PNAS RESEARCH ARTICLE | NEUROSCIENCE

Modeling and dissociation of intrinsic and input-driven neural population dynamics underlying behavior

Parsa Vahidi[a,1], Omid G. Sani[a,1], and Maryam M. Shanechi[a,b,c,2]

Edited by Terrence Sejnowski, Salk Institute for Biological Studies, La Jolla, CA; received July 28, 2022; accepted December 3, 2023

Neural dynamics can reflect intrinsic dynamics or dynamic inputs, such as sensory inputs or inputs from other brain regions. To avoid misinterpreting temporally structured inputs as intrinsic dynamics, dynamical models of neural activity should account for measured inputs. However, incorporating measured inputs remains elusive in joint dynamical modeling of neural-behavioral data, which is important for studying neural computations of behavior. We first show how training dynamical models of neural activity while considering behavior but not input or input but not behavior may lead to misinterpretations. We then develop an analytical learning method for linear dynamical models that simultaneously accounts for neural activity, behavior, and measured inputs. The method provides the capability to prioritize the learning of intrinsic behaviorally relevant neural dynamics and dissociate them from both other intrinsic dynamics and measured input dynamics. In data from a simulated brain with fixed intrinsic dynamics that performs different tasks, the method correctly finds the same intrinsic dynamics regardless of the task while other methods can be influenced by the task. In neural datasets from three subjects performing two different motor tasks with task instruction sensory inputs, the method reveals low-dimensional intrinsic neural dynamics that are missed by other methods and are more predictive of behavior and/or neural activity. The method also uniquely finds that the intrinsic behaviorally relevant neural dynamics are largely similar across the different subjects and tasks, whereas the overall neural dynamics are not. These input-driven dynamical models of neural-behavioral data can uncover intrinsic dynamics that may otherwise be missed.

intrinsic dynamics | input dynamics | behavior | neural encoding | dynamical systems

Significance

Neural dynamics emerge either intrinsically within the recorded brain regions or due to inputs to those regions, such as sensory inputs or neural inputs from other regions. Further, recorded neural dynamics may or may not be related to a specific measured behavior of interest. We first show how intrinsic neural dynamics that underlie a behavior can be confounded by both measured inputs and other intrinsic neural dynamics. To address this challenge, we develop methods that dissociate the intrinsic neural dynamics related to specific behaviors from other intrinsic dynamics and measured input dynamics simultaneously. We show the success of these methods in simulations and real data from three subjects in two independent neural datasets recorded during two distinct motor tasks.

Neural population activity exhibits rich temporal structures (1–26). Investigating these temporal structures, i.e., dynamics, can reveal the neural computations that underlie behavior (5, 6, 12, 15, 16, 19, 20). Much progress has been made in developing models that can describe the dynamics of neural population activity using a low-dimensional latent state (2–4, 7, 8, 10–14, 16, 19). However, a major challenge in such investigations is that neural dynamics can arise due to two distinct sources that reflect distinct computations (12, 15, 27). The first source consists of the intrinsic dynamics within a given brain region. Intrinsic dynamics arise due to the recurrent connections within a region's neuronal population as it responds in a temporally structured manner to any excitations from within or outside that region (6, 12, 15, 18, 27, 28). The second source consists of input dynamics, which are temporal structures that already exist in inputs to the recorded brain region, including sensory inputs or inputs from other brain regions (1, 9, 12, 15, 27–31). While measuring all inputs is infeasible experimentally, measurements of sensory inputs such as task instructions or partial measurements of neural inputs into a brain region are often possible. As such, correctly interpreting how neural computations in a given brain region give rise to a specific behavior can greatly benefit from simultaneously achieving two objectives, which remains elusive.

First, given the above two sources, neural dynamics that are intrinsic to a given brain region need to be dissociated from those that are simply due to temporally structured measured inputs to that region. Second, within intrinsic neural dynamics, those that are relevant to the specific behavior of interest need to be dissociated from other intrinsic neural dynamics. This latter dissociation is important because neural dynamics of a specific behavior often constitute a minority of the total variance in the recorded neural activity (5, 6, 19, 32–39). Indeed, recent work has shown that learning dynamical models of neural-behavioral data together and in a way that dissociates and prioritizes their shared dynamics can unmask behaviorally relevant neural dynamics that may otherwise not be found (19, 20). We refer to such prioritized learning approach for neural-behavioral data as preferential dynamical modeling because it preferentially models the behaviorally relevant neural dynamics with priority instead of non-preferentially modeling prevalent dynamics in neural data as is typically done. However, prior methods for preferential dynamical

Author contributions: P.V., O.G.S., and M.M.S. designed research; performed research; contributed new analytic tools; analyzed data; and wrote the paper.

Competing interest statement: USC has a patent related to modeling and decoding of shared dynamics between signals.

This article is a PNAS Direct Submission.

[1]P.V. and O.G.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: shanechi@usc.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2212887121/-/DCSupplemental.

Published February 9, 2024.

modeling of neural-behavioral data do not account for the effect of measured inputs to a given brain region. Thus, the dissociation of intrinsic and input-driven neural population dynamics that underlie specific behaviors has remained challenging.

Here, we first show how misinterpretation and incorrect identification of intrinsic behaviorally relevant dynamics could result from modeling neural activity while considering behavior but not input or while considering input but not behavior. Indeed, modeling neural activity without considering the measured input could result in a model that mistakes the temporal structure in the input as part of the intrinsic dynamics within the recorded brain region (9, 27) and consequently confounds scientific conclusions. For non-preferential modeling of neural activity on its own, while not commonly done, various methods can be adapted to fit models with measured inputs (40) but they cannot account for behavior. Thus, as we show, despite considering input, these non-preferential methods can miss those intrinsic neural dynamics that are behaviorally relevant. Further, as stated above, methods for preferential dynamical modeling that consider the neural-behavioral data together do not consider measured inputs. Here we aim to formulate and solve a learning problem that involves neural activity, behavior, and measured inputs simultaneously.

To do so, we develop a preferential modeling approach, termed input preferential subspace identification (IPSID) that can consider both measured inputs and behaviors in the training set while learning linear dynamical models of neural population activity. By doing so, IPSID provides the capability to learn the intrinsic behaviorally relevant neural dynamics with priority and dissociate them both from other intrinsic neural dynamics and from the dynamics of measured inputs. We also develop a version of IPSID that achieves this capability when some input dynamics influence the behavior through pathways that are neither recorded nor downstream of the recorded neural activity. Compared with our prior preferential dynamical modeling method (i.e., PSID) (19, 41), which does not incorporate input or dissociate intrinsic and input dynamics, IPSID requires distinct mathematical operations and additional steps (*SI Appendix*, Note S1). We show that two capabilities introduced by IPSID are critical for accurate dissociation of intrinsic behaviorally relevant neural dynamics: prioritized learning of these dynamics in the presence of input and ensuring all learned dynamics are directly present in the neural recordings even when inputs affect behavior.

We validate IPSID and its capabilities in extensive numerical simulations of diverse dynamical systems and in two independent motor cortical datasets from three non-human primates (NHP) recorded during two different tasks with task instruction sensory inputs. First, we simulate a brain with fixed intrinsic dynamics that performs different behavioral tasks. IPSID correctly learns the same intrinsic behaviorally relevant neural dynamics regardless of which specific task is used to collect the simulated training neural data. In contrast, other methods learn intrinsic dynamics that are inaccurate and influenced by the specific task. Second, we apply IPSID to motor cortical population activity recorded from three NHPs in two independent datasets with two different 2-dimensional (2D) cursor-control tasks. IPSID finds intrinsic behaviorally relevant dynamics that not only predict motor behavior better than non-preferential methods even with input, but also predict neural activity better than preferential methods, which cannot consider task instruction inputs. Further, IPSID reveals that intrinsic behaviorally relevant neural dynamics are largely similar across the three animals despite differences in the two cursor-control tasks and animals, while other methods miss these similar dynamics. By dissociating intrinsic behaviorally relevant dynamics from both other intrinsic dynamics and measured input dynamics, IPSID can help explore unanswered questions regarding how intrinsic and input-driven neural computations give rise to behavior across subjects and tasks.

Methods

Modeling Intrinsic Neural Dynamics Underlying Behavior in the Presence of Inputs. To see how measured inputs, if unaccounted for, can be misinterpreted as intrinsic neural dynamics, consider a task where a subject is instructed to follow an on-screen target with their hand while motor cortical activity that represents the hand movements is recorded (Fig. 1*A*). Here, movements of the target would result in corresponding movements in the hand that follows the target and thus would also introduce corresponding dynamics in the neural activity that represents hand movements. Consequently, any arbitrary movement of the target will be, to some extent, reflected in the recorded neural activity. An example is shown in a numerical simulation in Fig. 1 *A* and *B*. As another example, if the target moves up and down with a 1-s period, one would expect the neural activity to also include similar periodic patterns with a 1-s period. If the period of target movements changes to 2 s, so would the period of the patterns in neural activity that represent the hand movements. Any neural modeling that is not informed by target movements, which serve as task instruction sensory inputs, cannot distinguish between such input dynamics and intrinsic dynamics that originate in the recorded brain region. Thus, modeling without considering this input may incorrectly conclude that there exist intrinsic dynamics originating in the recorded brain area that are periodic with a 1-s period. The reflection of input dynamics in neural dynamics can also be seen in terms of the frequency domain spectrum of these signals (Fig. 1*B*). In this view, the correct dissociation of intrinsic dynamics from input dynamics requires the correct learning of the transfer function from inputs to neural signals, in a way that does not incorrectly attribute the input dynamics that appear in neural activity to having originated from the transfer function (Fig. 1*B*).

To formulate the goal of IPSID, we represent the dynamical state of the recorded brain regions as a high-dimensional vector. Each state dimension may or may not contribute to generating the specific behavior of interest, i.e., be behaviorally relevant (Fig. 1*A*). As discussed in the Introductory paragraphs, two major factors can confound the learning of intrinsic behaviorally relevant neural dynamics: 1) the dynamics of the measured input and 2) other intrinsic neural dynamics. IPSID removes both confounding factors by accounting for neural activity, behavior, and measured inputs simultaneously during learning. Unlike IPSID, prior methods address only one or the other confound but not both. First, non-preferential neural dynamic modeling (NDM) with input (*SI Appendix, Methods*), which we term INDM, accounts for neural activity and measured input but not behavior during learning. As such, INDM may miss or confound the intrinsic neural dynamics that are behaviorally relevant. Second, a dynamical method termed PSID (19, 41) addresses the second confound by accounting for neural activity and behavior during learning but not input. As such, PSID does not dissociate intrinsic and input dynamics. We thus use this naming convention for ease of exposition but the algebraic operations in IPSID are different from those in both PSID and INDM and further IPSID includes additional steps compared with these prior methods (*SI Appendix*, Notes S1 and S2).

