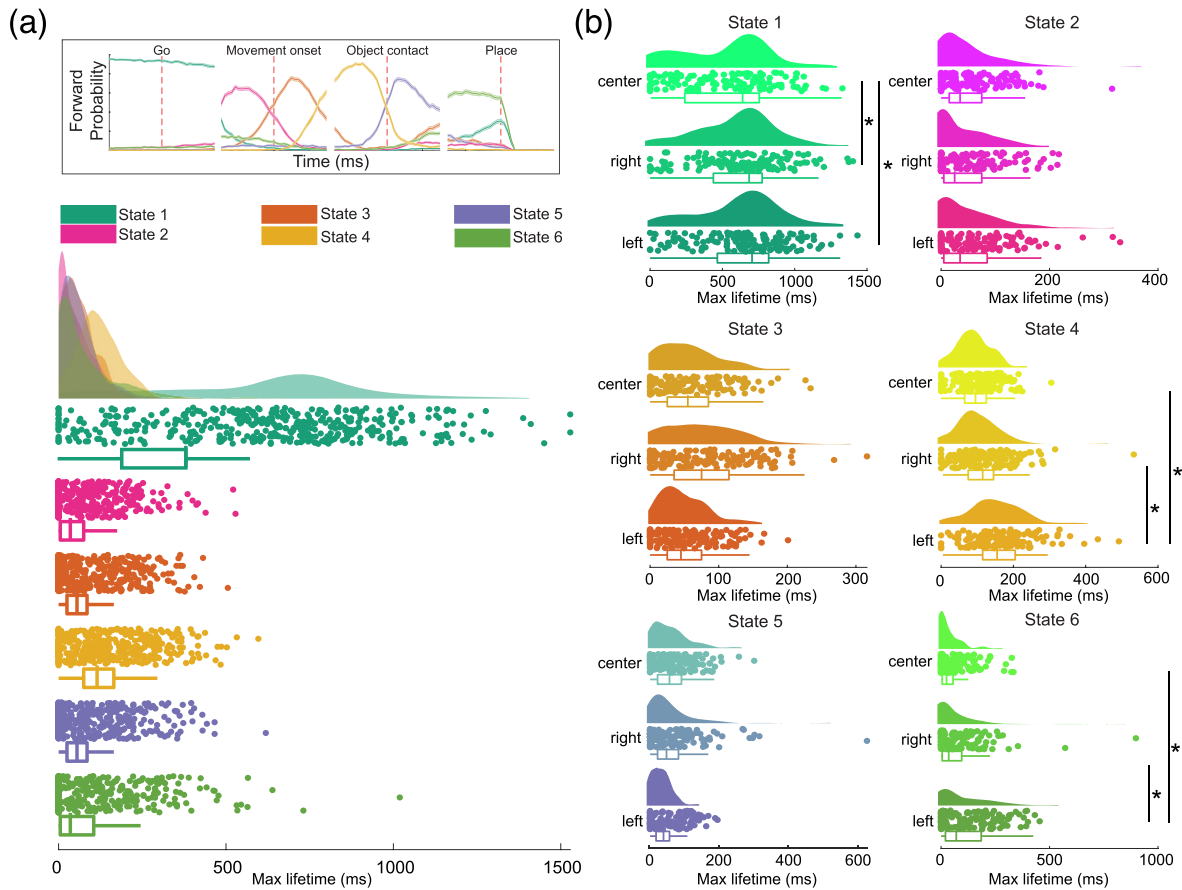
FIGURE 6 Maximum state lifetime compared between conditions. (a) The maximum lifetime (duration of the longest activation per trial) for each state identified by the model. Shaded areas indicate the distribution density, scatter plots show the maximum state lifetime for individual trials (ordered by trial in the y dimension, and jittered in the x dimension), and box plots show descriptive statistics: median (in the box), interquartile interval (the box) and outliers (above whiskers). The inset shows the event-aligned mean state forward probabilities from Figure 2. (b) Maximum lifetime for each state compared by conditions: centre, right, left. Asterisks indicate significant pairwise comparisons.

4 had longer duration in the left condition ($M = 160.2$, $SE = 11.2$ ms) than in the centre ($M = 97.7$, $SE = 11.4$ ms, $t[2616] = -3.92$, $p < .001$) and right ($M = 114.4$, $SE = 11.2$, $t[2616] = 2.89$, $p = .011$) conditions. In contrast, state 5 was activated fewer times in the left condition ($M = .11$, $SE = .08$) than the centre ($M = .61$, $SE = .06$, $z = 5.10$, $p < .001$) and right ($M = .68$, $SE = .06$, $z = -5.81$, $p < .001$) conditions. State 6 was also activated more times and had longer lifetime in the left condition (activations: $M = .77$, SE $= .06$; lifetime: $M = 114.8$, SE $= 11.2$ ms) than the centre (activations: $M = .26$, SE $= .07$, z $= -5.59$, $p < .001$, lifetime: $M = 43.1$, SE $= 11.4$ ms, $t[2616] = -4.50$, $p < .001$) and right (activations: $M = .30$, SE $= .07$, z $= 5.23$, $p < .001$; lifetime: $M = 67.7$, SE $= 11.2$ ms, $t[2616] = 2.97$, $p = .008$) conditions. The number of activations per trial and the lifetime of state 2 did not vary by condition. In monkey 2, there was only a main effect of state on the number of state activations per trial and their lifetime;

however, there was almost no difference between conditions in response time, reach duration and placing duration for this monkey.

