

Inferring intrinsic neural timescales using optimal control theory

Corresponding Author: Dr Linden Parkes

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

My evaluation of the manuscript and my detailed comments are reported in the attached Report.

(Remarks on code availability)

The GitHub repository seems very well organized and complete. It should allow reproducing the manuscript's results and perform new analyses.

Reviewer #2

(Remarks to the Author)

The authors couple network-control theory with automatic differentiation to estimate region-specific self-inhibition (decay-rate) parameters directly from resting-state fMRI time series, using structural connectomes derived from diffusion MRI as the coupling matrix. Optimised decay rates (i) lower the energy and dimensionality of control inputs needed to steer whole-brain dynamics, (ii) correlate with intrinsic neural timescales (INTs) obtained from autocorrelation analyses, (iii) follow cortical gene-expression and cell-type gradients, and (iv) predict several cognitive performance measures. Findings are replicated across two large human cohorts and an independent mouse dataset, underscoring robustness and cross-species generalisability. Taken together, the work advances control-theoretic models toward greater biophysical plausibility and provides a principled bridge between macroscopic dynamics and microscale physiology.

MAJOR COMMENTS

1. The study assumes that “neurobiological effects are already encoded in the connectome,” yet diffusion MRI tractography cannot resolve (i) axonal direction, (ii) cell-type specificity ($E \rightarrow E$, $E \rightarrow I$, $I \rightarrow E$, $I \rightarrow I$), or (iii) laminar termination. A single streamline reconstructed by dw-MRI trackography between regions A and B therefore conflates multiple, often opposing, pathways. Because the optimised decay rate (or inhibitory strength) is highly sensitive to the balance of large-scale excitation versus feed-forward and feedback inhibition, these unknowns can bias both the optimisation and its biological interpretation. Explicitly acknowledge these limitations regarding the model's biological realism and associated diffusion-MRI constraints, and temper claims implying that decay-rate optimisation directly “uncovers” true biological self-inhibition.

2. A more negative d is presented as simultaneously indexing “stronger self-inhibition” and “faster intrinsic timescales.” These are not interchangeable. Synaptic self-inhibition strength (a circuit-level parameter) and membrane/ensemble decay constants (a biophysical time constant) arise from different mechanisms, operate on different timescales, and are treated as separate parameters in biophysically informed models. By collapsing both effects into a single scalar d , the current formulation blurs two distinct sources of variance, making it impossible to discern whether observed associations stem from changes in synaptic strength, passive decay, or some hybrid “effective” quantity. My recommendation is to explicitly acknowledge this conflation in the text, emphasising that d should be interpreted as a lumped representation of multiple mechanisms rather than a direct measure of neither synaptic inhibition nor decay rates, and to temper causal language linking d to “biological self-inhibition.” Clarifying this point is essential, as the manuscript's stated aim is to enhance biological realism.

3. Decay rates are optimised for each of 42 empirical state transitions, yielding large within-region variance (Fig. 2B). In vivo, neither inhibitory synaptic strength nor intrinsic decay constants plausibly fluctuate at this pace. Explain and discuss which

biological mechanisms could account for such variability.

4. Given the well-known non-uniqueness of optimal solutions in linear control models, a single optimisation run per participant is insufficient to claim that the regional decay rate d is uniquely identifiable. My recommendation is to run the optimiser from many random initialisations and report the spread of the resulting d estimates. Convergence toward a tight cluster would support identifiability; dispersion would signal model degeneracy. In addition, I recommend to use bootstrapping to derive confidence intervals for each region's d and present these intervals alongside point estimates. Providing these diagnostics will lend empirical weight to causal statements such as "we estimate the structural form of regional self-inhibition."

5. The Abstract highlights the use of automatic differentiation, yet this aspect is only briefly mentioned in the Methods, with no further detail provided. Please clarify how automatic differentiation contributed to obtaining the results—specifically, what role it played in the optimisation process—and explain how it offered advantages over conventional gradient estimation or optimisation techniques. This would help justify its prominent mention in the Abstract and clarify its methodological significance.

MINOR COMMENTS

1. In Fig. 1 it is unclear where the decay parameter d enters the displayed system equation. Consider displaying the full equation.

2. Replace "innervate eigenmodes" with "excite" or similar to avoid confusion with physiological "innervation".

3. Correct multiple references to non-existent panel "Fig. 2F".

4. Missing Data Availability Statement: include direct links or DOIs for all raw and derived datasets (human and mouse) and note any access restrictions.

(Remarks on code availability)

The GitHub repo (https://github.com/LindenParkesLab/nct_xr) lacks a minimal working example and instructions for obtaining and preprocessing the data. A short README outlining (i) data-access links, (ii) the exact preprocessing pipeline, and (iii) an end-to-end demo would greatly improve reproducibility.

Reviewer #3

(Remarks to the Author)

The paper, "Inferring macroscopic intrinsic neural timescales using optimal control theory" by Kim et al. extends the widely used network control theory (NCT) framework to estimate regional self-inhibition, providing greater biophysical realism. The main analyses used data from 960 HCP-YA subjects, the authors confirm their main findings in a second dataset (MICA-MICs) of 50 individuals, and demonstrate translational convergence of their main findings in mice using the Allen Mouse Brain Connectivity Atlas. The authors show: 1) that including regional estimates of self-inhibition in the NCT model leads to reduced control energy; 2) the self-inhibition parameters correlate with intrinsic neural timescales (INTs) estimated from resting-state fMRI data; 3) the self-inhibition parameters correlate positively with AHBA-SST and PVALB-Chandelier, and negatively with AHBA-PVALB and PVALB-Basket; and 4) control energy from the model including regional self-inhibition parameters better predicts cognitive performance on several tasks. The manuscript is well written, easy to read, and has a logical flow to it. The replications across human and mouse datasets are strengths. I found the conclusions to be well supported by the data and believe the results will be of great interest to a wide audience. I have a few comments that I believe would increase the significance of the manuscript.

1. The control set analyses are very interesting and help move NCT to more biologically realistic and relevant uses such as TMS, tDCS, and DBS (as mentioned by the authors). I think it would be of great interest to readers if the authors could more thoroughly investigate the subset of nodes that were necessary to achieve successful state transitions. Biologically, I would expect certain brain regions to have a greater influence, likely determined by their pattern of connectivity.

2. I understand that there are separate self-inhibition parameters estimated for each state transition, but it is unclear to me why the authors chose to look at the principal components of the self-inhibition parameters as opposed to the average over all state transitions. This is especially relevant given that the average self-inhibition parameters correlated most strongly with INTs measure from resting-state fMRI data ($\rho = 0.70$ for the average vs $\rho = 0.42$ for PC1 and $\rho = 0.22$ for PC2).

3. The analyses with gene expression are very interesting and provide important biological context for their results. I think it would be of great interest to readers if the authors could also investigate relationships with excitatory gene expression and pyramidal cells. This is of particular interest given the finding that the hierarchy of INTs is believed to be driven by dendritic spine density, i.e. recurrent excitation (Chaudhuri, et al., doi:10.1016/j.neuron.2015.09.008).

4. Relatedly, why did the authors not investigate relationships with VIP if their focus was on interneurons?

5. The authors very nicely show using brain-based predictive modeling (BPM) that including regional self-inhibition parameters in the NCT model improves the ability to predict individuals' cognitive performance using control energy. I would be interested to see if using the regional self-inhibition parameters improves predictive performance compared to control energy or if the combination (both self-inhibition and control energy) improves performance over using just control energy. This would speak to the functional importance of INTs as they relate to cognition which is something that has been investigated using both resting-state fMRI and ECoG measured INTs.

6. Related to the BPM analyses, I am not sure why the authors chose to report results for the correlation between the true y and predicted y as opposed to the R-squared, given they say the correlation is the secondary outcome measure. Given

current best practices (e.g., Poldrack et al., doi:10.1001/jamapsychiatry.2019.3671), I suggest reporting R-squared in the main text for the measurement of predictive performance and would also suggest including a measure of error such as mean absolute error or root mean squared error.

7. Further related to the BPM analyses, I'm not sure that comparing the distributions of cross-validated predictive performance between the INTuniform and INToptimized models is the most appropriate method for determining significant differences in performance. I would suggest generating a null distribution of the difference in predictive performance using the null models for each individual method and comparing the difference in the mean cross-validated predictive performance between the two methods to this null distribution. This should be relatively straight forward to do since the authors already have the null distributions for each method.

8. How correlated are the regional estimates of self-inhibition across the HCP-YA and MICA-MICs datasets? This would help demonstrate the stability of the proposed NCT method.

Minor comments:

1. It would be nice to add a plot for this finding even if in the Supplement: "Note, we found that the mean of optimized decay rates (taken from Fig. 2F) also correlated positively with ACFlag ($\rho = 0.70$, $-\log_{10}(\text{psmash}) > 25$)."

2. Typo: "That is, optimized decay rates correlated positively with AHBA-SST, AHBA-PVALB, and PVALB-Chandelier, and negatively with PVALB-Basket," should be: "That is, optimized decay rates correlated positively with AHBA-SST and PVALB-Chandelier, and negatively with AHBA-PVALB and PVALB-Basket."

3. Typo: "Briefly, this approach challenges another long-held assumption of NCT that each control signal exclusively impacts a signal node," should be: "Briefly, this approach challenges another long-held assumption of NCT that each control signal exclusively impacts a single node;"

(Remarks on code availability)

The code is well-documented and should be usable by the community. I was unable to run the code because the necessary data are not provided (which is reasonable given the nature of the data). One minor comment is that in one of the python notebooks (results_control_energy_subjects.ipynb) there is output of individual subject data including restricted data elements (Age_in_Yrs) which should not be present.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

In their revised manuscript, the Authors have thoroughly and convincingly addressed all points raised in my previous report, as well as points raised by the other reviewers.

The new version of the manuscript is clearer and considerably more effective in conveying the relevance of the main findings. Overall, the major finding that optimized timescales allow for a more efficient exploration of the neural state space and correlate with neurobiological markers is worthy of attention to the neuroscience community. The article is likely to inspire several works aimed at assessing the relevance of timescale-optimized control for healthy and pathological brain function, as well as neuromodulation experiments.

I thus think that the work deserves publication in Nature Communications, and I would recommend publication of the manuscript as is.

(Remarks on code availability)

The GitHub repository allows reproducing the manuscript's results and perform new analyses.

Reviewer #2

(Remarks to the Author)

I have now reviewed the revised manuscript and author responses and am pleased with the authors' thorough and careful revisions. All of my previous concerns have been addressed, and I have no further comments. The manuscript is now in excellent shape, and I enthusiastically recommend it for publication in Nature Communications.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I would like to thank the authors for their thorough responses to the reviewers' comments. I have three remaining comments:

1. Related to R3 comment #1: A figure showing the "iteration removed" metric as a brain map to provide spatial information regarding the results would be helpful. It would also be beneficial to discuss these findings in the "Controlling brain dynamics effectively from fewer brain regions" section of the discussion in conjunction with the findings from Fig. 3 that are currently discussed.

2. Related to R3 comments #3 & #4: I appreciate the addition of the VIP, AMPA, and NMDA analyses. I believe these results should be mentioned more explicitly than simply stating: "see also Fig. S8 for correlations between PC 1 and an extended selection of gene transcripts from the AHBA." It should be noted that the relationships observed with SST and PVALB are not specific to those genes since similar relationships are observed for the other gene sets.

3. Related to R3 comment #5: The authors did not investigate the combination of control energy and decay rates to see if adding decay rates improves predictive performance. Since decay rates are also able to predict cognitive performance, it would be useful to understand if control energy and decay rates together better predict cognitive performance or if they are capturing similar underlying variability which results in similar predictive performance when used in isolation. If it is the former, this would suggest that control energy and decay rates are capturing distinct sources of variability that are both important for cognition.

(Remarks on code availability)

It does not look like they added the code for running the cognitive performance prediction using the decay rates

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Authors' Response to Reviews of

Inferring macroscopic intrinsic neural timescales using optimal control theory

Jason Z. Kim, Richard F. Betzel, Ahmad Beyh, Amber Howell, Amy Kuceyeski, Bart Larsen, Caio Seguin, Xi-Han Zhang, Avram Holmes, Linden Parkes
Nature Communications

R: Reviewer's Comment, A: Authors' Response, □ Manuscript Text

1. Reviewer #1

- R:** *Since the seminal paper by Gu et al. (2015), network control theory (NCT) has been frequently applied in “network neuroscience” studies focusing on large-scale human brain networks (accessible through functional, structural and diffusion MRI). In principle, brain NCT would allow designing an exogenous control signal driving the brain into a desired activity state, and hence it could be used to inform brain stimulation protocols. However, fine-grained control of the brain state would require simultaneously delivering stimulation to a large number of brain sites, which is largely beyond current limitations of brain stimulation protocols. Thus, in practice, most of the brain NCT literature has focused on more descriptive tasks such as estimating the ‘cost’ of transitions between different brain states, encapsulated by the control energy (the integrated amplitude of the control signal). Exploring the control energy landscape can highlight nodes that are putatively more central for brain dynamics, assess the difficulty of reaching specific target states, and altogether quantify the ease of transitioning between different activity states - and variations in all these properties due development, aging, pathology, etc.*
- R:** *The most commonly used brain NCT model is fully linear, and combines diffusive dynamics along structural connections (as given by diffusion MRI), with local self-inhibition to ensure stability. The self-inhibition parameters, which are akin to local timescales, are usually taken to be uniform. In their manuscript entitled “Inferring macroscopic intrinsic neural timescales using optimal control theory”, the Authors augment NCT with a parameter optimization scheme that numerically optimizes the self-inhibition terms (‘timescales’) to ensure that trajectories remain close to the target state. Their methodology is applied to two human data sets and a rodent data set.*
- R:** *Their main results are the following:*
- the proposed scheme allows systematically reducing the control energy, and reducing the number of driver nodes needed for accurate control*
 - across brain regions, the optimized timescales correlate with empirically measured timescales (as revealed by fMRI autocorrelation), as well as with spatial variation in the abundance of different types of inhibitory neurons and related gene expression (revealed by gene expression maps).*
 - results of the human and rodent model are fully consistent*
 - the control energy landscape emerging from the timescale-optimized model allows better prediction of performance scores in a range of cognitive tasks*
- R:** *Overall, the study is methodologically clear, statistically well-powered, and findings are robust. The method-*

ology proposed is certainly an original development of NCT applied to large-scale neuroscience. The results are surely of large interest to the small community focusing on brain NCT. However, I think that the current presentation is not suitable to capture the attention of the wide neuroscience community. Therefore, I believe that substantial changes in presentation are needed to make it suitable for Nature Communications. The main problem lies in the overall ‘take home messages’, as they emerge from the Abstract, Introduction, Discussion and (partially) Results sections. They tend to overemphasize comparatively weak findings, while not placing enough emphasis on other, potentially relevant ones. In what follows, I will expound my main concerns on this point.

A: We thank the reviewer for their thoughtful summary of the key findings of our manuscript, and for their interest in our work as well as their deep review.

1.1. R1 comment #1

R: *Technical development.* The manuscript introduces a novel scheme to couple NCT with automatic differentiation. As the Authors correctly state, “So far no study has sought to optimize this constraint by modifying the connectivity to minimize the delta between $x(t)$ and x_r . Here, we extend NCT to allow for this minimization.” (Discussion) In fact, I think that this relevant technical development could be emphasized more. This scheme allows in principle asking new questions not only about timescales, but potentially also about connectivity. It could be, e.g., used to optimize non-diagonal entries in the matrix, too, and thus provide a link between optimal control and effective connectivity.

A: We thank the reviewer for emphasizing the broad applicability of our framework, including beyond uncovering timescales to much more general links between optimal control and effective connectivity. We agree with the reviewer that the broad applicability of our method warrants stronger emphasis. As such, in the revised manuscript, we have made several adjustments to the text to increase this emphasis. For brevity, we have not reproduced all of these changes here, but we refer the reviewer to the edits made in blue typeface throughout the revised manuscript. We do, however, list several notable changes below. Most notably, we now include a new paragraph in the future directions subsection of the Discussion that emphasizes this point.

From the revised Discussion:

A final future direction is to generalize our method to optimize inter-regional connectivity within the connectome. While the goal of the present paper was to uncover region-specific INTs—which we achieved through targeted optimization of the diagonal elements of A —our method can, in principle, be applied to optimize the off-diagonal elements of A as well. Doing so would allow NCT to uncover directed (i.e., effective) structural connectivity from an otherwise fundamentally undirected imaging modality (diffusion-weighted MRI). Such an approach is consistent with other methods from network science that are sensitive to asymmetries in inter-regional connectivity [1], but would provide the novel capacity to uncover context-specific shifts in directed communication on an undirected network. However, optimizing the off-diagonal elements of A carries with it a key challenge of model identifiability. Specifically, here we perform model optimization using the midpoint of the state trajectory, $x(t)$, to adjust the decay rates, d , that are embedded along the diagonal of A . Critically, both $x(t)$ and d have the same dimensionality, which is equal to N brain regions. Thus, our cost function has the same dimensionality as the parameters we seek to optimize. This parity enables strong identifiability in our model solutions (see Section ?? in the Supplementary Materials), which in turn facilitates reliable large-scale model fitting (e.g., to hundreds of subject-level connectomes). By contrast, if we were to also optimize the off-diagonals of A , we would be seeking to optimize N^2 parameters from an N -dimensional cost

function, which would lead to identifiability issues. Such identifiability issues may be mitigated via additional cost function terms such as optimizing more empirically-derived states along the trajectory, or incentivizing biophysically realistic conditions such as distance-based constraints [2] based upon known hierarchical organization [3, 4]. This an important avenue for future research.

From the revised Methods:

and we optimize D to minimize deviations between the state trajectory, $x(t)$, and the reference state, $\bar{x}_r$, at the midpoint of the simulation, $t = T/2$. For this optimization, we implement optimal control in the automatic differentiation framework (see Section ?? below) to perform backpropagation on d using the *PyTorch* library [5] first by normalizing the adjacency matrix (Eq. ??), then by numerically solving for the optimal trajectory $x^*(t)$, and finally by penalizing deviations of the optimal trajectory at the midpoint. **Note, while here we apply our algorithm to selectively modify the values of d —which are set along the diagonals of A during normalization—a strength of our approach is that it can, in principle, be applied more generally to optimize any part of A , including the off-diagonal entries.**

We also note that this comment ties directly in with R2C4, which focuses on potential non-uniqueness and identifiability issues of our model solutions. As discussed below (see R2C4), our model solution is highly identifiable precisely because it only optimizes the diagonals of A and not the off-diagonals as well. If we chose to optimize the off-diagonals, we would likely need to incorporate additional biophysical constraints, which opens the door for many exciting future extensions. As discussed above, such extensions are well within the scope of our algorithm in general, but we leave its implementation and exploration to future projects.

1.2. R1 comment #2

R: *Inferring biologically relevant parameters. From the very title, the Authors place a large emphasis on using their framework to infer biologically relevant timescales. In fact, all the evidence for a relation between NCT-optimized timescales and biology depends on the correlation with known neurobiological metrics. But what is special, then, about NCT-optimized timescales? I suppose that also ACF-based timescales show a similar degree of correlation with neurobiological metrics, and they are easier to infer from fMRI. One could argue that NCT-optimized timescales are more interpretable, as they represent ‘self-inhibitory’ terms in a linear dynamic model. However, one could obtain such terms by directly fitting a linear dynamic model to fMRI data (see, e.g., Nozari, Erfan, et al. "Macroscopic resting-state brain dynamics are best described by linear models." Nature biomedical engineering 8.1 (2024): 68-84.), without need to invoke NCT. Moreover, to what extent these terms in the linear model would really correspond to a macroscopic manifestation of microscopic inhibition is unclear. Thus, the claim that “we estimate the structural form of regional self-inhibition—a quantity that is experimentally difficult to measure—from MRI data” is not fully warranted in my opinion.*

A: We thank the reviewer for this comment, and elaborate on our reasoning for the claim. What is special about NCT-optimized timescales is that they are based on a generative dynamical model while being informed by known biological structural connectivity. Precisely as the reviewer has stated, as compared to ACF-based metrics, NCT-optimized timescales represent self-inhibitory terms in a linear dynamical model, providing numerous advantages such as interpretability. Hence, we focus on the advantage of our method compared to autoregressive approaches.

In contrast to directly fitting a linear dynamic model to fMRI data, our approach has two main advantages.

First, a purely autoregressive approach assumes that neural dynamics are homogeneous and do not receive directed targeted input, which fails to account for changes in neural activity induced by non-homogeneous inputs. Second, a purely autoregressive approach does not incorporate any information about known inter-regional interactions supported by structural connectivity. As such, a purely autoregressive approach will produce a linear model whose interactions between brain regions deviate from empirically measured structural connectivity. In contrast, our method assumes dynamics unfolding on top of known structural connectivity, allowing for the explicit incorporation of known empirical data.

