

# **Cortical Representation of Movement Across the Developmental Transition to Continuous Neural Activity**

**Ryan Glanz<sup>1</sup>, Greta Sokoloff<sup>1,2</sup>, and Mark S. Blumberg<sup>1,2,\*</sup>**

<sup>1</sup> Department of Psychological & Brain Sciences, University of Iowa, Iowa City, IA 52242 USA

<sup>2</sup> Iowa Neuroscience Institute, University of Iowa, Iowa City, IA 52242 USA

\* Correspondence: [mark-blumberg@uiowa.edu](mailto:mark-blumberg@uiowa.edu) (M.S.B.)

## Summary

Primary motor cortex (M1) exhibits a protracted period of development that includes the establishment of a somatosensory map long before motor outflow emerges. In rats, the sensory representation is established by postnatal day (P) 8 when cortical activity is still “discontinuous.” Here, we ask how the representation survives the sudden transition to noisy “continuous” activity at P12. Using neural decoding to predict forelimb movements based solely on M1 activity, we show that a linear decoder is sufficient to predict limb movements at P8, but not at P12; in contrast, a nonlinear decoder effectively predicts limb movements at P12. The change in decoder performance at P12 reflects an increase in both the complexity and uniqueness of kinematic information available in M1. We next show that the representation at P12 is more susceptible to the deleterious effects of “lesioning” inputs and to “transplanting” M1’s encoding scheme from one pup to another. We conclude that the emergence of continuous cortical activity signals the developmental onset in M1 of more complex, informationally sparse, and individualized sensory representations.

## INTRODUCTION

Perhaps the most striking feature of infant cortical activity is its discontinuity: Periods of silence are punctuated by bursts of population-level activity [1, 2]. This early phase of discontinuity ends with the sudden and dramatic onset of continuous cortical activity. Whereas in humans this transition occurs around the time of birth [3], in infant rats and mice it occurs at the end of the second postnatal week [4-8]. In fact, the onset of continuous cortical activity is but one aspect of a more global reorganization of brain dynamics that includes a shift in GABAergic functioning [9, 10], the increase and diversification of inhibitory interneurons [11-14], accelerated myelin deposition [15], and the onset of brain rhythms such as delta [16-18] and theta [19].

In primary motor cortex (M1), continuous activity emerges in rats between postnatal days (P) 8 and P12 [6]. Despite its name, M1 at these ages does not produce movement, but instead functions exclusively as a somatosensory structure several weeks before the emergence of motor outflow [20-22]. Recently, we showed that neurons within M1's nascent somatosensory map are tuned to limb kinematics, as is the case for M1's adult motor map [23, 24]. Specifically, we found precise sensory tuning to movement amplitude at P8—especially for the limb twitches that occur during active (REM) sleep [6]. However, upon the emergence of continuous activity at P12, this tuning disappeared.

Was M1's kinematic tuning truly lost, or was it only obscured within the noise of continuous activity? Here, we address this and related questions using a computational technique—*neural decoding*—that allows us to predict the timing and amplitude of forelimb movements based solely on M1 activity (see [25]). Our findings demonstrate that

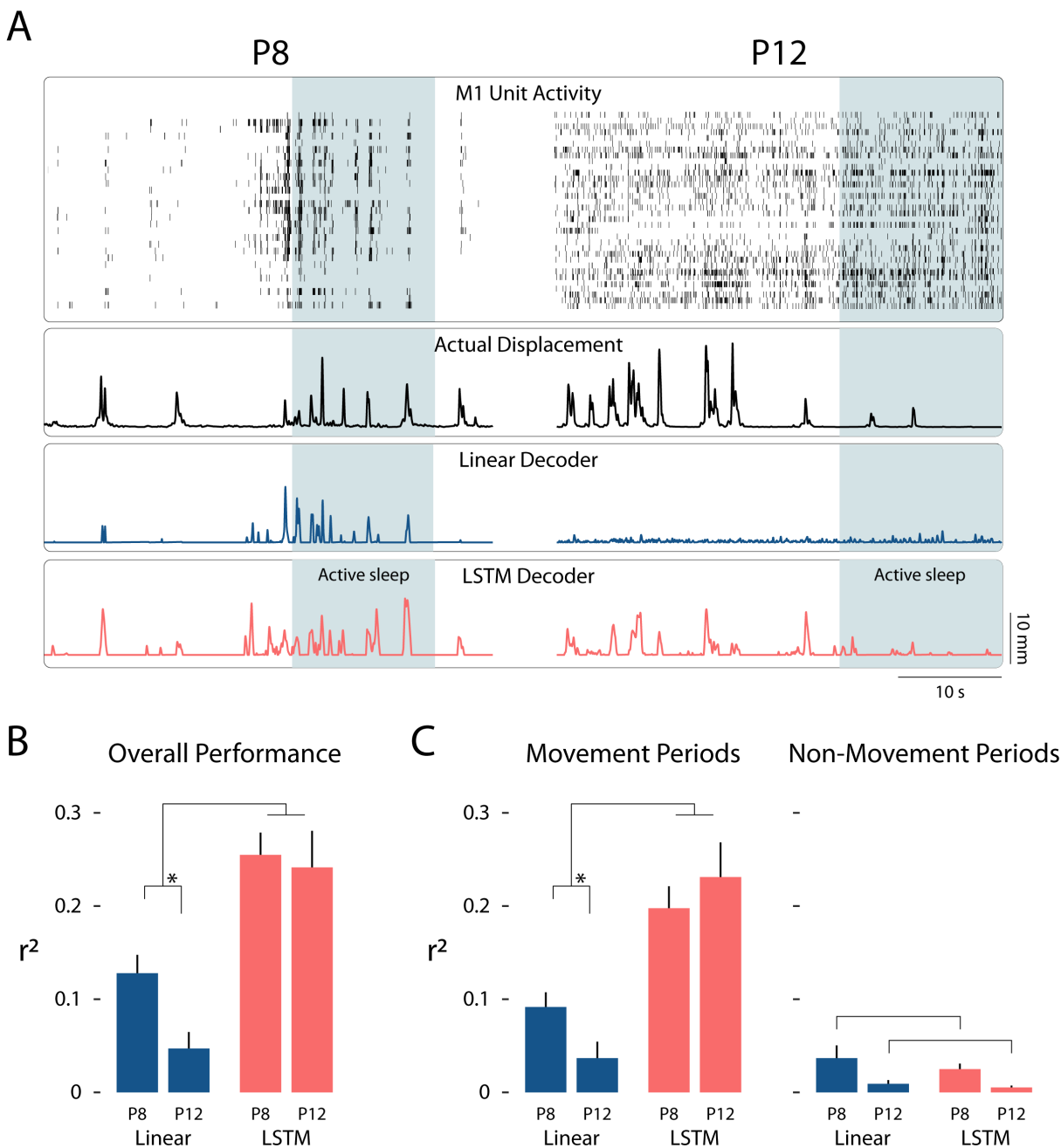
M1's processing of sensory information is intact but substantially transformed after this fundamental transition in cortical dynamics.

## RESULTS

We performed neural decoding using a previously published dataset from P8 and P12 rats ( $n = 8$  at each age; [6]). This dataset consists of one-hour recordings of M1 unit activity and video-based records of 3-D forelimb displacement. For all decoding procedures, M1 unit activity and forelimb displacement served as the predictor and target variables, respectively. Each recording was divided into training (36 min; 60%) and testing (9 min; 15%) datasets; a validation dataset (15 min; 25%) was held back until the final model parameters were established. All analyses were performed on the validation dataset to ensure an unbiased assessment of decoder performance [25].