Finally, we looked at relationships between state statistics and response time, reach duration and placing duration. For monkey 1, response time was correlated with the number of activations and lifetime of state 1 (Table S4). Reach duration was correlated with the number of activations of states 3, 4, and 6. Placing duration was correlated with the number of activations of states 1 and 4, as well as the number of activations and lifetime of state 6. For monkey 2, response time was correlated with the number of activations and lifetime of state 1, and the number of activations of state 2. Reach duration was correlated with the number of activations of states 2, 3, and 5, and the number of activations and lifetime of state 4. Placing duration was correlated with the number of activations of state 2, the lifetime of state 4, and the number of activations and lifetimes of states

3 and 5. The number of activations and lifetime of identified states was therefore related to response time, reach duration and placing duration in a state-specific way, with earlier occurring states related to response time, states occurring in the middle of the trial to reach duration, and states occurring later to placing duration.

3.4 | States represent distinct spatiotemporal patterns of neural population activity

Having determined that the timing, number of activations, and lifetime of the HMM-identified states are tightly coupled to the behavioural phases of reaching, grasping, and placing an object, we next examined neural dynamics associated with state transitions and occurring within states. We aligned multi-unit firing rates to the onset and offset times of each state in every trial, and baseline-corrected them using activity from the 100 ms prior to the state onset (monkey 1: Figure 7; monkey 2: Figure S6). State transitions tended to be associated by rapid changes in firing rates followed by, for each state, a different pattern of activity across electrodes. For each electrode, we compared the effect of state on the multi-unit firing rate in the 100 ms prior to state onset to that in the first half of the state duration, and the multi-unit firing rate in the last half of the state duration with that in the 100 ms following the state offset. For monkey 1, all electrodes had a significant interaction between state and time period, and follow-up comparisons showed that

states differed in the number of electrodes in which multi-unit firing changed with the onset (state 1: $N = 0$, state 2: $N = 14$, state 3: $N = 18$, state 4: $N = 19$, state 5: $N = 19$, state 6: $N = 16$) and offset (state 1: $N = 16$, state 2: $N = 20$, state 3: $N = 15$, state 4: $N = 19$, state 5: $N = 14$, state 6: $N = 18$) of the state. The same was true for monkey 2 for all electrodes for both state onset (state 1: $N = 0$, state 2: $N = 18$, state 3: $N = 24$, state 4: $N = 15$, state 5: $N = 20$) and offset (state 1: $N = 20$, state 2: $N = 19$, state 3: $N = 21$, state 4: $N = 21$, state 5: $N = 12$).

Although the model states represent a particular distribution of multi-unit firing rate across electrodes, these firing rates are not stationary and thus may change during the lifetime of a state. We therefore examined electrode firing rates within each state in order to determine if they systematically varied over the state's lifetime. With this aim, we compared the effect of state on firing rate in the first half of the state duration to that in the second half. A significant interaction between state and time period was found in 19 out the 21 electrodes for monkey 1, with the fewest number of electrodes exhibiting within-state dynamic activity in state 1 and the greatest number in state 4 (state 1: $N = 0$, state 2: $N = 5$, state 3: $N = 9$, state 4: $N = 19$, state 5: $N = 9$, state 6: $N = 6$). For monkey 2, there was a significant state by time period interaction in 24 out of the 25 electrodes, with the least within-state dynamics in state 1 and the most in state 5 (state 1: $N = 5$, state 2: $N = 15$, state 3: $N = 14$, state 4: $N = 16$, state 5: $N = 14$). In both monkeys, state 1 therefore represents a static period of low activity in all electrodes, but each other state represents a transition to a
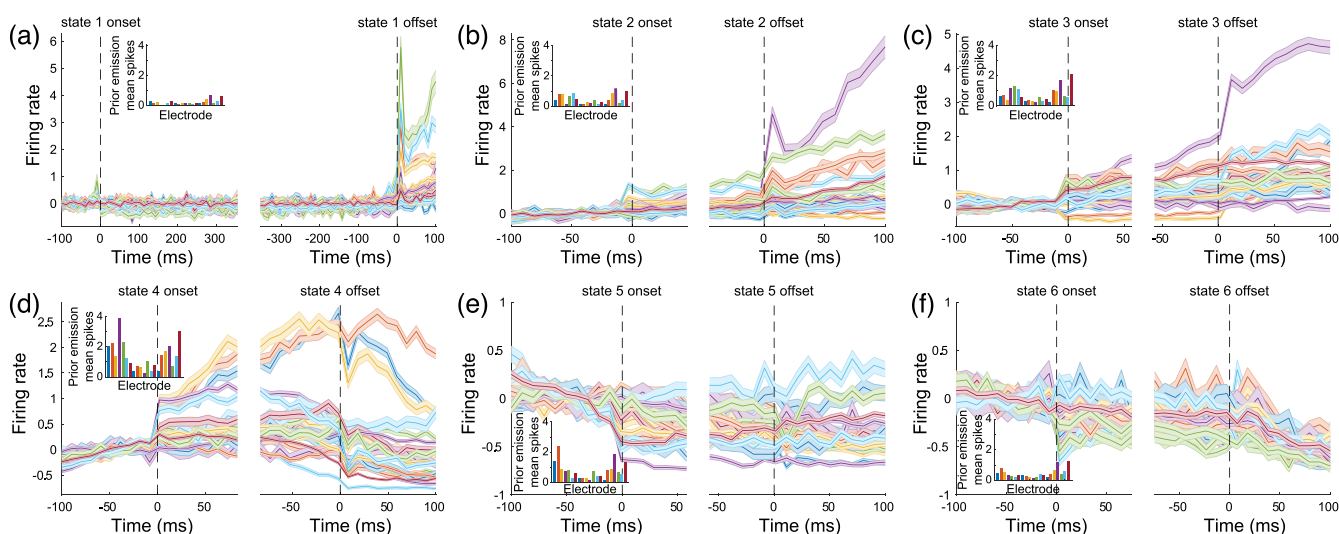