Nevertheless, we agree with the reviewer that our choice of expression in the original manuscript may have been misleading, especially given that neural timescales can be straight forwardly estimated from fMRI data. As such, we have modified the sentence in question located in the Abstract, as well as generalized revised the Abstract for greater clarity and precision:

Abstract

The temporal evolution of **whole-brain activity is contingent** upon complex interactions within and between brain regions that are mediated by neurobiology and connectivity, **respectively**. To understand these interactions, many large-scale **modeling** efforts have measured structural connectivity, **brain** activity, gene expression, and cognition across multiple modalities and species. However, **accurate modeling of the brain’s traversal through distinct brain states remains difficult. Uncovering new insights requires data-driven** frameworks for estimating the interplay between brain structure and function while preserving biophysical **detail**. Here, we provide such a framework by integrating network control theory (NCT) with automatic differentiation to estimate model parameters with greater biophysical realism from data. Specifically, we estimate **regions’ intrinsic neural timescales (INTs) using NCT, and find that doing so broadens the range of dynamics supported by the connectome and improves the alignment between the brain’s connectivity and its traversal through state-space**. Next, we demonstrate that the resulting model-based **INTs** correlate with INTs **measured empirically from functional neuroimaging data**, neurobiological measures of gene expression and cell-type densities, **as well as** behavioral measures of cognition. We demonstrate consistent results across multiple datasets and species. Finally, we demonstrate that our **model-based INTs** enable the efficient control of brain **states** from fewer brain regions. Taken together, our results provide a simple and flexible quantitative framework that more accurately captures the interplay between brain structure, function, and dynamics with greater biophysical realism.

An additional point regarding the uniqueness of NCT-optimized timescales is that they are fit to specific state transitions. This capacity allows us to not only model regions’ intrinsic timescales, but to also model context-dependent modulations in those timescales. Critically, this transition-specific approach to model fitting is unique compared to alternative dynamical systems approaches. For example, while large-scale coupled neural mass models (NMMs) can also infer neural inhibition from MRI data, they do so by fitting highly parameterized non-linear models to time-resolved functional connectivity estimates. This produces regional neural inhibition parameters designed to produce neural time series that, when correlated pairwise, capture statistical properties of empirical FC matrices. While NMMs provide their own important insights, they are unable to model regional self-inhibition parameters that support transitions between specific whole-brain patterns of activity defined by the researcher. Our NCT model fills this gap. We discussed this advantage in the Results and Discussion sections of the original manuscript, but we appreciate that its emphasis may not have been as strong as it could have been. As such, we have strengthened this emphasis in the revised manuscript, as well as made references to future work where we plan to explore the benefits of our approach further.

From the revised Results section:

Collectively, the above results demonstrate that our model-based INTs emerge as a product of jointly optimizing on real biological connectivity and activity, and that these effects are in tension with one another. Specifically, our results show that the brain’s macroscale neural state-space is relatively difficult to traverse, and that the connectome is wired to support this traversal as efficiently as possible. Critically, our findings suggest that mesoscale INTs act as a conduit that manage this tension. That is, while the spatial patterning of our model-based INTs is clearly anchored by the connectome, they remain able to flexibly update to meet the needs of a specific state transition. Thus, our model-based INTs have both a stable and a flexible component, and the brain leverages these two components to orchestrate the coupling between macroscale structure and function. This conclusion illustrates a critical advantage of our approach, namely the ability to model context-dependent shifts in regions’ INTs in response to specific state transitions defined *a priori*.

From the revised Discussion section:

Here, we observed that modifying the self-inhibition parameters in NCT consistently reduced the control energy needed to move between brain activity states. In the standard implementation of NCT (INT_{uniform}), control signals are tasked with controlling brain activity by jointly leveraging the diverse connectivity between nodes as well as the fixed internal dynamics within nodes [6-8]. By contrast, our approach enables control signals to leverage both diverse extrinsic connectivity and diverse intrinsic dynamics to control **communication between regions. Thus, our results demonstrate that brain regions’ intrinsic dynamics, not just their inter-connectivity, play an important role in enabling the brain to traverse its neural state-space.** Notably, unlike the connectivity between nodes, these diverse intrinsic dynamics are not set as inputs, rather they are **learned** by NCT during model fitting. In turn, **compared to the INT_{uniform} model, the added parameter flexibility in the $INT_{\text{optimized}}$ model allows it** to find more efficient state transitions by uncovering, and then harnessing, transition-specific regional INTs. We found that this added model flexibility yielded reduced control energy for the $INT_{\text{optimized}}$ model, indicating that our approach narrowed the gap between **macroscale** brain structure and function. This observation is consistent with previous studies documenting increased structure-function coupling in network models that embrace more complexity in their assumptions about neural signaling [1, 9, 10].

We found that the difference in control energy between the two models was driven jointly by the topology of the connectome and the spatial patterning of the state transitions. Specifically, using a pair of null models [11, 12], we observed that the effects of connectome topology and the state transitions were in tension with one another. **In the $INT_{\text{optimized}}$ model, control energy was significantly lower** than expected when compared to a connectome with edges that were rewired, and significantly **higher** than expected compared to when the target states were permuted. This effect demonstrates that uncovering INTs using our approach required balancing the competing influences of connectome topology and the spatial properties of the brain states. **Furthermore, this effect delivers two key insights. First, from a methodological perspective, this effect shows that the outputs of our algorithm were not a trivial consequence of either of the key model inputs (i.e., the connectome or the brain states). Rather, uncovering regional INTs in our model required both inputs. Second, and more importantly, this effect suggests that our model-based INTs exhibit both a stable and a flexible component, where the former is rooted in the human connectome and the latter is able to respond to the demands of specific state transitions. Previous research suggests that empirically-measured INTs exhibit a stable component that varies systematically across the cortex [13] as well as throughout**

development [14], and **that this component** likely explains similarities in INTs between rest and task contexts [15, 16]. Our results indicate that the stable component is tied to the topology of the human connectome. However, while regions' INTs are tightly constrained by a relatively static connectome topology, they are also inherently dynamic, which allows them to flexibly respond to different contexts (i.e., state transitions). **Our results demonstrate that our model is sensitive to these dynamic shifts in regional INTs, thus giving rise to a variability in regions' INTs across state transitions.** This **flexible component** is consistent with previous work showing that while INTs exhibit similar global structure across rest and task conditions, they also show clear modulation in response to specific task demands [16, 17].

From the revised future directions subsection in the Discussion:

A third future direction is to use our $INT_{\text{optimized}}$ model to study how the brain transitions from being at rest to engaging in cognitively demanding tasks. As discussed above, a key advantage of our model is its capacity to uncover both stable and flexible components of regions' INTs. Here, using only a set of transitions between resting-state activity patterns, we observed unique contributions of both of these components in our results. However, as noted above, previous work has shown that empirically-measured INTs change systematically in response to task demands [15, 16]. Thus, an important avenue for future work will be to examine how our model-based INTs dynamically reorganize during cognitive tasks.

R: *Also not fully supported is the claim that “These results robustly validate our approach as a means of uncovering biologically-grounded whole-brain INTs directly from the connectome in vivo.” In fact, NCT-optimized timescales do not solely depend on the connectome, but also from the chosen set of transitions, which in turn is derived from functional data.*

A: We thank the reviewer for this comment, and agree that NCT-optimized timescales depend both on the connectome and the chosen transitions. We note that we acknowledged the joint influence of the connectome and functional data on INTs in the original manuscript. Indeed, results from our twin null models, as well as our discussion of stable and flexible components of INTs, explicitly acknowledged this. In addition, we have made several revisions to the text (see above) that expand this discussion. Nevertheless, we acknowledge that the language here may have overemphasized the connectome's contribution. As such, we have revised this statement in the revised manuscript:

Together, these results robustly validate our approach as a means of uncovering biologically-grounded whole-brain INTs from the human connectome *in vivo*.

Above we have removed the word "directly" so as to avoid implying that the connectome is the only ingredient needed in our model to uncover INTs. However, we note here that there are circumstances where a researcher would be able to estimate INTs using our model even if all the way had was their own connectome data. Strictly speaking, all researchers need is a connectome and brain states, but the latter needn't be derived from functional MRI data. Instead, a researcher could model INTs using brain states derived by other means (e.g., a canonical system-level organization of the brain such as the Yeo atlas).

R: *Finally, if one claims a strong relation with neurobiology, the dependency of the timescales on the chosen transition appears problematic. In fact, the argument “A key advantage of our data-driven approach is that it enables us to uncover both transition-specific and subject-specific INTs, and to use them as proxies for measures of neurobiology that are otherwise unobservable in vivo.” (Discussion) is not fully clear: it would*

seem that the transition-dependence of optimized NCT rates actually undermines their usage as ‘proxies’ for (static, transition-independent) neurobiological measures.

- A: We thank the reviewer for this comment. We appreciate the reviewer’s point that there is apparent tension between transition-specific INTs and correlations with static measures of neurobiology, and are happy to clarify our intention here. Central to our argument is that INTs estimated for any given state transition always encode a stable component mixed with the transition-specific (i.e., "flexible") component. This mixing of stable and flexible components is illustrated by the fact that our twin null models exert opposing effects on control energy, which we discussed in the original manuscript but have significantly reworked in the revised manuscript (see below response to R1C7). That is, if INTs only encoded a stable component, then they would only be driven by the connectome. Likewise, if INTs only encoded the transition-specific flexible component, then they would only be driven by the specific state transitions. Neither of these situations are the case. Instead, we observe that our model-based INTs strike a balance between stable and flexible components during model fitting, hence giving rise to the opposing null effects we report. Furthermore, we only demonstrate correlations with static neurobiological measures using a PCA of the INTs performed across all state transitions. By design, this PCA extracts the dominant sources of variations in INTs, which we interpret as broadly capturing the stable component. To avoid confusion for readers, we have adjusted the aforementioned sentence in the revised manuscript:

While both approaches are viable, a key advantage of our data-driven approach is that it enables us to uncover both transition-specific and subject-specific INTs, and to use **their common (i.e., stable) component as a proxy** for measures of neurobiology that are otherwise unobservable *in vivo*.

1.3. R1 comment #3

- R: *Applicability to brain stimulation. The applicability of the NCT paradigm (and its timescale-optimized version) for brain stimulation is likely overstated. “These results indicate that our approach—which seeks to improve the biophysical realism of NCT—enables the successful control of brain dynamics from relatively focal inputs. In turn, this improvement to model fitting reduces the gap between NCT and real-world applications of brain stimulation.” “this added parameter flexibility allows the INToptimized model to find more efficient state transitions by uncovering, and then harnessing, transition-specific regional dynamics” (Discussion) Control energies required for fine-grained activity control are exceedingly high unless a sizable fraction of the nodes is controlled with external inputs (see Tu et al, NeuroImage 176 (2018): 83-91, but also Fig. 3C in the manuscript). In practice, control of more than 50 nodes is required, when current techniques allow at best for 1-2 nodes. The Authors refer to “relatively few control nodes”, but 50 is not few, in the light of current technical limitations. In addition, the timescale-optimized version of NCT does not clearly contribute to closing the gap, because, contrary to the input signals, the timescales are not externally controllable parameters. Thus, while transition-optimized timescales would allow to enact the transition more easily, one does not see a simple way to externally manipulate the timescales to facilitate a preferred transition. This point is not clearly discussed in the manuscript. (Perhaps one could think of using pharmacology to manipulate local inhibition, thus altering timescales locally to facilitate specific transitions?)*

- A: We thank the reviewer for this comment. We agree with the reviewer on the number of nodes that are practically controllable, and would first like to clarify the intended meaning of our statements.

To clarify, when we state that “this added parameter flexibility allows the INToptimized model to find more efficient state transitions by uncovering, and then harnessing, transition-specific regional dynamics,” we do not propose to modify the biological timescales through pharmacology. Rather, we are making a comparison between the INToptimized NCT model and the INTuniform NCT model, the latter of which does not allow for heterogeneity in the self-inhibition. We apologize for the confusing phrasing, and modify the statement in

the main text:

In turn, **compared to the INT_{uniform} model, the added parameter flexibility in the $INT_{\text{optimized}}$ model allows it to find more efficient state transitions by uncovering, and then harnessing, transition-specific regional INTs.**

Additionally, we consider our statement of “relatively few control nodes” to be accurate in its context, which was with respect to the standard NCT model that uses a full control set. However, we agree with the reviewer that, given the current practical limitations in brain stimulation, more subtly is warranted. As such, we revised the statement:

To address this gap, here—as well as in our previous work [18]—we developed a novel extension to NCT that allowed us to model state transitions successfully from **a relatively small proportion of control nodes compared to the standard NCT model.**

Finally, we have added a limitation to the Discussion section that specifically acknowledges this issue:

Third, while our $INT_{\text{optimized}}$ model improved the control of brain dynamics from sparse control sets, we were still only able to complete state transitions successfully using a relatively large number of control nodes (here, ~ 64) compared to that which would be stimulated via neurostimulation techniques such as TMS. Finally, ...

1.4. R1 comment #4

R: *Timescales determine the control energy landscape. In my opinion, one of the most relevant findings of the manuscript is the robust evidence that timescales, not just connectivity, significantly contribute to the control energy landscape. This finding is not totally surprising, since diagonal terms significantly impact on the spectrum, and hence on the general controllability of the system. However, this was not discussed by previous literature, and it may deserve more stress. While the correlation between optimized timescales from NCT and empirical timescales is highlighted as evidence that the optimized timescales are biologically grounded, it may be also (more convincingly, to me) be brought as evidence that observed empirical timescales might be optimized to ensure easier transitions between frequent activity states in spontaneous brain dynamics. In other words, the NCT results might contribute to explaining the spatial structure of observed timescales, and its changes during development/aging/pathology, in terms of optimal control. The Authors seem to recognize this when they say “Our approach has the potential to connect these entwined phenomena to better understand how the multi-scale relationship between brain structure and function emerges throughout development.” (Discussion)*

A: We thank the reviewer for this insightful comment, and wholeheartedly agree that the contribution of timescales to the control energy landscape—while not completely unexpected—is a significant and broadly-relevant frontier, and that our methods now enable the quantitative exploration of relationships between entwined phenomena such as transition frequency and physiology.

R: *In fact, additional analyses may even strengthen this perspective, with possible impact on the general understanding of large-scale brain function. Possible analysis steps could be (in order of feasibility):*

1. *One may conjecture that timescales in single subjects are optimized to facilitate transitions between the most frequently occurring CAPs. One should then observe a stronger similarity between the empirical timescales in single subjects and the NCT timescales optimized for the CAP transitions occurring most frequently in the single subject.*

2. One could conjecture that individuals whose empirical timescales are more ‘optimal’ would display enhanced cognitive abilities. One could then use the HCP dataset to compute the correlation between the average (over transitions) optimized timescales and the empirical timescales in a single subject basis, and in turn test whether this correlation is predictive of cognitive performance
3. Using the task data in the HCP data set, one could compute the optimized timescales to transition from an average ‘rest’ state to the average ‘task’ state and back, and check whether the empirical timescales during the task correlate with these optimal timescales. This would provide evidence that the brain can flexibly adapt timescales to facilitate specific transitions.
4. Using data from the HCP aging data set, one could analyze changes in the optimized timescales during aging. These optimal timescales would adapt to, and potentially compensate for, aging-related changes in the connectome. If aging-related changes in optimal timescales are reflected in aging-related changes in empirical timescales, this would again constitute relevant evidence that timescales are a key mechanism to confer flexibility and resilience to brain dynamics.
5. In neurological disease, the connectome is often irreparably damaged. Can a timescale adaptation partially compensate for this? An analysis of empirical and optimized timescale changes in, e.g., a stroke dataset, would be of significant clinical interest.

R: Of course I am not asking the Authors to perform all these analyses. But I think that adding one or two of them, or analogous ones, might strengthen the article.

A: We thank the reviewer for these excellent suggestions. Given the scale and length of this project, we have chosen to focus primarily on the first suggestion in the reviewer’s list and to leave the remaining suggestions for future studies; though, we touch on the second suggestion as well (see below). To that end, we have added a new analysis to the revised manuscript examining how empirically-observed subject-specific transition probabilities (i.e., from the rs-fMRI data) correlate with the coupling between subject-specific optimized decay rates (model-based INTs) and subject-specific ACF_{lag} .

From the revised Results:

Optimizing regions’ internal dynamics leads to improved prediction of cognition

Above, we demonstrated that our $INT_{\text{optimized}}$ model gives rise to more efficient state transitions (Fig. R2 Fig. R3) and yields interpretable maps of brain regions’ intrinsic dynamics that correlate with empirical INTs (Fig. ??), gene expression and cell-type distributions (Fig. ??). However, all of these analyses were performed on a group-averaged connectome. One of the key advantages of NCT is that it is highly amenable to fitting to single-subject data [6]. Thus, in our final analysis, we fit our NCT model to individual connectomes. Briefly, we fit both our INT_{uniform} and $INT_{\text{optimized}}$ models to each of 960 subject-specific HCP-YA connectomes and **performed the following two analyses. First, we examined whether the subject-specific coupling between model-based INTs and empirically-measured INTs (ACF_{lag}) varied as a function of the transition probabilities observed in subjects’ rs-fMRI data (see Materials and Methods). Second, we examined the correlation between subject-specific control energy and subjects’ scores on a battery of behavioral tasks (see Materials and Methods). Note, in both of these analyses, we used the same CAPs extracted at the group-level (see Fig. R2A) as brain states for subject-specific NCT analysis. This approach allowed us to isolate inter-individual differences in NCT outputs to individual variability in connectome topology.**

Regarding the first analysis, we correlated subject-specific optimized decay rates with each subject’s whole-brain map of ACF_{lag} . For each state transition, this correlation coefficient quantified the extent to which model-based INTs coupled to empirical INTs measured at rest. Next, using each subject’s rs-fMRI timeseries data, we computed the probabilities associated with transitioning between each of our brain states (cf. Fig. R2A). When averaged over state transitions, we observed a positive correlation between transition probabilities and the coupling between model-based and empirical INTs (Fig. R9A, top). Additionally, we found that this effect held at the within-subjects level (Fig. R9A, bottom). Specifically, Fig. R9A bottom shows the coupling between model-based and empirical INTs averaged separately over the most frequent state transitions (upper 50%) and the least frequent (lower 50%). A paired samples t -test indicated that coupling was significantly stronger for the most frequent state transitions ($t(959) = 9.92, p = 3 \times 10^{-22}$). These results demonstrate that subjects who exhibit more frequent state transitions during rest also show stronger coupling between their empirical INTs and their (stable) model-based INTs, and that model-based INTs selectively align more strongly with more frequent transitions at the single-subject level. In turn, this result suggests that subjects’ resting-state INTs may be spatially configured to facilitate efficient transitions between distinct brain states, and that our model is sensitive to individual differences in the optimality of this spatial pattern.

Regarding the second analysis, using a cross-validated penalized regression model, we examined whether control energy from our $INT_{\text{optimized}}$ model outperformed the INT_{uniform} model at out-of-sample prediction of behavior (see Materials and Methods for more details, as well as Fig. R8 for results using optimized decay rates to predict cognition). While we examined a broad range of cognitive variables present in the HCP-YA dataset, here we only report results for variables that yielded positive explained variance (R^2) scores in our regression model. Across 100 repeats of 5-fold cross-validation, we found that control energy from the $INT_{\text{optimized}}$ model outperformed the INT_{uniform} model across all behavioral measures that met this criteria (Fig. R9B). Additionally, empirical null models (see Materials and Methods) revealed that prediction performance for both variants of control energy exceeded chance levels (Fig. R9B, insets; all $ps < 0.05$). These results show that optimizing the internal decay parameters of our NCT model at the single-subject level leads to a significant improvement in the out-of-sample prediction of cognition in the HCP-YA.