In IPSID, we use the following linear state-space model to jointly describe the dynamics of neural activity (y_k) and behavior (z_k) in the presence of measured input (u_k)

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + w_k \\ y_k = C_y x_k + D_y u_k + v_k, \quad x_k = \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} \\ z_k = C_z x_k + D_z u_k + \epsilon_k \end{cases} \quad [1]$$

where $x_k \in \mathbb{R}^{n_x}$ is the latent state in the recorded neural activity and composed of two parts: 1) $x_k^{(1)} \in \mathbb{R}^{n_1}$, which is the behaviorally relevant states and 2) $x_k^{(2)} \in \mathbb{R}^{n_x - n_1}$, which is the other states. In this model, $y_k \in \mathbb{R}^{n_y}$, $z_k \in \mathbb{R}^{n_z}$, and $u_k \in \mathbb{R}^{n_u}$ represent the recorded neural activity, the measured behavior, and the measured input, respectively. Here, $x_k^{(1)}$ being behaviorally relevant means

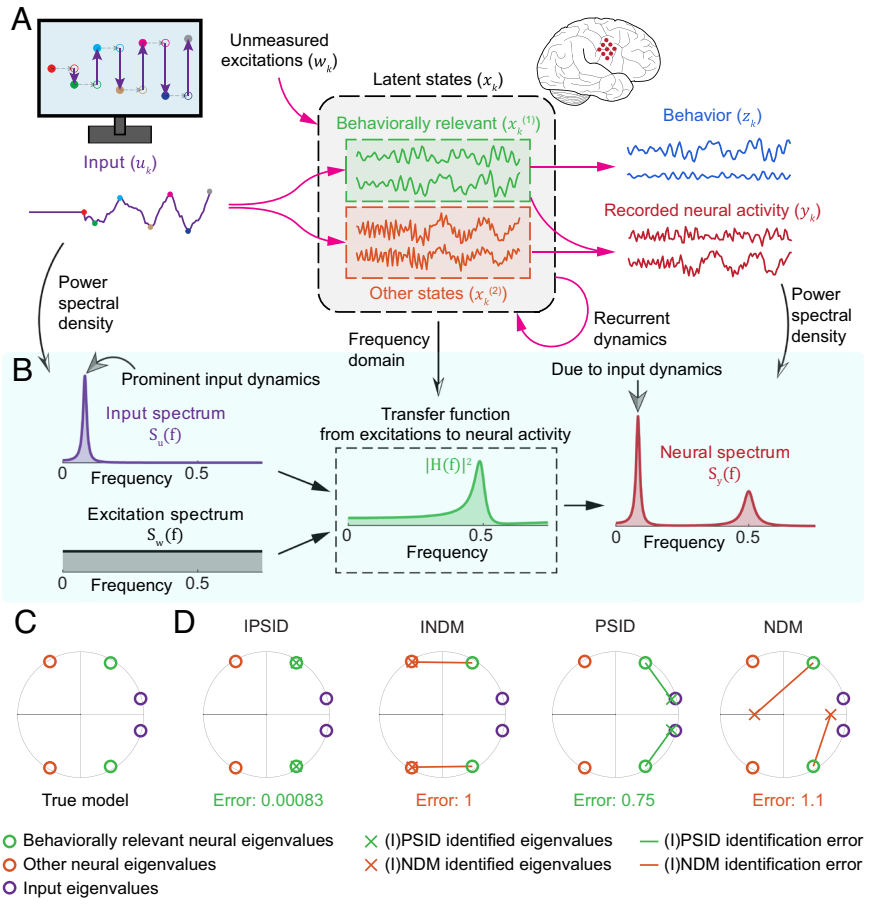

Fig. 1. Intrinsic behaviorally relevant neural dynamics may be confounded by other intrinsic neural dynamics as well as by measured input dynamics, a challenge that the IPSID method resolves. (*A*) Data generated from a simulated brain following Eq. **1** with a 1D input and a 4D latent state out of which only 2 dimensions (green) drive behavior. The input is taken as the sensory input such as target position moving up and down on a screen as depicted, but input can also consist of measured activity from other upstream brain regions. Neural dynamics that arise from the recurrent dynamics of neuronal networks within the brain region constitute the intrinsic neural dynamics. Oscillating temporal patterns of the input (*Left*) constitute the input dynamics and clearly also appear in the neural activity (*Right*) in a way that is mixed with the intrinsic neural dynamics. (*B*) Appearance of input dynamics in neural dynamics can also be clearly seen in the frequency domain representation of (*A*), showing: the power spectral density (PSD), or spectrum, of input time series $S_u(f)$ (*Top-Left*); PSD of unmeasured excitations $S_w(f)$ modeled as white Gaussian noise (*Bottom-Left*); transfer function from inputs to the neural activity (*Middle*); and PSD of neural activity (*Right*). Neural activity exhibits two dominant frequency components. In this simulation, the lower-frequency component is the reflection of input dynamics while the higher-frequency component represents intrinsic neural dynamics (as also present in the transfer function). Horizontal axes show the normalized frequency with 1 being the maximum, i.e., π. (*C*) The eigenvalues of the state transition matrix A in the simulated brain model in Eq. **1**. (*D*) Learned eigenvalues using (I)PSID or (I)NDM and their error (red lines). The normalized error value—average line length normalized by the average true eigenvalue magnitude—is noted below each plot.

that only those dimensions of x_k corresponding to $x_k^{(1)}$ contribute to generating behavior (z_k) in the third row of Eq. **1**. Finally, w_k and v_k are zero mean white Gaussian noises (*SI Appendix, Methods*), and ϵ_k is a general Gaussian random process representing any behavior dynamics not encoded in the recorded neural activity (i.e., not driven by x_k).

Prior works have not addressed the problem of fitting this model in a way that dissociates and prioritizes the learning of behaviorally relevant latent states, which is achieved by IPSID. Operationally, dissociation is the process of differentiating two subtypes of neural dynamics from each other and returning both to the user. Prioritization is the process of dedicating model capacity (e.g., latent state dimensions) to explaining one subtype first and dedicating model capacity to other subtypes only if some model capacity is left, which results in the learning of the former subtype taking priority over the learning of the second subtype. To enable such preferential/prioritized learning, IPSID introduces a two-stage learning procedure that incorporates input as follows. In the first stage of IPSID, we develop algebraic operations that extract the behaviorally relevant latent states with priority via an oblique (non-orthogonal) projection of future behavior onto past neural activity and past inputs along the subspace spanned by future inputs (*SI Appendix*, Fig. S1 and *Methods*). Then, in an optional second stage, we devise algebraic operations that extract any other latent neural states by another oblique projection from any residual/unexplained future neural activity onto past neural activity and past inputs along future inputs (*SI Appendix*, Fig. S1). Model parameters are then learned via least squares based on the extracted latent states and their relation in Eq. **1**.

IPSID's two-stage learning introduces the capability for prioritized learning of the intrinsic behaviorally relevant neural dynamics over other intrinsic neural dynamics in the presence of inputs, because the former dynamics are learned first, i.e., in the first stage. Specifically, IPSID can learn a minimally complex model of those intrinsic neural dynamics that are behaviorally relevant in the first stage (i.e., a model with low-dimensional states), instead of having to learn a more complex model that includes all of the intrinsic neural dynamics simultaneously. As learning less complex models can be more accurate for a given number of training samples, this two-stage learning can lead to learning more accurate models of intrinsic behaviorally relevant dynamics for a given dataset as shown in simulations and in real data analyses below. Moreover, IPSID achieves dissociation of behaviorally relevant dynamics because the two sets of states learned by the two stages are placed in predetermined and distinct dimensions of the latent state: the first n_1 dimensions versus the rest. After the model is learned, in the test set, extraction of intrinsic behaviorally relevant neural dynamics is done without using behavior and via a Kalman filter associated with the learned model (*SI Appendix, Methods*). Details of IPSID are provided in *SI Appendix, Methods* and Notes S1 and S2.

To assess the methods, we look at the eigenvalues of the latent state transition matrix A, which quantify the dynamics (*SI Appendix, Methods* and Fig. 1 *C* and *D*). We also compute the accuracy in decoding behavior from neural activity as well as in neural self-prediction–defined as predicting neural activity one step ahead from its own past (*SI Appendix, Methods*).

Results

IPSID Correctly Learns All Model Parameters in the Presence of Inputs. We first validated the accurate learning of intrinsic behaviorally relevant neural dynamics using IPSID in a simulated model (Fig. 1*A*). The eigenvalues of the state transition matrix A affect the transfer function from the input to the states and neural activity (Fig. 1*B*), characterize the state response to excitations, and describe the dynamics (Fig. 1*C* and *SI Appendix, Methods*). We thus use these eigenvalues to quantify the intrinsic neural dynamics (*SI Appendix, Methods*). We found that IPSID was the only method that correctly learned the eigenvalues associated with the intrinsic behaviorally relevant neural dynamics (Fig. 1*D*). In contrast, NDM or PSID that do not consider inputs learned models that were confounded by input dynamics (eigenvalues were deflected toward input eigenvalues); INDM that does not

consider behavior was confounded by other intrinsic neural dynamics beyond the behaviorally relevant ones (Fig. 1*D*).

To more comprehensively validate IPSID, we applied it to data generated from 100 random models in the form of Eq. **1** with random parameters and dimensions (*SI Appendix, Methods*). To provide input to these models, we independently simulated another 100 models without input (Eq. 3 from *SI Appendix, Methods*) with random parameters and passed their output as the input to the original models—these inputs are thus generated by an independent dynamical system and can be thought of as activity of other brain regions or as structured sensory inputs. IPSID correctly learned all model parameters in the presence of inputs (*SI Appendix*, Fig. S2). Moreover, the rate of convergence of parameters as a function of training samples was similar to INDM (*SI Appendix*, Fig. S2*B*); this suggests that despite its additional capability in dissociating intrinsic behaviorally relevant dynamics, IPSID does not require more training data than INDM even when modeling all dynamics.

IPSID Prioritizes the Learning of Intrinsic Behaviorally Relevant Dynamics in the Presence of Inputs. In another numerical simulation, we found that IPSID correctly prioritizes the learning of intrinsic behaviorally relevant neural dynamics in the presence of inputs (Fig. 2). We simulated 100 random models formulated by Eq. **1** with a 6D latent state, out of which only 2 dimensions were behaviorally relevant (*SI Appendix, Methods*). To get the input to these models, we independently simulated 100 random models without input (Eq. 3 from *SI Appendix, Methods*) with 2D latent states and passed their output as the input to the original models. We then learned and evaluated models using (I)PSID and (I)NDM with varying latent state dimensions (n_x). In each case, we computed the error in learning the intrinsic behaviorally relevant eigenvalues, which quantifies how accurately intrinsic behaviorally relevant dynamics are learned (Fig. 2*B* and *SI Appendix*, Fig. S3).

We found that only IPSID could learn all the intrinsic behaviorally relevant neural dynamics/eigenvalues using the minimal latent state dimension of 2, which is their true dimension (Fig. 2*B* and *SI Appendix*, Fig. S4). Thus, IPSID could simultaneously dissociate the intrinsic behaviorally relevant dynamics from other intrinsic dynamics and input dynamics by considering both input and behavior during learning. In contrast, even though INDM considers inputs, it does not consider behavior during learning and thus it required a much larger latent state dimension of 6 (true total model dimension) to learn the intrinsic behaviorally relevant eigenvalues (Fig. 2*B*). This higher required dimension also led to INDM's higher eigenvalue error with the same training sample size as IPSID (Fig. 2*C*) because models with higher dimensional states are more complex and difficult to learn. Indeed, IPSID required orders of magnitude fewer training samples to learn the intrinsic behaviorally relevant dynamics in the presence of inputs (Fig. 2*C*).

We next found that NDM and PSID models, which do not consider input, were unable to dissociate the intrinsic versus input dynamics, leading to a high intrinsic eigenvalue error (Fig. 2*B*). This error was high even when increasing NDM/PSID's state dimensions to learn a mixture of all intrinsic neural dynamics and input dynamics first. When we reduced these high-dimensional models to only keep the two dimensions that were best in decoding behavior (as we did with INDM above, *SI Appendix, Methods*), the associated eigenvalues were still much less accurate than low-dimensional models learned with IPSID (see Fig. 2*B* at high dimensions).