### **Movement encoding is obscured by continuous activity at P12**

Discontinuous M1 activity presents a strong contrast between periods of movement (when reafference triggers neural activity) and rest (when there is little to no activity). Moreover, at P8, we previously found that forelimb twitches trigger rate-coded M1 responses that correlate with twitch amplitude [6]. Accordingly, we predicted that a linear decoder would accurately predict the temporal and spatial properties of forelimb movements at P8. In contrast, because continuous activity at P12 occludes the temporal and spatial relationships between M1 activity and forelimb movements, we predicted that a linear decoder would no longer predict forelimb movements at that age. Indeed, by comparing actual with predicted limb displacement using a linear decoder (**Figure 1A**, blue lines), we confirmed both predictions: A linear decoder is sufficient to predict forelimb movements at P8, but not at P12.



**Figure 1.** Neural decoding in M1 predicts forelimb movements across the transition to continuous cortical activity.

(A) Representative data from a P8 (left) and P12 (right) rat. From top to bottom: M1 unit activity where each row denotes an individual unit and each vertical tick denotes an action potential; actual forelimb displacement (black lines) in mm, representing the Euclidean distance traveled by the forelimb in 3-dimensional space; forelimb displacement as predicted by a linear decoder (blue lines); forelimb displacement as predicted by an LSTM nonlinear decoder (orange lines). Shaded blue regions represent periods of active sleep. (B) Mean (+SEM) overall performance of the linear (blue bars) and LSTM (orange bars) decoders for P8 and P12 rats, as measured by  $r^2$  (the proportion of variance in forelimb displacement explained by the decoder). Brackets denote a significant interaction between age and decoder ( $p < .05$ ). Asterisk denotes significantly better performance of the linear decoder at P8 compared to P12 ( $p < .05$ ). (C) Left: Same as in (B), but for periods of forelimb movement. Right: Same as in (B), but for non-movement periods; brackets denote that both decoders performed significantly better at P8 than at P12 ( $p < .05$ ).

Although we observed a loss of decoding accuracy at P12, M1 activity is nonetheless highly correlated with limb kinematics in adults [23, 24]. Accordingly, we hypothesized that M1 activity continues to represent movement kinematics after the emergence of continuous activity, but that this representation is too complex to be revealed using a linear decoder. We tested this hypothesis using a nonlinear decoder (long short-term memory decoder, LSTM; [26]) and found that it successfully decoded forelimb movements at P12, as well as P8 (**Figure 1A**, orange lines).

To quantify decoder performance, we computed the proportion of variance explained ( $r^2$ ) between the actual and predicted forelimb displacement for each decoder (**Figure 1B**). This metric represents the amount of temporal and spatial kinematic information available in M1 immediately after a forelimb movement. There was a significant interaction between pup age and the decoder used ( $F(1, 14) = 19.15$ ,  $p < .001$ , adj.  $\eta_p^2 = .548$ ; **Figure 1B**), indicating (1) that the linear decoder performed significantly better at P8 than P12 ( $F(1, 14) = 10.08$ ,  $p = .007$ , adj.  $\eta_p^2 = .376$ ), and (2) that the nonlinear decoder significantly outperformed the linear decoder at both ages. Thus, M1 activity indeed preserves kinematic information after continuous activity emerges at P12, though this relationship is only observable using a nonlinear decoder.

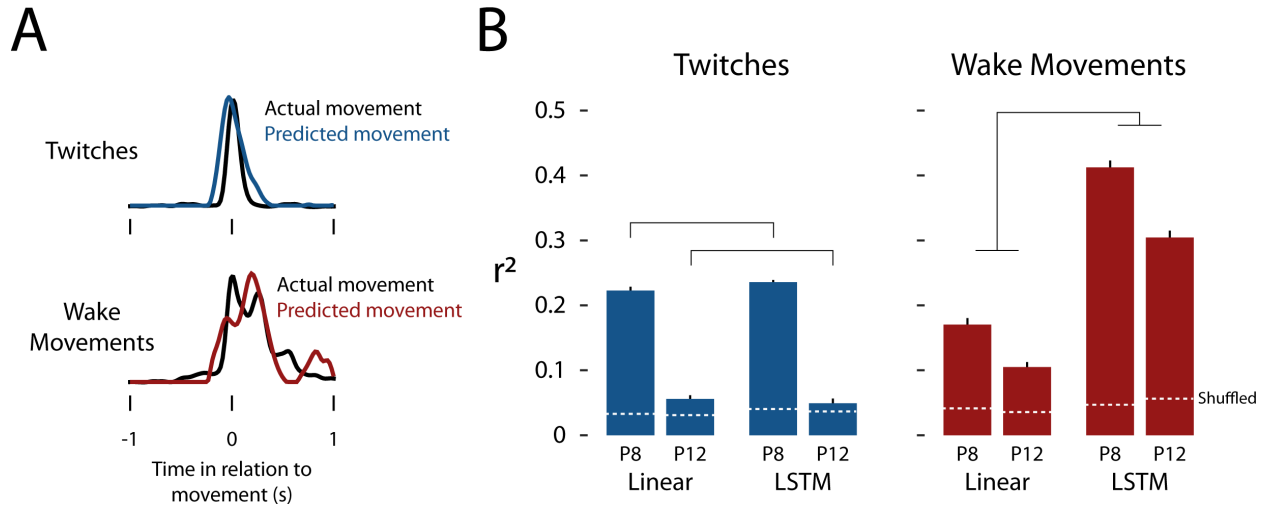
### **State-dependent encoding of movement-related information**

To confirm that decoder performance was driven by M1 activity specifically during periods of limb movement, the recordings were split into movement and non-movement periods, as described previously [6]. As expected, successful decoder performance was attributable to movement periods ( $F(1, 14) = 7.35$ ,  $p = .017$ , adj.  $\eta_p^2 = .297$ ; **Figure 1C**,

**left**) and not to non-movement periods ( $F(1, 14) = 0.06, p = .817, \text{adj. } \eta_p^2 < .001$ ; **Figure 1C, right**). Although there was a main effect of age on decoder performance during non-movement periods ( $F(1, 14) = 9.15, p = .009, \text{adj. } \eta_p^2 = .352$ ), this effect was negligible compared with the performance of the decoder during movement periods.

The last analysis did not distinguish between movements during sleep (i.e., twitches) and wake. Thus, we next compared decoding accuracy for twitches and wake movements, focusing on a 2-s window centered on each forelimb movement (**Figure 2A**). For twitches, a main effect of age indicated that decoding performance was significantly better at P8 than at P12, regardless of whether a linear or non-linear decoder was used ( $F(1, 3808) = 534.25, p < .001, \text{adj. } \eta_p^2 = .123$ ; **Figure 2B, left**). Unexpectedly, the nonlinear decoder's performance for twitches did not improve on the linear decoder's performance at either age ( $F(1, 3808) = 1.67, p = .197, \text{adj. } \eta_p^2 < .001$ ). This finding suggests that M1 activity no longer represents twitch kinematics after continuous activity emerges (although reafference from twitches continues to trigger M1 activity through at least P20; [27]).