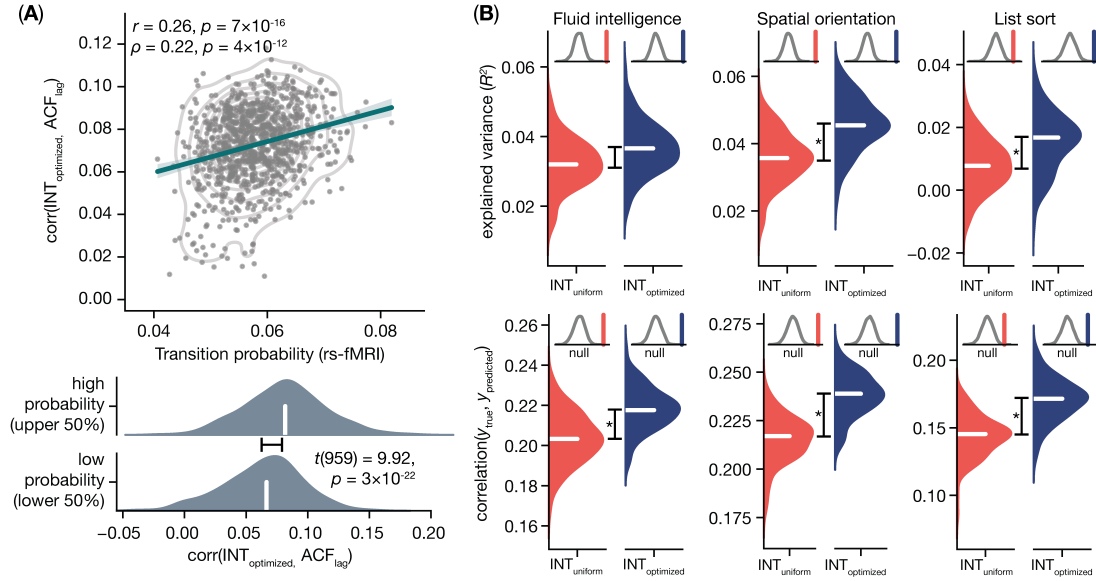


Figure R1: **Fitting the $\text{INT}_{\text{optimized}}$ model to subject-level data yields sensitivity to individual differences in empirical INTs and improves the out-of-sample prediction of cognition.** (A) **Correlations between subject-specific optimized decay rates (model-based INTs) and subject-specific ACF_{lag} (empirical INTs).** In the top plot, we calculated the transition probabilities for all state transitions as they occurred in each subject's rs-fMRI timeseries data (x-axis). Next, we calculated the correlation between model-based INTs and empirical INTs for each subject (y-axis). Then, we averaged these quantities over state transitions and correlated them. We found that subjects who exhibit more frequent state transitions during their rs-fMRI scan exhibit stronger coupling between their model-based INTs and their empirical INTs. In the bottom plot, the same analysis was performed focusing on within-subjects effects. Here, the correlations between model-based INTs and empirical INTs were averaged separately over the most frequent state transitions (upper 50%) and the least frequent (lower 50%). A paired samples t -test indicated that coupling was significantly stronger for the most frequent state transitions ($t(959) = 9.92$, $p = 3 \times 10^{-22}$). (B) **Results from a brain-based predictive modeling analysis that used subject-specific control energy to predict individuals' cognition.** We extracted subject-specific control energy from both $\text{INT}_{\text{optimized}}$ and $\text{INT}_{\text{uniform}}$ models and used those features as predictors in separate cross-validated penalized regression models. Control energy from the $\text{INT}_{\text{optimized}}$ model outperformed that of the $\text{INT}_{\text{uniform}}$ model at predicting subjects' cognitive scores. **The top row shows prediction models scored according to out-of-sample explained variance (R^2), while the bottom row shows models scored according to the correlation between true and predicted scores.** See Fig. R10 for models scored according to root mean squared error (RMSE). Permutation testing revealed that energy from both models performed above chance levels (insets). $*$ = $p < 0.05$ for an exact test of differences between paired performance distributions.

From the revised Methods:

Transition probabilities

As mentioned in the main text, we examined how the transition probabilities occurring in subjects' rs-fMRI data correlated with the coupling between their empirical INTs and model-based INTs. To conduct this analysis, we computed the transition probabilities between each of our brain states (see Fig. R2A) separately for each subject using their rs-fMRI data. This analysis was possible because the k -means clustering analysis we performed above to extract the brain states (see Section ??) also labeled each time point in the rs-fMRI data as belonging to one of our seven brain states. We used these labels to count the frequency with which the time points assigned to a given brain state transitioned to a different brain state at the next step. Repeating this process between all pairs of brain states yielded a transition probability matrix for all of our state transitions. We calculated this matrix for each subject separately.

As can be seen from these new analyses, we find support for R1's conjecture that single subjects are optimized to facilitate transitions between the most frequently occurring CAPs.

Note, while we did not explicitly perform analysis for the second suggestion from R1's list, we did add a new BPM analysis that used subject-specific optimized decay rates instead of control energy to predict cognition. This new analysis was added in response to R3C5, but can also be seen in the above revised text. This new BPM analysis speaks to R1's query (suggestion #2) about whether subject-specific optimized decay rates explain variance in cognition.

1.5. R1 comment #5

R: *Methodology.* While the methodology is clearly detailed, some additional information could be useful. First, many brain NCT articles focus on minimizing just the control energy (with no constraint on the trajectory). Could one try to directly minimize the control energy with respect to timescales? After all, the control energy can be expressed analytically in terms of the initial and final states and the controllability Gramian, so it would not seem impossible to use this as a cost function. Would this approach lead to significantly different results? Secondly, can the approach (which leverages the Hamiltonian framework) be used for asymmetric connectomes? Third, what was the time horizon chosen for control?.

A: We thank the reviewer for raising these important methodological questions and clarifications. We address the reviewer's multi-part question in order.

1. "Could one try to directly minimize the control energy with respect to timescales?"

Yes, the control energy can be directly minimized with respect to timescales. This minimization is most easily done in the minimum control energy framework, where the minimum control energy can indeed be expressed analytically in terms of the initial and final states and the controllability Gramian,

$$E^* = (e^{AT} \mathbf{x}_0 - \mathbf{x}_T)^\top W_c^{-1} (e^{AT} \mathbf{x}_0 - \mathbf{x}_T).$$

However, the optimal control framework requires the simultaneous construction and evolution of the states and co-state lagrange multipliers, the conversion of the co-states to the control inputs, and the integration of the inner product of those inputs through the time horizon. This inner product integral may not be able to be written analytically as an algebraic expression, mainly due to the fact that the state/co-state equation includes a matrix exponential that couples the state and co-state dynamics, and

the integrand would require an inner product in just the co-state variables of the form

$$e^{M^T t} \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} e^{Mt},$$

which may not have an integral that can be evaluated into an algebraic form. However, one can still minimize the control energy in the optimal control framework by numerically sampling the co-state trajectories. Given the length and methodological complexity of our paper currently, we leave exploration of this suggestion to follow-up work.

2. "Would this approach lead to significantly different results?"

While we do not know whether minimizing the control energy directly would lead to significantly different results from all of those presented in the paper, it would reproduce one of the main findings of the paper; that control energy is lower for the optimized model compared to the uniform model. The reason we do not explicitly minimize the control energy is because, consistent with other literature applying network control theory in neuroscience, we use control energy as a dependent variable and a quantity of merit. Specifically, our goal is to study whether optimizing the model's self-inhibition according to empirically-derived fMRI states results in lower control energy, and thus a tighter alignment between the network's structure (connectivity) and function (empirical state transition) due to requiring less external compensation (energy). Optimizing NCT on the state trajectory (and not energy) allows us to preserve our use of control energy as a quantity of merit, while explicitly minimizing it would undermine that goal.

3. "Secondly, can the approach (which leverages the Hamiltonian framework) be used for asymmetric connectomes?"

Yes, the optimal control framework leveraging the control Hamiltonian can be used with no loss of generality to asymmetric connectomes. In fact, our application to the mouse uses a directed mouse connectome with asymmetric connections. We mentioned this in the original manuscript:

Results replicate across datasets and species.

In this section, we examined whether our model replicated in a second human dataset and generalized across species. Specifically, we replicated our key results using an undirected human connectome taken from the MICA-MICs dataset as well as a **directed mouse connectome** taken from the Allen Mouse Brain Connectivity Atlas (see Materials and Methods).

4. "Third, what was the time horizon chosen for control?"

Throughout, we chose a default time horizon of $T = 1$. We apologize for failing to include this detail, and now include it in the revised methods:

where $\mathbf{x}(t)$ is an N -dimensional vector of activity across N brain regions, $\mathbf{u}(t)$ is a k -dimensional vector of inputs, A is an $N \times N$ matrix of connectivity strength between brain regions, and B is an $N \times k$ matrix of connections from control inputs to brain regions. Here, we take $B = I$ to be the identity matrix where $k = N$ such that each region receives its own independent control input. This setup is known as a *uniform full control set* [6]. **As per our previous work [6], we set the time horizon of our model to $T = 1$.** The goal of linear optimal control theory in NCT

is to find the input $u(t)$ that transitions the system from an initial state, x_0 , to a final state, x_T , while minimizing the following cost,

1.6. R1 comment #6

R: *A flexible structure-function coupling. In the Introduction and Discussion, the Authors often refer to ‘structure-function coupling’ (SFC), but the usage of this term, and the related discussion, are not fully clear. In the literature, SFC generally indicates the correlation between structural and functional connectivity (at the nodal or whole-brain level). The Authors seem to use ‘SFC’ in a different meaning, as the general relation between ‘multiscale structural parameters’ (the connectome and local neurobiology) and the generation of common activity states, “NCT has been able to flexibly probe how brain structure constrains brain function” “However, these applications of NCT have relied upon a set of simplifying assumptions that lack biophysical realism, limiting the capacity for NCT to yield multi-scale insights into SFC.” “using optimized decay rates yields lower control energy associated with transitions between empirical brain states, demonstrating stronger SFC” (Introduction) The terminology should be rigorously defined (what is intended exactly for ‘SFC’, and for ‘multiscale’), as the authors emphasize this point several times: “Understanding this multi-scale structure-function coupling (SFC) in the brain is a core goal of network neuroscience” “These effects give rise to a spatial pattern of SFC that varies systematically across the brain” (Introduction)*

A: We thank the reviewer for this comment, and we agree that our use of the term ‘structure-function coupling’ (SFC) was overly broad and, in some cases, underspecified in the original manuscript. We have gone through the entire manuscript and improved our precision of language for this term. Additionally, we have made numerous revisions to the text to clarify our meaning and use of the term ‘multi-scale’. For brevity, we have not reproduced our text edits here, since they are numerous and distributed. Instead, we refer R1 to the revised manuscript, where text revisions are present in blue typeface.

1.7. R1 comment #7

R: *Null models for the timescales. In my opinion, the discussion of null models in §2.3 is not very clear. The Authors present two null models, one that randomly rewires the connectome (preserving key network features) and one that randomizes the initial/target states (preserving spatial autocorrelation). The control energy gain afforded by optimizing timescales is significantly lower for the first model, and significantly larger for the second one. To me, it is not fully clear how these results can be interpreted, and the Authors’ interpretation does not seem fully convincing. “These results provide robust evidence that our optimized decay rates are jointly influenced by the topology of the connectome and the spatial patterning of the target states; as opposed to being driven exclusively by one or the other” “The impact of the connectome and the brain states on the optimized decay rates are in tension with one another, and that our model balances these competing influences throughout optimization.” “Together, these results demonstrate our model-based INTs emerge as a product of jointly optimizing on real biological connectivity and activity” In practice, the Authors deduce that there is a “tug of war” effect, where the connectome and the transitions have opposite effects on the timescales. Actually, to demonstrate such an effect, one should observe two opposing “forces” acting on the timescales, e.g., with the connectome ‘pushing’ for lower values of a given timescale and the transitions ‘pushing’ for higher values (or vice versa). This effect is not observed (to me, it is not even clear which analysis could reveal it). What is observed is only an effect on the control energy gain (the control energy reduction gained by optimizing timescales), which just implies that in presence of random transition states (which will generally be ‘difficult-to-reach states’) the control energy gain is larger than with frequent transition states, while in presence of a randomized connectome (which will probably be a slightly more regular, exponential-decreasing-rule connectome) the energy gain is lower (perhaps because the randomized connectome is quite conducive to transitions?). These results are difficult to interpret, unless also the bare control energy values are shown in addition to the energy gains.*

- A: We thank the reviewer for this comment, and appreciate that our original analysis was not as clear as it could have been. We also agree that the original *tug of war* language was not helpful. As such we have removed it from the manuscript everywhere. Critically, we have now substantially reworked this section of the results. This reworking has involved changing the way we use our twin nulls. Whereas before we applied the null to the difference in energy between our two models, now we only apply it to the $INT_{\text{optimized}}$ model. This has allowed us to draw clearer insights into our results, and articulate them more plainly.

From the revised Results:

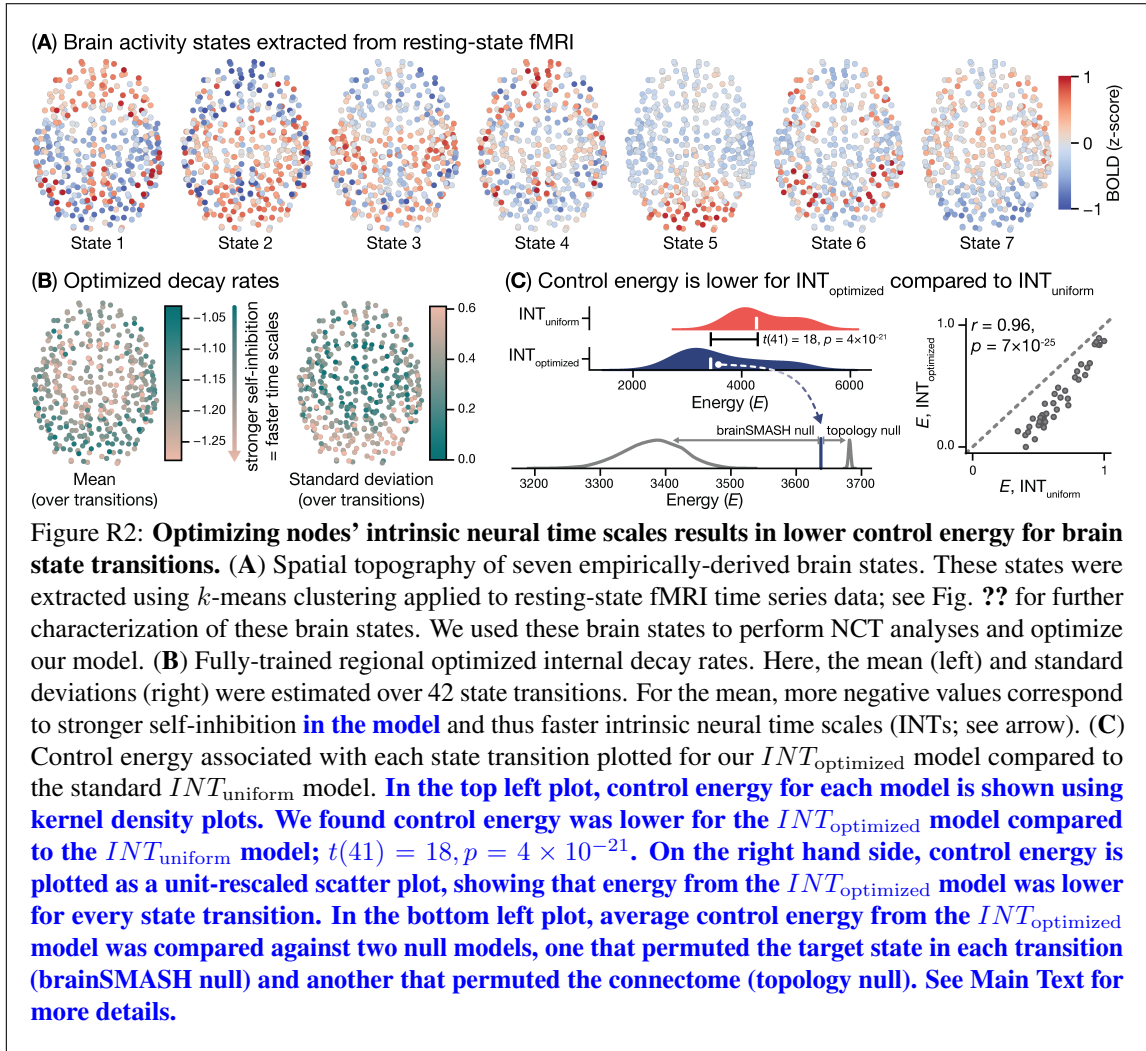
Our $INT_{\text{optimized}}$ model is designed to increase the biophysical realism of NCT by modeling biological variation in the node's decay rates. **Once our model was trained** (see Fig. ?? for model training), we plotted the mean and standard deviation (**across all state transitions**) of the optimized decay rates, and **found** that they **were both** spatially-patterned **and transition-specific** (Fig. R2B; see Fig. ?? for transition-specific optimized decay rates). The increased realism of our $INT_{\text{optimized}}$ model should more accurately **capture mesoscale intrinsic** brain dynamics, which we predict will lead to stronger **coupling between macroscale connectivity and activity and thus** lower control energy. To examine this prediction, we calculated control energy, E , separately using both our INT_{uniform} and $INT_{\text{optimized}}$ models. To control for differences in the overall level of self-inhibition between the two models, we set all values of d for the INT_{uniform} model to the mean of d from the $INT_{\text{optimized}}$ model ($d_{\text{uniform}} = \bar{d}_{\text{optimized}}$). As predicted, we found that control energy was lower for $INT_{\text{optimized}}$ compared to INT_{uniform} (Fig. R2C, top left; $t(41) = -18, p = 4 \times 10^{-21}$). Furthermore, this reduction in control energy was highly consistent across state transitions (Fig. R2C, right; $r = 0.96, p = 7 \times 10^{-25}$). Critically, these effects are not-trivial because the decay rates are optimized to minimize the distance between the controlled state trajectory and a reference state, and *not* the control energy. **Control energy is only calculated after training. Thus,** these results demonstrate that our extension to NCT allows for more efficient control of **macroscale** neural dynamics from connectome topology, which in turn narrows the gap between brain structure and function.

To examine whether optimized decay rates conferred a significant energetic advantage specifically for real biological connectivity and activity, we compared **mean control energy from the $INT_{\text{optimized}}$ model** against the following pair of null models [I9]. First, to quantify the influence of connectome topology on energy, we randomly rewired the connectome 5,000 times subject to constraints on nodes' spatial embedding and the network's strength distribution [I1], and for each we retrained our $INT_{\text{optimized}}$ model. Second, to quantify the influence of the specific state transition on energy, we randomly permuted the values of the target states—and thus the reference states—while preserving the spatial autocorrelation of the fMRI data (brainSMASH [I2]), and for each we retrained our $INT_{\text{optimized}}$ model. We found that energy was significantly **higher** than expected **by chance** under the brainSMASH null (Fig. R2C, brainSMASH null) and significantly **lower** than expected **by chance** under the topology null (Fig. R2C, topology null). **Furthermore, we observed that these opposing effects on control energy increased gradually as a function of the distance between the permuted data and the empirical data across permutations (Fig. ??), indicating a systematic effect.**

These results provide several key insights. First, they provide robust evidence that our optimized decay rates are jointly influenced by the topology of the connectome and the spatial patterning of the brain states; as opposed to being driven exclusively by one or the other. **Second, the spatial patterning of our brain states, as well as their relationship to each other (see Fig. ??B), confers a lower limit on the control energy achievable by our model in these data. Evidence for this insight comes the fact that energy was reduced under the brainSMASH null, which suggests that transitioning**

to permuted target states was easier than between the empirically-observed ones. Thus, the state-space traversed by the brain at rest requires more effort than traversing between states not observed in the fMRI data. Third, the topology of the human connectome is optimized to complete these relatively difficult state transitions as efficiently as possible. Evidence for this insight comes the fact that energy increased under the topology null. Thus, the opposing null effects reported here show that our model balances the competing influences of the connectome and the state-space while optimizing nodes' internal decay rates.

Collectively, the above results demonstrate that our model-based INTs emerge as a product of jointly optimizing on real biological connectivity and activity, and that these effects are in tension with one another. Specifically, our results show that the brain's macroscale neural state-space is relatively difficult to traverse, and that the connectome is wired to support this traversal as efficiently as possible. Critically, our findings suggest that mesoscale INTs act as a conduit that manage this tension. That is, while the spatial patterning of our model-based INTs is clearly anchored by the connectome, they remain able to flexibly update to meet the needs of a specific state transition. Thus, our model-based INTs have both a stable and a flexible component, and the brain leverages these two components to orchestrate the coupling between macroscale structure and function. This conclusion illustrates a critical advantage of our approach, namely the ability to model context-dependent shifts in regions' INTs in response to specific state transitions defined *a priori*.



1.8. R1 comment #8

R: *Other models of large-scale activity.* The Authors say that their proposal comes in a situation where biologically-grounded, data-driven models of whole-brain activity are lacking: “However, data-driven discovery of large-scale activity models remains difficult owing to the lack of flexible quantitative frameworks for estimating the interplay between brain structure and function while preserving biophysical realism” (Introduction). I do not fully share this view. Recent literature features several hybrid biophysical-data driven models. Typical examples are Refs. 67, 80, 81 in the manuscript. It is not clear what would be the specific advantages of a NCT approach compared to whole-brain modeling approaches in these works. It is also not clear how the flexible component of timescale could be incorporated into dynamical models, that typically feature ‘static’ parameters “This flexible component has implications for studies seeking to integrate spatially-patterned features of neurobiology into dynamical systems models”. By the way, the approach in Ref. 67 could also inspire a different version of timescale optimization in NCT. One could assume that the timescales are proportional to some neurobiological quantities (e.g., the densities of certain neurons), and let the diagonal terms in the matrix be parametrized by these maps as $A_{ii} = \alpha_1 b_{i1} + \alpha_2 b_{i2} + \dots + \beta$ (where $b_{i1}, b_{i2}, \dots$ are the neurobiological maps and $\alpha_1, \alpha_2, \dots, \beta$ are coefficients) and use automatic differentiation to optimize the linear combination parameters.