IPSID Can Dissociate the Effects of Input on Behavior that Are Reflected in the Recorded Neural Activity from those that Are Not. In Eq. **1**, all the effects of input on behavior happen through latent states that are reflected in the recorded neural activity. In this scenario, all the downstream regions of the input are either covered in the recordings or reflected in them (e.g., are downstream of the recorded regions). In addition to this scenario, we now show that IPSID can also apply to a more general scenario where inputs may also influence behavior through pathways/regions that are neither recorded nor reflected in the recorded activity (Fig. 3*A*). We formulate this scenario with the following model

$$\begin{cases} \begin{bmatrix} x^{(1)}_{k+1} \\ x^{(2)}_{k+1} \\ x^{(3)}_{k+1} \end{bmatrix} = \begin{bmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & A_{33} \end{bmatrix} \begin{bmatrix} x^{(1)}_{k} \\ x^{(2)}_{k} \\ x^{(3)}_{k} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} u_k + \begin{bmatrix} w^{(1)}_{k} \\ w^{(2)}_{k} \\ w^{(3)}_{k} \end{bmatrix} \\ y_k = \begin{bmatrix} C_{y_1} & C_{y_2} & 0 \end{bmatrix} \begin{bmatrix} x^{(1)}_{k} \\ x^{(2)}_{k} \\ x^{(3)}_{k} \end{bmatrix} + D_y u_k + v_k \\ z_k = \begin{bmatrix} C_{z_1} & 0 & C_{z_3} \end{bmatrix} \begin{bmatrix} x^{(1)}_{k} \\ x^{(2)}_{k} \\ x^{(3)}_{k} \end{bmatrix} + \epsilon_k \end{cases}, \quad [2]$$

where compared with Eq. **1**, an additional segment $x^{(3)}_k$ is added to the latent state x_k to represent the effects of input u_k on behavior z_k that are not reflected in the recorded neural activity y_k. In this formulation, IPSID dissociates the latent state into three segments: 1) $x^{(1)}_k \in \mathbb{R}^{n_1}$, which is the behaviorally relevant latent state that is reflected in neural activity y_k, 2) $x^{(2)}_k \in \mathbb{R}^{n_2}$, which is the latent state that describes any other neural dynamics, and 3) $x^{(3)}_k \in \mathbb{R}^{n_x - n_1 - n_2}$, which is the behaviorally relevant latent state not reflected in the recorded neural activity y_k. These three types of latent states are shown in an example in Fig. 3*A*. Note that in this case, only $x^{(1)}_k$ and $x^{(2)}_k$ constitute the intrinsic latent states because only these latent states drive the recorded neural activity. To add support for dissociation of these three types of latent states to IPSID, we developed two additional optional steps for IPSID (*SI Appendix*, Fig. S5 and Note S2).

In the first additional step, before the initial oblique projection of behavior onto neural activity and input, we project behavior onto the subspace of latent states in neural activity (i.e., neural states) irrespective of the relevance of these states to behavior; these neural states are obtained using only the second stage of IPSID (*SI Appendix, Methods*, Note S2 and Figs. S5 and S6*A*). We then apply IPSID as before (*SI Appendix*, Note S1) but now use the results of this additional projection as the behavior signal. This additional projection ensures that behavior dynamics that are not encoded in the recorded neural activity are not included in the first set of states $x^{(1)}_k$.

In the second additional step, we optionally extract $x^{(3)}_k$, which represents any behavior dynamics that are driven by the input but are not encoded in the recorded neural activity—e.g., due to processing in the downstream regions of input that are not recorded/reflected as part of neural activity. In this step, after performing the first additional step above and subsequently both stages of IPSID to extract $x^{(1)}_k$ and $x^{(2)}_k$, we compute the residual behavior that is still not predictable using $x^{(1)}_k$ and $x^{(2)}_k$. Then, using the second stage of IPSID, we build a model that predicts these residual behavior dynamics purely using the input (*SI Appendix, Methods*, Note S2 and Fig. S5)—this gives $x^{(3)}_k$. Together, these two additional steps enable IPSID to learn a model as in Eq. **2**. If extraction of $x^{(3)}_k$ is not of interest, the second step can be skipped and solely the first step can be added to IPSID.

We simulated models in the form of Eq. **2** and confirmed that with the above additional steps, again only IPSID correctly dissociates

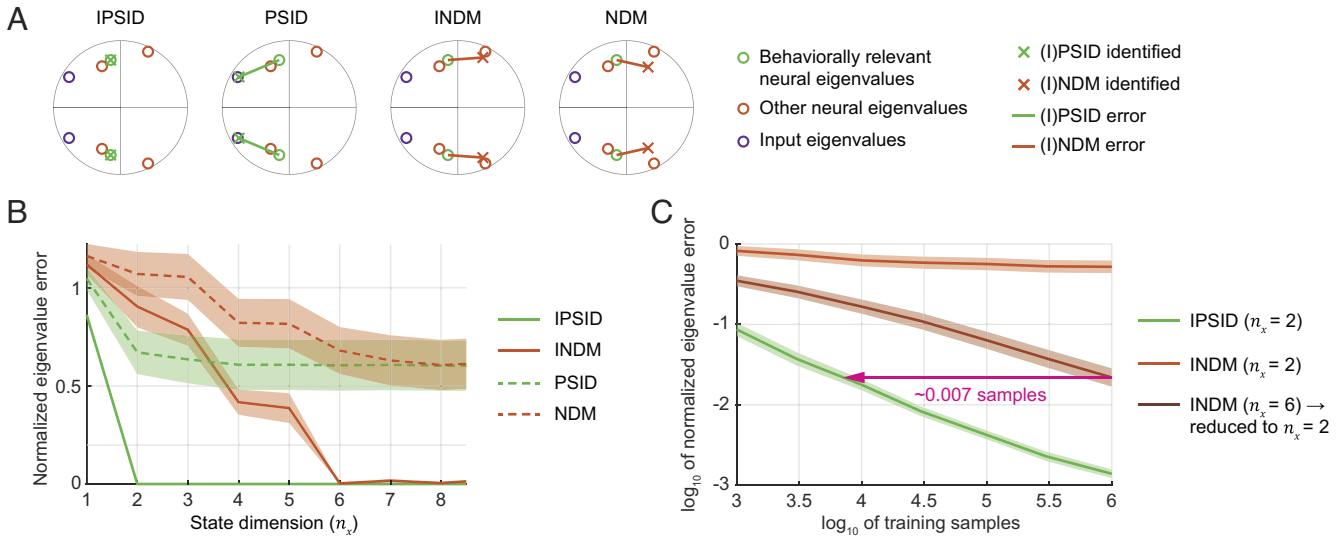

Fig. 2. IPSID prioritizes the learning of intrinsic behaviorally relevant neural dynamics thus achieving preferential neural-behavioral modeling even in the presence of input. (*A*) For one simulated model (Eq. **1**), the identified intrinsic behaviorally relevant eigenvalues are shown for (I)PSID and (I)NDM using a 2D latent state. Eigenvalues of the state transition matrix *A* in the true model are shown as colored circles. Crosses show the identified behaviorally relevant eigenvalues when modeling the neural activity. (*B*) Normalized error of learning the intrinsic behaviorally relevant eigenvalues vs. state dimension given 10^6 training samples is shown. Results are averaged over 100 random models each with total latent state dimension of $n_x = 6$ and behaviorally relevant state dimension of $n_1 = 2$. For all models, an independent random model with state dimension of 2 generated the input (*SI Appendix*, Methods). Solid lines show the average and shaded areas show the SEM ($n = 100$ random models). For all methods, we vary the state dimension n_x from 1 to 8; for $n_x < 2$, we find the 2 state dimensions that best predict behavior and evaluate their 2 associated eigenvalues (*SI Appendix, Methods*). (*C*) Normalized error of learning the intrinsic behaviorally relevant eigenvalues vs. training samples for 100 random models. For INDM, we try i) directly learning a model with a 2D latent state and ii) first learning a model with a high enough dimension to achieve almost zero error in *B* and then reducing the model to keep the top 2 dimensions with the best behavior decoding (indicated by dimension → 2) (*SI Appendix, Methods*).

intrinsic behaviorally relevant neural dynamics (i.e., $x_k^{(1)}$) from other dynamics—i.e., from other intrinsic neural dynamics, input dynamics, and behavior dynamics not encoded in the recorded neural activity (Fig. 3*C*). Moreover, across 100 random models, IPSID correctly learned all model parameters in Eq. **2** (*SI Appendix*, Fig. S7). Finally, by learning $x_k^{(3)}$, IPSID also achieved ideal prediction of behavior from input and neural activity (*SI Appendix*, Fig. S8).

These results demonstrate that IPSID is applicable to scenarios where the recorded neural activity does not cover all the downstream regions of the measured input. IPSID can also dissociate the influences of input on behavior that are reflected in the recorded neural activity from those that are not. Without this capability, some of the learned dynamics may not be present in the recorded region (Fig. 3*C*, top row comparisons). Thus, this is another capability by IPSID that is important for accurately dissociating intrinsic behaviorally relevant dynamics in neural recordings.

IPSID's Prioritized Modeling of Intrinsic Behaviorally Relevant Neural Dynamics Is Important for their Accurate Learning. Using its two-stage learning procedure in the presence of inputs, IPSID enables prioritized learning of intrinsic behaviorally relevant neural dynamics. To show the importance of two-stage learning, we also implemented an alternative block-structured numerical optimization approach to solve our formulation; in this approach, we fit a model with the same block structure as the IPSID model in Eq. 6 from *SI Appendix, Methods* but do so in a single stage by simultaneously maximizing the neural-behavioral data log-likelihood (*SI Appendix, Methods*). When applied to the same simulated data as in Fig. 2*C*, IPSID's two-stage approach was significantly more accurate than this single-stage block-structured numerical optimization in learning the intrinsic behaviorally relevant eigenvalues. Also, IPSID required orders of magnitude fewer training samples to achieve comparable accuracy (*SI Appendix*, Fig. S9*A*). Consistent with its more accurate intrinsic behaviorally relevant eigenvalues, IPSID also outperformed this single-stage method and INDM in terms of achieving higher behavior data likelihood (*SI Appendix*, Fig. S9*B*) while achieving comparable neural data likelihood (*SI Appendix*, Fig. S9*C*). These results highlight the benefit of two-stage (i.e., prioritized) learning of intrinsic behaviorally relevant dynamics over their single-stage learning (see also *SI Appendix, Methods*). Finally, IPSID was also significantly faster in model learning than the numerical optimization method, given that IPSID involves a fixed set of linear algebraic operations whereas numerical optimization involves iterative gradient descent (*SI Appendix*, Fig. S10).

Realistic Motor Task Simulations Show How Sensory Inputs Can Confound Models of Neural Activity. Sensory inputs to the brain such as task instructions can have different dynamics from task to task, even if the intrinsic neural dynamics remain unchanged (Fig. 1*A*). Developing a method that can learn the correct intrinsic neural dynamics regardless of the task would allow experimenters to study any behavioral task of interest or compare intrinsic dynamics across different tasks without worrying about confounding the results and without limiting the task design. We hypothesized that even when the intrinsic neural dynamics remain unchanged, *methods* that do not consider the task sensory inputs may learn different and incorrect intrinsic dynamics depending on the exact task, whereas IPSID can learn the same intrinsic dynamics regardless of the task. Here, we confirm this hypothesis by simulating a brain performing multiple different realistic cursor control tasks during which simulated neural data for model training is observed (Fig. 4 and *SI Appendix, Methods*).