In contrast with twitches, for wake movements the nonlinear decoder performed significantly better than the linear decoder at P8 and P12 ( $F(1, 1340) = 240.91, p < .001, \text{adj. } \eta_p^2 = .202$ ; **Figure 2B, right**). This finding was particularly surprising at P8 because reafference from wake movements is blunted at this age [20, 21]. Thus, wake movements appear to be represented in a nonlinear fashion even as early as P8, and this representation survives the transition to continuous activity at P12. For both decoders, there was a small but statistically significant decrease in performance at P12 compared with P8 ( $F(1, 1340) = 18.24, p < .001, \text{adj. } \eta_p^2 = .012$ ), suggesting that the emergence of



**Figure 2.** Decoding performance for twitches and wake movements varies by age.

(A) Forelimb displacement for actual movements (black lines) compared to a predicted twitch (blue line, top) and a predicted wake movement (red line, bottom). A single representative twitch and wake movement is shown. (B) Mean (+SEM) decoder performance ( $r^2$ ) for twitches (left; blue bars) and wake movements (right; red bars). For each type of movement, the linear decoder is compared to the LSTM decoder across P8 and P12 rats. For twitches, brackets denote that both decoders performed significantly better at P8 than at P12 ( $p < .05$ ). For wake movements, brackets denote that the LSTM decoder performed significantly better than the linear decoder at both ages ( $p < .05$ ). Horizontal dashed white lines indicate chance performance using a shuffling procedure.

continuous activity modestly interferes with M1's representation of wake movements. Together, these results indicate that although M1's representation of twitch kinematics diminishes between P8 and P12, its representation of wake-movement kinematics is robust at both ages.

### Continuous activity increases the uniqueness of M1 information

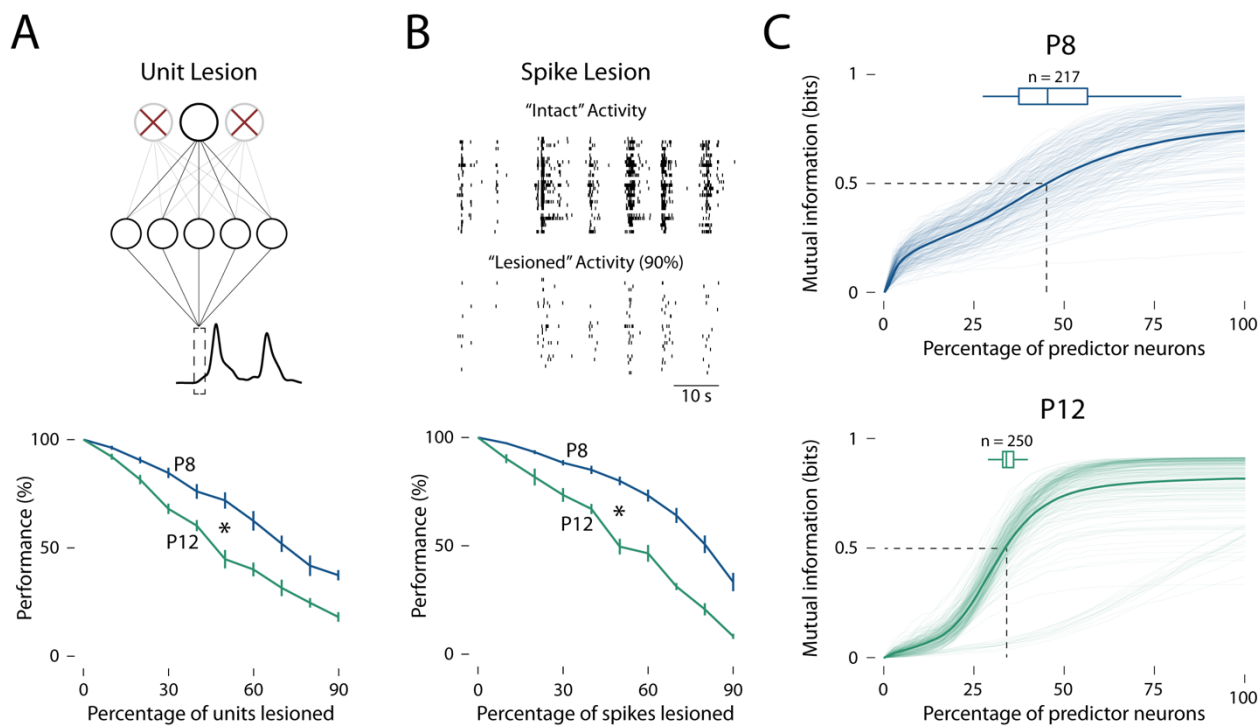
The presence of decorrelated, continuous M1 activity at P12 represents a transition from a “dense” (i.e., redundant) encoding of information to a more energy- and information-efficient “sparse” code [28]. Accordingly, we predicted that removal of redundant neural input at P8 would not degrade decoding performance as much as removal of sparse neural input at P12.



To test this hypothesis, two “lesion” experiments were performed. In the first “unit lesion” experiment, a varying percentage of M1 units (up to 90%) recorded from a given pup was randomly discarded from the decoding process (**Figure 3A, top**). The resulting decoding performance was then compared with the decoding performance of the “intact” model (i.e., 0% of units removed). (Only the nonlinear decoder was used for these experiments.) As predicted, although unit lesions degraded decoder performance at both ages, the loss was significantly greater at P12 than at P8 ( $F(1.50, 20.94) = 18.60, p < .001, \text{adj. } \eta_p^2 = .539$ ; **Figure 3A, bottom**). For example, lesioning 50% of units led to a decoder performance of  $70.4 \pm 3.9\%$  at P8, compared with  $41.9 \pm 4.4\%$  at P12. Such differences in decoder performance were statistically significant across the range of unit lesions (all  $p_s \leq .005$ ).

The second “spike lesion” experiment was performed in a similar fashion. Here, a varying percentage of spikes (i.e., action potentials) across all M1 units (up to 90%) was randomly discarded from the decoding process (**Figure 3B, top**). As predicted, spike lesions resulted in significantly greater declines in decoder performance at P12 than at P8 ( $F(5.05, 70.63) = 8.41, p < .001, \text{adj. } \eta_p^2 = .330$ ; **Figure 3B, bottom**). For example, lesioning 50% of spikes led to a decoder performance of  $79.5 \pm 1.8\%$  at P8, compared with  $51.0 \pm 3.2\%$  at P12. Again, such differences in decoder performance were statistically significant across the range of spikes removed (all  $p_s \leq .03$ ).

That both lesions decreased decoder performance more at P12 than P8 implies that M1 activity contains more *unique* information about limb kinematics at the older age. To test this implication, we measured the mutual information between a target M1 unit and a



**Figure 3.** M1 activity contains more unique information at P12 than at P8.

(A) Top: Representation of the “unit lesion” experiment. The top row of nodes represents the individual M1 units input to the LSTM decoder. Red X’s denote units that were “lesioned” (i.e., removed from the decoding process). The second row of nodes represents the hidden layer of the LSTM decoder. The black trace at the bottom represents forelimb displacement as predicted by the LSTM decoder. Bottom: Mean ( $\pm$ SEM) decoder performance (measured as the percentage change in the “lesioned”  $r^2$  relative to the “intact”  $r^2$ ) for P8 (blue line) and P12 (green line) rats as a function of the percentage of units lesioned. Asterisk denotes significantly worse decoder performance at P12 than at P8 across the range of units lesioned ( $p < .05$ ).

(B) Top: Representation of the “spike lesion” experiment. “Intact” activity represents baseline M1 activity, where each row represents an individual unit and each vertical tick denotes an action potential. “Lesioned” activity represents M1 activity after 90% of individual spikes were removed from the decoding process. Bottom: Same as in (A), but for the spike lesion experiment. Asterisk denotes significantly worse decoder performance at P12 than at P8 across the range of spikes lesioned ( $p < .05$ ).