FIGURE 7 States represent different patterns of neural activity. (a–f) Each panel shows mean, baseline-corrected firing rates from each electrode from monkey 1 aligned to the onset (left plot) and offset (right plot) of states 1–6. Dashed vertical lines indicate the onset and offset time of the state. Insets depict, for each state, the group-level (prior) emission mean spike counts for each electrode (without baseline correction). State transitions are associated with rapid changes in state-specific patterns of neural activity.

specific combination of sustained and dynamic activity across electrodes.

We then evaluated the spatiotemporal specificity of these activity patterns by shuffling the data, either temporally (i.e. shuffling time points; Figure S7) or spatially (i.e. shuffling electrodes; Figure S8) and using the HMM to obtain forward probabilities and sequences of the most likely state. For monkey 1, both temporal and spatial shuffling abolished the relationships between the movement onset time and the onset (temporal shuffled $\rho = .14$; compared to unshuffled: $p = .01$; spatial shuffled $\rho = .08$; compared to unshuffled: $p = .01$) and offset (temporal shuffled $\rho = .09$; compared to unshuffled: $p = .01$; spatial shuffled $\rho = .04$; compared to unshuffled: $p = .01$) of state 2. The correlations between the object contact event and the onset of state 3 (temporal shuffled $\rho = -.02$, compared to unshuffled: $p = .02$; spatial shuffled $\rho = .10$, compared to unshuffled: $p = .02$) and the onset of state 5 (temporal shuffled $\rho = .10$, compared to unshuffled: $p = .01$; spatial shuffled $\rho = .13$, compared to unshuffled: $p = .01$) were also destroyed by temporal and spatial shuffling. Only spatial shuffling had an effect on the correlation between the placing event and the onset of state 6 (spatial shuffled $\rho = -.08$, compared to unshuffled: $p = .02$), and only temporal shuffling disrupted the relationship between object contact and the offset of state 4 ($\rho = -.02$, $p = .01$). For monkey 2, both temporal and spatial shuffling abolished the relationships between the movement onset time and the onset (temporal shuffled $\rho = -.05$; compared to unshuffled: $p = .01$; spatial shuffled $\rho = .06$; compared to unshuffled: $p = .01$) and offset (temporal shuffled $\rho = -.05$; compared to unshuffled: $p = .01$; spatial shuffled $\rho = 004$; compared to unshuffled: $p = .01$) of state 2. The correlations between the object contact event and the onset of state 3 were not destroyed by either temporal or spatial shuffling, but those between object contact and the onset (temporal shuffled $\rho = -.03$, compared to unshuffled: $p = .01$; spatial shuffled $\rho = .10$; compared to unshuffled: $p = .02$) and offset (temporal shuffled $\rho = -.02$, compared to unshuffled: $p = .01$; spatial shuffled $\rho = .06$, compared to unshuffled: $p = .01$) of state 4 were affected by both forms of shuffling. Both temporal and spatial shuffling had an effect on the correlation between the placing event and the offset of state 4 (temporal shuffled $\rho = -.03$, compared to unshuffled: $p = .01$; spatial shuffled $\rho = .04$, compared to unshuffled: $p = .01$), but neither had an effect on that between the placing event and the onset of state 5. The states identified by the HMM therefore represent rapid transitions between distinct spatiotemporal patterns of neural population activity which are generated during specific phases of the reach, grasp and place task.