Specifically, we modeled the brain as an optimal feedback controller (42–44) (OFC), which controls a part of its state that represents the 2D cursor kinematics toward targets presented via task instructions (*SI Appendix, Methods* and Fig. 4 *A* and *B*). For generality, as part of the simulated brain, we included two latent states (similar to $x_k^{(3)}$ in Eq. **2**) that are driven by input and affect the measured motor behavior but are not reflected in neural dynamics (*SI Appendix,*

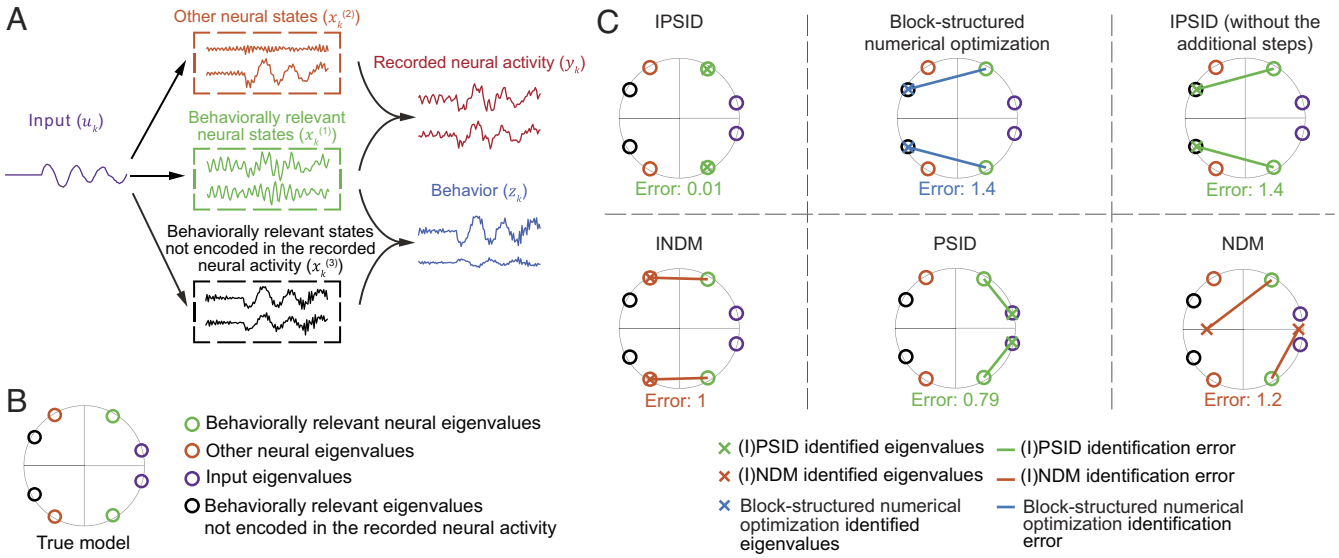

Fig. 3. IPSID also applies to scenarios where the recorded regions do not cover all downstream regions of the input. (*A*) A simulated brain (as in Eq. **2**) with a 6D latent state out of which only 4 dimensions drive the recorded neural activity and the other 2 dimensions just drive the behavior. (*B*) The eigenvalues of the state transition matrix *A* in the simulated model. The 4 eigenvalues associated with the 4 state dimensions that drive the recorded neural activity are shown as green and orange circles, depending on whether they drive behavior (green) or not (orange). Eigenvalues associated with the two additional state dimensions that only drive the behavior but not recorded neural activity are shown as black circles. (*C*) Eigenvalues of the models learned using IPSID, block-structured numerical optimization, IPSID (without additional steps), PSID, and (I)NDM. A simplified schematic of key operations for each method is in *SI Appendix*, Fig. S6. The block-structured numerical optimization learns the model parameters via gradient descent (*SI Appendix, Methods*). Notation is as in Fig. 1. IPSID can also address this scenario and its additional steps are needed to avoid the black eigenvalues/dynamics (behaviorally relevant dynamics not reflected in the recordings; see the top row comparisons).

Methods). As the first task, we simulated 8 equally spaced targets around a circle and instructed the simulated brain to move the cursor to the targets in order (Fig. 4 *C*, *Left*). As the second task, we simulated a standard center-out-and-back task where in each trial the cursor needs to move from the center to a randomly specified target among 8 targets and then return back to the center (Fig. 4 *C*, *Middle*). Last, we simulated a 10-by-10 grid of targets where in each trial a random target within a limited distance of the most recent target needs to be visited (Fig. 4 *C*, *Right*) similar to the tasks in the NHP datasets (*SI Appendix, Methods*). For each task, we used (I)PSID and (I)NDM to learn models of neural dynamics (Fig. 4*D*).

We found that regardless of the task, IPSID correctly learned the intrinsic behaviorally relevant neural dynamics. This is evident from comparing the IPSID eigenvalues and flow fields for every task with their ground truth (first row of Fig. 4*D* vs. Fig. 4*B*). INDM, which considers input but not behavior during training, learned an approximation of some intrinsic behaviorally relevant neural dynamics with error, and also mistakenly included some intrinsic neural dynamics that were not relevant to behavior (Fig. 4*D*, second row). PSID, which considers behavior and neural activity but not input during training, learned biased intrinsic neural dynamics that were influenced by task instruction inputs (Fig. 4*D*, third row). Finally, NDM, which only considers neural activity during training, not only learned neural dynamics that were not related to behavior but also learned inaccurate intrinsic behaviorally relevant neural dynamics that were influenced by task instruction inputs (Fig. 4*D*, fourth row). For example, in the first task, the biased dynamics learned by NDM and PSID were very close to the dominant frequency of the task instructions, which was around 0.2 Hz (Fig. 4*D*, left column). These results demonstrate that by considering both behavior and sensory inputs such as task instructions during model training, IPSID can learn models of neural dynamics that are not confounded by the specific behavioral task during which neural data are collected. Avoiding these confounds is critical for comparing intrinsic neural dynamics across tasks in neuroscience investigations, as we also show in our NHP data analyses below (Figs. 5–7).

Consistent with the above results, models trained by IPSID on data from one task had minimal drop in behavior decoding performance when tested on data from a different task, thus achieving generalization from task to task. In contrast, models learned by all other methods had significantly larger drops in behavior decoding performance in the other task (*SI Appendix*, Fig. S11; $P < 0.001$; one-sided signed-rank; $n = 10$ simulations).

Modeling Task Instructions as Inputs Reveals Distinct Intrinsic Behaviorally Relevant Neural Dynamics in Non-human Primate Neural Population Activity. We next used IPSID to study intrinsic behaviorally relevant neural dynamics in two independent motor cortical datasets recorded from three monkeys (monkeys I and L from the first datasets and monkey T from the second dataset) during two distinct behavioral motor tasks with planar cursor movements (Figs. 5*A* and 6*A*). In the first dataset, which was made publicly available by the Sabes lab (45), primary motor cortical (M1) population activity was recorded while two monkeys controlled a 2D cursor to reach random targets on a grid (Fig. 5*A* and *SI Appendix, Methods*). The 3D position of the monkey's fingertip was tracked and its horizontal elements controlled the cursor (*SI Appendix, Methods*). In the second dataset, which was made publicly available by the Miller lab (46, 47), population activity from the dorsal premotor cortex (PMd) was recorded while the monkey performed sequential reaches to random target positions on a plane (Fig. 6*A* and *SI Appendix, Methods*). The cursor was controlled via a manipulandum that only allowed horizontal movements. For all subjects, we modeled the smoothed spike counts (3, 13, 39, 48) as neural signals (*SI Appendix, Methods*). We took the 2D position and velocity of the cursor as the behavior signal, and the time series of target positions as the input task instructions (Figs. 5*A* and 6*A*).

First, we found that IPSID revealed distinct intrinsic behaviorally relevant neural dynamics/eigenvalues that were not found by other methods. This could be seen from the learned eigenvalues by IPSID that were different from those found by other methods (Figs. 5*B* and 6*B* and *SI Appendix*, Fig. S12*B*). Second, eigenvalues found by

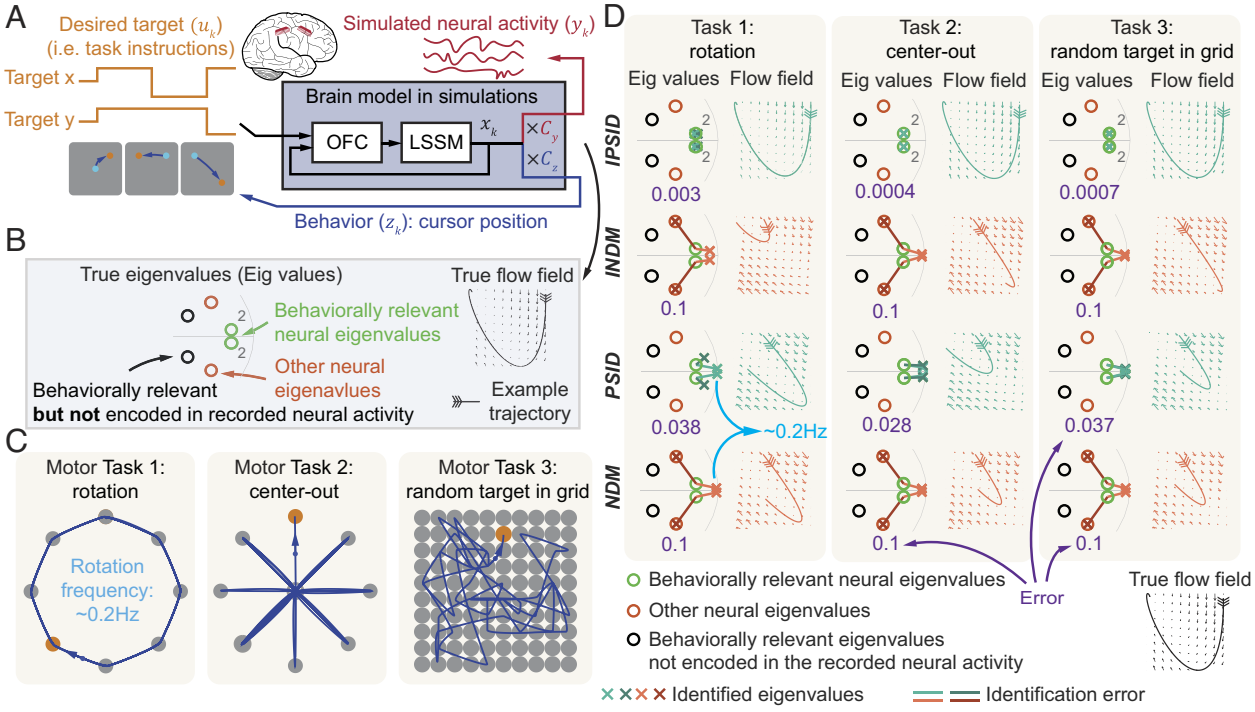

Fig. 4. By considering task instruction inputs, IPSID learns the correct intrinsic behaviorally relevant neural dynamics regardless of the task unlike other methods. (*A*) The brain model consists of an optimal feedback controller (OFC) combined with a linear state space model (LSSM). Four of the 8 latent state dimensions of the LSSM encode the 2D position and velocity of the cursor (*SI Appendix, Methods*). OFC controls these four state dimensions such that cursor position reaches the target shown on the screen while cursor velocity goes to zero (i.e., cursor stops at the target). (*B*) Eigenvalues of the state transition matrix in the full brain model (i.e., OFC together with the LSSM) and the flow field associated with the behaviorally relevant neural eigenvalues. Flow fields show the direction in which the state would change starting from various initial values. In this brain model, there are two sets of behaviorally relevant complex conjugate eigenvalues that are at the same location and thus overlapping. Each set is associated with one movement direction, horizontal and vertical, respectively, per Eq. 13, *SI Appendix, Methods*. The fact that there are two overlapping sets of eigenvalues is indicated by writing a 2 next to these eigenvalues. Since horizontal and vertical directions have identical dynamics, the flow field is only shown for one of them. In addition to the four states representing position and velocity in the 2D space, there are two states that only drive the neural activity, whose associated eigenvalues are depicted as orange circles. There are also two states that only drive the behavior, whose associated eigenvalues are depicted as black circles. (*C*) Tasks performed by the simulated brain. (*D*) Identified eigenvalues for each task using each method with a state dimension of 4. The flow field for one of the two sets of eigenvalues identified by each method (the one with the lighter green/red color) is also shown as an example. Only IPSID correctly learns the intrinsic behaviorally relevant neural eigenvalues regardless of the behavioral task used during training.