(C) For P8 (top, blue) and P12 (bottom, green) rats, mutual information (in bits) is shown as a function of increasing size of the predictor subset (expressed as a percentage of the maximum size). Solid lines represent the mean increase in mutual information in the target unit; translucent lines represent individual target units (P8:  $n = 217$  units; P12:  $n = 250$  units). Dashed lines show the mean percentage of predictor units required to achieve 0.5 bits (half the theoretical maximum); horizontal boxplots show the percentage of predictor units required to achieve 0.5 bits for individual target units.

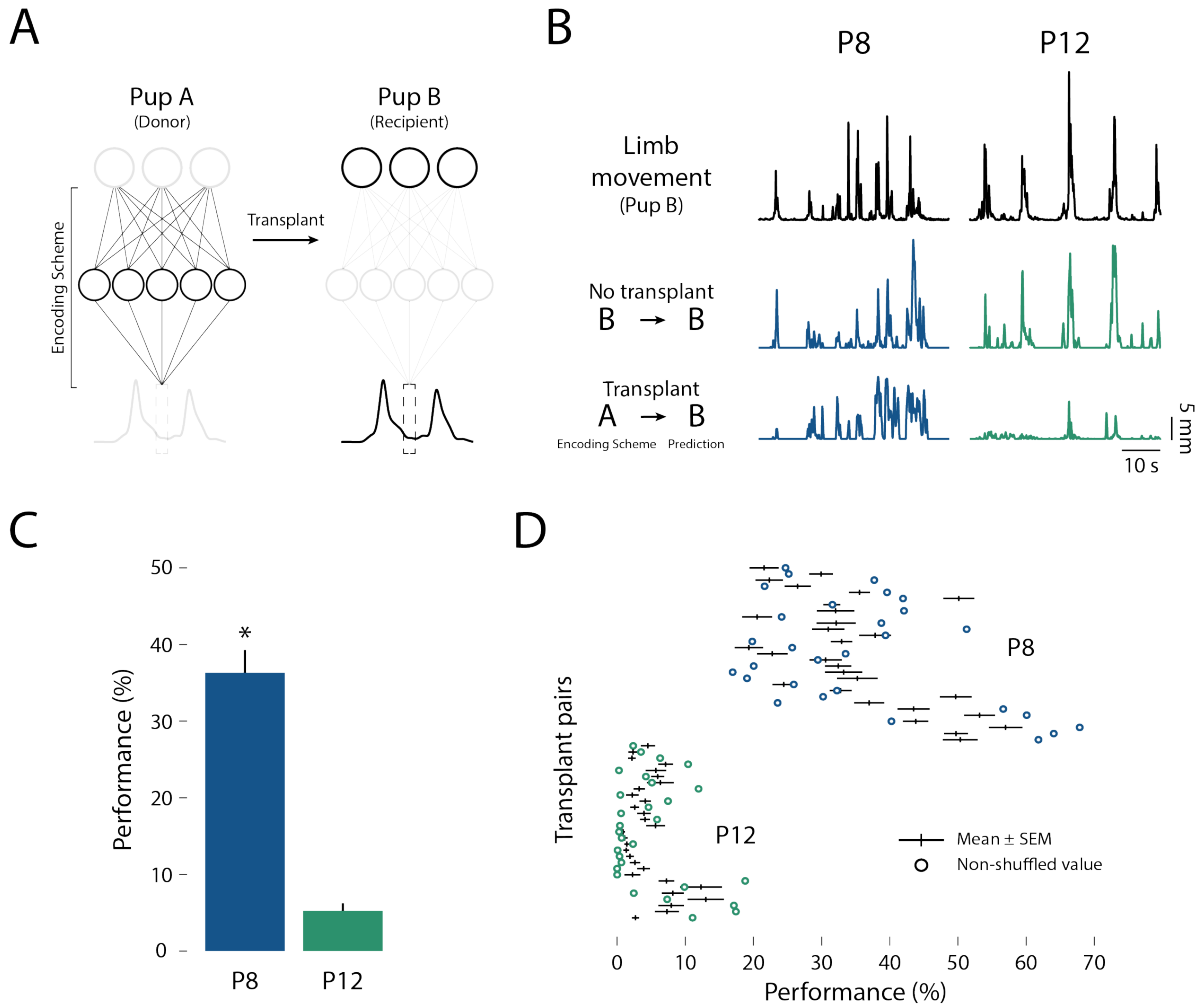
randomly selected subset of units [29]; the size of the subset varied from 1 additional unit to all available units for a given pup. Here, mutual information refers to the predictability of the target unit’s response (to a movement) when the responses of the subset are known. By definition, mutual information *only* increases when the subset contains non-

redundant (i.e., unique) information about the target unit [30]. We found that adding units increased mutual information significantly more quickly at P12 ( $38.7 \pm 1.0\%$  units to reach 0.5 bits) than at P8 ( $50.9 \pm 1.2\%$  units to reach 0.5 bits;  $t(436.01) = 7.64$ ;  $p \leq .001$ , adj.  $\eta^2 = .116$ ; **Figure 3C**; note that 1 bit is the theoretical maximum value). Together, these results support the notion that the onset of continuous activity at P12 increases the uniqueness of information available in M1.

### **M1's encoding scheme is interchangeable at P8, but not at P12**

The finding that M1's representation of movement at P8, but not at P12, is so redundant as to be robust to the removal of neural activity raises an intriguing hypothesis: Namely, that discontinuous activity is not only redundant, but perhaps is also generic to the point of being interchangeable between individual pups. Such a property of M1 activity could reflect a gross encoding scheme that is present during this period of development when somatotopic relations among M1 units and forelimb muscles are still being established. Accordingly, we predicted that after "transplanting" the encoding scheme of one pup into that of another, decoding performance would remain high at P8, but would be degraded at P12 (**Figure 4A**).

To test this prediction, M1 activity of a given P8 or P12 rat (the recipient) was decoded using the model weights (inferred encoding scheme) transplanted from another P8 or P12 rat, respectively (the donor). M1 units were aligned between the donor and recipient rats according to firing rate (i.e., sorted from highest to lowest firing rates). For both P8 and P12 rats, 29 such donor-recipient pairs were generated. (Again, only the nonlinear decoder was used.) As predicted, decoder performance was significantly better across



**Figure 4.** Transplanting M1’s encoding scheme between pups leads to successful decoding at P8, but not at P12.

(A) Representation of the “model transplant” experiment. Pup A (“donor”; left) donates its encoding scheme to Pup B (“recipient”; right). Thus, Pup B’s M1 unit activity (top row) is used to predict Pup B’s forelimb displacement (black trace at bottom) using Pup A’s encoding scheme. (B) Representative traces of actual limb movement (top) compared to limb movements predicted by Pup B’s (original) decoder model (middle) and limb movement predicted by Pup B’s (transplanted) decoder model (bottom) for P8 (left, blue lines) and P12 (right, green lines) rats. (C) Mean (+SEM) decoder performance (measured as the ratio of the “transplant”  $r^2$  to the original  $r^2$ , expressed as a percentage) for P8 (blue) and P12 (green) rats. Asterisk denotes significantly better decoder performance after transplantation at P8 than at P12 ( $p < .05$ ). (D) Decoder performance (now on the x-axis) is shown for each donor-recipient pair across 30 random shuffles. Black lines indicate the mean ( $\pm$ SEM) decoder performance of the 30 random shuffles for each donor-recipient pair. Blue and green circles indicate decoder performance of the unshuffled models at P8 and P12, respectively.