A: We appreciate the reviewer’s comments and have made several revisions to the text that we believe should alleviate concerns. First, we have revised the abstract to make our claims more specific to our approach and its deliverable insights:

Abstract

The temporal evolution of **whole-brain activity is contingent** upon complex interactions within and between brain regions that are mediated by neurobiology and connectivity, **respectively**. To understand these interactions, many large-scale **modeling** efforts have measured structural connectivity, **brain** activity, gene expression, and cognition across multiple modalities and species. However, **accurate modeling of the brain’s traversal through distinct brain states remains difficult. Uncovering new insights requires data-driven** frameworks for estimating the interplay between brain structure and function while preserving biophysical **detail**. Here, we provide such a framework by integrating network control theory (NCT) with automatic differentiation to estimate model parameters with greater biophysical realism from data. Specifically, we estimate **regions’ intrinsic neural timescales (INTs) using NCT, and find that doing so broadens the range of dynamics supported by the connectome and improves the alignment between the brain’s connectivity and its traversal through state-space**. Next, we demonstrate that the resulting model-based **INTs** correlate with INTs **measured empirically from functional neuroimaging data**, neurobiological measures of gene expression and cell-type densities, **as well as** behavioral measures of cognition. We demonstrate consistent results across multiple datasets and species. Finally, we demonstrate that our **model-based INTs** enable the efficient control of brain **states** from fewer brain regions. Taken together, our results provide a simple and flexible quantitative framework that more accurately captures the interplay between brain structure, function, and dynamics with greater biophysical realism.

Second, we have briefly expanded our discussion on the implications of our flexible component to studies that use neural mass models, such as that of Deco et al. (2021; Science Advances):

Fourth, related to the above future direction, our stable and flexible INTs have implications for studies seeking to integrate spatially-patterned features of neurobiology into other dynamical systems models, such as non-linear neural mass models (NMMs) [20-22]. For example, previous

work has incorporated regional variations to gene expression into NMMs, and examined the impact of incorporating these static features of neurobiology on simulated functional connectivity [20]. Our results highlight that incorporating variable context-specific patterns of mesoscale dynamics are important for modeling large-scale neural systems. Given this, an interesting direction for future research would be to jointly model the effects of static and context-dependent timescales on NMMs.

1.9. R1 comment #9

R: *Correlation with neurobiological markers.* The Authors put a lot of emphasis on the correlation with neurobiological maps. However, many cortical maps (including all neurobiological maps analyzed) broadly align with the cortical hierarchy (e.g., the principal connectivity gradient), as the authors correctly mention: “Consistent with this idea, previous research has shown that higher-order association regions exhibit relatively high SST expression and slower INTs while primary sensorimotor regions exhibit relatively high PVALB expression and faster INTs” “We observed a negative correlation between PC 1 and intracortical myelin content, which is consistent with the observation that highly myelinated sensorimotor regions exhibit faster INTs” Thus, the correlation of the ‘fixed’ component of the timescales with these maps is hardly evidence of a direct link between the timescales and specific neurobiological maps. When the Authors state that “Collectively, these results demonstrate that parameters from our INTOptimized model track the divergent spatial patterning of distinct inhibitory interneurons across the cortex.” “Critically, our effects tracked known divergences” It is not fully clear what they mean. ‘Tracking the divergent patterning’ would imply that by looking at the timescales, one could predict when the neurobiological patterns are more divergent. This would be clear if, e.g., one of the principal components of the timescales aligned with the difference between two neurobiological maps, or at least if the correlation between timescales and a given neurobiological map would persist even after regressing out the effect of another map.

A: We apologize for the lack of clarity on our use of the term ‘diverging’ and are happy to clarify. Our use of this term was just to highlight the known anti-correlated nature of the PVALB and SST genes that we studied. Previous work has shown that SST and PVALB are anti-correlated across the cortex (Anderson et al., 2020, Nature Communications). As such, the spatial patterning of these two genes *diverge* across the cortex; regions with higher expression of one have low expression of the other, and vice versa. This is all that we meant when we used the term *divergent*. We can see that this language caused confusion, and have thus revised the corresponding sections of the manuscript. For example:

Moreover, we showed that our **model-based** INTs correlated systematically with diverse markers of inhibitory interneurons, including expression of SST and PVALB genes from the AHBA [23] as well as the density of PVALB-expressing basket and chandelier cells [24]. Critically, our effects tracked known **differences in the spatial patterning of** these interneuron markers consistently across the human and the mouse brain, illustrating that our model-based INTs track variation in genes and cell-types that separate INTs across the cortex [25-27].

We have removed all other uses of the terms *diverging* and *divergent* in the manuscript in this context and replaced them with terms such as *differing* and *different*.

1.10. R1 minor comments

R: “Of particular biological relevance is optimal control”. Why is it of particular biological relevance?

A: See revised text below:

Of particular biological relevance is *optimal control*, which **enables modeling of specific point-to-point transitions through neural state-space (macroscale function), and allows researchers to examine how the brain’s structural connectivity (macroscale structure) enables those transitions. Optimal control achieves this outcome by finding** the exogenous inputs, $u(t)$, that control the brain to transition between two activity states while minimizing (i) the magnitude of those inputs (i.e. the *control energy*), and (ii) deviations of the state trajectory from a user-defined reference brain state [6].

R: *In Fig. 2B, a better color map should be used to allow for better visualization.*

A: We apologize but are unclear what the issue with our chosen colormap is. Additionally, since no alternative colormap is suggested by the reviewer that would improve visualization, we have elected to maintain our original choice.

R: *I guess a kernel density was used in Fig 2C to smooth out a histogram. This should be mentioned*

A: This is correct. We have now clarified our use of kernel density plots in the corresponding figure caption.

R: *In Fig. 3A, one sees that, in addition to enhancing precision and lowering the energy, the timescale optimization allows to reduce the dependence on the specific choice of input nodes. This point could be stressed.*

A: We agree that this is a point worth placing emphasis on. We have revised the text:

Specifically, for both models, we observed correlations between x_T and $x(T)$ that were > 0.9 (Fig. R3A), and reconstruction errors $< 1 \times 10^{-2}$ (Fig. R3B), when $k \geq 124$. This results indicates that state transitions completed successfully using control sets that were only $\approx 1/3$ the size of our system (here, $N_{\text{system}} = 400$). Critically, the $INT_{\text{optimized}}$ model yielded this same performance when $k \geq 64$, indicating that our model achieved successful transitions with control sets that were $\approx 48\%$ smaller than those required for the INT_{uniform} model. **Additionally, Figure R3A also demonstrates that correlations for the $INT_{\text{optimized}}$ model were highly consistent across randomly sampled control nodes at a given k , whereas correlations for the INT_{uniform} model showed marked variation (as indicated by variability in the red traces). This result indicates that the successful completion of state transitions was less contingent on the specific set of partial control nodes in the $INT_{\text{optimized}}$ model. Finally, for state transitions that complete successfully, smaller control sets produced lower control energy for the $INT_{\text{optimized}}$ model compared to the INT_{uniform} model (Fig. R3C). Thus, the optimized decay rates in our model allowed us to control connectome dynamics more efficiently (i.e., with fewer control nodes), more easily (i.e., lower control energy), and with greater stability (i.e., lower variability in state completion) from a partial control set. These results indicate that our approach—which seeks to improve the biophysical realism of NCT—enables the successful control of brain dynamics from relatively sparse inputs. In turn, this improvement to model fitting reduces the gap between NCT and real-world applications of brain stimulation.**

R: *In Fig. 3C, one sees that when the number of control nodes is below 80, there is no energy reduction when optimizing timescales. The Authors state that “smaller control sets also produced lower control energy for the $INT_{\text{optimized}}$ model compared to the INT_{uniform} model (Fig. 3C)”. A more nuanced statement would be in order.*

A: We agree with the reviewer, and have revised that sentence:

... Finally, for state transitions that complete successfully, smaller control sets also produced lower control energy for the $INT_{\text{optimized}}$ model compared to the INT_{uniform} model

R: In Fig. 3D, which transition and which control node sets were used? Was this a typical case?

A: We apologize for the confusion here. For the analysis in Figure 3D, a uniform full control set used. This is because this analysis was just an eigen decomposition of the adjacency matrix with the optimized decay rates set along the diagonal; and these optimized decay rates were taken from when the model was trained using a uniform full control set. We have clarified this in the revised text:

To provide intuition for the above effects, we performed eigenmode decomposition on the connectome, wherein the diagonals were set either to the transition-specific decay rates from the $INT_{\text{optimized}}$ model (estimated using a uniform full control set) or to their mean; see also Section ?? in the Supplementary Materials for further characterization of nodes' importance to control.

R: When saying that "single nodes innervate only relatively small subsets of the eigenmodes" what do they mean by 'innervate'? That the eigenmodes have nonzero projection (or projection above a threshold) on the nodes?

A: Yes, we mean that nodes only project to a relatively sparse subset of eigenmodes, as indicated by only a small fraction of non-zero eigenvectors. We have revised the text for clarity:

This result indicates that single nodes excite (i.e., project to) only relatively small subsets of the eigenmodes in the $INT_{\text{optimized}}$ model;

R: The Authors say "we performed eigenmode decomposition on the connectome, wherein the diagonals were set either to the transition-specific decay rates from $INT_{\text{optimized}}$ model or to their mean". In Fig. 3D there are results for a specific transition. Results for the mean do not appear - not that they should be shown, but if they aren't, it would be better to say it: "we performed eigenmode decomposition on the connectome, wherein the diagonals were set either to the transition-specific decay rates from $INT_{\text{optimized}}$ model (Fig. 3) or to their mean (not shown to avoid redundancy)"

A: We apologize for the confusion, but this isn't the case. Throughout the manuscript, when we present results for the INT_{uniform} , we do so by setting the diagonal of the adjacency matrix to the mean of the optimized decay rates from the $INT_{\text{optimized}}$ for a given transition. This ensures that results are not driven trivially by a baseline difference in regions' self-inhibition between our two models. This was mentioned in the original manuscript:

To control for differences in the overall level of self-inhibition between the two models, we set all values of d for the INT_{uniform} model to the mean of d from the $INT_{\text{optimized}}$ model ($d_{\text{uniform}} = \bar{d}_{\text{optimized}}$).

As such, when we report eigen decomposition in Figure 3D for the INT_{uniform} model, we are reporting a decomposition of the adjacency matrix where the mean of the optimized decay rates (for a given transition) was set along the diagonals. Thus, these results are shown.

R: Do the timescale-optimized eigenmodes more strongly align with functional connectivity eigenmodes, thus bridging structure and function?

A: This is an interesting suggestion from the reviewer, but since the breadth and depth of the manuscript is already quite expansive, we have chosen to pursue this suggestion in subsequent work.

R: *In Fig. S3, a better color map should be used to facilitate visualization*

A: As above, since we are unsure of the issue with our chosen colormap, or what would constitute a better one in this context, we have chosen to retain our original colormap.

R: *From Fig. S4, one sees that PC1 projects more strongly on transition, involving CAP 4, which is the most similar to the default mode network. One could speculate that the average timescales are set to optimize transitions to/from the DMN.*

A: This is a good point raised by the reviewer. We have now incorporated this speculation into the revised text:

For each of these PCs, greater scores correspond to slower INTs in our model across diverse state transitions, **but capture different subsets (see Fig. ??, as well as Fig. ??). For example, PC 1 predominantly tracks variations in optimized decay rates that capture transitions in and out of the default mode system, while PC 2 tracks variations that capture transitions in and out of the visual system, the frontoparietal control network, and the ventral attention network.**

R: *Fig. 4 reports an extremely small P-value. How can this emerge from a permutation approach (where the smallest P-value is set by the number of permutations)?*

A: Apologies, this was a typo. We thank the reviewer for catching it. We have corrected the typo in the revised manuscript.

R: *“we also correlated PC 1, PC 2, as well as the mean of optimized decay rates with regional degree and strength estimated from the structural connectome and found no significant relationships”. Empirical timescales tend to correlate with degree, as also stated by the Authors, “INTs correlate tightly with the strength of regions’ structural connectivity to the rest of the brain”, so in principle it may be hard to disentangle the two contributions. I then do not think that the (weak) the correlation with degree is problematic.*

A: We agree with the reviewer that the weak correlation with degree is not a concern or problematic. Nevertheless, we have opted to keep it in, since the broader network neuroscience community will likely be interested to know about it.

R: *What is the correlation between the biological metrics in Fig. 4 and empirical timescales from fMRI? Are the optimized timescales more or less correlated than the empirical ones?*

A: We thank the reviewer for this comment, but unfortunately are not clear on the reviewer’s query. There are no other biological metrics presented in Figure 4 except for the empirical timescales from fMRI. Indeed, Figure 4 only presents the correlation between (PCs of) the model-based timescales inferred from NCT and empirically-measured timescales captured using fMRI. As such, we are not sure what correlation the reviewer is requesting to see here.

R: *“the dominant source of variance in optimized decay rates correlated negatively with in vivo microstructure”. What is the single microstructural metric used?*

A: We apologize for the missing explanation of these data, and thank the reviewer for bringing it to our attention. This microstructure measure was previously estimated by Royer et al. (2022), and we extracted it from their database and used it as-is. We have added a new paragraph to the Methods section that briefly outlines their

measure:

Quantitative T1-weighted Imaging (qT1).] We examined microstructural profile covariance (MPC) maps processed by Royer *et al.* [28]. Specifically, we extracted parcellated MPC data directly from the MICA-MICs database, pulling from the `group_ses-01_space-fsnative_atlas-schaefer400_desc-mpcGradient.txt` file. Briefly, Royer *et al.* [28] used quantitative T1-weighted imaging (qT1) to estimate layer-specific intracortical microstructural composition across the cortical sheet. Then, these layer-specific profiles were correlated across cortical regions to index pairwise similarity in regions’ microstructure. Finally, the resulting distance matrix was projected down to the region level, creating a gradient that captures the dominant source of variation in MPC across the cortex. We refer the reader to Royer *et al.* [28] for more details.

R: “This result suggests that the exogenous control of large-scale brain dynamics from the connectome is rooted in microscale cellular processes”. Propagation of control signals along the connectome would in principle suffice to allow for exogenous control by an external agent. I would say, “the exogenous control of large-scale brain dynamics from the connectome can leverage microscale cellular processes”

A: We thank the reviewer for this comment and have revised the text accordingly:

The results presented in this section demonstrate that the key outputs from the $INT_{optimized}$ model replicate across multiple datasets and are conserved across species. Critically, our findings demonstrate convergent multi-scale linkages across humans and mice **spanning the brain’s microscale neurobiology, mesoscale regional dynamics, and macroscale connectivity**. In both organisms, genes that track **distinct** inhibitory processes across the cortex—and that are thought to underpin regions’ myelin content and inter-connectivity—correlate consistently with regions’ INTs estimated from NCT. This result suggests that **the exogenous control of large-scale brain dynamics from the connectome can leverage microscale cellular processes**, and that these effects are phylogenetically conserved.

R: “we fit our NCT model to individual connectomes (see Materials and Methods)”. Were the 7 CAPs also recomputed at a single subject level?

A: This is an excellent question, and we are happy to clarify here and in the revised text. No, we used the same CAPs extracted from the group analysis for the subject-specific NCT analysis. This was a deliberate choice, since it allowed us to isolate inter-individual differences in NCT outputs to inter-individual differences in connectome topology. Of course, we could have extracted and used subject-specific CAPs for this analysis, but this would introduce two sources of inter-individual variability into our analyses, and variability in the definition of brain states (as introduced by subject-specific CAPs) would confound our research goals. We have clarified our use of group-averaged CAPs in the revised text:

Thus, in our final analysis, we fit our NCT model to individual connectomes. Briefly, we fit both our $INT_{uniform}$ and $INT_{optimized}$ models to each of 960 subject-specific HCP-YA connectomes and **performed the following two analyses. First, we examined whether the subject-specific coupling between model-based INTs and empirically-measured INTs (ACF_{lag}) varied as a function of the transition probabilities observed in subjects’ rs-fMRI data (see Materials and Methods). Second, we examined the correlation between subject-specific control energy and subjects’ scores on a battery of behavioral tasks (see Materials and Methods). Note, in both of these analyses, we used the same CAPs extracted at the group-level (see Fig. R2A) as brain states for subject-specific NCT analysis.**

This approach allowed us to isolate inter-individual differences in NCT outputs to individual variability in connectome topology.

R: “Critically, our findings demonstrate convergent multi-scale linkages between brain structure and function across humans and mice.” Clarify what you mean by ‘multiscale’ (e.g., the model combined the macro-scale connectome with the meso-scale regional dynamics)

A: We thank the reviewer for this comment and have clarified the text:

Critically, our findings demonstrate convergent multi-scale linkages across humans and mice **spanning the brain’s microscale neurobiology, mesoscale regional dynamics, and macroscale connectivity.**

R: “Our approach corrects a long-held assumption embedded within the NCT framework: that brain regions’ intrinsic dynamics are fixed and homogeneous across the brain.” Other Authors have used non-uniform timescales. See e.g. Manjunatha et al. 2024, where effective connectivity was used instead of structural connectivity, with node-dependent diagonal terms in the coupling matrix.

A: We have revised this sentence to soften its implicit claim:

Our approach corrects a **commonly adopted but oversimplified** assumption embedded within the NCT framework: that brain regions’ intrinsic dynamics are fixed and homogeneous across the brain.

R: “This assumption has limited the biophysical realism of NCT, in turn hampering its application to neuroscience”. This sentence contradicts the previous assertion that “The use of NCT in neuroscience is well-established” (§ 2.1). I would say, “This assumption has limited the biophysical realism of NCT, raising some concern for its widespread application to neuroscience”

A: We have revised this sentence in line with the reviewer’s request:

... that brain regions’ intrinsic dynamics are fixed and homogeneous across the brain. **This assumption has limited the biophysical realism of NCT, raising concern for its widespread application to neuroscience.**

R: “As discussed in our recent protocol paper [23], controlling neural dynamics from very few nodes is often intractable in NCT”. This was first clearly discussed in Tu et al, NeuroImage 176 (2018): 83-91.

A: We have revised this sentence in line with the reviewer’s request:

As noted in our recent protocol paper [23], **as well as originally discussed by [82]**, controlling neural dynamics from very few nodes is often intractable in NCT

R: In Eq. (5), $(x_r - xT)$ should be probably replaced by $(x - x_r)$

A: We thank the reviewer for catching this typo. We have corrected it in the revised manuscript.

R: “Briefly, this approach challenges another long-held assumption of NCT that each control signal exclusively impacts a signal node; an assumption that is inconsistent with the spatial continuity of the cerebral cortex and its effects on dynamics”. While most papers applying NCT to the brain make this assumption, this is not

a standard in NCT in general. See e.g. Liu, Y. Y., Slotine, J. J., Barabási, A. L. (2011). Controllability of complex networks. nature, 473(7346), 167-173.

A: We have revised this sentence to soften its implicit claim:

Briefly, this approach challenges another **commonly adopted** assumption of NCT **in neuroscience** that each control signal exclusively impacts a single node

1.11. R1 remarks on code availability

R: *The GitHub repository seems very well organized and complete. It should allow reproducing the manuscript's results and perform new analyses.*

A: Thank you

2. Reviewer #2

R: *The authors couple network-control theory with automatic differentiation to estimate region-specific self-inhibition (decay-rate) parameters directly from resting-state fMRI time series, using structural connectomes derived from diffusion MRI as the coupling matrix. Optimised decay rates (i) lower the energy and dimensionality of control inputs needed to steer whole-brain dynamics, (ii) correlate with intrinsic neural timescales (INTs) obtained from autocorrelation analyses, (iii) follow cortical gene-expression and cell-type gradients, and (iv) predict several cognitive performance measures. Findings are replicated across two large human cohorts and an independent mouse dataset, underscoring robustness and cross-species generalisability. Taken together, the work advances control-theoretic models toward greater biophysical plausibility and provides a principled bridge between macroscopic dynamics and microscale physiology.*

A: We thank the reviewer for the careful review of our manuscript, and address the reviewer's comments below.