PSID were far from those found by IPSID, whereas eigenvalues found by NDM were close to those found by INDM (Figs. 5*B* and 6*B* and *SI Appendix*, Fig. S12*B*). Note that IPSID/PSID focus on explaining the behaviorally relevant neural dynamics whereas INDM/NDM focus on explaining the overall neural dynamics regardless of relevance to behavior. Thus, the aforementioned result suggests that task instructions, which are taken as inputs in IPSID/INDM, are highly informative of behaviorally relevant neural dynamics (seen from their effect on PSID vs. IPSID), but are not very informative of the overall neural dynamics (seen from NDM and INDM results being similar). This is consistent with the vast body of work suggesting that neural dynamics relevant to any specific behavior may constitute a minority of the overall neural variance (5, 6, 19, 32–39).

In these analyses, we used the additional steps in IPSID that were designed for scenarios in which some input dynamics may affect behavior through unrecorded regions/pathways (*SI Appendix*, Fig. S5). However, we found that even without these additional steps, the average learned eigenvalues remained almost unchanged in one subject (*SI Appendix*, Fig. S13*B*) and remained relatively similar in the other two subjects (*SI Appendix*, Fig. S13 *A* and *C*). This result could suggest, particularly in the former (*SI Appendix*, Fig. S13*B*), that behaviorally relevant neural dynamics that were downstream of visual task instruction inputs were largely reflected in, or downstream of, the motor cortical recordings here. Having established the distinction of eigenvalues found by IPSID, we next explored whether these eigenvalues better describe the data.

IPSID Learns more Accurate Intrinsic Behaviorally Relevant Neural Dynamics in Non-human Primate Neural Population Activity. We hypothesized that as in simulations (Fig. 4), the eigenvalues learned by IPSID are more accurate descriptions of the true intrinsic behaviorally relevant neural dynamics. As a measure of closeness of two sets of dynamics, we computed the Kullback–Leibler (KL) divergence between the distribution of their associated eigenvalues (*SI Appendix, Methods*). We performed multiple evaluations to test this hypothesis.

First, we showed that IPSID's algebraic operations can mitigate the problem of learning intrinsic dynamics that are confounded by input dynamics, unlike NDM and PSID. We characterized the input dynamics by modeling the time series of task instructions as a linear state-space model and finding the associated input eigenvalues (Eq. 3 in *SI Appendix, Methods*). We found that in all three subjects and in the two tasks, the input eigenvalues were close to those learned using NDM and PSID but not to those learned using IPSID (Figs. 5*B* and 6*B* and *SI Appendix*, Fig. S12*B*). Also, in all subjects, the KL-divergence between the input dynamics and learned dynamics was much larger for IPSID compared with PSID, which does not consider inputs during learning (Figs. 5*C* and 6*C* and *SI Appendix*, Fig. S12*C*).

Second, we demonstrated the success of preferential neural-behavioral modeling in the presence of input enabled by IPSID by comparing with INDM and NDM. In all three subjects, IPSID learned the intrinsic behaviorally relevant neural dynamics significantly more accurately than both INDM and NDM (Figs. 5*D*

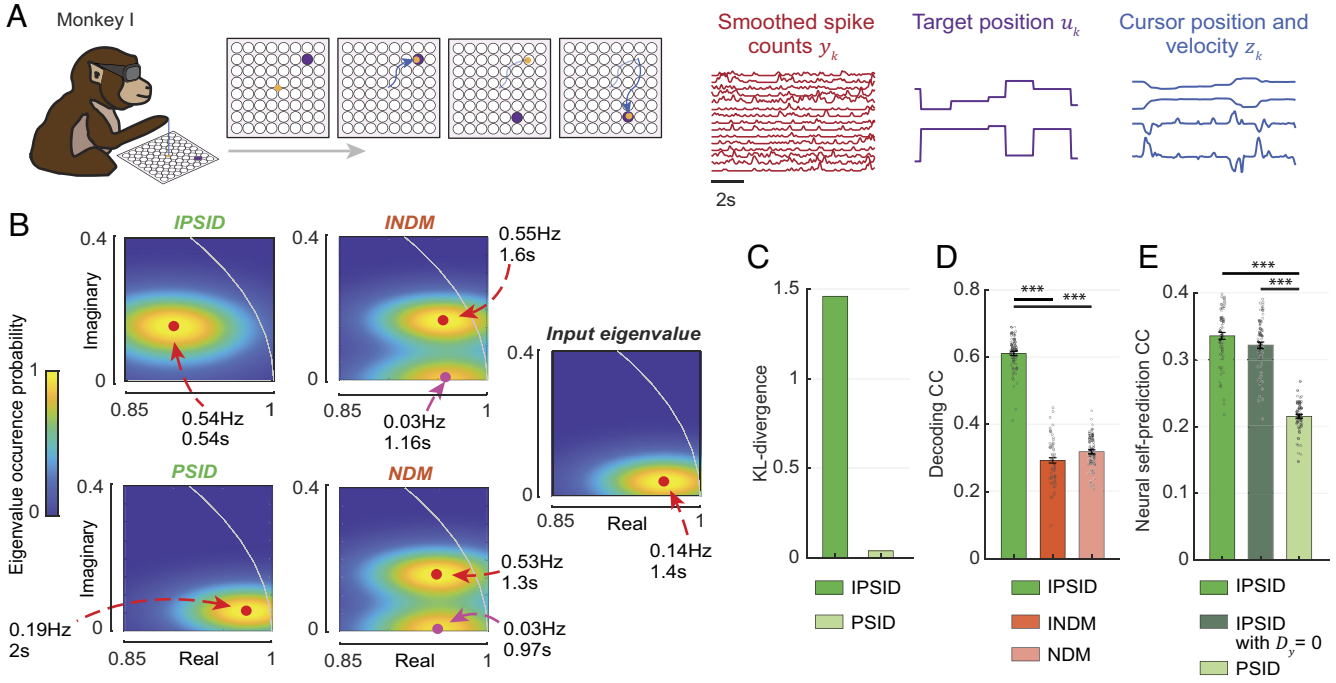

Fig. 5. IPSID uncovers distinct and more accurate intrinsic behaviorally relevant neural dynamics in motor cortical population activity by considering task instructions as inputs to the brain. (*A*) We modeled the population spiking activity in a monkey (monkey I) performing a 2D cursor control task (*SI Appendix, Methods*). See *SI Appendix*, Fig. S12 for results from a second monkey in this task and Fig. 6 for results in a second dataset recorded from a different monkey in a different task. Spike counts are smoothed using a Gaussian kernel with SD of 50 ms (*SI Appendix, Methods*). The 2D position and velocity of the cursor were taken as the behavior signal and the instructed target position time series was taken as the input to the brain. (*B*) Distribution of the eigenvalues of the state transition matrix for models learned using (I)PSID and (I)NDM across datasets. Input eigenvalue was found by applying NDM to the time-series of instructed targets. Models were learned with a latent state dimension of $n_x = 4$, which is sufficient for capturing most behavior dynamics (*SI Appendix*, Fig. S14). We estimated the probability of an eigenvalue occurring at each location on the complex plane by adding Gaussian kernels centered at locations of all identified eigenvalues ($n = 70$ cross-validation folds across two channel subsets and seven recording sessions, *SI Appendix, Methods*). Red dots indicate the location that has the maximum estimated eigenvalue occurrence probability, with the associated frequency and decay rate (*SI Appendix, Methods*) noted. When the occurrence probability map has more than one local maximum (i.e., for NDM or INDM), pink dots indicate the location of the second local maximum. (*C*) KL-divergence between the probability mass function of input eigenvalues (panel B, *Right*) and that of eigenvalues learned by IPSID/PSID (panel *B*, *Top* and *Bottom Left*). The eigenvalues learned by PSID were much closer to input eigenvalues than the eigenvalues learned by IPSID, showing the success of IPSID's distinct algebraic operations in accounting for inputs in neural-behavioral modeling. (*D* and *E*) Cross-validated behavior decoding (panel *D*) and neural self-prediction (panel *E*) when modeling data with dimension $n_x = 4$ and corresponding to models in *B*. Triple asterisks indicate $P < 0.0005$ for a one-sided signed-rank test.

and 6*D* and *SI Appendix*, Fig. S12*D*). This was evident from comparing the cross-validated behavior decoding from neural activity for these methods (Figs. 5*D* and 6*D* and *SI Appendix*, Fig. S12*D*).

Third, we showed the success of IPSID's algebraic operations in accounting for inputs in preferential neural-behavioral modeling by comparing it to PSID, which is preferential yet does not consider inputs. We found that IPSID learned models that were significantly more predictive of neural dynamics compared to PSID in all three subjects, as evident by comparing the cross-validated neural self-prediction accuracy across the two methods (Figs. 5*E* and 6*E* and *SI Appendix*, Fig. S12*E*). These results held even if the feedthrough term $D_y u_k$ in Eq. **2**—which reflects the effect of input on neural activity directly and not through the latent states x_k—was discarded when predicting neural activity using IPSID (Figs. 5*E* and 6*E* and *SI Appendix*, Fig. S12*E*). This analysis demonstrates that the better prediction in IPSID is due to its latent states being more predictive of neural dynamics rather than due to a static feedthrough effect of input on neural dynamics.

Overall, these consistent results from three NHPs in two independent neural datasets with two different tasks suggest that IPSID can successfully dissociate intrinsic behaviorally relevant neural dynamics from other intrinsic neural dynamics and from measured input dynamics. Moreover, these results demonstrate that not considering task instruction sensory inputs when modeling neural activity can result in less accurate models of neural dynamics and confound conclusions about intrinsic dynamics, a problem that IPSID addresses (see also next section).

IPSID Uniquely Revealed that Intrinsic Behaviorally Relevant Dynamics Were Similar across the Different Subjects and Tasks. While the specific task instructions are different in the two behavioral tasks in the independent datasets here—reaches to random targets on a grid vs. sequential reaches to random targets—the two datasets also have similarities; they both have recordings from the motor cortical areas and involve cursor control tasks with targets on a 2D plane. We thus hypothesized that there may be similarities in the intrinsic behaviorally relevant neural dynamics across the two tasks and three subjects. To test this hypothesis, we compared the distribution of eigenvalues learned using IPSID across all pairs of the three subjects (Fig. 7) and quantified their average difference with three metrics: 1) symmetric KL divergence between eigenvalue distributions (*SI Appendix, Methods* and Fig. 7*D*), 2) correlation coefficient (CC) between the probability mass functions of the eigenvalue distributions (Fig. 7*E*) 3) distance between the modes of the eigenvalue distributions, i.e., most probable locations (Fig. 7*F*).