P8 pairings than P12 pairings ( $t(56) = 11.71$ ,  $p < .001$ ,  $\text{adj. } \eta^2 = .705$ ; **Figure 4B–C**). To ensure that this finding was not simply due to our method of aligning firing rates across

pups, we performed 30 additional trials of the experiment using random unit-unit pairings; this randomization had no substantive effect on the age-related difference in decoding performance after transplantation (**Figure 4D**). Thus, continuous activity appears to contribute to the emergence of individualized M1 encoding schemes at P12.

## DISCUSSION

We previously reported that M1's encoding of forelimb movements at P8 disappears with the onset of continuous activity at P12 [6]. That finding raised the possibility that M1's somatosensory representations are “reset” at P12 by the sudden change in cortical dynamics. An alternative possibility was that the somatosensory representation persists at P12 but is obscured by the noisy continuous activity. We considered the second possibility more parsimonious than the first, in part because twitch-related reafference continues to trigger M1 activity through at least P20 [27]. Accordingly, we predicted that M1 continues to encode movement-related information at P12, but through a more complex encoding scheme. This prediction was confirmed using a nonlinear neural decoder.

We expected the nonlinear decoder to reveal M1 encoding of twitch movements at P12. To our surprise, however, this was not the case; instead, only movement-related information during wake was encoded at P12. Also, wake movements were better reconstructed by the nonlinear decoder at P8, indicating that a significant component of M1's representation of wake movements is nonlinear even *before* the onset of continuous activity. This last finding is particularly surprising given that the brainstem selectively dampens wake-related reafference in M1 at P8 [20, 21, 31]; moreover, our previous analysis using linear methods led us to conclude that M1 units are not sensitive to wake-

related kinematics at either age [6]. Thus, the present finding that M1 encodes twitches and wake movements through linear and nonlinear means, respectively, adds another dimension to our understanding of M1's state-dependent sensory representation in early development.

It must be stressed that although individual twitches were not reliably decoded at P12, this age does not mark the end of twitches' significance for sensorimotor development. For example, in P20 rats, twitches are implicated in the developmental emergence of a cerebellar-dependent internal model of movement [27]. To what extent twitching contributes to other forms of plasticity in adults remains an open question [32-34].

### **Continuous activity corresponds with more unique M1 information**

In adults, continuous activity is associated with complex functions such as sparse coding [28, 35, 36] and predictive coding [35, 37, 38]. Continuous activity is also thought to enhance reafference by providing contextual information related to ongoing behaviors, cognitive processes, and cortical dynamics [39-41]. Accordingly, one might expect the emergence of continuous activity at P12 to immediately sharpen reafference in M1. However, this was not the case (see **Figure 2B**), suggesting that continuous activity *per se* does not enhance reafference and lead to better neural decoding outcomes.

Although continuous activity did not immediately enhance reafference at P12, it did correspond with an increase in the *uniqueness* of information in M1. This unique information could be due, in part, to the developmental narrowing of M1 receptive fields. At P8, M1 units respond across a range of movement amplitudes, suggesting that M1 receptive fields are broadly tuned to multiple forelimb muscles [6]. This broad tuning is no

longer apparent at P12, suggesting a progression toward the narrow receptive fields displayed by M1 units in adulthood (Stefanis & Jasper 1964, Nudo et al. 1992, Georgopoulos & Stefanis 2007). Similar receptive-field narrowing has been demonstrated in other cortical (Hubel & Wiesel 1963, DeAngelis et al. 1993) and subcortical (Chen & Regehr 2000, Tschetter et al. 2018) areas and is an activity-dependent process (Chakrabarty & Martin 2005).

The noise that characterizes continuous activity may have a functional benefit related to the concept of “regularization.” Regularization is the process of introducing noise to a system so as to prevent overfitting (i.e., forming inappropriately strong relationships too early in the learning process). Regularization through noise is a conventional technique in the creation of robust computational networks [42-44]. Accordingly, we propose that regularization contributes to receptive-field narrowing by ensuring that connections between M1 units and muscle fibers are only strengthened when they fire together with high-temporal precision exceeding the noise band. Connections with less-precise temporal relations (i.e., at the edges of the receptive field) would be weakened and pruned.

### **M1’s encoding scheme individualizes at P12**

We found that when the encoding scheme of one P8 rat was “transplanted” to another P8 rat, decoding outcomes were significantly better than after a similar “transplant” between P12 rats. In other words, M1’s encoding scheme becomes significantly more individualized, suggesting that the complex encoding schemes at P12 are more specific to individual pups than are the encoding schemes at P8.

Developing animals face the problem of matching cortical connections to a moving target: their rapidly growing bodies. Discontinuous activity solves this problem before P12, as nearly all M1 activity occurs in response to movement-related reafference or external stimulation [6, 8, 45]. This pattern of activity maximizes the correlation between behavior and neural activity. Similarly, before P12, a 10–20 Hz corticothalamic rhythm—known as a *spindle burst*—amplifies reafference and promotes the development of sensorimotor pathways [2, 46-49]. This strong correlation between limb movements and M1 activity is thought to strengthen those M1 connections early in development that will serve as the foundation for later-emerging motor control [6, 21, 45].

## **Conclusion**

In adults, M1 contributes in many complex ways to motor control and motor learning [50-52]. Also, M1 integrates information arising from the other senses [53, 54], neuromodulatory systems [55, 56], and ongoing behaviors [39-41]. Although little is known about the development of these higher-level functions of M1, it is now clear that these functions—like M1's most basic motor capacities—rest upon an early-developing sensory foundation. But even this early "sensory phase" of M1 development is protracted and complex: It begins during the discontinuous period of M1 activity as gross somatotopic relations are formed and, as shown here, is transformed through the transition to continuous activity. Thus, the present findings add new dimensions to our growing understanding of M1's sensory development before it assumes its more familiar role in motor control.



## **Acknowledgments**

Preparation of this article was made possible by a grant from the National Institute of Child Health and Human Development (R37-HD081168) to M.S.B.

## **Author Contributions**

Conceptualization: R.G. and M.S.B.; Methodology: R.G.; Software: R.G.; Formal Analysis: R.G.; Investigation: R.G.; Data Curation: R.G.; Writing – Original Draft: R.G.; Writing – Review & Editing: R.G., G.S., and M.S.B.; Visualization: R.G. and M.S.B.; Funding Acquisition: G.S. and M.S.B.; Resources: G.S. and M.S.B.; Supervision: M.S.B.

## **Declaration of Interests**

The authors declare no competing interests.

## STAR METHODS

### KEY RESOURCES TABLE

REGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Experimental models: Organisms/strains</b>		
Sprague-Dawley Norway Rats	Envigo	RRID: RGD_10401918
<b>Software and algorithms</b>		
MATLAB, version 2020b	Mathworks	RRID: SCR_001622
Python Programming Language, version 3.8	Python Software Foundation	RRID: SCR_008394
Tensorflow, version 2.9	Google	RRID: SCR_016345
Adobe Illustrator Creative Cloud 2022	Adobe	RRID: SCR_010279
SPSS 28	IBM	RRID:SCR_019096

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources should be directed to, and will be fulfilled by, the lead contact, Dr. Mark Blumberg ([mark-blumberg@uiowa.edu](mailto:mark-blumberg@uiowa.edu)).

#### Materials availability

This study did not generate new unique reagents.

### **Data and code availability**

Timestamps of actional potentials, twitches, wake movements, and sleep-wake transitions, as well as the position of the forelimb in Cartesian coordinates are available for download at [https://github.com/rglanz/Glanz\\_et\\_al\\_2023](https://github.com/rglanz/Glanz_et_al_2023). All custom software is available upon request. (Also see [https://github.com/KordingLab/Neural\\_Decoding](https://github.com/KordingLab/Neural_Decoding) for neural decoding software and <https://github.com/nmtimme/Neuroscience-Information-Theory-Toolbox> for information theory software.)