2.1. R2 comment #1

R: *The study assumes that “neurobiological effects are already encoded in the connectome,” yet diffusion MRI tractography cannot resolve (i) axonal direction, (ii) cell-type specificity ($E \rightarrow E$, $E \rightarrow I$, $I \rightarrow E$, $I \rightarrow I$), or (iii) laminar termination. A single streamline reconstructed by dw-MRI trackography between regions A and B therefore conflates multiple, often opposing, pathways. Because the optimised decay rate (or inhibitory strength) is highly sensitive to the balance of large-scale excitation versus feed-forward and feedback inhibition, these unknowns can bias both the optimisation and its biological interpretation. Explicitly acknowledge these limitations regarding the model's biological realism and associated diffusion-MRI constraints, and temper claims implying that decay-rate optimisation directly “uncovers” true biological self-inhibition.*

A: We thank the reviewer for this important comment, and agree that MRI tractography cannot resolve axonal direction, cell-type specificity, and laminar termination, which could potentially bias the optimized weights. In response, we first clarify our intended usage of the word “encoded,” and subsequently incorporate the reviewer's suggestion of explicitly acknowledging the conflation in the main text.

First, we clarify our statement that “effects of neurobiology are already encoded in the connectome,” was not intended to imply that these features can be read-out from the connectome directly and plainly. Rather, our intention was to state that these complex and entwined features of neurobiology (e.g., cell-type distributions) exert a known effect on brain organization that in turn shapes the human connectome. For example, given different cell-type specificity, the dynamics and evolutionary pressures would likely result in a different diffusion-MRI measurement and streamline reconstruction. In this way, the word “encoded” was meant to reflect the idea that the topology of the connectome is the way that it is because of neurobiology, at least in part. In fact, as the reviewer states: “the optimised decay rate (or inhibitory strength) is highly sensitive to the balance of large-scale excitation versus feed-forward and feedback inhibition.” We set out to develop a data-driven method that uncovers INTs that should be sensitive to the balance of neural excitation to inhibition.

Second, we acknowledge that our intention was not clear in the original manuscript. Thus, we have revised this sentence:

Rather, our approach assumes that the **complex multi-scale** effects of neurobiology **on the brain** are already **implicitly** encoded in **both the topology of the connectome and the topography of the brain** states, and thus **that** optimizing decay rates will uncover **regional** INTs that reflect underlying neurobiology.

Furthermore, as requested by the reviewer, we now explicitly acknowledge these limitations in the Discussion:

This study has several limitations. **First, we quantified extrinsic connectivity using tractography estimated from diffusion-weighted MRI data. Tractography is unable to resolve the axonal direction, cell-type specificity, or laminar origination and termination of white matter pathways. This insensitivity limits our capacity to tease apart the contributions of each of these effects on our optimized decay rates. As such, these distinct neurobiological processes likely all contribute to our model outputs in complex and potentially complementary ways. Second, we validated our optimized INTs ...**

2.2. R2 comment #2

R: *A more negative d is presented as simultaneously indexing “stronger self-inhibition” and “faster intrinsic timescales.” These are not interchangeable. Synaptic self-inhibition strength (a circuit-level parameter) and membrane/ensemble decay constants (a biophysical time constant) arise from different mechanisms, operate on different timescales, and are treated as separate parameters in biophysically informed models. By collapsing both effects into a single scalar d , the current formulation blurs two distinct sources of variance, making it impossible to discern whether observed associations stem from changes in synaptic strength, passive decay, or some hybrid “effective” quantity. My recommendation is to explicitly acknowledge this conflation in the text, emphasising that d should be interpreted as a lumped representation of multiple mechanisms rather than a direct measure of neither synaptic inhibition nor decay rates, and to temper causal language linking d to “biological self-inhibition.” Clarifying this point is essential, as the manuscript’s stated aim is to enhance biological realism.*

A: We thank the reviewer for this comment, and agree that synaptic self-inhibition and ensemble decay constants are not interchangeable. In response, we first clarify our usage of “self-inhibition” and “intrinsic timescales” in the mesoscale context of our data, and then incorporate the reviewer’s suggestion to explicitly acknowledge the conflation.

First, we note that we do not claim that our model parameters reflect synaptic self-inhibition at the neuronal level. Instead, we interpret self-inhibition in a relatively broad mesoscale context. In turn, we contend that, in this context, the connection between the two concepts is reasonable; regions in our model with stronger self-inhibition (more negative “ d ”) exhibit mesoscale dynamics that decay more quickly, and mesoscale dynamics that decay more quickly can fluctuate over shorter time scales. To demonstrate this effect more explicitly in the revised manuscript, we now show that regions with stronger self-inhibition (more negative “ d ”) do indeed participate in dynamical connectome modes with faster decays.

From the revised Results:

Additionally, the corresponding eigenvectors were noticeably more sparse for the $INT_{\text{optimized}}$ model compared to those from the INT_{uniform} model (Fig. R3E). This result indicates that single nodes **excite (i.e., project to)** only relatively small subsets of the eigenmodes in the $INT_{\text{optimized}}$ model; nodes **excited**, on average, $\sim 48\%$ of the eigenmodes for the $INT_{\text{optimized}}$ model compared to 92% for the INT_{uniform} model. Hence, our $INT_{\text{optimized}}$ model gives rise to a greater ability to control targeted subsets of dynamical modes from single nodes. **Finally, we observed close correspondence between the eigenvalues of these sparse eigenvectors (columns of matrices shown in Fig. R3E, top) and the optimized decay rates of the nodes that excited them (rows of the same matrices). Specifically, nodes with slower model-based INTs (i.e., less negative optimized decay rates, d) tended to excite eigenmodes that govern slower modes of connectome dynamics (i.e., less negative**

eigenvalues, λ). This result shows direct linkage between the mesoscale regional timescales from our $INT_{\text{optimized}}$ model and the patterns of changes in large-scale brain activity that emerge from the connectome in our model.

Taken together, the results presented in this section demonstrate that, under the INT_{uniform} model, all nodes of the connectome tend to **excite** the majority of the eigenmodes of the system, and that the effect of these modes on whole-brain dynamics are largely consistent over time (i.e., homogenous eigenvalues). By contrast, under the $INT_{\text{optimized}}$ model, each node **excites** only a sparse set of eigenmodes that exert a more diverse impact on whole-brain dynamics over time (i.e., heterogenous eigenvalues). In turn, this difference leads to the efficient completion of state transitions from fewer control nodes **that have diverse INTs**.

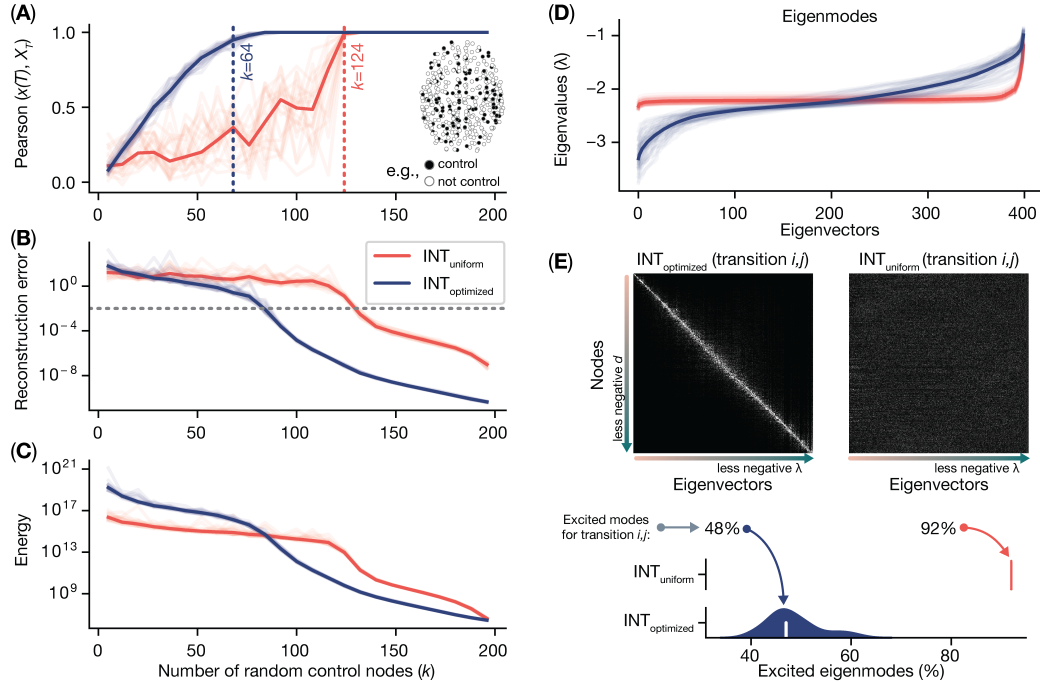


Figure R3: Optimizing nodes' intrinsic neural time scales results in more efficient control of connectome dynamics from fewer control nodes. We estimated control energy using random partial control sets of increasing size, k . For each k , we generated 20 random control sets and calculated control energy for all state transitions. **(A)** The correlation between $x(t)$ and x_T for both the $INT_{\text{optimized}}$ and INT_{uniform} models averaged over the 20 control sets at each k . **(B)** The reconstruction error for both models averaged over the 20 control sets at each k . **(C)** The control energy for both models averaged over the 20 control sets at each k . Here, control energy was rescaled by multiplying by dt , which was set to 0.001 in our simulations. Together, these plots show that state transitions were completed more efficiently (i.e., with fewer **control** nodes), more easily (i.e., lower control energy), **and with greater stability (i.e., lower variability in state completion)** when using the $INT_{\text{optimized}}$ model. **(D)** The connectome eigenmodes for both the $INT_{\text{optimized}}$ and INT_{uniform} models. The eigenvalues for each eigenvector are shown for all state transitions, overlaid by their average. This data shows that the eigenvectors of the system have more diverse eigenvalues when using the $INT_{\text{optimized}}$ model. **(E)** The corresponding eigenvector matrices for an example transition (top), alongside their density averaged over state transitions (bottom). This data shows that the nodes of the connectome (**rows**) **excite** different subsets of eigenmodes (**columns**) in the $INT_{\text{optimized}}$ model compared to the INT_{uniform} model. On average, **single** nodes **excited** only 48% of the eigenmodes for the $INT_{\text{optimized}}$ model. By contrast, **single** nodes **excited** 92% of the eigenmodes for the INT_{uniform} model. Thus, under the $INT_{\text{optimized}}$ model, each node **excites** only a relatively sparse set of eigenmodes that exert a more diverse impact on whole-brain dynamics over time, and this leads to more efficient control over state transitions.

Second, in accordance with the reviewer’s suggestion, we now explicitly acknowledge the conflation of these two effects in the scalar “ d ” and clarify our language linking “ d ” to biological self-inhibition in several places throughout the revised manuscript.

From the revised Introduction:

Here, we present a simple and flexible extension to the optimal control framework that optimizes the parameters that govern nodes’ intrinsic **mesoscale** dynamics. In NCT, each node’s neural dynamics are encoded by an internal decay rate [6], which governs that node’s INTs and can be thought of as its **general** propensity towards self-inhibition; greater self-inhibition yields faster **decay of mesoscale dynamics**, which equates to faster **regional** INTs. Conventionally, decay rates are set uniformly across the connectome (*uniform INTs*). However, this approach is at odds with our understanding of how **regions’ mesoscale INTs** vary across the brain in tandem with spatially-patterned properties of neurobiology [13, 14, 29–31]. Thus, to improve the biological realism of NCT, we developed a data-driven approach to optimizing nodes’ internal decay rates. Specifically, we implement learnable decay rates that minimize the distance between the model’s controlled state trajectory and an empirical reference state, thereby yielding *optimized regional INTs*.

From the revised Results:

Our extension to NCT involves modifying the decay rates, d , of nodes’ internal dynamics (see Materials and Methods) (Fig. R5C-E). In NCT, prior to simulating a state transition, each node’s dynamics is set to decay over time in the absence of external input, which ensures system stability [6, 32]. **Here, since nodes represent relatively large brain regions in a whole-brain connectome, these internal decay rates encode regions’ mesoscale intrinsic dynamics as well as their rate of fluctuation of time.** In a continuous-time system, **these decay rates are** set uniformly across nodes by subtracting the identity matrix (I) from the connectome (A) during normalization (see Materials and Methods). This approach establishes a baseline decay rate of $d = -1$ for each node, which we refer to as *uniform INTs* (Fig. R5C; INT_{uniform}). Here, we optimize these decay rates to minimize the distance between the controlled state trajectory and the reference state. This approach is data-driven and results in a set of heterogenous decay rates that give rise to variable INTs across the brain; an approach we refer to as *optimized INTs* (Fig. R5D; $INT_{\text{optimized}}$). For a given state transition, we achieved this optimization by performing gradient descent *via* automatic differentiation on the difference between the controlled state trajectory, $x(t)$, and the reference state, x_r , while adjusting d (Fig. R5E). Once converged, this algorithm provides a unique decay rate for each node and for each state transition. Then, these optimized decay rates can be (i) used to calculate control energy for their associated transition and (ii) validated against patterns of **microscale** neurobiology known to influence empirical **mesoscale neural dynamics and macroscale** connectivity.

From the revised Materials and Methods:

This optimization procedure comprises our $INT_{\text{optimized}}$ model. **As discussed in the main text, this optimization procedure produces a set of heterogenous decay rates that give rise to variable intrinsic neural timescales (INTs) across the brain. Importantly, here we interpret these optimized decay rates as representing mesoscale INTs governed by a region’s general self-inhibition. Our model is not setup to infer finer-scale and distinct neuronal mechanisms such as strength of synaptic self-inhibition or membrane decay constants. Instead, we are modeling a unitary construct that jointly encodes a region’s general self-inhibition and, by extension, the rate at**

which it's dynamics fluctuate. We compared this $INT_{\text{optimized}}$ model to a standard model where A was normalized according to Eq. ?? . As noted above, we refer to this standard model as INT_{uniform} .

2.3. R2 comment #3

R: *Decay rates are optimised for each of 42 empirical state transitions, yielding large within-region variance (Fig. 2B). In vivo, neither inhibitory synaptic strength nor intrinsic decay constants plausibly fluctuate at this pace. Explain and discuss which biological mechanisms could account for such variability.*

A: We thank the reviewer for this comment, and apologize for the confusion. Together with the above comment (R2C2), we believe that confusion stems from us (the authors) and the reviewer interpreting self-inhibition under different sets of assumptions. As discussed in response to the previous comment (R2C2), our model's inhibition parameters do not code for inhibitory synaptic strength or intrinsic decay constants at the neuronal level; we agree that these do not plausibly fluctuate at the pace reflected in our model's outputs and our results. Instead, our model outputs code for inhibition at the mesoscale regional level, akin to that which is represented by the activity of populations of neurons acting in concert (e.g., "neural masses"). Furthermore, here, the term *inhibition* primarily refers to model-space, where the definition of the system equations requires that nodes have internal inhibitory processes that stabilize dynamics. Indeed, this is the primary reason that we contextualize our model's parameters neuroscientifically using the term *intrinsic neural timescales* (INTs) instead of inhibition. In this way, we argue that our model-based inhibition parameters code for properties of mesoscale activity, not neuronal-level inhibitory processes like synaptic strength or intrinsic decay constants. In this context, there is a long history of recognition by the field that region's INTs fluctuate in a state dependent manner. For example, as discussed in the original manuscript, previous fMRI work has shown that INTs change systematically from rest to task, which is akin to how our model-based INTs adapt to specific state transitions. Other work has shown that INTs change under a variety of different contexts as well, including in response to neurostimulation, sleep, and development (see Wolff, Trends in Cognitive Sciences, 2022; Bradley, Nature Reviews Neuroscience, 2022; Sydnor, Nature Neuroscience, 2023).

To minimize confusion on this point, we have combed through the manuscript and tightened our expression and phrasing around these issues. This has led to numerous changes in the revised manuscript that are presented in blue typeface. For brevity, we have not reproduced them here.

Furthermore, for completeness, below we include explanation for how our model connects with concepts from non-linear dynamical systems. Specifically, what could account for variability in inhibition in our model is the changing linear response of a nonlinear system at different states. Specifically, NCT posits linear dynamics and interactions between regions such that

$$\dot{x} = Ax + Bu.$$

Hence, the dynamics, eigenmodes, and timescales of the system are independent of the specific state of the system, x^* . In contrast, the true dynamics are most likely nonlinear such that

$$\dot{x} = f(A, B, x, u),$$

and linearizing the dynamics at two different states, x_1^* and x_2^* , yields different effective linear response at each point:

$$\dot{x} = A_1x + B_1u \quad \text{or} \quad \dot{x} = A_2x + B_2u$$

where $A_1 = \nabla_x f(A, B, x, u)|_{x_1^*}$, and $A_2 = \nabla_x f(A, B, x, u)|_{x_2^*}$ are the linear responses of the system about two different operating states. Hence, by estimating "d" for each transition, we are circumventing the difficulty of identifying the true nonlinear dynamics by estimating the transition-specific differences between A_1 and A_2 using the unresolvable diagonal weights.

2.4. R2 comment #4

R: *Given the well-known non-uniqueness of optimal solutions in linear control models, a single optimisation run per participant is insufficient to claim that the regional decay rate d is uniquely identifiable. My recommendation is to run the optimiser from many random initialisations and report the spread of the resulting d estimates. Convergence toward a tight cluster would support identifiability; dispersion would signal model degeneracy. In addition, I recommend to use bootstrapping to derive confidence intervals for each region's d and present these intervals alongside point estimates. Providing these diagnostics will lend empirical weight to causal statements such as “we estimate the structural form of regional self-inhibition.”*

A: We thank the reviewer for this important comment, and agree that there are various challenges of uniqueness broadly across identifying linear systems. In response, we first discuss why our approach does not suffer from issues with identifiability. Second, to demonstrate this fact empirically, we perform the reviewer's suggestion of running the optimizer from different random initializations, and report a nearly exact consistency between estimated d .

Regarding the first point, here and in the revised manuscript, we discuss why we expect strong identifiability in our optimization. The most generic form of the system identification problem for a linear time-invariant system involves estimating every entry of the system matrix, A . For $A \in \mathbb{R}^{N \times N}$, the ability to modify all N^2 entries enables the complete optimization of all of the system's eigenvalues and eigenvectors, which would lead to degeneracies in the solution for our relatively sparse set of optimization costs (penalizing only N values of the system states at the midpoint of the controlled state trajectory). Put differently, in the generic case, we would be seeking to optimize a relatively large number of variables (i.e., all N^2 entries in A) while penalizing a relatively small number of deviations (i.e., N values of the state trajectory, x). In contrast, our optimization only allows for the modification of the N diagonal entries of A . As such, we have only as many variables to optimize as we have constraints in our cost function, which enables identifiability.

Regarding the second point, to support our above assertion, we demonstrate that optimizing d from random initial conditions converges to the same value. Specifically, we computed the optimized decay rates, d , under two different conditions. For the first condition, we initialized all decay rates uniformly at $d_{ii} = 1$ (just like we did in the manuscript), and computed the optimized decay rates separately for all state transitions. For the second condition, we initialized all decay rates randomly and uniformly between $d_{ii} \in [0, 1]$, and again computed the optimized decay rates for all state transitions. Finally, we correlated all optimized decay rates from all transitions across the two initialization conditions (Fig. R4, left). Additionally, for completeness, we also computed and correlated optimized decay rates for two separate runs of the random uniform initialization strategy (Fig. R4, right). As can be seen, the optimized decay rates, whether initialized uniformly or randomly, converged to nearly the same value with a Pearson correlation coefficient of $\rho \approx 1.0$. Hence, the optimized decay rates exhibit almost no spread based on initialization strategy, providing strong evidence for identifiability.

We now include these analyses and results in the revised Supplementary Materials:

Identifiability and robustness of optimized decay rates to random initial conditions

In the main text, we presented optimized diagonal weights, d , that minimize the distance between the controlled state trajectory and the reference state. Here, we provide evidence for the uniqueness and identifiability of the reported optimized weights. Specifically, we demonstrate that optimizing d from random initial conditions converges to the same value by computing the optimized decay rates under two different conditions. For the first condition, we initialized all decay rates uniformly at $d_{ii} = 1$ (just as we did in the Main Text), and computed the optimized

decay rates separately for all state transitions. For the second condition, we initialized all decay rates randomly and uniformly between $d_{ii} \in [0, 1]$, and again computed the optimized decay rates for all state transitions. Finally, we correlated all optimized decay rates from all transitions across the two initialization conditions (Fig. R4, left). Additionally, for completeness, we also computed and correlated optimized decay rates for two separate runs of the random uniform initialization strategy (Fig. R4, right). As can be seen, whether initialized uniformly or randomly, the optimized decay rates converged to nearly the same value with a Pearson correlation coefficient of $\rho \approx 1.0$. Hence, the optimized decay rates exhibit almost no spread based on initialization strategy, providing strong evidence for identifiability.