3.5 | The multilevel HMM outperforms a basic HMM

To investigate the potential advantages of multilevel HMMs over single-level HMMs, we conducted a comparative analysis by fitting single-level HMMs to the same data used to train the multilevel model, and evaluating their decoding accuracy for simulations. The results of the Monte Carlo simulation indicate that the balanced accuracy of state decoding was significantly higher for the multilevel HMM than the single-level HMM for both monkey 1 and monkey 2. In monkey 1, the percent mean balanced state decoding accuracy was 8.11 (CI95[6.43–12.2]) points higher for the multilevel HMM compared to the single-level HMM over the simulation repetitions, which was statistically significant (paired t-test, $t[99] = 40.27$, $p < .001$). Similarly, in monkey 2, the percent mean balanced state decoding accuracy was 7.07 (CI95 [5.60–8.51]) points higher for the multilevel HMM compared to the single-level HMM, which was also statistically significant (paired t-test, $t[99] = 99.30$, $p < .001$). These findings were consistent over the four evaluation metrics assessed (see Table S1), which indicates that the multilevel HMM is a more effective model for state decoding compared to the single-level HMM in both monkeys.

Subsequently, we examined the relationship between state transitions and behavioural events, and between state lifetime and number of activations and behavioural state durations for both monkeys using the single-level HMM. However, the low number of trials in which states were active precluded the computation of partial correlations between state onsets and offsets and behavioural events, and between state lifetimes and behavioural state durations for both monkeys, specifically for state 5 from monkey 1, which was only activated in 17.54% of the trials, and state 4 from monkey 2, which was only activated in 26.37% of the trials.

Moreover, we found that the number of activations of states from the single-level HMM were no longer correlated with the behavioural state durations, except for state 1 with reaction time ($\rho = .18$, $p < .001$) and placing duration ($\rho = .10$, $p = .035$), state 4 and placing duration ($\rho = .15$, $p = .001$), and state 6 and placing duration ($\rho = .21$, $p < .001$). No correlations between the number of activations of any state and reaching duration were significant. For monkey 2, only the correlations between the number of activations of state 1 and response time ($\rho = .36$, $p < .001$), state 2 and reaching duration ($\rho = .24$, $p < .001$), state 2 and placing duration ($\rho = .31$, $p < .001$), and state 3 and reaching duration ($\rho = .12$, $p = .023$) were significant.

These results suggest that the inability of the single-level HMM to account for trial-to-trial variability led to states, particularly state 5 from monkey 1 and state 4 from monkey 2, that only represented patterns of neural population activity in certain trials or days of recording. Consequently, the single-level HMM could not capture the relationship between neural activity and behaviour across all trials.

4 | DISCUSSION

In this paper, we present a new multilevel Bayesian HMM, designed for analysing longitudinal neural recordings from chronic multi-electrode arrays through the use of trial-specific random effects in transition and emission probabilities, multivariate Poisson log-normal emission probability distributions, and trial-specific condition covariates. We trained instantiations of the model on long-term data recorded from the primary motor cortex of the macaque monkey during a task involving reaching, grasping and placing a target object. We showed that although the model was not trained with any information about the task structure, the timing of discrete state transitions was tightly linked to behavioural events (i.e. the timing of the movement onset, object contact and placing events), and state statistics were closely related to the duration of intervals between events (i.e. response time, reach duration and placing duration). The states identified by the HMM represented fast transitions between distinct spatiotemporal patterns of neural population activity which are generated during specific phases of the reach, grasp and place task, likely corresponding to different motor control processes.

At the single trial level, various states were coupled to different phases of the task. In both monkeys, state 1 represented a static period of low activity across all electrodes, and the go signal event did not elicit any state transition. However, sometime after the go signal, there was a transition to state 2, whose onset and offset correlated with movement onset time. State 2 therefore likely represents movement preparation or initiation mechanisms. States 3–5 in monkey 1 and states 3 and 4 in monkey 2 were correlated with the object contact, the reach duration, and the placing duration and therefore may be differentially involved in the acceleration and deceleration phases of reaching (Kadmon Harpaz et al., 2019), or the coordination of the wrist orientation. Model comparison resulted in a model with six states for monkey 1, and five states for monkey 2. Analysis of the kinematic data revealed that monkey 2 had faster response times and made faster reaches than monkey 1, suggesting that monkey 2 was making more ballistic reaching movements

with less reliance on feedback motor control. Monkey 2 also exhibited poorer performance in both the posterior predictive checks and the decoding accuracy when compared to monkey 1. One explanation for this difference could be attributed to monkey 2's more rapid reaching movements. However, it is also plausible that the smaller number of trials for monkey 2 ($N = 383$), as opposed to monkey 1's larger number of trials ($N = 439$), could have contributed to the observed contrast. The difference in the number of states could therefore be because state 4 in monkey 1 corresponds to feedback control mechanisms used during the deceleration phase (Kadmon Harpaz et al., 2019), but these mechanisms are less implicated in monkey 2's reaching movements. States 5 (monkey 2) and 6 (monkey 1) are activated before, and aligned to, the placing event, and therefore likely reflects control of a reaching movement in the opposite direction from the initial reach-to-grasp (Kadmon Harpaz et al., 2019). The model was trained solely on the multi-unit spike data from each trial without any event timing information, and yet state transitions and lifetimes were tightly linked to kinematic events and movement durations, suggesting that each state represents distinct motor control mechanisms.