### **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

All experiments were conducted in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals (NIH Publication No. 80–23) and were approved by the Institutional Animal Care and Use Committee of the University of Iowa (Protocol #: 0021955).

As described previously [6], the data used in this study were collected from Sprague-Dawley Norway rats (*Rattus norvegicus*) at P8–9 (hereafter referred to as P8;  $n = 8$ ) and P12–13 (hereafter referred to as P12;  $n = 8$ ). Equal numbers of males and females were used, and all subjects were selected from different litters. See our previous report for further details related to experimental subjects.

### **METHOD DETAILS**

As described previously [6], high-speed (100 fps) video and M1 extracellular unit activity were collected from unanesthetized pups as they cycled through sleep and wake. Forelimb displacement was quantified using DeepLabCut [57, 58] and forelimb

movements were identified using custom software and were visually confirmed. See our previous report for further details related to surgeries, electrophysiological recordings, video data collection, histology, spike sorting, and determination of behavioral state and forelimb movements.

## **QUANTIFICATION AND STATISTICAL ANALYSIS**

### **Data preparation**

For all decoding procedures, M1 unit activity and forelimb displacement served as the predictor and target variables, respectively. Both variables were binned in 20-ms increments. Forelimb displacement was calculated as the absolute value of zero-centered limb position along the x-, y-, and z-dimensions and was smoothed using a 20-ms half-width Gaussian kernel.

To prevent model overfitting, the one-hour recordings were split into training (36 min; 60%) and testing (9 min; 15%) datasets, which were used to train and test model parameters for best performance. A validation dataset (15 min; 25%) was held back until the final model parameters were established. The validation dataset was then used to perform all present analyses in order to ensure an unbiased final assessment of decoder performance [25].

### **Data scaling**

M1 activity was z-scored prior to decoding. The scaling factors (mean and standard deviation) were calculated using only the training dataset to avoid data leakage (i.e., positive bias in performance estimates due to information from the testing dataset leaking into the training dataset; [59]). Forelimb position was normalized prior to decoding. The

scaling factors (minimum and maximum) were again calculated using only the training dataset to avoid data leakage.

### **Linear model**

The linear model was composed of a single-layer Tensorflow model [60] with no hidden layer or activation function. This design forms a linear perceptron [61] and is equivalent to an ordinary least-squares (i.e., linear) regression. All models were assembled in Keras, a software package that assists in building Tensorflow models [62]. A 240-ms time window surrounding each timepoint was flattened (i.e., crossed with the feature of individual units) to provide additional temporal context to the model.

The decoder's learning parameters were as follows: Mean-square error was selected as the loss function (i.e., a function used to determine the error between the real limb displacement and the predicted limb displacement). Adam [63] was selected as an optimizer (i.e., an algorithm that updates the internal parameters of the model across training iterations). The learning rate (i.e., the degree to which the internal parameters are updated) was set to 0.001 arbitrary units, which is the default learning rate for Adam in the Keras package. Importantly, we tested a variety of different combinations of loss functions, optimizers, and learning rates and did not find meaningful impact on decoder performance (data not shown).

The model was allowed to train until the validation loss function stopped decreasing, which typically occurred after 3–5 iterations. Three target variables were independently predicted for each time bin, corresponding to x-, y-, and z-displacement of the forelimb.

The three predictions were combined using the Pythagorean theorem prior to comparison with actual limb displacement:

**Equation 1:**  $\sqrt{x^2 + y^2 + z^2}$

### **Nonlinear model**

Several nonlinear models were compared before settling on a long-short-term memory (LSTM) neural network [26]. The LSTM model was chosen for its superior performance on the current dataset and its successful application in similar neural decoding applications [25, 64, 65]. The number of nodes in the input layer was identical to the number of units in each specific M1 recording (i.e., 14–38 nodes). Larger numbers of nodes were tested (up to 500) but did not result in significantly better decoding performance (data not shown). A 240-ms time window, similar to that added to the linear model, was included as a recurrent feature.

To further prevent overfitting, a penalty was added to the recurrent dimension of the model to prevent rapid changes to the model's internal parameters across training iterations. The penalty selected was L2 (ridge regression; [66]):

**Equation 2:**  $\lambda \cdot \|\beta\|$

where  $\lambda$ , which represents the magnitude of the penalty, was set to 0.001 arbitrary units.

The optimizer, learning rate, and loss function were identical to the linear model. The nonlinear model was also allowed to run until the validation error stopped decreasing (typically within 3–5 iterations). Similar to the linear model, three independent predictions

(x-, y- and z-displacement) were made for each time bin and summed using the Pythagorean theorem (Equation 1) prior to comparison with actual limb displacement.

### Decoder performance

Decoder performance was evaluated using the square of the Pearson correlation coefficient ( $r^2$ ) with actual (x) and predicted (y) limb displacement as the two variables:

$$\text{Equation 3: } r^2 = \left( \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \right)^2$$

In typical neural decoding applications, regression performance is measured using the Coefficient of Determination ( $R^2$ )

$$\text{Equation 4: } R^2 = 1 - \frac{\sum(y_i - x_i)^2}{\sum(x_i - \bar{x})^2}$$

which is sensitive to the scale of the errors. The scale-invariant definition of model performance used here (Equation 3) solves the “intermittent demand” problem [67]. Briefly, because the limb’s displacement is near zero (i.e., at rest) across the vast majority of sampled time points,  $R^2$  disproportionately rewards predictions equal to the mean limb displacement and punishes predictions that vary from baseline, which is not the case with  $r^2$ .

### Lesion experiment

Two “lesion” experiments were performed in which individual M1 units or action potentials (spikes) were randomly selected to be removed from the decoding process. A given percentage of units or spikes, in a range from 0% to 90% (in increments of 10%), was randomly selected and removed without replacement using the NumPy software package

in Python [68]. The resulting units or spike trains were then used as the predictor variable and given to the decoder model that was trained on the “intact” dataset. The performance of the decoding model after a lesion was calculated as the ratio of the lesion  $r^2$  to the “intact”  $r^2$ , converted to a percentage. The randomization procedure was repeated ten times and the ten  $r^2$  values were averaged (per animal).

### **Mutual Information**

Mutual information between a target unit ( $X_0$ ) and a subset of predictor units ( $\{X_k\}$ ) of size  $k$  was defined as:

$$\text{Equation 5: } I(X_0; \{X_k\}) = S(X_0) - S(X_0 | \{X_k\})$$

where  $I(X_0; \{X_k\})$  represents the mutual information between the target unit and subset of units [69-71]. This is equivalent to the entropy of the target unit ( $S(X_0)$ ) minus the entropy of the target unit explained by the subset ( $S(X_0 | \{X_k\})$ ). Software from [72] was used to calculate mutual information. Random subsets of units were randomly selected from all possible combinations (10 selections per subset of size  $k$ ).

### **Model transplant experiment**

To test the interchangeability of decoder models between animals of a particular age, a “model transplant” experiment was performed. The M1 activity (predictor variable) of an individual rat (referred to as Pup A; “recipient”) was used to predict Pup A’s forelimb displacement using the model weights of a different rat (Pup B; “donor”) of the same age. Two animals of the same age were paired when Pup A had the same number of M1 units, or a greater number, than Pup B. The two animals’ units were matched by descending firing rate. The excess units (from Pup A) were discarded from the analysis. In total, 29



such donor-recipient pairs were selected for both P8 and P12 animals. To test whether firing-rate sorting introduced bias, the experiment was repeated 30 times with random unit-unit pairings. The performance of the “model transplant” was measured as the ratio of the transplanted  $r^2$  (Pup A  $\times$  Pup B) to the original  $r^2$  (Pup A  $\times$  Pup A).