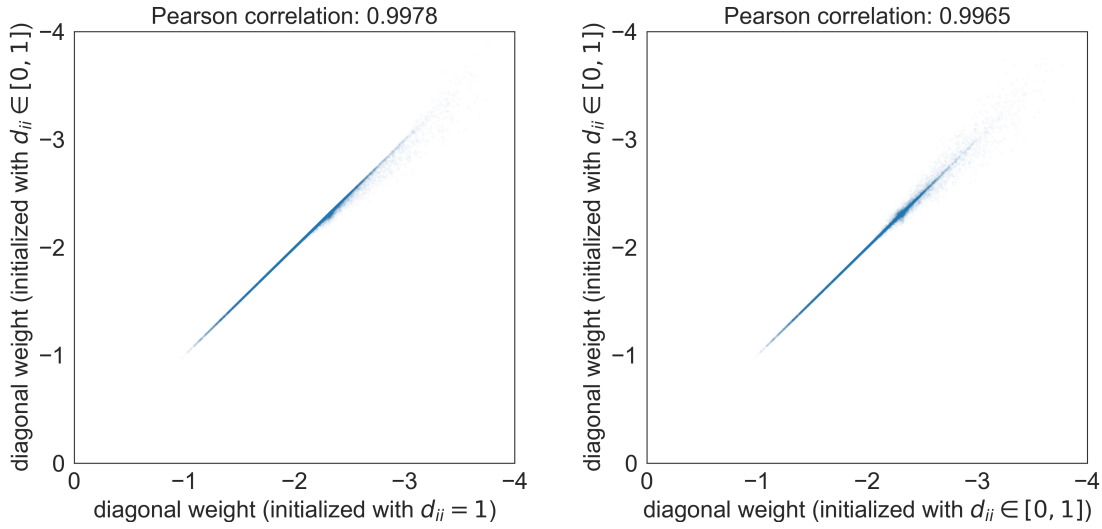


Figure R4: **Identifiability of optimized decay rates across random initializations.** (left): A scatterplot of the optimized decay rates between a uniform initialization (initial $d_{ii} = 1$) on the x-axis, and a random initialization (initial $d_{ii} \in [0, 1]$) on the y-axis, showing almost identical converged values and strong identifiability. (right): Same scatterplot between two random initializations also showing almost identical converged values and strong identifiability.

2.5. R2 comment #5

R: *The Abstract highlights the use of automatic differentiation, yet this aspect is only briefly mentioned in the Methods, with no further detail provided. Please clarify how automatic differentiation contributed to obtaining the results—specifically, what role it played in the optimisation process—and explain how it offered advantages over conventional gradient estimation or optimisation techniques. This would help justify its prominent mention in the Abstract and clarify its methodological significance.*

A: We thank the reviewer for this comment. In response, we now include a paragraph in the revised Methods emphasizing the advantages of automatic differentiation in our framework for scale, speed, and numerical precision:

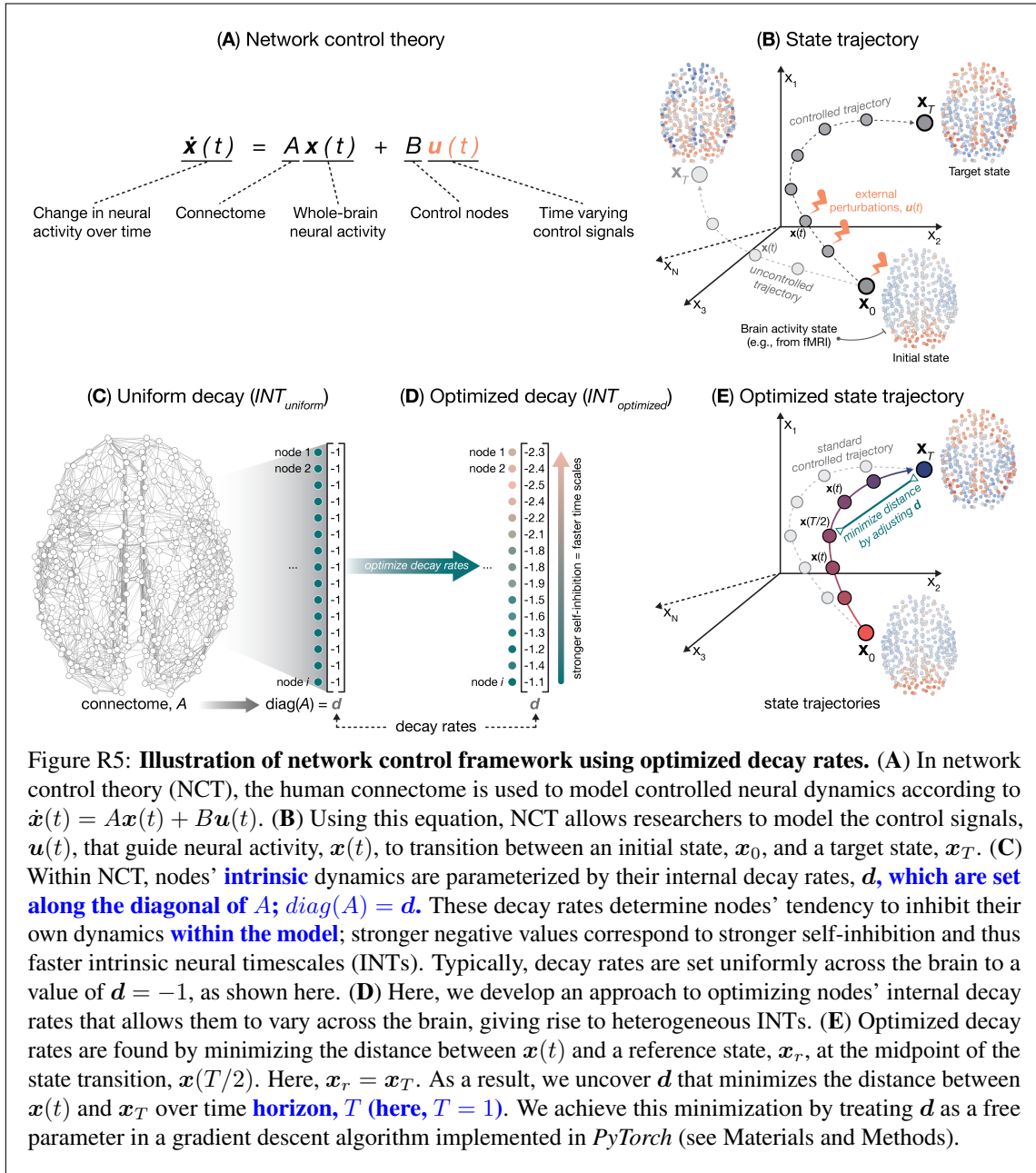
Automatic differentiation

To accomplish model optimization at scale, automatic differentiation plays a key role in computational scale, speed, and numerical accuracy [33]. Compared to standard numerical differentiation which requires the evaluation of the cost function for each optimized variable and suffers from roundoff and truncation error, reverse-mode automatic differentiation only requires one forward and one reverse pass to evaluate the gradient to machine precision. Such advantages are especially relevant considering that our optimization is not simply evolving and fitting an ODE, but requires the evaluation, inversion, and eigendecomposition of very large (order 1000×1000) dimensional matrices, with the potential to accomodate much larger and more complex systems and biophysical constraints.

2.6. R2 minor comments

R: *In Fig. 1 it is unclear where the decay parameter d enters the displayed system equation. Consider displaying the full equation.*

A: We apologize for this confusion. The vector of optimized decay rates, d , is set along the diagonals of A . Hence, including d in the equation $\dot{x} = Ax + Bu$ would be redundant. However, we agree this was not clear from Figure 1 and the associated caption. As such, we have updated Figure 1 by specifying that d is set along the diagonal of A , and outlined this in the revised figure caption:



R: Replace "innervate eigenmodes" with "excite" or similar to avoid confusion with physiological "innervation".

A: We have revised the text in accordance with the reviewer's suggestion:

Additionally, the corresponding eigenvectors were noticeably more sparse for the $INT_{\text{optimized}}$ model compared to those from the INT_{uniform} model (Fig. R3E). This result indicates that single nodes **excite (i.e., project to)** only relatively small subsets of the eigenmodes in the $INT_{\text{optimized}}$ model; nodes **excited**, on average, $\sim 48\%$ of the eigenmodes for the $INT_{\text{optimized}}$ model compared to 92% for the INT_{uniform} model. Hence, our $INT_{\text{optimized}}$ model gives rise to a greater ability to control targeted subsets of dynamical modes from single nodes.

R: *Correct multiple references to non-existent panel "Fig. 2F".*

A: We thank the reviewer for catching this typo. The reference should have been to Fig. 2B, not Fig. 2F. We have corrected this mistake.

R: *Missing Data Availability Statement: include direct links or DOIs for all raw and derived datasets (human and mouse) and note any access restrictions.*

A: We thank the reviewer for noting this omission. We have added a Data Availability Statement to the revised manuscript:

Data availability

This study made use of multiple datasets. The primary human neuroimaging dataset, known as the Human Connectome Project (HCP), is available at <https://db.humanconnectome.org>. The secondary human neuroimaging dataset, known as Microstructure-Informed Connectomics (MICA-MICs), is available at <https://osf.io/j532r/>. Human gene expression and cell-type data are available at <https://portal.brain-map.org/> and <https://cellxgene.cziscience.com/collections/d17249d2-0e6e-4500-abb8-e6c93fa1ac6f>, respectively. Finally, data from the mouse brain is also available at <https://portal.brain-map.org/>.

2.7. R2 remarks on code availability

R: *The GitHub repo (https://github.com/LindenParkesLab/nct_xr) lacks a minimal working example and instructions for obtaining and preprocessing the data. A short README outlining (i) data-access links, (ii) the exact preprocessing pipeline, and (iii) an end-to-end demo would greatly improve reproducibility.*

A: We thank the reviewer for this comment. We have provided example code for generating connectomes and parcellating fMRI data in the HCP-YA dataset that should aid users wishing to replicate our results. Our future goals are to incorporate our new algorithm into our existing NCT Python package (<https://github.com/LindenParkesLab/nctpy>), which is hosted separately from the repository associated with the present study. There, we will include an end-to-end demo as part of a new release of that software that will come at a later date.

3. Reviewer #3

R: *The paper, “Inferring macroscopic intrinsic neural timescales using optimal control theory” by Kim et al. extends the widely used network control theory (NCT) framework to estimate regional self-inhibition, providing greater biophysical realism. The main analyses used data from 960 HCP-YA subjects, the authors confirm their main findings in a second dataset (MICA-MICs) of 50 individuals, and demonstrate translational convergence of their main findings in mice using the Allen Mouse Brain Connectivity Atlas. The authors show: 1) that including regional estimates of self-inhibition in the NCT model leads to reduced control energy; 2) the self-inhibition parameters correlate with intrinsic neural timescales (INTs) estimated from resting-state fMRI data; 3) the self-inhibition parameters correlate positively with AHBA-SST and PVALB-Chandelier, and negatively with AHBA-PVALB and PVALB-Basket; and 4) control energy from the model including regional self-inhibition parameters better predicts cognitive performance on several tasks. The manuscript is well written, easy to read, and has a logical flow to it. The replications across human and mouse datasets are strengths. I found the conclusions to be well supported by the data and believe the results will be of great interest to a wide audience. I have a few comments that I believe would increase the significance of the manuscript.*

A: We thank the reviewer for the thorough and positive assessment of our manuscript, and we address the reviewer’s comments below.

3.1. R3 comment #1

R: *The control set analyses are very interesting and help move NCT to more biologically realistic and relevant uses such as TMS, tDCS, and DBS (as mentioned by the authors). I think it would be of great interest to readers if the authors could more thoroughly investigate the subset of nodes that were necessary to achieve successful state transitions. Biologically, I would expect certain brain regions to have a greater influence, likely determined by their pattern of connectivity.*

A: We thank the reviewer for this excellent suggestion. To address this comment, using our $INT_{optimized}$ model, we performed a greedy node removal analysis to quantify each node’s importance for control. Specifically, starting from a uniform full control set, for each state transition, we removed one node at a time and re-computed control energy. This gave us a vector of length N that quantified how much energy increased when a given node was removed from the control set. Next, we selected the node that caused the smallest increase in energy and removed it permanently from the control set. We repeated this entire process starting from the new control set that excluded the control node removed in the previous run. We repeated this greedy search algorithm iteratively, permanently removing one control node at a time until only 50 control nodes were left. For each transition, this greedy algorithm yielded a list of length $N - 50$ that stored the iteration (integer values) that each node was removed from the control set. In this list, lower values represent nodes that were less important for control (since they were removed earlier in the greedy search) and higher values represent nodes that were more important for control (since they were removed later). We averaged these lists across all state transitions and then correlated the corresponding mean node importance values with empirical INTs (ACF_{lag}) as well as nodes’ intracortical myelin and unweighted degree. We report the results of this analysis in the revised Supplement:

Assessing control node important using greedy search

In the main text, we demonstrated the ability of networks with optimized decay rates to better achieve state transitions from partial control sets (see Fig. R3). Here, to further explore the biological significance and correlates of these partial control sets, we implemented a greedy search

algorithm that identified an ordering of control nodes from least to greatest impact on control. Specifically, using the adjacency matrix with the optimized decay rates, and beginning with a full control set ($B = I$), for each node we computed the change in control energy, ΔE_i , required to achieve a given state transition when control node i was removed. We then identified the node whose removal caused the smallest increase in control energy and permanently removed that node from the control set. We repeated this node removal process iteratively 350 times, resulting in an ordered list of 350 nodes (out of 400) that ranked nodes from the smallest to largest impact on control. The remaining 50 nodes had the greatest impact on control energy. Note, we stopped the procedure at 350 nodes because the computation of optimal control was no longer numerically stable after this point. We performed this whole procedure for all 42 state transitions separately, resulting in 42 lists of node importance. Finally, we averaged these lists over state transitions, and correlated nodes' mean control importance with their empirical INTs (ACF_{lag}), myelin, and binary degree (Fig. R6). We found a strong positive correlation between node importance (iteration removed) and ACF_{lag} (Fig. R6A; $r = 0.69, p = 3 \times 10^{-57}$), as well as weak negative correlations between node importance and myelin (Fig. R6B; $r = -0.14, p = 5 \times 10^{-3}$) and degree (Fig. R6C; $r = -0.12, p = 2 \times 10^{-2}$). These results indicate that nodes that are more important for control exhibit longer empirical INTs. More broadly, these results highlight that nodes' importance for control is coupled to a much greater extent with brain function than with brain structure (i.e., myelin, degree).

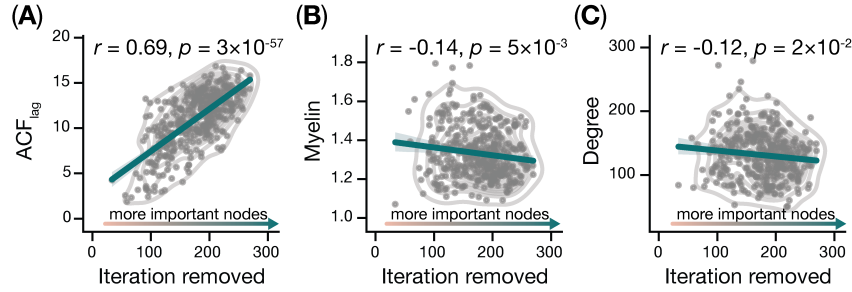


Figure R6: **Nodes' importance for control couples tightly to brain function.** (A) Correlation between node importance (iteration removed) and empirical INTs (ACF_{lag}). (B) Correlation between node importance and intracortical myelin. (C) Correlation between node importance and unweighted degree.

As can be seen from the above results, the nodes that are most important for control have the longest empirical timescales, as measured by resting-state fMRI. Notably, this coupling is very strong, yielding a correlation of $r = 0.69$. Finally, this effect is clearly stronger than the coupling between node importance and measures of regional brain structure (myelin) and connectivity (degree).

3.2. R3 comment #2

R: *I understand that there are separate self-inhibition parameters estimated for each state transition, but it is unclear to me why the authors chose to look at the principal components of the self-inhibition parameters as opposed to the average over all state transitions. This is especially relevant given that the average self-inhibition parameters correlated most strongly with INTs measure from resting-state fMRI data ($\rho = 0.70$ for the average vs $\rho = 0.42$ for PC1 and $\rho = 0.22$ for PC2).*

A: We thank the reviewer for this comment and apologize for the confusion. We are happy to clarify. The choice to primarily study PCs of optimized decay rates, rather than the mean, was made intentionally to respect the fact that our optimized decay rates are clearly sensitive to specific state transitions. We reasoned that a PCA respected this between-transition variance to a greater extent than analyzing the mean (which collapses between-transition variance in a greedy way). We believe that our results highlight the utility of this choice. For example, we find that both PC 1 and 2 exhibit weaker correlations with empirical timescales than the mean does, but this is unsurprising given that PCs encode separate sources of variance into orthogonal components while the mean mixes those components together. In this way, the correlation of 0.70 represents a ceiling on what the PCA results can capture (since mean and PCA are both linear operations). The value of the PCA here is that, while the correlations are weaker, it demonstrates that resting-state empirical INTs encode a complex mixing of separate sources of variance in regional dynamics. Furthermore, our results show that only one of these dominant sources of variance (PC 1) tracks inter-regional differences in static features of neurobiology (e.g., gene expression). Again this highlights the utility of our PCA because it tells us not only that resting-state INTs represent a complex mixing of signals, but also that some of these signals connect with microscale neurobiology while some do not. We have clarified our intentions briefly in the revised manuscript:

Note, we found that the mean of the optimized decay rates (taken from Fig. R2B, left) also correlated positively with ACF_{lag} (Fig. ??; $\rho = 0.70$, $p = 4 \times 10^{-59}$), but we focused here on results from the PCA because of the clear between-transition variance present in our optimized decay rates (Fig. R2B, right). These results ...

3.3. R3 comment #3

R: *The analyses with gene expression are very interesting and provide important biological context for their results. I think it would be of great interest to readers if the authors could also investigate relationships with excitatory gene expression and pyramidal cells. This is of particular interest given the finding that the hierarchy of INTs is believed to be driven by dendritic spine density, i.e. recurrent excitation (Chaudhuri, et al., doi:10.1016/j.neuron.2015.09.008).*

A: We thank the reviewer for this suggestion and are happy to include correlations with excitatory genes. Drawing on Deco et al. (Science Advances, 2021 doi:10.1126/sciadv.abf4752), we extracted two subgroups of excitatory gene transcripts comprising AMPA and NMDA receptors. The AMPA group comprised transcripts for GRIA1, GRIA2, GRIA3, and GRIA4, while the NMDA group comprised transcripts for GRIN1, GRIN2A, GRIN2B, and GRIN2C. We averaged over the transcripts in each group and correlated them with PC 1 of our optimized decay rates. We have added a new Supplementary Figure that displays these effects:

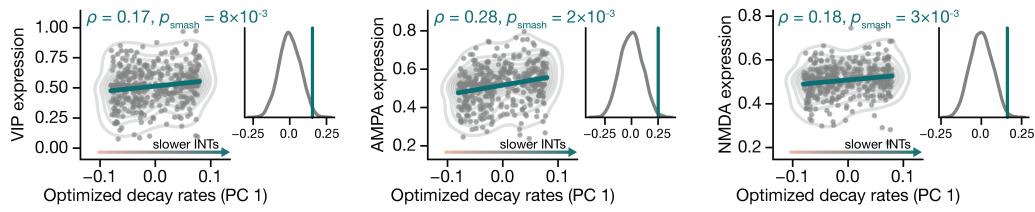


Figure R7: **Model-based intrinsic neural time scales correlate with whole-brain maps of gene expression.** Correlations between PC 1 of optimized decay rates (taken from Fig. ??) and whole-brain maps of gene expression extracted from the Allen Human Brain Atlas [23]. We correlated PC 1 with gene transcripts for vasoactive intestinal peptide interneurons (VIP, left) as well as groups of excitatory AMPA (middle) and NMDA (right) receptor subunits. Following [20], we averaged over the expression of GRIA1, GRIA2, GRIA3, and GRIA4 to estimate AMPA expression, as well as over the expression of GRIN1, GRIN2A, GRIN2B, and GRIN2C to estimate NMDA expression. We assigned p-values to these correlations using brainSMASH [12].