This paper presents three modelling enhancements that improve the effectiveness of the basic hidden Markov model in various ways. Firstly, the trial-specific random effects in transition and emission probabilities enable the model to account for differences in the underlying dynamics of the system between trials. This captures individual variability in neural activity that cannot be captured by a single-level model, resulting in significant decoding accuracy improvement over a single-level model as measured in the Monte Carlo simulation for the two monkeys. Secondly, the use of multivariate Poisson log-normal emission probability distributions allows for flexible modelling of emission probabilities, taking into account the correlation between different neural signals and skewness in neural activity data. Lastly, the inclusion of trial-specific covariates on the transition distribution facilitates the exploration of the effect of experimental conditions on the neural dynamics of the system across trials. This enhances the interpretability of the model and enables statistical assessment of the influence of experimental conditions (i.e. whether the null value zero is excluded from the 95% credibility intervals for regression coefficients of the experimental conditions).

There are a few limitations to this framework which include the basic assumptions of HMMs, longitudinal generalisability for predictive decoding, task conditions, zero-inflation and the use of multi-unit, rather than single-unit, activity. All HMMs make the assumption that only one state can be active at any given time, and that

state transition probabilities only depend on the current state rather than the recent history of states (the Markov assumption). These may not be reasonable assumptions for neural population activity, and there have been recent efforts to develop a similar framework without, or with relaxed versions of, these assumptions (Gohil et al., 2022). A promising future direction might be to extend the benefits of multilevel Bayesian HMMs to such frameworks, for example, by implementing an explicit-duration hidden semi-Markov model that relaxes the Markov assumption, decoupling the state duration from the transition probabilities by explicitly assigning distributions to the duration of the states (Yu, 2010). One advantage of the multilevel approach is the ability to model between-trial differences in activity that might undermine attempts to identify the population state. However, as would also be the case for single level models, new trials would have to use group-level mean model parameters. One possibility for future work might be to take advantage of the trial-specific parameters in the multilevel model and extrapolate them for decoding future activity. The task used involves only three reach directions and one grasp type. This was a necessary restriction for the current study to establish and test the framework, but future applications should use a wider range of reach and grasp movements along with kinematic tracking in order to identify the specific motor control mechanisms represented by each state. Model checking with PPCs revealed a minor underestimation of the proportion of zeros in the spike count data, consistent with a zero-inflated generative process (i.e. a generative process assuming a 'background' component that generates zero spikes ['structural zeros'] and an 'active' component that generates a number of zero or more spikes). Future versions of the model could be extended with a zero-inflated Poisson emission distribution to model the excess of zeros beyond what a common random-effect Poisson distribution can accommodate (Aguero-Valverde, 2013). Finally, we trained the model on multi-unit spiking data, because of the inherent difficulty in tracking single units over long-term recordings. However, this is also a strength of this approach, as multi-unit activity has been shown to be sufficient to capture neural population dynamics (Trautmann et al., 2019), and therefore circumvents the need for longitudinal spike sorting.

Multilevel Bayesian HMMs provide a framework for analysing neural population activity within relatively long trials and over extended periods of time. However, while these models are able to identify discrete segments of activity, they do not directly provide insight into the neural dynamics occurring within each segment. A promising avenue for future research is to combine multilevel HMMs with PCA-based dimensionality reduction

techniques (Duncker & Sahani, 2021; Gallego et al., 2017). Such techniques can extract continuous latent variables from neural population activity but are typically applied to very short-time windows (200–700 ms; Gallego, 2018; Mazor & Laurent, 2005; Russo et al., 2020). These latent variable dynamics have been shown to be stable over long periods of recording by aligning, day-to-day, the low-dimensional manifolds that they are embedded in (Gallego et al., 2020). These techniques could thus be used to probe the dynamic neural activity within the states identified by a multilevel HMM during longer trial periods. It may also be possible to take advantage of the group- and trial-level HMM parameters to guide manifold alignment.

5 | CONCLUSION

The multilevel Bayesian HMM presented here has several unique advantages for analysing longitudinal neural population dynamics from multi-electrode arrays compared to current approaches. The emission probabilities are multivariate Poisson distributions, capturing the statistical distribution of neuronal spikes. The model is trained trial-by-trial, without the need for epochs of the same length, making it applicable to motor tasks which typically have variable trial duration because of movement variability. The inclusion of covariates allows the model to be fit on data from multiple experimental conditions and therefore simplifies comparison between them. In this study, each monkey was pre-trained on the reaching, grasping and placing task, but because the multilevel Bayesian framework easily handles longitudinal data, it is especially well suited for studies of longer-term neural population plasticity in the context of motor learning.