We found that IPSID identified intrinsic behaviorally relevant dynamics that were strikingly similar across the two tasks and three subjects both qualitatively (Fig. 7 *A*–*C*) and quantitatively (Fig. 7 *D*–*F*). This similarity was despite the fact that the task instruction sensory inputs were distinct between the two tasks and that these

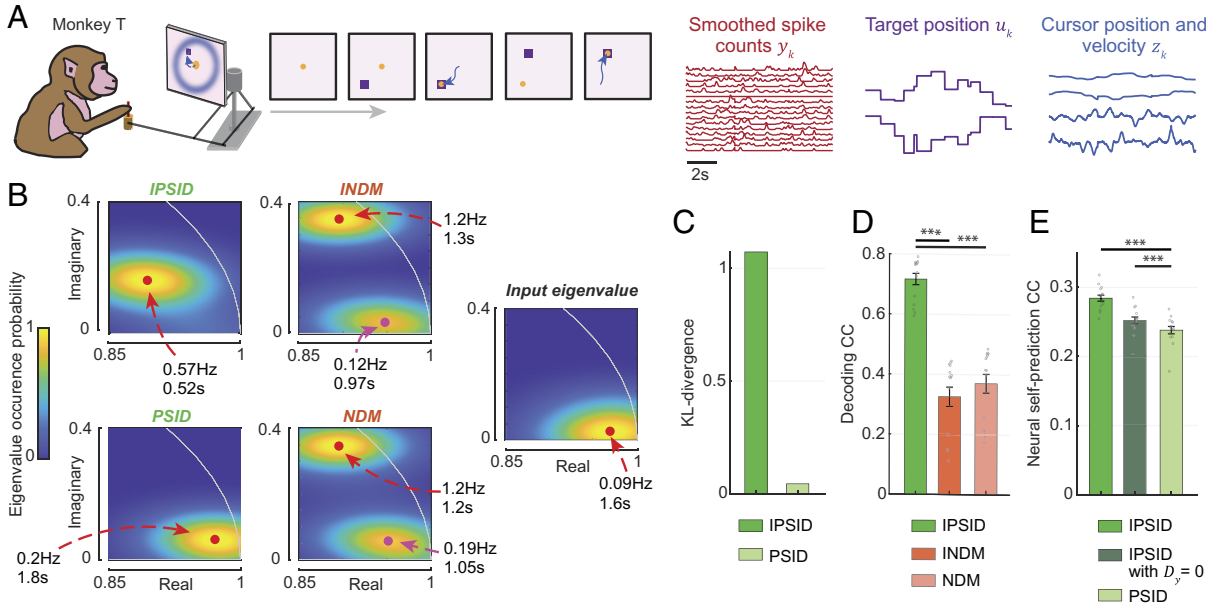

Fig. 6. In a second dataset recorded from a different monkey and during a different task, IPSID again uncovers distinct and more accurate intrinsic behaviorally relevant neural dynamics by considering task instructions as inputs. (*A–E*) Similar to Fig. 5, for the second subject (monkey T, $n = 15$ cross-validation folds across three recording sessions, *SI Appendix, Methods*) during a different second task with sequential reaches to random targets (*SI Appendix, Methods*).

recordings were from three different animals across two independent datasets. Also, even without its additional steps (*SI Appendix*, Fig. S5 and Note S2), IPSID still found largely similar eigenvalues across tasks and monkeys showing the robustness of this result, but the additional steps helped it reveal this similarity slightly more strongly (*SI Appendix*, Fig. S13 *D–F*).

We next studied the dynamics found by INDM. INDM aims to learn the overall intrinsic neural dynamics while IPSID aims to prioritize the learning of intrinsic behaviorally relevant neural dynamics. Interestingly, unlike IPSID, the dynamics found by INDM were much more distinct across the three monkeys both visually (Fig. 7 *A–C*) and quantitatively (Fig. 7 *D–F*). Moreover, as shown in the previous section, the more similar dynamics found by IPSID were also a more accurate description of intrinsic behaviorally relevant neural dynamics in each monkey (Figs. 5*D* and 6*D* and *SI Appendix*, Fig. S12*D*). Together, these results suggest that while the overall intrinsic neural dynamics (as found by INDM) were different across these two planar motor tasks and three animals, the intrinsic behaviorally relevant neural dynamics were similar as revealed by IPSID. We propose that this similarity may suggest that similar neural computations in the motor cortex underlie these planar cursor control tasks despite the differences between task instruction inputs and animals.

IPSID was the only method that revealed the above similarity of dynamics because it not only accounts for inputs (task instructions) but also prioritizes the learning of intrinsic behaviorally relevant dynamics over other neural dynamics in the presence of input—which is something INDM cannot do. Interestingly, this result is also consistent with our simulation study in Fig. 4 in which IPSID was the only method that correctly found the fixed intrinsic behaviorally relevant dynamics regardless of task while other methods were confused by the task instructions and/or overall intrinsic dynamics. Thus, IPSID can help researchers compare the intrinsic neural dynamics across different behavioral tasks by mitigating the confound that similarity or lack thereof in dynamics may simply be due to task instruction/input comparisons across tasks.

Together, these results highlight that the algebraic operations in IPSID can lead both to more accurate models and to useful scientific insight. These results also demonstrate that while measuring all inputs to a given brain region is typically experimentally infeasible, even incorporating partial input measurements (task instruction sensory inputs in this case) can already yield insights into neural computations across different tasks and subjects that may otherwise be missed.

Discussion

We developed IPSID, a method that introduces the capability to perform preferential dynamical modeling of neural-behavioral data in the presence of measured inputs. In the IPSID formulation, a dynamical model of neural activity is learned by accounting for measured input, neural, and behavioral data simultaneously, and the learning of intrinsic behaviorally relevant neural dynamics is prioritized over other intrinsic dynamics. By doing so, IPSID can dissociate intrinsic behaviorally relevant dynamics not only from other intrinsic dynamics but also from the dynamics of measured inputs such as task instructions or recorded activity of upstream regions. We demonstrated that without IPSID, dynamics in measured inputs to a given brain region or other intrinsic neural dynamics may be incorrectly identified as intrinsic behaviorally relevant neural dynamics within that brain region and thus confound conclusions. Indeed, in the neural data from monkeys, we showed that task instructions can act as such confounding inputs. IPSID can analytically account for such measured inputs to reveal more accurate intrinsic behaviorally relevant neural dynamics compared with alternative approaches even when they considered input (as in INDM). IPSID also provided useful scientific insights by revealing the similarity of intrinsic neural dynamics of behavior across different tasks and animals, which was not found by other methods.

IPSID could allow future studies to more easily compare across tasks without worrying about the temporal structure of task instruction inputs and how their reflection in neural activity may be misinterpreted as intrinsic neural dynamics. We first showed this potential with experiments where a simulated brain with fixed intrinsic dynamics performed different cursor control tasks. Here,

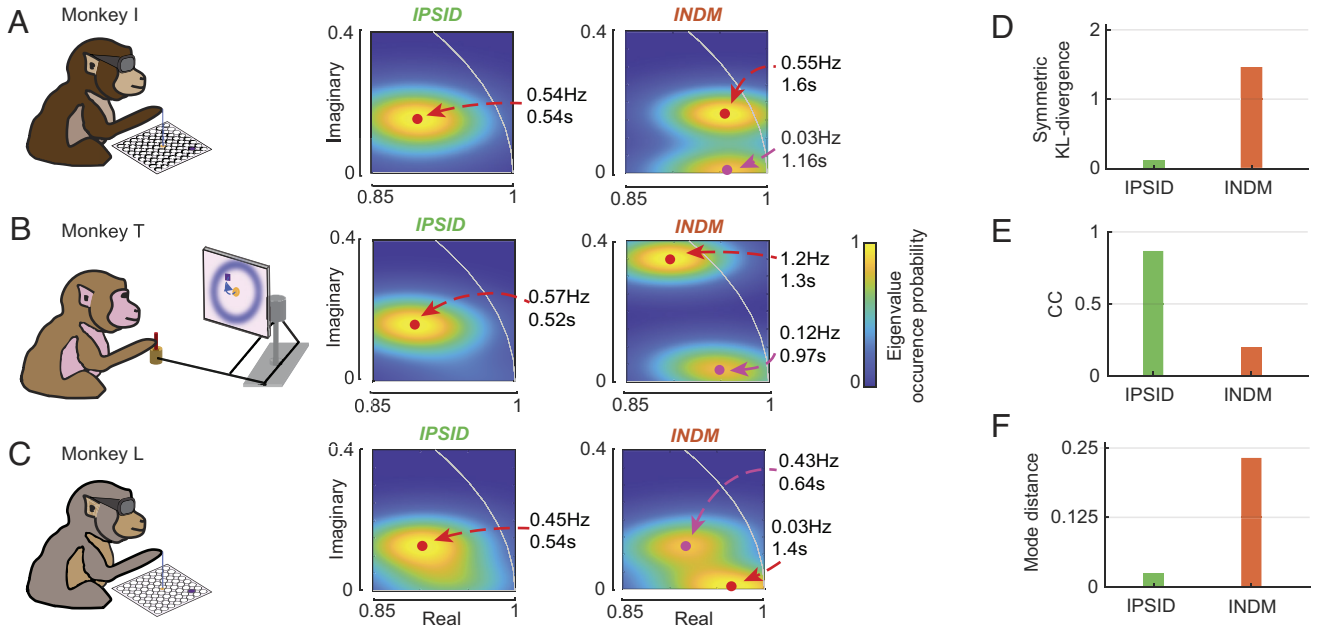

Fig. 7. IPSID reveals largely similar intrinsic behaviorally relevant neural dynamics across three monkeys and two tasks from two independent datasets while INDM identifies different overall intrinsic neural dynamics. (*A*) Same as Fig. 5*B*, showing the eigenvalues learned for IPSID and INDM. (*B* and *C*) Similar to *A* for the second and third monkeys, respectively (taken form Fig. 6 and *SI Appendix*, Fig. S12). (*D–F*) Average pairwise symmetric KL-divergence between the eigenvalue probability mass functions of the three monkeys (*D*), average pairwise Pearson correlation coefficient (CC) between these probability mass functions (*E*), and average pairwise distance between the modes (i.e., most probable eigenvalue location) of these probability mass functions (*F*). Lower KL-divergence/mode distance implies more similarity across monkeys, with a minimum possible value of 0. Higher CC implies more similarity across monkeys, with a maximum possible value of 1. Based on all three metrics, IPSID finds largely similar eigenvalues across tasks and animals whereas INDM finds eigenvalues that are different across tasks and animals.

sensory inputs in the form of task instructions could lead to learning intrinsic dynamics that incorrectly appeared different across tasks. IPSID addressed this issue and was the only approach that found the correct intrinsic behaviorally relevant neural dynamics regardless of the task. Consistently, in the real motor cortical datasets and by modeling the task instructions as sensory inputs, IPSID not only learned the intrinsic behaviorally relevant neural dynamics more accurately but also was the only method that revealed their similarity across tasks and animals.