3.4. R3 comment #4

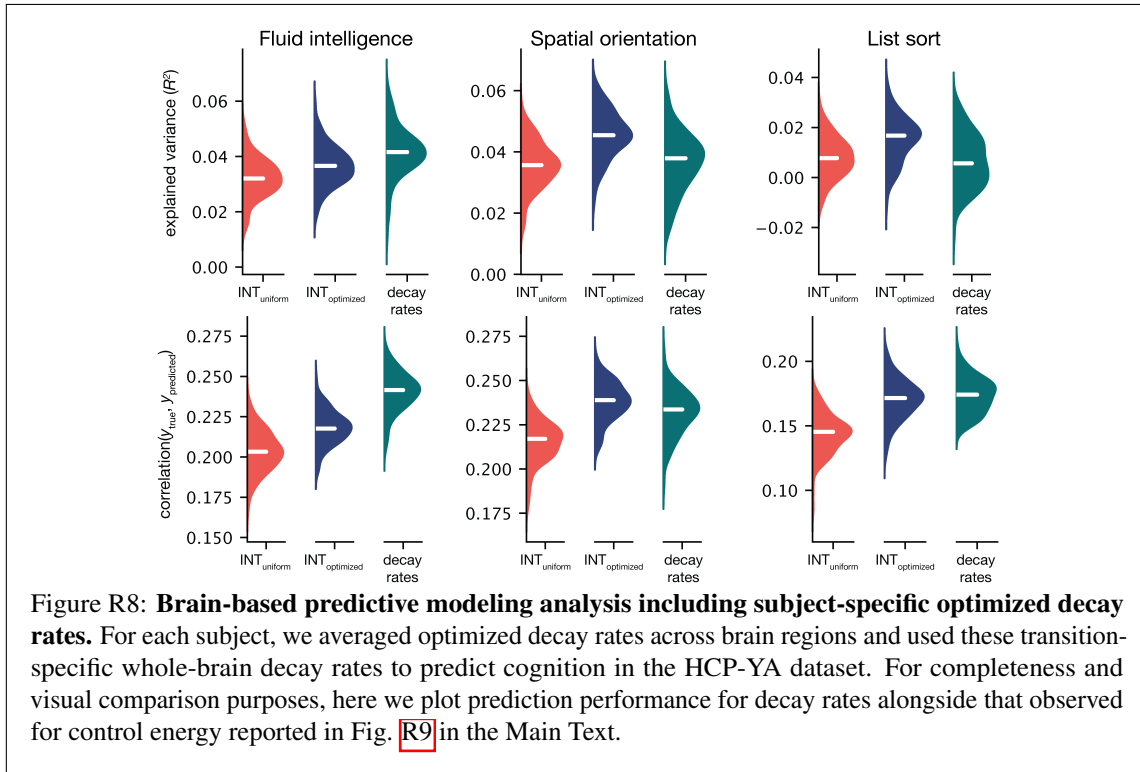
R: *Relatedly, why did the authors not investigate relationships with VIP if their focus was on interneurons?*

A: We thank the reviewer for noting this omission. We have now included a correlation with VIP in the above mentioned new Supplementary Figure (see response to R3C3 above).

3.5. R3 comment #5

R: *The authors very nicely show using brain-based predictive modeling (BPM) that including regional self-inhibition parameters in the NCT model improves the ability to predict individuals' cognitive performance using control energy. I would be interested to see if using the regional self-inhibition parameters improves predictive performance compared to control energy or if the combination (both self-inhibition and control energy) improves performance over using just control energy. This would speak to the functional importance of INTs as they relate to cognition which is something that has been investigated using both resting-state fMRI and ECoG measured INTs.*

A: We thank the reviewer for this suggestion. Originally, we chose to only use control energy in the BPM analysis because it is the only quantity that we are able to estimate across both models. That is, while we obtain decay rates that exhibit between-subject variance for the $INT_{\text{optimized}}$ model, no such variance exists for the INT_{uniform} model. Nevertheless, in the revised supplement we have now included a new figure that includes predictive performance estimated from optimized decay rates:



3.6. R3 comment #6

R: *Related to the BPM analyses, I am not sure why the authors chose to report results for the correlation between the true y and predicted y as opposed to the R -squared, given they say the correlation is the secondary outcome measure. Given current best practices (e.g., Poldrack et al., doi:10.1001/jamapsychiatry.2019.3671), I suggest reporting R -squared in the main text for the measurement of predictive performance and would also suggest including a measure of error such as mean absolute error or root mean squared error.*

A: We thank the reviewer for this comment. We have updated Figure 7 to report explained variance (R^2) and correlation between true and predicted for cognitive variables that yielded positive explained variance in our BPM analysis:

Regarding the second analysis, using a cross-validated penalized regression model, we examined whether control energy from our $INT_{\text{optimized}}$ model outperformed the INT_{uniform} model at out-of-sample prediction of behavior (see Materials and Methods for more details, as well as Fig. R8 for results using optimized decay rates to predict cognition). While we examined a broad range of cognitive variables present in the HCP-YA dataset, here we only report results for variables that yielded positive explained variance (R^2) scores in our regression model. Across 100 repeats of 5-fold cross-validation, we found that **control energy from the $INT_{\text{optimized}}$ model outperformed the INT_{uniform} model across all behavioral measures that met this criteria** (Fig. R9B). Additionally, empirical null models (see Materials and Methods) revealed that prediction performance for both variants of control energy exceeded chance levels (Fig. R9B, insets; all p s < 0.05). These results show that optimizing the internal decay parameters of our NCT model at the single-subject level leads to a significant improvement in the out-of-sample prediction of cognition in the HCP-YA.

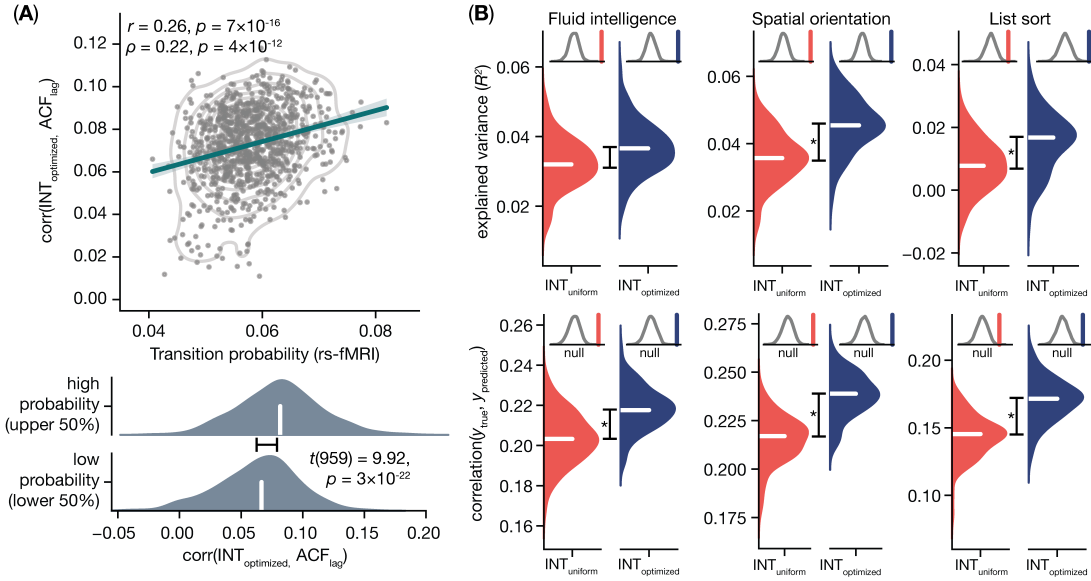


Figure R9: Fitting the $\text{INT}_{\text{optimized}}$ model to subject-level data yields sensitivity to individual differences in empirical INTs and improves the out-of-sample prediction of cognition. (A) Correlations between subject-specific optimized decay rates (model-based INTs) and subject-specific ACF_{lag} (empirical INTs). In the top plot, we calculated the transition probabilities for all state transitions as they occurred in each subject's rs-fMRI timeseries data (x-axis). Next, we calculated the correlation between model-based INTs and empirical INTs for each subject (y-axis). Then, we averaged these quantities over state transitions and correlated them. We found that subjects who exhibit more frequent state transitions during their rs-fMRI scan exhibit stronger coupling between their model-based INTs and their empirical INTs. In the bottom plot, the same analysis was performed focusing on within-subjects effects. Here, the correlations between model-based INTs and empirical INTs were averaged separately over the most frequent state transitions (upper 50%) and the least frequent (lower 50%). A paired samples t -test indicated that coupling was significantly stronger for the most frequent state transitions ($t(959) = 9.92, p = 3 \times 10^{-22}$). (B) Results from a brain-based predictive modeling analysis that used subject-specific control energy to predict individuals' cognition. We extracted subject-specific control energy from both $\text{INT}_{\text{optimized}}$ and $\text{INT}_{\text{uniform}}$ models and used those features as predictors in separate cross-validated penalized regression models. Control energy from the $\text{INT}_{\text{optimized}}$ model outperformed that of the $\text{INT}_{\text{uniform}}$ model at predicting subjects' cognitive scores. The top row shows prediction models scored according to out-of-sample explained variance (R^2), while the bottom row shows models scored according to the correlation between true and predicted scores. See Fig. R10 for models scored according to root mean squared error (RMSE). Permutation testing revealed that energy from both models performed above chance levels (insets). $*$ = $p < 0.05$ for an exact test of differences between paired performance distributions.

Additionally, we have added a new supplementary figure that reports BPM results scored according to (negative) root mean squared error:

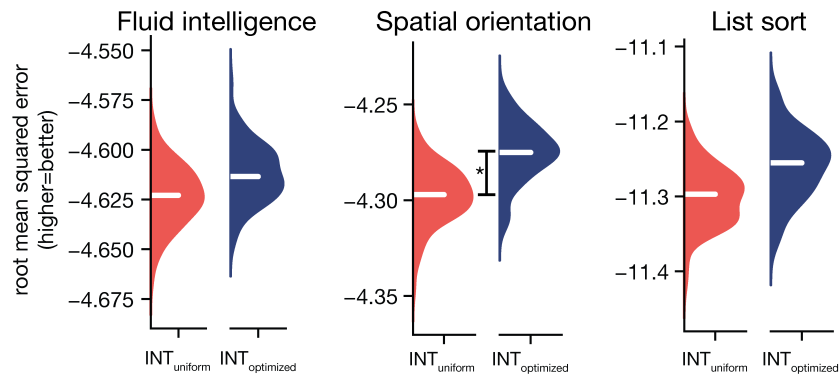


Figure R10: **Brain-based predictive modeling analysis scored using root mean squared error (RMSE).** Here, RMSE is shown with a negative sign so that higher scores equal better performance.

3.7. R3 comment #7

R: *Further related to the BPM analyses, I'm not sure that comparing the distributions of cross-validated predictive performance between the INTuniform and INToptimized models is the most appropriate method for determining significant differences in performance. I would suggest generating a null distribution of the difference in predictive performance using the null models for each individual method and comparing the difference in the mean cross-validated predictive performance between the two methods to this null distribution. This should be relatively straight forward to do since the authors already have the null distributions for each method.*

A: We thank the reviewer for this comment. We recomputed performance differences using the reviewer's suggestion and found that all reported significant differences in prediction performance were maintained. We also note that the approach we took originally is consistent with our several previous work that also included comparative BPMs. As such, and in light of consistency of results between methods, we have chosen to retain our original approach in the revised manuscript.

3.8. R3 comment #8

R: *How correlated are the regional estimates of self-inhibition across the HCP-YA and MICA-MICs datasets? This would help demonstrate the stability of the proposed NCT method.*

A: We thank the reviewer for this suggestion. We find that the group-averaged optimized decay rates correlate positively across the brain between the HCP-YA and MICA-MICs datasets at $r = 0.50$. We have now included this result in the supplement alongside the MICA-MICs replication analysis:

We found that our results replicated, with the $INT_{\text{optimized}}$ model yielding lower control energy across all state transitions. **We also observed that average optimized decay rates from the HCP-YA dataset (taken from Fig. 2B) correlated positively with average optimized decay rates from the MICA-MICs dataset ($r = 0.50, p = 1 \times 10^{-26}$; $\rho = 0.50, p = 4 \times 10^{-26}$).** These replication efforts demonstrate that our findings were robust to differences in sample characteristics, data acquisition, data processing, and state definition.

3.9. R2 minor comments

R: *It would be nice to add a plot for this finding even if in the Supplement: "Note, we found that the mean of optimized decay rates (taken from Fig. 2F) also correlated positively with ACFlag ($\rho = 0.70$, $-\log_{10}(\text{psmash}) > 25$)."*

A: We agree, and as such have add a new Supplemental Figure to the revised manuscript:

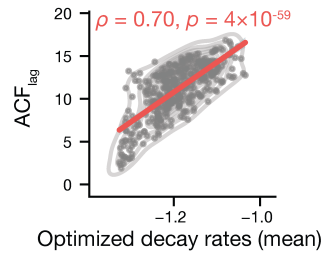


Figure R11: **Model-based intrinsic neural time scales (INTs) correlate with empirically-measured INTs.** Here, we show the correlation between optimized decay rates averaged over all transitions and empirical INTs estimated using ACF_{lag} (see Main Text).

R: *Typo: “That is, optimized decay rates correlated positively with AHBA-SST, AHBA-PVALB, and PVALB-Chandelier, and negatively with PVALB-Basket,” should be: “That is, optimized decay rates correlated positively with AHBA-SST and PVALB-Chandelier, and negatively with AHBA-PVALB and PVALB-Basket.”*

A: Thank you, we have fixed this typo.

R: *Typo: “Briefly, this approach challenges another long-held assumption of NCT that each control signal exclusively impacts a signal node,” should be: “Briefly, this approach challenges another long-held assumption of NCT that each control signal exclusively impacts a single node;”*

A: Thank you, we have fixed this typo.

3.10. R3 remarks on code availability

R: *The code is well-documented and should be usable by the community. I was unable to run the code because the necessary data are not provided (which is reasonable given the nature of the data). One minor comment is that in one of the python notebooks there is output of individual subject data including restricted data elements (Age in Yrs) which should not be present.*

A: Thank you for pointing this out. We have corrected this issue and updated the GitHub

Authors' Response to Reviews of

Inferring intrinsic neural timescales using optimal control theory

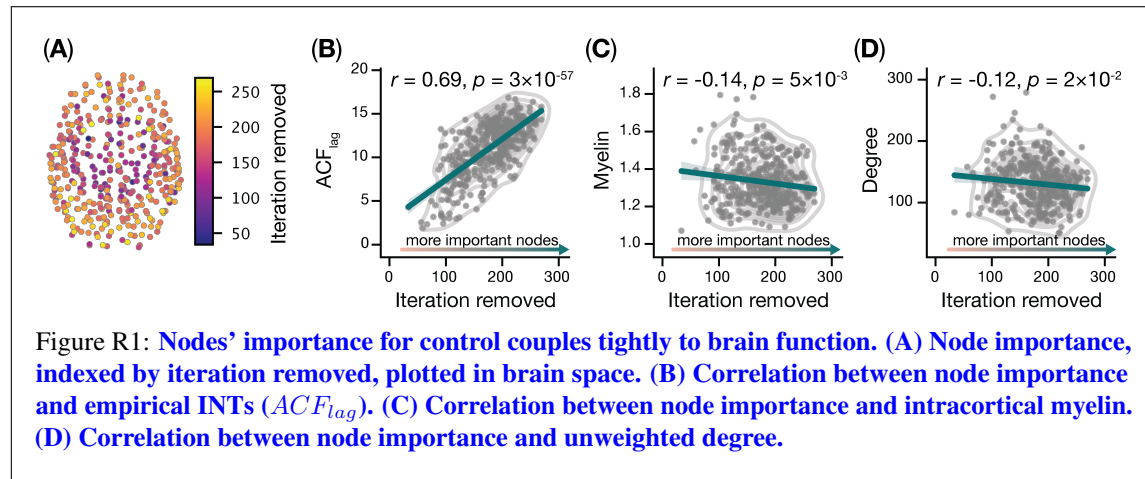
Jason Z. Kim, Richard F. Betzel, Ahmad Beyh, Amber Howell, Amy Kuceyeski, Bart Larsen, Caio Seguin, Xi-Han Zhang, Avram Holmes, Linden Parkes
Nature Communications

R: Reviewer's Comment, A: Authors' Response, □ Manuscript Text

1. Reviewer #3

1.1. R3 comment #1

- R:** A figure showing the “iteration removed” metric as a brain map to provide spatial information regarding the results would be helpful. It would also be beneficial to discuss these findings in the “Controlling brain dynamics effectively from fewer brain regions” section of the discussion in conjunction with the findings from Fig. 3 that are currently discussed.
- A:** We thank the reviewer for this suggestion, and have updated the corresponding supplementary figure accordingly:



We have also referenced these results in the Discussion section of the revised manuscript:

We found that the $INT_{\text{optimized}}$ model enabled successful low-energy state transitions from fewer regions than were required for the INT_{uniform} model (Fig. 3), and that the most important regions for control had the slowest model-based INTs (Fig. R1). This result demonstrates that expanding the biophysical realism of NCT has the potential to close the gap between the historical application of NCT and one of its clinical use cases (modeling neurostimulation).

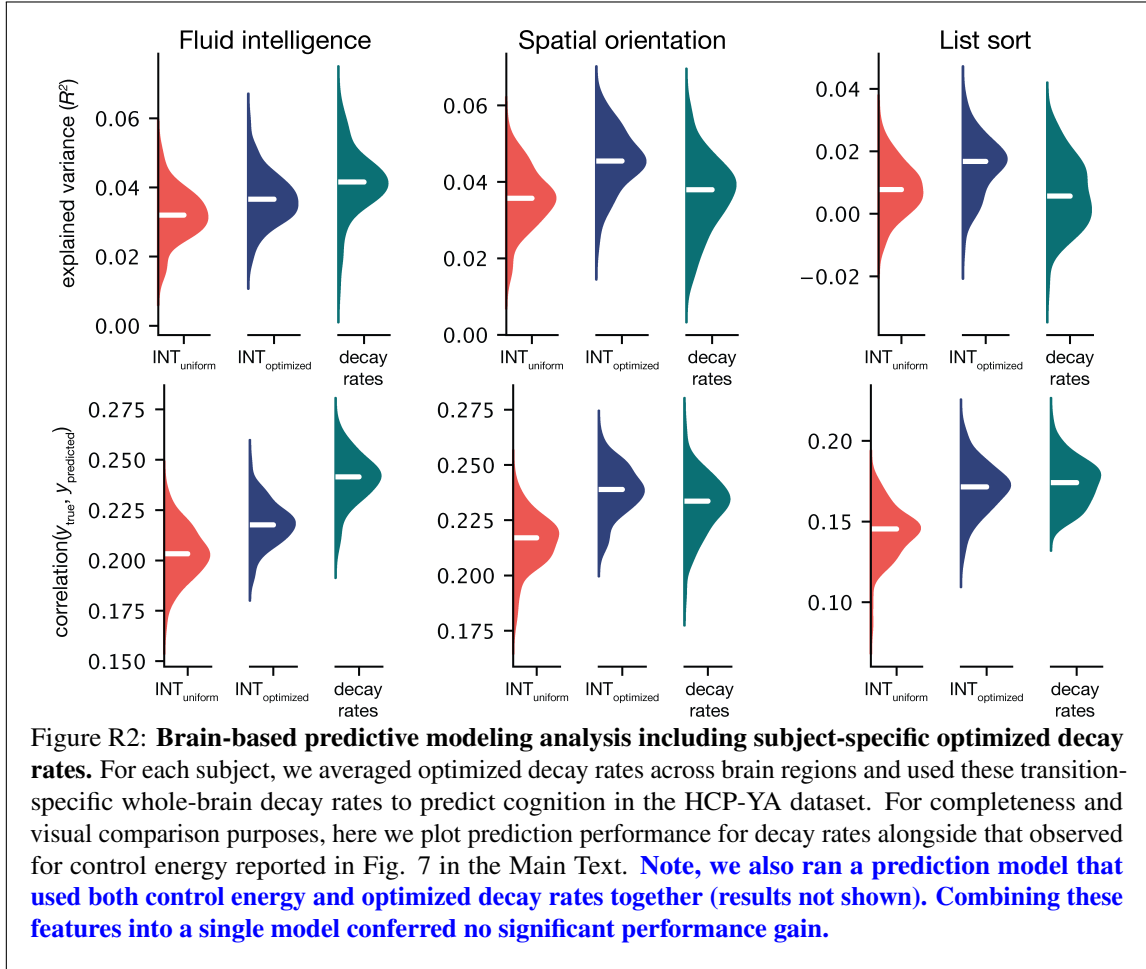
- R:** I appreciate the addition of the VIP, AMPA, and NMDA analyses. I believe these results should be mentioned more explicitly than simply stating: “see also Fig. S8 for correlations between PC 1 and an extended selection

of gene transcripts from the AHBA.” It should be noted that the relationships observed with SST and PVALB are not specific to those genes since similar relationships are observed for the other gene sets.

- A: We have expanded the corresponding section of the revised Results to more specifically articulate the points raised by R3:

Next, we observed a negative correlation between PC 1 and intracortical myelin content (Fig. S7; $\rho = -0.54, p_{smash} < 1 \times 10^{-5}$), which is consistent with the observation that highly myelinated sensorimotor regions exhibit faster INTs [1]. **See also Fig. S8 for correlations between PC 1 and an extended selection of gene transcripts from the AHBA, including expression of VIP, AMPA, and NMDA genes; these results demonstrate that our model-based INTs are sensitive to the spatial patterning of genes other than PVALB and SST.** Finally, we observed no significant correlations between PC 2 of optimized decay rates and any of the above brain maps. Thus, while there are two dominant sources of variance in our model-based INTs that correlate with empirical INTs (see Fig. 4), only the first source of variance connects with maps of neurobiology that reflect cortical inhibition. Collectively, these results demonstrate that parameters from our $INT_{\text{optimized}}$ model track the differing spatial patterns of distinct inhibitory interneurons across the cortex.