AUTHOR CONTRIBUTIONS
Sébastien Kirchherr: Data curation; formal analysis; investigation; visualization; writing—original draft; writing—review and editing. **Sebastian Mildiner Moraga:** Conceptualization; formal analysis; investigation; methodology; software; validation; visualization; writing—original draft; writing—review and editing. **Gino Coudé:** Conceptualization; data curation; methodology; resources; supervision. **Marco Bimbi:** Methodology; software. **Pier F Francesco:** Funding acquisition; project administration; resources; supervision; writing—review and editing. **Emmeke Aarts:** Conceptualization; formal analysis; funding acquisition; methodology; software; supervision; validation; writing—original draft; writing—review and editing. **James J Bonaiuto:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration;

CONFLICT OF INTEREST STATEMENT
The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT
The datasets used and analysed in this study are freely available in Zenodo (10.5281/zenodo.7737574). The source code for both the basic and the multilevel HMM is freely available on the developer branch of the R package mHMMbayes on GitHub (see functions HMM_plnorm and mHMM_plnorm, respectively; https://github.com/smildiner/mHMMbayes/tree/develop). Code for analysing the trained models is freely available on GitHub (https://github.com/danclab/motor_mea_mbhmm).

PEER REVIEW
The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ejn.16065.

ORCID
Sebastien Kirchherr https://orcid.org/0000-0002-2855-5329
Sebastian Mildiner Moraga https://orcid.org/0000-0001-6282-7394
Marco Bimbi https://orcid.org/0000-0002-7541-3710
Pier F. Ferrari https://orcid.org/0000-0003-1987-8748
Emmeke Aarts https://orcid.org/0000-0002-2432-7564
James J. Bonaiuto https://orcid.org/0000-0001-9165-4082

REFERENCES
Aarts, E. (2019). mHMMbayes: Multilevel Hidden Markov Models Using Bayesian Estimation (0.1.1.9002) [R]. https://cran.r-project.org/web/packages/mHMMbayes/mHMMbayes.pdf

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & van der Sluis, S. (2014). A solution to dependency: Using multilevel analysis to accommodate nested data. *Nature Neuroscience*, *17*(4), 4. https://doi.org/10.1038/nn.3648

Aguero-Valverde, J. (2013). Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. *Accident Analysis & Prevention*, *50*, 289–297. https://doi.org/10.1016/j.aap.2012.04.019

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Controzl*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Altman, R. M. (2007). Mixed Hidden Markov Models. *Journal of the American Statistical Association*, *102*(477), 201–210. https://doi.org/10.1198/016214506000001086

Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature Reviews Neuroscience*, *22*(6), 6. https://doi.org/10.1038/s41583-021-00448-6

Barrese, J. C., Aceros, J., & Donoghue, J. P. (2016). Scanning electron microscopy of chronically implanted intracortical microelectrode arrays in non-human primates. *Journal of Neural Engineering*, *13*(2), 026003. https://doi.org/10.1088/1741-2560/13/2/026003

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious Mixed Models (arXiv:1506.04967). https://doi.org/10.48550/arXiv.1506.04967

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, *37*(6), 1554–1563. https://doi.org/10.1214/aoms/1177699147

Bollimunta, A., Totten, D., & Ditterich, J. (2012). Neural Dynamics of Choice: Single-Trial Analysis of Decision-Related Activity in Parietal Cortex. *The Journal of Neuroscience*, *32*(37), 37. https://doi.org/10.1523/JNEUROSCI.5752-11.2012

Brooks, S. P., & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455. https://doi.org/10.1080/10618600.1998.10474787

Congdon, P. D. (2019). *Bayesian hierarchical models: With applications using R* (Second Edition 2e éd. ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429113352

Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, *17*(11), 11. https://doi.org/10.1038/nn.3776

Dayan, E., & Cohen, L. G. (2011). Neuroplasticity Subserving Motor Skill Learning. *Neuron*, *72*(3), 3. https://doi.org/10.1016/j.neuron.2011.10.008

de Haan-Rietdijk, S., Kuppens, P., Bergeman, C. S., Sheeber, L. B., Allen, N. B., & Hamaker, E. L. (2017). On the Use of Mixed Markov Models for Intensive Longitudinal Data. *Multivariate Behavioral Research*, *52*(6), 6. https://doi.org/10.1080/00273171.2017.1370364

Diomedi, S., Vaccari, F. E., Galletti, C., Hadjidimitrakis, K., & Fattori, P. (2021). Motor-like neural dynamics in two parietal areas during arm reaching. *Progress in Neurobiology*, *205*, 102116. https://doi.org/10.1016/j.pneurobio.2021.102116

Duncker, L., & Sahani, M. (2021). Dynamics on the manifold: Identifying computational dynamical activity from neural

population recordings. *Current Opinion in Neurobiology*, *70*, 163–170. https://doi.org/10.1016/j.conb.2021.10.014

Forney, G. D. (1973). The viterbi algorithm. *Proceedings of the IEEE*, *61*(3), 268–278. https://doi.org/10.1109/PROC.1973.9030

Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., & Bolker, B. (2019). car: Companion to Applied Regression. R package version 3.0-2. Website. https://cran.r-project.org/package=car [Accessed 17 March 2020].