Unexpectedly, despite differences in animals and in motor tasks/instructions across the motor cortical datasets, we found similar intrinsic behaviorally relevant dynamics in all three animals across both tasks using IPSID. In contrast, INDM found that the dominant overall intrinsic dynamics were different across tasks and animals. This result may suggest that motor cortical regions across different animals could have different intrinsic dynamics overall, but the part of their intrinsic dynamics that is engaged in arm movements to control 2D planar cursors may have similarity. These similar dynamics may suggest that similar intrinsic neural computations in the motor cortex underlie the performance of these two different planar cursor-reaching tasks. Prior work has found similarities in static projections of neural activity (49, 50) across subjects (50) or tasks (49), but these prior works have not modeled temporal dynamics (e.g., eigenvalues) and have not disentangled the effect of task instruction input dynamics on the observed similarity. Thus, IPSID provides a useful tool to explore whether such observed similarities reflect input dynamics or are intrinsic.

When the activity of some upstream brain regions that have inputs to the recorded region (27, 31, 51–53) is not measured, the learned intrinsic dynamics could also partly originate from these other regions. In the motor cortical datasets here for example, neural dynamics in upstream regions such as the visual cortex—which is involved in processing the sensory input and passing it to other regions along the visual-motor pathway—may also be reflected in the learned intrinsic motor cortical dynamics. Taking the sensory instructions as input can, to some extent, account for the dynamics of inputs from these upstream visual areas. Similarly, a sensory input that is not measured or accounted for, for example, the sunrise-sunset cycles during chronic recordings, may confound the modeled neural dynamics of a specific behavioral or mental state such as mood (e.g., in the form of circadian rhythms) (54, 55). Thus, recording activity from more upstream regions and measuring more sensory inputs can allow IPSID to consider more comprehensive inputs during modeling to better discover intrinsic behaviorally relevant dynamics.

As it is mostly experimentally infeasible to identify and record all inputs to a given brain region, a complete disentanglement of intrinsic dynamics from all input dynamics to a region becomes impractical. This experimental limitation is thus a fundamental limit on methodological efforts aimed at disentanglement. Thus, one still needs to interpret the results cautiously by noting that only dynamics of measured inputs are being disentangled from intrinsic dynamics. Nevertheless, our results show that even this partial disentanglement can lead to more accurate models and to useful scientific insights compared to alternative models which either do not consider measured inputs, or consider measured input but not behavior during learning.

Here, we address the challenge of preferential modeling of neural-behavioral data with measured inputs, which has been unresolved. For non-preferential modeling of neural data on its own and when inputs are not measured, prior studies have looked at the distinct problem of separating the recorded neural dynamics into intrinsic dynamics and a dynamic input that is inferred (12, 56, 57). This decomposition is typically done by making certain a priori assumptions about the input such that inputs can be inferred, for example, that input is constrained to be considerably less dynamic than intrinsic neural dynamics, or that input is sparse or spatiotemporally independent (12, 56). In addition to preferential neural-behavioral modeling with measured inputs, which is addressed here, future work can extend preferential modeling to also incorporate similar input

inference approaches, which could be complementary to IPSID. For example, such input inference approaches can help further interpret the intrinsic behaviorally relevant dynamics extracted by IPSID and hypothesize which parts of them could be due to unmeasured inputs. The results from such input inference efforts can depend on the a priori assumptions made regarding the input, since mathematically both extremes are plausible when inputs are not measured: All neural dynamics could be due to input from another area or they could all be intrinsic. For this reason, validating the inferred inputs from these inference approaches against actually measured inputs is an important step (12, 53, 56, 57). Such validation is also important because the underlying dynamics and inputs can have potential nonlinearities, thus making the inference of unmeasured inputs challenging or infeasible due to the potential unidentifiability in nonlinear systems (58).

One main contribution here is to formulate and highlight the problem of how intrinsic neural dynamics underlying a specific behavior can be confounded by both input dynamics and other intrinsic neural dynamics. We formulated this disentanglement problem that simultaneously involves measured input, neural, and behavioral data during learning and derived IPSID as an analytical solution based on subspace identification. By comparing with INDM and implementing a block-structured numerical optimization approach (Fig. 3 and *SI Appendix*, Fig. S9), we showed that two capabilities in IPSID are critical for disentanglement: first, prioritized learning of intrinsic behaviorally relevant dynamics via the two-stage learning operations with inputs; second, dissociating those behavior dynamics that are due to input but not reflected in the neural recordings from those that are, via the additional analytical steps (*SI Appendix*, Figs. S1 and S5). Prior works have proposed enforcing block-structure on linear dynamic models and developed Expectation–Maximization algorithms for fitting them (59, 60). But these studies have distinct goals and thus do not address the input disentanglement problem, or the behaviorally relevant dissociation problem addressed here. As such, they also do not enable the above two capabilities enabled by IPSID that are important for solving these problems. Future work can utilize the ideas developed here for enabling the IPSID capabilities in order to develop alternative numerical optimization solutions to the formulated disentanglement problem.

In addition to sensory inputs or activity in other brain regions, the input could also be any external electrical or optogenetic brain stimulation, for example in a brain-machine interface (BMI). Developing closed-loop stimulation treatments for mental disorders such as depression (61, 62) hinges on building dynamic models of neural activity that satisfy two criteria: i) describe how mental states are encoded in neural activity (61, 62); ii) describe the effect of electrical stimulation on the neural activity (28, 62, 63). The approach developed here enables learning of models that satisfy both criteria. First, by prioritizing behaviorally relevant dynamics, models could accurately learn the neural dynamics relevant to behavioral measurements of mental states [e.g., mood reports in depression (61)]. Further, this prioritization enables the learned models to have lower-dimensional latent states, which is important in developing robust controllers (64). Second, the models could explicitly learn the effect of external electrical stimulation parameters on neural activity (28, 63).

Here, we used continuous-valued variables with Gaussian distributions to model neural activity, as has been done extensively in prior works modeling local field potentials (LFP) (14, 16, 19, 30, 44, 61, 65, 66) and spike counts (7, 19, 67, 68). However, recent works suggest that modeling spike counts as Poisson distributed variables (8, 12, 16, 69–72) can improve BMI performance (70, 71). Thus, an interesting direction is to extend the method to support Poisson distributed neural observations or support simultaneous Gaussian and Poisson neural observations for multiscale modeling of neural modalities such as LFP and spikes together (16, 44, 65, 73–75). We also focused on learning linear dynamical models given their interpretability for neuroscience investigations (e.g., eigenvalue analyses in Figs. 5–7), as well as their computational efficiency and their tractability for real-time and/or closed-loop control systems applications such as BMIs (7, 28, 62, 63, 67, 70, 71, 76, 77). Further, linear dynamical models could approximate neural dynamics well given enough latent state dimensions (14, 20, 78). Nevertheless, capturing nonlinearities in models of intrinsic dynamics is another interesting future direction, which may be facilitated by incorporating a two-stage learning approach similar to that of IPSID into a numerical optimization learning framework. Moreover, similar to nonlinear dynamical models, linear dynamical models with input can have multiple fixed points because the fixed point can change with input. Thus, it would be interesting to investigate whether neural dynamics that can be explained by multiple fixed points can be alternatively explained with linear dynamical models with measured input for example from other brain regions or whether nonlinear models are essential for explaining these dynamics even with input. Finally, developing adaptive extensions that update the dynamical latent state model to adapt to non-stationarities in neural signals or to stimulation-induced plasticity (43, 79–82) will be important for BMIs and for studying learning and plasticity and their effect on intrinsic behaviorally relevant dynamics.

In conclusion, we develop an analytical method for preferential dynamical modeling of neural-behavioral data that can account for measured inputs—whether sensory input, neural input from other regions, or external stimulation. We show the importance of doing so for correct interpretation and modeling of neural computations/dynamics that underlie behavior and for gaining useful scientific insights about them across different tasks and subjects. These results and the developed preferential modeling approach have important implications for future neuroscientific and neuroengineering studies.

Data, Materials, and Software Availability. Datasets used in this work are publicly available online (45–47). The code for IPSID is available online at https://github.com/ShanechiLab/PSID (Matlab) (83) and https://github.com/ShanechiLab/PyPSID (Python) (84).

ACKNOWLEDGMENTS. This work was partly supported by NIH R01MH123770, NIH DP2MH126378, and NSF CRCNS Award IIS 2113271 (M.M.S.), and USC Annenberg Fellowship (O.G.S.). We sincerely thank the Sabes lab at the University of California San Francisco and the Miller lab at Northwestern University for making the NHP datasets that we used here publicly available.

Author affiliations: [a]Ming Hsieh Department of Electrical and Computer Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089; [b]Neuroscience Graduate Program, University of Southern California, Los Angeles, CA 90089; and [c]Thomas Lord Department of Computer Science and Alfred E. Mann Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089

1. D. V. Buonomano, W. Maass, State-dependent computations: Spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* **10**, 113–125 (2009).
2. W. Wu, J. E. Kulkarni, N. G. Hatsopoulos, L. Paninski, Neural decoding of hand motion using a linear state-space model with hidden states. *IEEE Trans. Neural Syst. Rehabil. Eng.* **17**, 370–378 (2009).
3. B. M. Yu *et al.*, Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* **102**, 614–635 (2009).
4. J. H. Macke *et al.*, Empirical models of spiking in neuronal populations. *Adv. Neural Inf. Process. Syst. NIPS* **24**, 1–9 (2011).
5. M. M. Churchland *et al.*, Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).