### **Statistical analyses**

Prior to analysis, all data were tested for normality using the Shapiro-Wilk test, for equal variance using Levene’s test (for between-subjects variables), and for sphericity using Mauchly’s test (for within-subjects variables with >2 groups). For analyses in which the variance between groups was not equal, a pooled error term was not used when generating simple main effects and post-hoc tests. For analyses in which sphericity was violated, a Huynh-Feldt correction was applied to the degrees of freedom. All  $r^2$  values were arc-sin transformed prior to analysis. The mean and standard error of the mean (SEM) were used throughout as measures of central tendency and dispersion, respectively.

All analyses were performed as independent t tests or two-way mixed-design ANOVAs. Simple main effects were only tested if the interaction term was significant. In all two-way ANOVAs, an adjusted partial eta-squared was used as an estimate of effect size that corrects for positive bias due to sampling variability [73]. For t tests, an adjusted eta-squared estimate of effect size was reported.

## REFERENCES

1. Vanhatalo, S. and Kaila, K (2006). Development of neonatal EEG activity: from phenomenology to physiology. *Seminars in Fetal and Neonatal Medicine*, **11**(6): p. 417-478.
2. Colonnese, M. T. and Phillips, M. A. (2018). Thalamocortical function in developing sensory circuits. *Current Opinion in Neurobiology*. **52**: p. 72-79.
3. Vanhatalo, S., Palva, J. M., Andersson, S., Rivera, C., Voipio, J., and Kaila, K. (2005). Slow endogenous activity transients and developmental expression of K<sup>+</sup>-Cl<sup>-</sup> cotransporter 2 in the immature human cortex. *European Journal of Neuroscience*. **22**(11): p. 2799-2804.
4. Golshani, P., Goncalves, J. T., Khoshkhoo, S., Mostany, R., Smirnakis, S., and Portera-Cailliau, C. (2009). Internally mediated developmental desynchronization of neocortical network activity. *Journal of Neuroscience*. **29**(35): p. 10890-10899.
5. Rochefort, N. L., Garaschuk, O., Milos, R.-I., Narushima, M., Marandi, N., Pichler, B., Kovalchuk, Y., and Konnerth, A. (2009). Sparsification of neuronal activity in the visual cortex at eye-opening. *Proceedings of the National Academy of Sciences*. **106**(35): p. 15049-15054.
6. Glanz, R. M., Dooley, J. C., Sokoloff, G., and Blumberg, M. S. (2021). Sensory coding of limb kinematics in motor cortex across a key developmental transition. *Journal of Neuroscience*. **41**(32): p. 6905-6918.
7. Riyahi, P., Phillips, M. A., and Colonnese, M. T. (2021). Input-independent homeostasis of developing thalamocortical activity. *eNeuro*. **8**(3).
8. Gómez, L. G., Dooley, J. C., and Blumberg, M. S. (2022). Activity in developing prefrontal cortex is shaped by sleep and sensory experience. *bioRxiv*.
9. Virtanen, M. A., Uvarov, P., Mavrovic, M., Poncer, J. C., and Kaila, K. (2021). The multifaceted roles of KCC2 in cortical development. *Trends in Neurosciences*. **44**(5): p. 378–392.
10. Rivera, C., Voipio, J., and Kaila, K. (2005). Two developmental switches in GABAergic signalling: the K<sup>+</sup>-Cl<sup>-</sup> cotransporter KCC2 and carbonic anhydrase CAVII. *The Journal of Physiology*. **562**(1): p. 27-36.
11. Ben-Ari, Y., Gaiarsa, J.-L., Tyzio, R., and Khazipov, R. (2007). GABA: a pioneer transmitter that excites immature neurons and generates primitive oscillations. *Physiological Reviews*. **87**: p. 1215-1284.
12. Colonnese, M. T. (2014). Rapid developmental emergence of stable depolarization during wakefulness by inhibitory balancing of cortical network excitability. *Journal of Neuroscience*. **34**(16): p. 5477-5485.

13. Che, A., Babij, R., Iannone, A. F., Fetcho, R. N., Ferrer, M., Liston, C., Fishell, G., and De Marco Garcia, N. V. (2018). Layer I interneurons sharpen sensory maps during neonatal development. *Neuron*. **99**(1): p. 98-116 e117.
14. Dard, R. F., LePrince, E., Denis, J., Balappa, S. R., Suchkov, D., Boyce, R., Lopez, C., Giorgi-Kurz, M., Szwagier, T., and Dumont, T. (2022). The rapid developmental rise of somatic inhibition disengages hippocampal dynamics from self-motion. *eLife*. **11**: p. e78116.
15. Downes, N. and Mullins, P. (2014). The development of myelin in the brain of the juvenile rat. *Toxicologic Pathology*. **42**(5): p. 913-922.
16. Jouvet-Mounier, D., Astic, L., and Lacote, D. (1970). Ontogenesis of the states of sleep in rat, cat, and guinea pig during the first postnatal month. *Developmental Psychobiology*. **2**(4): p. 216-239.
17. Gramsbergen, A. (1976). The development of the EEG in the rat. *Developmental Psychobiology*. **9**(6): p. 501-515.
18. Seelke, A. M. and Blumberg, M. S. (2008). The microstructure of active and quiet sleep as cortical delta activity emerges in infant rats. *Sleep*. **31**(5): p. 691-699.
19. Del Rio-Bermudez, C., Kim, J., Sokoloff, G., and Blumberg, M. S. (2017). Theta oscillations during active sleep synchronize the developing rubro-hippocampal sensorimotor network. *Current Biology*. **27**(10): p. 1413-1424.
20. Tiriác, A., Del Rio-Bermudez, C., and Blumberg, M. S. (2014). Self-generated movements with "unexpected" sensory consequences. *Current Biology*. **24**(18): p. 2136-2141.
21. Dooley, J. C. and Blumberg, M. S. (2018). Developmental 'awakening' of primary motor cortex to the sensory consequences of movement. *eLife*. **7**: p. e41841.
22. Singleton, A. C., Brown, A. R., and Teskey, G. C. (2021). Development and plasticity of complex movement representations. *Journal of Neurophysiology*. **125**(2): p. 628-637.
23. Ashe, J. and Georgopoulos, A. P. (1994). Movement parameters and neural activity in motor cortex and area 5. *Cerebral Cortex*. **4**(6): p. 590-600.
24. Georgopoulos, A. P., Kettner, R. E., and Schwartz, A. B. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. *Journal of Neuroscience*. **8**(8): p. 2928-2937.
25. Glaser, J. I., Benjamin, A. S., Chowdhury, R. H., Perich, M. G., Miller, L. E., and Kording, K. P. (2020). Machine learning for neural decoding. *eNeuro*. **7**(4).

26. Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*. **9**(8): p. 1735-1780.
27. Dooley, J. C., Sokoloff, G., and Blumberg, M. S. (2021). Movements during sleep reveal the developmental emergence of a cerebellar-dependent internal model in motor thalamus. *Current Biology*. **31**(24): p. 5501-5511.
28. Olshausen, B. A. and Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*. **14**(4): p. 481-487.
29. Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*. **27**(3): p. 379-423.
30. Bettencourt, L. M. (2009). The rules of information aggregation and emergence of collective intelligent behavior. *Topics in Cognitive Science*. **1**(4): p. 598-620.
31. Tiriac, A. and Blumberg, M. S. (2016). Gating of reafference in the external cuneate nucleus during self-generated movements in wake but not sleep. *eLife*. **5**: p. e18749.
32. Herman, J. H. and Roffwarg, H. P. (1983). Modifying oculomotor activity in awake subjects increases the amplitude of eye movements during REM sleep. *Science*. **220**(4601): p. 1074-1076.
33. De Gennaro, L., Ferrara, M., Urbani, L., Bertini, M., and Neuroscienze, S. D. (2000). A complementary relationship between wake and REM sleep in the auditory system: a pre-sleep increase of middle-ear muscle activity (MEMA) causes a decrease of MEMA during sleep. *Experimental Brain Research*. **130**(1): p. 105-112.
34. Eckert, M., McNaughton, B., and Tatsuno, M. (2020). Neural ensemble reactivation in rapid eye movement and slow-wave sleep coordinate with muscle activity to promote rapid motor skill learning. *Philosophical Transactions of the Royal Society B*. **375**(1799): p. 20190655.
35. Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network: Computation in Neural Systems*. **3**(2): p. 213-251.
36. Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., and Harris, K. D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*. **364**(6437).
37. van Hateren, J. H. (1992). A theory of maximizing sensory information. *Biological Cybernetics*. **68**(1): p. 23-29.
38. Keller, G. B. and Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron*. **100**(2): p. 424-435.

39. Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*. **273**(5283): p. 1868-1871.
40. Goris, R. L., Movshon, J. A., and Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*. **17**(6): p. 858-865.
41. Lin, I.-C., Okun, M., Carandini, M., and Harris, K. D. (2015). The nature of shared cortical variability. *Neuron*. **87**(3): p. 644-656.
42. Camuto, A., Willetts, M., Simsekli, U., Roberts, S. J., and Holmes, C. C. (2020). Explicit regularisation in gaussian noise injections. *Advances in Neural Information Processing Systems*. **33**: p. 16603-16614.
43. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*. **15**(1): p. 1929-1958.
44. Poole, B., Sohl-Dickstein, J., and Ganguli, S. (2014). Analyzing noise in autoencoders and deep networks. *arXiv.1406.1831*.
45. Gómez, L. J., Dooley, J. C., Sokoloff, G., and Blumberg, M. S. (2021). Parallel and serial sensory processing in developing primary somatosensory and motor cortex. *Journal of Neuroscience*. **41**(15): p. 3418-3431.
46. Khazipov, R., Sirota, A., Leinekugel, X., Holmes, G. L., Ben-Ari, Y., and Buzsáki, G. (2004). Early motor activity drives spindle bursts in the developing somatosensory cortex. *Nature*. **432**(7018): p. 758-761.
47. Tolner, E. A., Sheikh, A., Yukin, A. Y., Kaila, K., and Kanold, P. O. (2012). Subplate neurons promote spindle bursts and thalamocortical patterning in the neonatal rat somatosensory cortex. *Journal of Neuroscience*. **32**(2): p. 692-702.
48. Yang, J.-W., An, S., Sun, J.-J., Reyes-Puerta, V., Kindler, J., Berger, T., Kilb, W., and Luhmann, H. J. (2013). Thalamic network oscillations synchronize ontogenetic columns in the newborn rat barrel cortex. *Cerebral Cortex*. **23**(6): p. 1299-1316.
49. Dooley, J. C., Glanz, R. M., Sokoloff, G., and Blumberg, M. S. (2020). Self-generated whisker movements drive state-dependent sensory input to developing barrel cortex. *Current Biology*. **30**(12): p. 2404-2410.
50. Sanes, J. N. and Donoghue, J. P. (2000). Plasticity and primary motor cortex. *Annual Review of Neuroscience*. **23**(1): p. 393-415.
51. Kawai, R., Markman, T., Poddar, R., Ko, R., Fantana, A. L., Dhawale, A. K., Kampff, A. R., and Ölveczky, B. P. (2015). Motor cortex is required for learning but not for executing a motor skill. *Neuron*. **86**(3): p. 800-812.

52. Peters, A. J., Liu, H., and Komiyama, T. (2017). Learning in the rodent motor cortex. *Annual Review of Neuroscience*. **40**: p. 77.
53. Ibrahim, L. A., Huang, S., Fernandez-Otero, M., Sherer, M., Qiu, Y., Vemuri, S., Xu, Q., Machold, R., Pouchelon, G., and Rudy, B. (2021). Bottom-up inputs are required for establishment of top-down connectivity onto cortical layer 1 neurogliaform cells. *Neuron*. **109**(21): p. 3473-3485.
54. Schuman, B., Dellal, S., Prönneke, A., Machold, R., and Rudy, B. (2021). Neocortical layer 1: an elegant solution to top-down and bottom-up integration. *Annual Review of Neuroscience*. **44**: p. 221-252.
55. Mechawar, N., Cozzari, C., and Descarries, L. (2000). Cholinergic innervation in adult rat cerebral cortex: a quantitative immunocytochemical description. *Journal of Comparative Neurology*. **428**(2): p. 305-318.
56. Allaway, K. C., Muñoz, W., Tremblay, R., Sherer, M., Herron, J., Rudy, B., Machold, R., and Fishell, G. (2020). Cellular birthdate predicts laminar and regional cholinergic projection topography in the forebrain. *eLife*. **9**: p. e63249.
57. Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* **21**(9): p. 1281-1289.
58. Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., and Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*. **14**(7): p. 2152-2176.
59. Varma, S. and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*. **7**(1): p. 1-8.
60. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., and Isard, M (2016). TensorFlow: a system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*.
61. Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*. **65**(6): p. 386.
62. Chollet, F. (2018). Keras: The python deep learning library. *Astrophysics Source Code Library*: p. ascl: 1806.1022.
63. Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv*. 1412.6980.
64. Wang, Y., Truccolo, W., and Borton, D. A. (2018). Decoding hindlimb kinematics from primate motor cortex using long short-term memory recurrent neural

networks. *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*.

65. Naufel, S., Glaser, J. I., Kording, K. P., Perreault, E. J., and Miller, L. E. (2019). A muscle-activity-dependent gain between motor cortex and EMG. *Journal of Neurophysiology*. **121**(1): p. 61-73.
66. Cortes, C., Mohri, M., and Rostamizadeh, A. (2012). L2 regularization for learning kernels. *arXiv*. 1205.2653.
67. Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*. **22**(4): p. 679-688.
68. Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., and Smith, N. J. (2020). Array programming with NumPy. *Nature*. **585**(7825): p. 357-362.
69. Schneidman, E., Still, S., Berry, M. J., and Bialek, W. (2003). Network information and connected correlations. *Physical Review Letters*. **91**(23): p. 238701.
70. Bettencourt, L. M., Gintautas, V., and Ham, M. I. (2008). Identification of functional information subgraphs in complex networks. *Physical Review Letters*. **100**(23): p. 238701.
71. Timme, N., Alford, W., Flecker, B., and Beggs, J. M. (2014). Synergy, redundancy, and multivariate information measures: an experimentalist's perspective. *Journal of Computational Neuroscience*. **36**(2): p. 119-140.
72. Timme, N. M. and Lapish, C. (2018). A tutorial for information theory in neuroscience. *eNeuro*. **5**(3): p. e0052.
73. Mordkoff, J. T. (2019). A simple method for removing bias from a popular measure of standardized effect size: Adjusted partial eta squared. *Advances in Methods and Practices in Psychological Science*. **2**(3): p. 228-232.