Gallego, J. A. (2018). Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature Communications*, *13*, 4233. https://doi.org/10.1038/s41467-018-06560-z

Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A., & Miller, L. E. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. *Nature Neuroscience*, *23*(2), 260–270. https://doi.org/10.1038/s41593-019-0555-4

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural manifolds for the control of movement. *Neuron*, *94*(5), 978–984. https://doi.org/10.1016/j.neuron.2017.05.025

Garcia S, Pouzat C. (2015). Tridesclous. https://github.com/tridesclous/tridesclous

Gat, I., Tishby, N., & Abeles, M. (1997). Hidden Markov modelling of simultaneously recorded cells in the associative cortex of behaving monkeys. *Network: Computation in Neural Systems*, *8*(3), 297–322. https://doi.org/10.1088/0954-898X_8_3_005

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2015). *Bayesian data analysis* ((3e éd.) ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Gelman, A., & Pardoe, I. (2006). Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models. *Technometrics*, *48*(2), 241–251.

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 8–38. https://doi.org/10.1111/j.2044-8317.2011.02037.x

Gohil, C., Roberts, E., Timms, R., Skates, A., Higgins, C., Quinn, A., Pervaiz, U., Amersfoort, J. van, Notin, P., Gal, Y., Adaszewski, S., & Woolrich, M. (2022). Mixtures of large-scale dynamic functional brain network modes (p. 2022.05.03.490453). bioRxiv. https://doi.org/10.1101/2022.05.03.490453

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge/Taylor & Francis Group.

Humphries, M. D. (2021). Strong and weak principles of neural dimension reduction. ArXiv:2011.08088 [q-Bio]. http://arxiv.org/abs/2011.08088

Kadmon Harpaz, N., Ungarish, D., Hatsopoulos, N. G., & Flash, T. (2019). Movement decomposition in the primary motor cortex. *Cerebral Cortex*, *29*(4), 1619–1633. https://doi.org/10.1093/cercor/bhy060

Kemere, C., Santhanam, G., Yu, B. M., Afshar, A., Ryu, S. I., Meng, T. H., & Shenoy, K. V. (2008). Detecting neural-state transitions using hidden Markov models for motor cortical prostheses. *Journal of Neurophysiology*, *100*(4), 2441–2452. https://doi.org/10.1152/jn.00924.2007

Lenth, R., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2020). *emmeans: Estimated marginal means, aka least-squares means, v1. 5.1*. R Core Team.

Ma, J., Kockelman, K. M., & Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention*, *40*(3), 964–975. https://doi.org/10.1016/j.aap.2007.11.002

Marcos, E., Londei, F., & Genovesio, A. (2019). Hidden Markov models predict the future choice better than a PSTH-based method. *Neural Computation*, *31*(9), 1874–1890. https://doi.org/10.1162/neco_a_01216

Mazor, O., & Laurent, G. (2005). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron*, *48*(4), 661–673. https://doi.org/10.1016/j.neuron.2005.09.032

Mazurek, K. A., Rouse, A. G., & Schieber, M. H. (2018). Mirror neuron populations represent sequences of behavioral epochs during both execution and observation. *The Journal of Neuroscience*, *38*(18), 4441–4455. https://doi.org/10.1523/JNEUROSCI.3481-17.2018

McClintock, B. T. (2021). Worth the effort? A practical examination of random effects in hidden Markov models for animal telemetry data. *Methods in Ecology and Evolution*, *12*(8), 1475–1497. https://doi.org/10.1111/2041-210X.13619

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

McMahon, D. B. T., Bondar, I. V., Afuwape, O. A. T., Ide, D. C., & Leopold, D. A. (2014). One month in the life of a neuron: Longitudinal single-unit electrophysiology in the monkey visual system. *Journal of Neurophysiology*, *112*(7), 1748–1762. https://doi.org/10.1152/jn.00052.2014

Muthmann, J.-O., Levi, A. J., Carney, H. C., & Huk, A. C. (2021). *A hardware/software system for electrophysiology "supersessions" in marmosets* (p. 2020.08.09.243279). bioRxiv. https://doi.org/10.1101/2020.08.09.243279

Pandarinath, C., O'Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., Trautmann, E. M., Kaufman, M. T., Ryu, S. I., Hochberg, L. R., Henderson, J. M., Shenoy, K. V., Abbott, L. F., & Sussillo, D. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, *15*, 805–815. https://doi.org/10.1038/s41592-018-0109-9

Ponce-Alvarez, A., Nacher, V., Luna, R., Riehle, A., & Romo, R. (2012). Dynamics of Cortical Neuronal Ensembles Transit from Decision Making to Storage for Later Report. *Journal of Neuroscience*, *32*(35), 35. https://doi.org/10.1523/JNEUROSCI.6176-11.2012

R Core Team. (2021). *R: A language and environment for statistical computing (version 4.1. 1,"Kick Things")*. R Foundation for Statistical Computing.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. https://doi.org/10.1109/5.18626

Radons, G., Becker, J. D., Dülfer, B., & Krüger, J. (1994). Analysis, classification, and coding of multielectrode spike trains with hidden Markov models. *Biological Cybernetics*, *71*(4), 359–373. https://doi.org/10.1007/BF00239623

Raffa, J. D., & Dubin, J. A. (2015). Multivariate longitudinal data analysis with mixed effects hidden Markov models. *Biometrics*, *71*(3), 821–831. https://doi.org/10.1111/biom.12296