- R: *The authors did not investigate the combination of control energy and decay rates to see if adding decay rates improves predictive performance. Since decay rates are also able to predict cognitive performance, it would be useful to understand if control energy and decay rates together better predict cognitive performance or if they are capturing similar underlying variability which results in similar predictive performance when used in isolation. If it is the former, this would suggest that control energy and decay rates are capturing distinct sources of variability that are both important for cognition.*
- A: We apologize for this omission. We did run the joint predictive model using control energy and decay rates but did not report it because we found that it did not improve performance. We have made a note of this in the revised manuscript:



1.2. R3 remarks on code availability

R: *It does not look like they added the code for running the cognitive performance prediction using the decay rates*

A: We apologize for this omission. We have now updated `results_control_energy_subjects.ipynb` to include the code for running predictive modeling using decay rates. To run the predictive models with decay rates, `include_weights` must be set to `True`.

References

- [1] Annemarie Wolff, Nareg Berberian, Mehrshad Golesorkhi, Javier Gomez-Pilar, Federico Zilio, and Georg Northoff. Intrinsic neural timescales: temporal integration and segregation. *Trends in Cognitive Sciences*, page S1364661321002928, January 2022. ISSN 13646613. doi: 10.1016/j.tics.2021.11.007. URL <https://linkinghub.elsevier.com/retrieve/pii/S1364661321002928>.

Since the seminal paper by Gu et al. (2015), network control theory (NCT) has been frequently applied in “network neuroscience” studies focusing on large-scale human brain networks (accessible through functional, structural and diffusion MRI). In principle, brain NCT would allow designing an exogenous control signal driving the brain into a desired activity state, and hence it could be used to inform brain stimulation protocols. However, fine-grained control of the brain state would require simultaneously delivering stimulation to a large number of brain sites, which is largely beyond current limitations of brain stimulation protocols. Thus, in practice, most of the brain NCT literature has focused on more descriptive tasks such as estimating the ‘cost’ of transitions between different brain states, encapsulated by the control energy (the integrated amplitude of the control signal). Exploring the control energy landscape can highlight nodes that are putatively more central for brain dynamics, assess the difficulty of reaching specific target states, and altogether quantify the ease of transitioning between different activity states - and variations in all these properties due to development, aging, pathology, etc.

The most commonly used brain NCT model is fully linear, and combines diffusive dynamics along structural connections (as given by diffusion MRI), with local self-inhibition to ensure stability. The self-inhibition parameters, which are akin to local timescales, are usually taken to be uniform. In their manuscript entitled “Inferring macroscopic intrinsic neural timescales using optimal control theory”, the Authors augment NCT with a parameter optimization scheme that numerically optimizes the self-inhibition terms (‘timescales’) to ensure that trajectories remain close to the target state. Their methodology is applied to two human data sets and a rodent data set.

Their main results are the following:

- the proposed scheme allows systematically reducing the control energy, and reducing the number of driver nodes needed for accurate control
- across brain regions, the optimized timescales correlate with empirically measured timescales (as revealed by fMRI autocorrelation), as well as with spatial variation in the abundance of different types of inhibitory neurons and related gene expression (revealed by gene expression maps).
- results of the human and rodent model are fully consistent
- the control energy landscape emerging from the timescale-optimized model allows better prediction of performance scores in a range of cognitive tasks

Overall, the study is methodologically clear, statistically well-powered, and findings are robust. The methodology proposed is certainly an original development of NCT applied to large-scale neuroscience. The results are surely of large interest to the small community focusing on brain NCT. However, I think that the current presentation is not suitable to capture the attention of the wide neuroscience community. Therefore, I believe that substantial changes in presentation are needed to make it suitable for Nature Communications.

The main problem lies in the overall ‘take home messages’, as they emerge from the Abstract, Introduction, Discussion and (partially) Results sections. They tend to overemphasize comparatively weak findings, while not placing enough emphasis on other, potentially relevant ones. In what follows, I will expound my main concerns on this point.

Technical development. The manuscript introduces a novel scheme to couple NCT with automatic differentiation. As the Authors correctly state,

“So far no study has sought to optimize this constraint by modifying the connectivity to minimize the delta between $x(t)$ and x_r . Here, we extend NCT to allow for this minimization.” (Discussion)

In fact, I think that this relevant technical development could be emphasized more. This scheme allows in principle asking new questions not only about timescales, but potentially also about connectivity. It could be, e.g., used to optimize non-diagonal entries in the matrix, too, and thus provide a link between optimal control and effective connectivity.

Inferring biologically relevant parameters. From the very title, the Authors place a large emphasis on using their framework to infer biologically relevant timescales. In fact, all the evidence for a relation between NCT-optimized timescales and biology depends on the correlation with known neurobiological metrics. But what is special, then, about NCT-optimized timescales? I suppose that also ACF-based timescales show a similar degree of correlation with neurobiological metrics, and they are easier to infer from fMRI. One could argue that NCT-optimized timescales are more interpretable, as they represent ‘self-inhibitory’ terms in a linear dynamic model. However, one could obtain such terms by directly fitting a linear dynamic model to fMRI data (see, e.g., Nozari, Erfan, et al. "Macroscopic resting-state brain dynamics are best described by linear models." *Nature biomedical engineering* 8.1 (2024): 68-84.), without need to invoke NCT. Moreover, to what extent these terms in the linear model would really correspond to a macroscopic manifestation of microscopic inhibition is unclear. Thus, the claim that

“we estimate the structural form of regional self-inhibition—a quantity that is experimentally difficult to measure—from MRI data”

is not fully warranted in my opinion. Also not fully supported is the claim that

“These results robustly validate our approach as a means of uncovering biologically-grounded whole-brain INTs directly from the connectome in vivo.”

In fact, NCT-optimized timescales do not solely depend on the connectome, but also from the chosen set of transitions, which in turn is derived from functional data.

Finally, if one claims a strong relation with neurobiology, the dependency of the timescales on the chosen transition appears problematic. In fact, the argument

“A key advantage of our data-driven approach is that it enables us to uncover both transition-specific and subject-specific INTs, and to use them as proxies for measures of neurobiology that are otherwise unobservable in vivo.” (Discussion)

is not fully clear: it would seem that the transition-dependence of optimized NCT rates actually undermines their usage as ‘proxies’ for (static, transition-independent) neurobiological measures.

Applicability to brain stimulation. The applicability of the NCT paradigm (and its timescale-optimized version) for brain stimulation is likely overstated.

“These results indicate that our approach—which seeks to improve the biophysical realism of NCT—enables the successful control of brain dynamics from relatively focal inputs. In turn, this improvement to model fitting reduces the gap between NCT and real-world applications of brain stimulation.”

“this added parameter flexibility allows the INTOptimized model to find more efficient state transitions by uncovering, and then harnessing, transition-specific regional dynamics”
(Discussion)

Control energies required for fine-grained activity control are exceedingly high unless a sizable fraction of the nodes is controlled with external inputs (see Tu et al, *NeuroImage* 176 (2018): 83-91, but also Fig. 3C in the manuscript). In practice, control of more than 50 nodes is required, when current techniques allow at best for 1-2 nodes. The Authors refer to “*relatively few control nodes*”, but 50 is not few, in the light of current technical limitations.

In addition, the timescale-optimized version of NCT does not clearly contribute to closing the gap, because, contrary to the input signals, *the timescales are not externally controllable parameters*. Thus, while transition-optimized timescales would allow to enact the transition more easily, one does not see a simple way to externally manipulate the timescales to facilitate a preferred transition. This point is not clearly discussed in the manuscript.

(Perhaps one could think of using pharmacology to manipulate local inhibition, thus altering timescales locally to facilitate specific transitions?)

Timescales determine the control energy landscape. In my opinion, one of the most relevant findings of the manuscript is the robust evidence that timescales, not just connectivity, significantly contribute to the control energy landscape. This finding is not totally surprising, since diagonal terms significantly impact on the spectrum, and hence on the general controllability of the system. However, this was not discussed by previous literature, and it may deserve more stress.

While the correlation between optimized timescales from NCT and empirical timescales is highlighted as evidence that the optimized timescales are biologically grounded, it may be also (more convincingly, to me) be brought as evidence that *observed empirical timescales might be optimized to ensure easier transitions between frequent activity states in spontaneous brain dynamics*. In other words, the NCT results might contribute to explaining the spatial structure of observed timescales, and its changes during development/aging/pathology, in terms of optimal control. The Authors seem to recognize this when they say

“Our approach has the potential to connect these entwined phenomena to better understand how the multi-scale relationship between brain structure and function emerges throughout development.” (Discussion)

In fact, additional analyses may even strengthen this perspective, with possible impact on the general understanding of large-scale brain function. Possible analysis steps could be (in order of feasibility):

- a) One may conjecture that timescales in single subjects are optimized to facilitate transitions between the most frequently occurring CAPs. One should then observe a stronger similarity between the empirical timescales in single subjects and the NCT timescales optimized *for the CAP transitions occurring most frequently in the single subject*.
- b) One could conjecture that individuals whose empirical timescales are more 'optimal' would display enhanced cognitive abilities. One could then use the HCP dataset to compute the correlation between the average (over transitions) optimized timescales and the empirical timescales in a single subject basis, and in turn test whether this correlation is predictive of cognitive performance
- c) Using the task data in the HCP data set, one could compute the optimized timescales to transition from an average 'rest' state to the average 'task' state and back, and check whether the empirical timescales during the task correlate with these optimal timescales. This would provide evidence that the brain can flexibly adapt timescales to facilitate specific transitions.
- d) Using data from the HCP aging data set, one could analyze changes in the optimized timescales during aging. These optimal timescales would adapt to, and potentially compensate for, aging-related changes in the connectome. If aging-related changes in optimal timescales are reflected in aging-related changes in empirical timescales, this would again constitute relevant evidence that timescales are a key mechanism to confer flexibility and resilience to brain dynamics.
- e) In neurological disease, the connectome is often irreparably damaged. Can a timescale adaptation partially compensate for this? An analysis of empirical and optimized timescale changes in, e.g., a stroke dataset, would be of significant clinical interest.

Of course I am not asking the Authors to perform all these analyses. But I think that adding one or two of them, or analogous ones, might strengthen the article.

Methodology. While the methodology is clearly detailed, some additional information could be useful. First, many brain NCT articles focus on minimizing just the control energy (with no constraint on the trajectory). Could one try to directly minimize the control energy with respect to timescales? After all, the control energy can be expressed analytically in terms of the initial and final states and the controllability Gramian, so it would not seem impossible to use this as a cost function. Would this approach lead to significantly different results?

Secondly, can the approach (which leverages the Hamiltonian framework) be used for *asymmetric* connectomes?

Third, what was the time horizon chosen for control?.

A flexible structure-function coupling. In the Introduction and Discussion, the Authors often refer to 'structure-function coupling' (SFC), but the usage of this term, and the related discussion, are not fully clear. In the literature, SFC generally indicates the correlation between structural and functional connectivity (at the nodal or whole-brain level). The Authors seem to use 'SFC' in a different meaning, as the general relation between 'multiscale structural

parameters' (the connectome and local neurobiology) and the generation of common activity states,

"NCT has been able to flexibly probe how brain structure constrains brain function"

"However, these applications of NCT have relied upon a set of simplifying assumptions that lack biophysical realism, limiting the capacity for NCT to yield multi-scale insights into SFC."

"using optimized decay rates yields lower control energy associated with transitions between empirical brain states, demonstrating stronger SFC" (Introduction)

The terminology should be rigorously defined (what is intended exactly for 'SFC', and for 'multiscale'), as the authors emphasize this point several times:

"Understanding this multi-scale structure-function coupling (SFC) in the brain is a core goal of network neuroscience"

"These effects give rise to a spatial pattern of SFC that varies systematically across the brain" (Introduction)

Null models for the timescales. In my opinion, the discussion of null models in §2.3 is not very clear. The Authors present two null models, one that randomly rewires the connectome (preserving key network features) and one that randomizes the initial/target states (preserving spatial autocorrelation). The control energy gain afforded by optimizing timescales is significantly lower for the first model, and significantly larger for the second one. To me, it is not fully clear how these results can be interpreted, and the Authors' interpretation does not seem fully convincing.

"These results provide robust evidence that our optimized decay rates are jointly influenced by the topology of the connectome and the spatial patterning of the target states; as opposed to being driven exclusively by one or the other"

"The impact of the connectome and the brain states on the optimized decay rates are in tension with one another, and that our model balances these competing influences throughout optimization."

"Together, these results demonstrate our model-based INTs emerge as a product of jointly optimizing on real biological connectivity and activity"

In practice, the Authors deduce that there is a "tug of war" effect, where the connectome and the transitions have opposite effects on the timescales. Actually, to demonstrate such an effect, one should observe two opposing "forces" acting *on the timescales*, e.g., with the connectome 'pushing' for lower values of a given timescale and the transitions 'pushing' for higher values (or vice versa). This effect is not observed (to me, it is not even clear which analysis could reveal it). What is observed is only an effect on the *control energy gain* (the control energy reduction gained by optimizing timescales), which just implies that in presence of random transition states (which will generally be 'difficult-to-reach states') the control energy gain is larger than with frequent transition states, while in presence of a randomized

connectome (which will probably be a slightly more regular, exponential-decreasing-rule connectome) the energy gain is lower (perhaps because the randomized connectome is quite conducive to transitions?). These results are difficult to interpret, unless also the bare control energy values are shown in addition to the energy gains.

Other models of large-scale activity. The Authors say that their proposal comes in a situation where biologically-grounded, data-driven models of whole-brain activity are lacking:

“However, data-driven discovery of large-scale activity models remains difficult owing to the lack of flexible quantitative frameworks for estimating the interplay between brain structure and function while preserving biophysical realism” (Introduction)

I do not fully share this view. Recent literature features several hybrid biophysical-data driven models. Typical examples are Refs. 67, 80, 81 in the manuscript. It is not clear what would be the specific advantages of a NTC approach compared to whole-brain modeling approaches in these works. It is also not clear how the flexible component of timescale could be incorporated into dynamical models, that typically feature ‘static’ parameters

“This flexible component has implications for studies seeking to integrate spatially-patterned features of neurobiology into dynamical systems models”

By the way, the approach in Ref. 67 could also inspire a different version of timescale optimization in NCT. One could assume that the timescales are proportional to some neurobiological quantities (e.g., the densities of certain neurons), and let the diagonal terms in the matrix be parametrized by these maps as $A_{ii} = \alpha_1 b_{i1} + \alpha_2 b_{i2} + \dots + \beta$ (where b_{i1} , b_{i2} , ... are the neurobiological maps and α_1 , α_2 , ..., β are coefficients) and use automatic differentiation to optimize the linear combination parameters.

Correlation with neurobiological markers. The Authors put a lot of emphasis on the correlation with neurobiological maps. However, many cortical maps (including all neurobiological maps analyzed) broadly align with the cortical hierarchy (e.g., the principal connectivity gradient), as the authors correctly mention:

“Consistent with this idea, previous research has shown that higher-order association regions exhibit relatively high SST expression and slower INTs while primary sensorimotor regions exhibit relatively high PVALB expression and faster INTs”

“We observed a negative correlation between PC 1 and intracortical myelin content, which is consistent with the observation that highly myelinated sensorimotor regions exhibit faster INTs”

Thus, the correlation of the ‘fixed’ component of the timescales with these maps is hardly evidence of a *direct* link between the timescales and specific neurobiological maps. When the Authors state that

“Collectively, these results demonstrate that parameters from our INTOptimized model track the divergent spatial patterning of distinct inhibitory interneurons across the cortex.”

“ *Critically, our effects tracked known divergences*”

it is not fully clear what they mean. ‘Tracking the divergent patterning’ would imply that by looking at the timescales, one could predict when the neurobiological patterns are more divergent. This would be clear if, e.g., one of the principal components of the timescales aligned with the *difference* between two neurobiological maps, or at least if the correlation between timescales and a given neurobiological map would persist even after regressing out the effect of another map.

Minor comments

- “*Of particular biological relevance is optimal control*”. Why is it of particular biological relevance?
- In Fig. 2B, a better color map should be used to allow for better visualization.
- I guess a kernel density was used in Fig 2C to smooth out a histogram. This should be mentioned
- In Fig. 3A, one sees that, in addition to enhancing precision and lowering the energy, the timescale optimization allows to *reduce the dependence on the specific choice of input nodes*. This point could be stressed.
- In Fig. 3C, one sees that when the number of control nodes is below 80, there is no energy reduction when optimizing timescales. The Authors state that “*smaller control sets also produced lower control energy for the INTOptimized model compared to the INTUniform model (Fig. 3C)*”. A more nuanced statement would be in order.
- In Fig. 3D, which transition and which control node sets were used? Was this a typical case?
- When saying that “*single nodes innervate only relatively small subsets of the eigenmodes*” what do they mean by ‘innervate’? That the eigenmodes have nonzero projection (or projection above a threshold) on the nodes?
- The Authors say “*we performed eigenmode decomposition on the connectome, wherein the diagonals were set either to the transition-specific decay rates from INTOptimized model or to their mean*”. In Fig. 3D there are results for a specific transition. Results for the mean do not appear - not that they should be shown, but if they aren’t, it would be better to say it: “*we performed eigenmode decomposition on the connectome, wherein the diagonals were set either to the transition-specific decay rates from INTOptimized model (Fig. 3) or to their mean (not shown to avoid redundancy)*”
- Do the timescale-optimized eigenmodes more strongly align with *functional connectivity* eigenmodes, thus bridging structure and function?
- In Fig. S3, a better color map should be used to facilitate visualization
- From Fig. S4, one sees that PC1 projects more strongly on transition, involving CAP #4, which is the most similar to the default mode network. One could speculate that the average timescales are set to optimize transitions to/from the DMN.
- Fig. 4 reports an extremely small P-value. How can this emerge from a permutation approach (where the smallest P-value is set by the number of permutations)?

- “we also correlated PC 1, PC 2, as well as the mean of optimized decay rates with regional degree and strength estimated from the structural connectome and found no significant relationships”. Empirical timescales tend to correlate with degree, as also stated by the Authors, “INTs correlate tightly with the strength of regions’ structural connectivity to the rest of the brain”, so in principle it may be hard to disentangle the two contributions. I then do not think that the (weak) the correlation with degree is problematic.
- What is the correlation between the biological metrics in Fig. 4 and empirical timescales from fMRI? Are the optimized timescales more or less correlated than the empirical ones?
- “the dominant source of variance in optimized decay rates correlated negatively with in vivo microstructure”. What is the single microstructural metric used?
- “This result suggests that the exogenous control of large-scale brain dynamics from the connectome is rooted in microscale cellular processes”. Propagation of control signals along the connectome would in principle suffice to allow for exogenous control by an external agent. I would say, “the exogenous control of large-scale brain dynamics from the connectome can leverage microscale cellular processes”
- “we fit our NCT model to individual connectomes (see Materials and Methods)”. Were the 7 CAPs also recomputed at a single subject level?

- “Critically, our findings demonstrate convergent multi-scale linkages between brain structure and function across humans and mice.” Clarify what you mean by ‘multiscale’ (e.g., the model combined the macro-scale connectome with the meso-scale regional dynamics)
- “Our approach corrects a long-held assumption embedded within the NCT framework: that brain regions’ intrinsic dynamics are fixed and homogeneous across the brain.” Other Authors have used non-uniform timescales. See e.g. Manjunatha et al. 2024, where effective connectivity was used instead of structural connectivity, with node-dependent diagonal terms in the coupling matrix.
- “This assumption has limited the biophysical realism of NCT, in turn hampering its application to neuroscience”. This sentence contradicts the previous assertion that “The use of NCT in neuroscience is well-established” (§ 2.1). I would say, “This assumption has limited the biophysical realism of NCT, raising some concern for its widespread application to neuroscience”
- “As discussed in our recent protocol paper [23], controlling neural dynamics from very few nodes is often intractable in NCT”. This was first clearly discussed in Tu et al, *NeuroImage* 176 (2018): 83-91.
- In Eq. (5), $(x_r - x_T)$ should be probably replaced by $(x - x_r)$
- “Briefly, this approach challenges another long-held assumption of NCT that each control signal exclusively impacts a signal node; an assumption that is inconsistent with the spatial continuity of the cerebral cortex and its effects on dynamics”. While most papers applying NCT to the brain make this assumption, this is not a standard in

NCT in general. See e.g. Liu, Y. Y., Slotine, J. J., & Barabási, A. L. (2011). Controllability of complex networks. *nature*, 473(7346), 167-173.