6. K. V. Shenoy, M. Sahani, M. M. Churchland, Cortical control of arm movements: A dynamical systems perspective. *Annu. Rev. Neurosci.* **36**, 337–359 (2013).
7. J. C. Kao *et al.*, Single-trial dynamics of motor cortex and their applications to brain-machine interfaces. *Nat. Commun.* **6**, 7759 (2015).
8. M. Aghagolzadeh, W. Truccolo, Inference and decoding of motor cortex low-dimensional dynamics via latent state-space models. *IEEE Trans. Neural Syst. Rehabil. Eng. Soc.* **24**, 272–282 (2016).
9. J. S. Seely *et al.*, Tensor analysis reveals distinct population structure that parallels the different computational roles of areas M1 and V1. *PLoS Comput. Biol.* **12**, e1005164 (2016).
10. S. Linderman *et al.*, "Learning and inference in recurrent switching linear dynamical systems" in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (PMLR, 2017), vol. 54, pp. 914–922.
11. A. Wu, N. A. Roy, S. Keeley, J. W. Pillow, Gaussian process based nonlinear latent structure discovery in multivariate spike train data. *Adv. Neural Inf. Process Syst.* **30**, 3496–3505 (2017).
12. C. Pandarinath *et al.*, Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* **15**, 805–815 (2018).
13. A. H. Williams *et al.*, Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron* **98**, 1099–1115.e8 (2018).
14. Y. Yang, O. G. Sani, E. F. Chang, M. M. Shanechi, Dynamic network modeling and dimensionality reduction for human ECoG activity. *J. Neural Eng.* **16**, 056014 (2019).
15. S. Vyas, M. D. Golub, D. Sussillo, K. V. Shenoy, Computation through neural population dynamics. *Annu. Rev. Neurosci.* **43**, 249–275 (2020).
16. H. Abbaspourazad, M. Choudhury, Y. T. Wong, B. Pesaran, M. M. Shanechi, Multiscale low-dimensional motor cortical state dynamics predict naturalistic reach-and-grasp behavior. *Nat. Commun.* **12**, 607 (2021).
17. M. Jazayeri, S. Ostojic, Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Curr. Opin. Neurobiol.* **70**, 113–120 (2021).
18. K. V. Shenoy, J. C. Kao, Measurement, manipulation and modeling of brain-wide neural population dynamics. *Nat. Commun.* **12**, 633 (2021).
19. O. G. Sani, H. Abbaspourazad, Y. T. Wong, B. Pesaran, M. M. Shanechi, Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nat. Neurosci.* **24**, 140–149 (2021).
20. O. G. Sani, B. Pesaran, M. M. Shanechi, Where is all the nonlinearity: Flexible nonlinear modeling of behaviorally relevant neural dynamics using recurrent neural networks. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.09.03.458628 (Accessed 10 October 2023).
21. Y. Chen, B. Q. Rosen, T. J. Sejnowski, Dynamical differential covariance recovers directional network structure in multiscale neural systems. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117234119 (2022).
22. A. Dubreuil, A. Valente, M. Beiran, F. Mastrogiuseppe, S. Ostojic, The role of population structure in computations through neural dynamics. *Nat. Neurosci.* **25**, 783–794 (2022).
23. J. A. Michaels, B. Dann, H. Scherberger, Neural population dynamics during reaching are better explained by a dynamical system than representational tuning. *PLoS Comput. Biol.* **12**, e1005175 (2016).
24. E. D. Remington, S. W. Egger, D. Narain, J. Wang, M. Jazayeri, A dynamical systems perspective on flexible motor timing. *Trends Cogn. Sci.* **22**, 938–952 (2018).
25. G. F. Elsayed, J. P. Cunningham, Structure in neural population recordings: An expected byproduct of simpler phenomena? *Nat. Neurosci.* **20**, 1310–1318 (2017).
26. C. Pandarinath *et al.*, Neural population dynamics in human motor cortex during movements in people with ALS. *ELife* **4**, e07436 (2015).
27. B. A. Sauerbrei *et al.*, Cortical pattern generation during dexterous movement is input-driven. *Nature* **577**, 386–391 (2020).
28. Y. Yang *et al.*, Modelling and prediction of the dynamic responses of large-scale brain networks during direct electrical stimulation. *Nat. Biomed. Eng.* **5**, 324–345 (2021).
29. S. Ardid *et al.*, Biased competition in the absence of input bias revealed through corticostriatal computation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8564–8569 (2019).
30. D. Susilaradeya *et al.*, Extrinsic and intrinsic dynamics in movement intermittency. *ELife* **8**, e40145 (2019).
31. R. Chen *et al.*, Songbird ventral pallidum sends diverse performance error signals to dopaminergic midbrain. *Neuron* **103**, 266–276.e4 (2019).
32. J. Reimer, N. G. Hatsopoulos, *Prog. Mot. Control Multidiscip. Perspect.*, D. Sternad, Ed. (Springer US, 2009), pp. 243–259.
33. V. Mante, D. Sussillo, K. V. Shenoy, W. T. Newsome, Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
34. M. T. Kaufman *et al.*, The largest response component in the motor cortex reflects movement timing but not movement type. *eNeuro* **3**, ENEURO.0085-16.2016 (2016).
35. P. Ramkumar, B. Dekleva, S. Cooler, L. Miller, K. Kording, Premotor and motor cortices encode reward. *PLoS One* **11**, e0160851 (2016).
36. K. Svoboda, N. Li, Neural mechanisms of movement planning: Motor cortex and beyond. *Curr. Opin. Neurobiol.* **49**, 33–41 (2018).
37. W. E. Allen *et al.*, Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science* **364**, eaav3932 (2019).
38. C. Stringer *et al.*, Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, eaav7893 (2019).
39. D. Kobak *et al.*, Demixed principal component analysis of neural population data. *ELife* **5**, e10989 (2016).
40. P. Van Overschee, B. De Moor, *Subspace Identification for Linear Systems* (Springer, US, 1996).
41. O. G. Sani, Modeling and control of behaviorally relevant brain states, PhD Thesis, University of Southern California, Los Angeles, CA (2020).
42. E. Todorov, M. I. Jordan, Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* **5**, 1226–1235 (2002).
43. H.-L. Hsieh, M. M. Shanechi, Optimizing the learning rate for adaptive estimation of neural encoding models. *PLoS Comput. Biol.* **14**, e1006168 (2018).
44. H.-L. Hsieh, Y. T. Wong, B. Pesaran, M. M. Shanechi, Multiscale modeling and decoding algorithms for spike-field activity. *J. Neural Eng.* **16**, 016018 (2018).
45. J. E. O'Doherty, M. M. B. Cardoso, J. G. Makin, P. N. Sabes, Nonhuman primate reaching with multichannel sensorimotor cortex electrophysiology. Zenodo. https://doi.org/10.5281/zenodo.583331. Accessed 10 October 2023.
46. P. N. Lawlor, M. G. Perich, L. E. Miller, K. P. Kording, Linear-nonlinear-time-warp-poisson models of neural activity. *J. Comput. Neurosci.* **45**, 173–191 (2018).
47. M. G. Perich, P. N. Lawlor, K. P. Kording, L. E. Miller, Extracellular neural recordings from macaque primary and dorsal premotor motor cortex during a sequential reaching task. *CRCNS.org*, (2018), https://dx.doi.org/10.6080/K0FT8J72. Accessed 10 October 2023.
48. J. A. Gallego, M. G. Perich, R. H. Chowdhury, S. A. Solla, L. E. Miller, Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* **23**, 260–270 (2020).
49. J. A. Gallego *et al.*, Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nat. Commun.* **9**, 1–13 (2018).
50. M. Safaie *et al.*, Preserved neural dynamics across animals performing similar behaviour. *Nature* **623**, 765–771 (2023).
51. T. Mao *et al.*, Long-range neuronal circuits underlying the interaction between sensory and motor cortex. *Neuron* **72**, 111–123 (2011).
52. A. Nashef, R. Mitelman, R. Harel, M. Joshua, Y. Prut, Area-specific thalamocortical synchronization underlies the transition from motor planning to execution. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2012658118 (2021).
53. H. T. Kalidindi *et al.*, Rotational dynamics in motor cortex are consistent with a feedback controller. *ELife* **10**, e67256 (2021).
54. C. A. Vadnie, C. A. McClung, Circadian rhythm disturbances in mood disorders: Insights into the role of the suprachiasmatic nucleus. *Neural Plast.* **2017**, 1504507 (2017).
55. R. W. Logan, C. A. McClung, Rhythms of life: Circadian disruption and brain disorders across the lifespan. *Nat. Rev. Neurosci.* **20**, 49–65 (2019).
56. M. Schimel, T.-C. Kao, K. T. Jensen, G. Hennequin, "iLQR-VAE: Control-based learning of input-driven dynamics with applications to neural data" in *International Conference on Learning Representations* (2022). https://openreview.net/forum?id=wRODLDHaAiW. Accessed 10 October 2023.
57. M. R. Keshtkaran *et al.*, A large-scale neural network training framework for generalized estimation of single-trial population dynamics. *Nat. Methods* **19**, 1572–1577 (2022).
58. M. Grewal, K. Glover, Identifiability of linear and nonlinear dynamical systems. *IEEE Trans. Autom. Control* **21**, 833–837 (1976).
59. J. Glaser, M. Whiteway, J. P. Cunningham, L. Paninski, S. Linderman, "Recurrent switching dynamical systems models for multiple interacting neural populations" in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2020), vol. 33, pp. 14867–14878.
60. J. Semedo, A. Zandvakili, A. Kohn, C. K. Machens, B. M. Yu, "Extracting latent structure from multiple interacting neural populations" in *Advances in Neural Information Processing Systems* (Curran Associates Inc., 2014), vol. 27.
61. O. G. Sani *et al.*, Mood variations decoded from multi-site intracranial human brain activity. *Nat. Biotechnol.* **36**, 954 (2018).
62. M. M. Shanechi, Brain–machine interfaces from motor to mood. *Nat. Neurosci.* **22**, 1554–1564 (2019).
63. Y. Yang, A. T. Connolly, M. M. Shanechi, A control-theoretic system identification framework and a real-time closed-loop clinical simulation testbed for electrical brain stimulation. *J. Neural Eng.* **15**, 066007 (2018).
64. G. Obinata, B. D. O. Anderson, *Model Reduction for Control System Design* (Springer Science & Business Media, 2012).
65. H. Abbaspourazad, H. Hsieh, M. M. Shanechi, A multiscale dynamical modeling and identification framework for spike-field activity. *IEEE Trans. Neural Syst. Rehabil. Eng.* **27**, 1128–1138 (2019).
66. S. D. Stavisky, J. C. Kao, P. Nuyujukian, S. I. Ryu, K. V. Shenoy, A high performing brain-machine interface driven by low-frequency local field potentials alone and together with spikes. *J. Neural Eng.* **12**, 036009 (2015).
67. M. M. Shanechi, Brain-machine interface control algorithms. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 1725–1734 (2017).
68. J. C. Kao, S. D. Stavisky, D. Sussillo, P. Nuyujukian, K. V. Shenoy, Information systems opportunities in brain–machine interface decoders. *Proc. IEEE* **102**, 666–682 (2014).
69. A. C. Smith, E. N. Brown, Estimating a state-space model from point process observations. *Neural Comput.* **15**, 965–991 (2003).
70. M. M. Shanechi, A. L. Orsborn, J. M. Carmena, Robust brain–machine interface design using optimal feedback control modeling and adaptive point process filtering. *PLoS Comput. Biol.* **12**, 1–29 (2016).
71. M. M. Shanechi *et al.*, Rapid control and feedback rates enhance neuroprosthetic control. *Nat. Commun.* **8**, 13825 (2017).
72. N. Sadras, B. Pesaran, M. M. Shanechi, A point-process matched filter for event detection and decoding from population spike trains. *J. Neural Eng.* **16**, 066016 (2019).
73. R. Bighamian, Y. T. Wong, B. Pesaran, M. M. Shanechi, Sparse model-based estimation of functional dependence in high-dimensional field and spike multiscale networks. *J. Neural Eng.* **16**, 056022 (2019).
74. C. Wang, M. M. Shanechi, Estimating multiscale direct causality graphs in neural spike-field networks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **27**, 857–866 (2019).
75. C. Wang, B. Pesaran, M. M. Shanechi, Modeling multiscale causal interactions between spiking and field potential signals during behavior. *J. Neural Eng.* **19**, 026001 (2022).
76. A. D. Degenhart *et al.*, Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity. *Nat. Biomed. Eng.* **4**, 672–685 (2020).
77. V. Gilja *et al.*, A high-performance neural prosthesis enabled by control algorithm design. *Nat. Neurosci.* **15**, 1752–1757 (2012).
78. E. Nozari *et al.*, Macroscopic resting-state brain dynamics are best described by linear models. *Nat. Biomed. Eng.* **8**, 68–84 (2024). 10.1038/s41551-023-01117-y.
79. K. V. Shenoy, J. M. Carmena, Combining decoder design and neural adaptation in brain-machine interfaces. *Neuron* **84**, 665–680 (2014).
80. P. Ahmadipour, Y. Yang, E. F. Chang, M. M. Shanechi, Adaptive tracking of human ECoG network dynamics. *J. Neural Eng.* **18**, 016011 (2020).
81. Y. Yang, P. Ahmadipour, M. M. Shanechi, Adaptive latent state modeling of brain network dynamics with real-time learning rate optimization. *J. Neural Eng.* **18**, 036013 (2020).
82. Y. Yang *et al.*, Developing a personalized closed-loop controller of medically-induced coma in a rodent model. *J. Neural Eng.* **16**, 036022 (2019).
83. O. G. Sani, P. Vahidi, M. M. Shanechi, PSID: The Matlab library for (I)PSID. GitHub. https://github.com/ShanechiLab/PSID. Deposited 1 January 2024.
84. O. G. Sani, P. Vahidi, M. M. Shanechi, PyPSID: The Python library for (I)PSID. GitHub. https://github.com/ShanechiLab/PyPSID. Deposited 1 January 2024.