Rueda, O. M., Rueda, C., & Diaz-Uriarte, R. (2013). A Bayesian HMM with random effects and an unknown number of states

for DNA copy number analysis. *Journal of Statistical Computation and Simulation*, *83*(1), 1. https://doi.org/10.1080/00949655.2011.609818

Russo, A. A., Khajeh, R., Bittner, S. R., Perkins, S. M., Cunningham, J. P., Abbott, L. F., & Churchland, M. M. (2020). Neural trajectories in the supplementary motor area and motor cortex exhibit distinct geometries, compatible with different classes of computation. *Neuron*, *107*(4), 745–758.e6. https://doi.org/10.1016/j.neuron.2020.05.020

Sadacca, B. F., Mukherjee, N., Vladusich, T., Li, J. X., Katz, D. B., & Miller, P. (2016). The Behavioral Relevance of Cortical Neural Ensemble Responses Emerges Suddenly. *Journal of Neuroscience*, *36*(3), 655–669. https://doi.org/10.1523/JNEUROSCI.2265-15.2016

Saxena, S., & Cunningham, J. P. (2019). Towards the neural population doctrine. *Current Opinion in Neurobiology*, *55*, 103–111. https://doi.org/10.1016/j.conb.2019.02.002

Scott, S. L. (2002). Bayesian Methods for Hidden Markov Models. *Journal of the American Statistical Association*, *97*(457), 337–351.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.

Seidemann, E., Meilijson, I., Abeles, M., Bergman, H., & Vaadia, E. (1996). Simultaneously recorded single units in the frontal cortex go through sequences of discrete and stable states in monkeys performing a delayed localization task. *The Journal of Neuroscience*, *16*(2), 752–768. https://doi.org/10.1523/JNEUROSCI.16-02-00752.1996

Shirley, K., Vasilaky, K. N., Greatrex, H. L., & Osgood, D. E. (2016). Hierarchical Bayes models for daily rainfall time series at multiple locations from heterogenous data sources. https://doi.org/10.7916/D8QF8SZ4

Shirley, K. E., Small, D. S., Lynch, K. G., Maisto, S. A., & Oslin, D. W. (2012). Hidden Markov models for alcoholism treatment trial data. *Annals of Applied Statistics*, *6*(1), 366–395. https://doi.org/10.1214/09-AOAS282

Schoot, R. van de, & Miočevic, M.(2020). *Small sample size solutions: A Guide for Applied Researchers andPractitioners*. https://doi.org/10.4324/9780429273872

Trautmann, E. M., Stavisky, S. D., Lahiri, S., Ames, K. C., Kaufman, M. T., O'Shea, D. J., Vyas, S., Sun, X., Ryu, S. I., Ganguli, S., & Shenoy, K. V. (2019). Accurate Estimation of Neural Population Dynamics without Spike Sorting. *Neuron*, *103*(2), 292–308.e4. https://doi.org/10.1016/j.neuron.2019.05.003

van Kempen, J., Gieselmann, M. A., Boyd, M., Steinmetz, N. A., Moore, T., Engel, T. A., & Thiele, A. (2021). Top-down coordination of local cortical state during selective attention. *Neuron*, *109*(5), 894–904.e8. https://doi.org/10.1016/j.neuron.2020.12.013

Visser, I., & Speekenbrink, M. (2010). depmixS4: An R package for hidden Markov models. *Journal of Statistical Software*, *36*, 1–21.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, *13*(2), 260–269. https://doi.org/10.1109/TIT.1967.1054010

Warwick, K. (2011). Revealing ensemble state transition patterns in multi-electrode neuronal recordings using hidden Markov models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, *19*(4), 4. https://doi.org/10.1109/TNSRE.2011.2157360

Welle, C. G., Gao, Y.-R., Ye, M., Lozzi, A., Boretsky, A., Abliz, E., & Hammer, D. X. (2020). Longitudinal neural and vascular structural dynamics produced by chronic microelectrode implantation. *Biomaterials*, *238*, 119831. https://doi.org/10.1016/j.biomaterials.2020.119831

Yu, S.-Z. (2010). Hidden semi-Markov models. *Artificial Intelligence*, *174*(2), 215–243. https://doi.org/10.1016/j.artint.2009.11.011

Zhang, Y., & Berhane, K. (2014). Bayesian mixed hidden Markov models: A multi-level approach to modeling categorical outcomes with differential misclassification. *Statistics in Medicine*, *33*(8), 1395–1408. https://doi.org/10.1002/sim.6039

Zhu, H., Li, X., Sun, L., He, F., Zhao, Z., Luan, L., Tran, N. M., & Xie, C. (2020). *Clustering with Fast, Automated and Reproducible assessment applied to longitudinal neural tracking* (arXiv:2003.08533). arXiv. https://doi.org/10.48550/arXiv.2003.08533

Zucchini, W., Macdonald, I. L., & Langrock, R. (2017). Hidden Markov models for time series: An introduction using R, second edition. In *Hidden Markov Models for Time Series: An Introduction Using R* (Second ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b20790

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.