

Quantifying Dysregulation of fMRI-Derived Control Circuits for Computational Psychiatry

Syed Sultan

State University of New York at Stony Brook

Steve Skiena

Stony Brook University

Lilianne Mujica-Parodi (✉ mujica@lcneuro.org)

State University of New York at Stony Brook

Article

Keywords: brain, circuit, dysregulation, fMRI, control system, trajectory, computational psychiatry, generative mo

Posted Date: March 17th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1413254/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Quantifying Dysregulation of fMRI-Derived Control Circuits for Computational Psychiatry

Syed Fahad Sultan^{1,✉}, Steven Skiena^{1,✉}, and Lillianne R. Mujica-Parodi^{2,3,✉}

¹Computer Science Department, Stony Brook University, Stony Brook, NY 11794

²Biomedical Engineering Department, Stony Brook University, Stony Brook, NY 11794

³Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA 02129

1 Psychiatric disorders are thought to result from *dysregulated*
2 brain circuits, yet human neuroimaging currently lacks stan-
3 dardized methods for quantifying neural dysregulation. Here,
4 we present a scalable framework for extracting fMRI-derived
5 (generative) control circuits, then use circuit trajectories to es-
6 timate their control error. Using synthetic circuits, we first
7 demonstrate that our framework accurately identifies each cir-
8 cuit’s architecture and models its dynamics by estimation of
9 transfer functions. As a use case, we then apply the frame-
10 work to human task-based functional MRI data (UK Biobank,
11 N=19,831). In a purely data-driven manner, without priors,
12 our framework identified thalamus-linked *prefrontal-limbic* and
13 *ventral stream* subcircuits, selectively engaged during sensori-
14 motor processing of affective and non-affective stimuli. Finally,
15 we demonstrate that circuit-wide dysregulation, defined by de-
16 gree of drift from healthy trajectories, tracks symptom sever-
17 ity for neuroticism (*ventral subcircuit*), depression (*prefrontal-*
18 *limbic subcircuit*), and bipolar disorder (*full circuit*).

19 brain | circuit | dysregulation | fMRI | control system | trajectory | computa-
20 tional psychiatry | generative model

21 Correspondence: mujica@icneuro.org

22 Introduction

23 Psychiatric disorders are commonly understood to reflect
24 *dysregulation* of one or more brain circuits. Yet, clinical
25 neuroscience generally conflates the term *circuits* with co-
26 activated brain regions, the latter of which are more accu-
27 rately described as *networks*. Because neuroimaging-derived
28 networks are normally defined by linear regressions ($y =$
29 $b_0 + b_1x_1 + \dots$), they are capable of reliably modeling only
30 a very narrow range of topologies, in which one or more in-
31 puts leads to a single output (1). This limitation excludes the
32 capacity for positive and negative feedback loops, as required
33 for regulation.

34 To quantify brain circuit dysregulation we exploit the gener-
35 ative aspect of data-derived control circuits, which allows us
36 to predict how a circuit’s output time series will evolve
37 over time. In a classic engineering control application, such
38 as autopilot (Figure 1a), a vehicle corrects for deviations from
39 its desired trajectory through negative feedback (e.g., *as the*
40 *vehicle starts to drift to the right, the control circuit corrects*
41 *the drift by steering to the left*). As such, the difference be-
42 tween the autopilot’s actual versus desired trajectories pro-
43 vides a measure of its *control error*, or dysregulation (Fig-

44 ure 1b). Here, we use trajectory drift as a measure of control
45 error. We calculate circuit-wide dysregulation across fMRI-
46 derived control circuits, and demonstrate its clinical utility as
47 applied to three psychiatric use cases: *neuroticism*, *depres-*
48 *sion*, and *bipolar disorder* (See Methods for definitions).

49 In developing this framework, we started from several
50 desiderata: the ability to test homeostatic regulation in re-
51 sponse to driving inputs (perturbation), a fundamental re-
52 quirement of control theory(2–4); the ability to conduct
53 whole-brain circuit discovery, free of priors; and the ability to
54 scale, thereby leveraging the marked increase in both mega-
55 scale neuroimaging datasets made possible through open-
56 science initiatives, as well as high resolution, fine-granularity
57 parcellations of the brain.

58 To date, the only standardized method capable of esti-
59 mating fMRI-derived generative circuits is Dynamic Causal
60 Modeling (DCM) (5). DCM is normally used to estimate cir-
61 cuit architecture, in the form of a directed, weighted graph.
62 However, circuit architecture by itself is not sufficient to
63 provide quantitative estimation of control parameters such
64 as dysregulation. Moreover, DCM’s computationally ex-
65 pensive algorithms, even for faster (resting-state only) vari-
66 ants such as spectral DCM (6), result in convergence times
67 so lengthy that they remain impractical for extracting cir-
68 cuits from mega-scale datasets, using purely data-driven ap-
69 proaches (>100 brain regions), with short (e.g., 5 minute)
70 time series. More recent (driving input-compatible) variants
71 such as regression DCM (rDCM) (7) and sparse rDCM (8) re-
72 place the hemodynamic forward model with a fixed hemody-
73 namic response function (HRF). As a result, they fail to allow
74 for heterogeneity in the blood oxygenation level dependent
75 (BOLD) signal across brain regions and individuals (9, 10).
76 This biophysical constraint can lead to confounds, particu-
77 larly when applied to neurodevelopmental, aging, and patient
78 populations (11, 12).

79 Thus, we introduce a generalized framework based on state
80 space systems that bridges the gap between network the-
81 ory and control theory, with the scalability required to mine
82 mega-scale datasets such as UK Biobank (N=19,831) (13).
83 First, we confirm that using time-series to estimate systems
84 of differential equations without biophysical constraints still
85 permits recovery of complex causal relationships and feed-
86 back loops that characterize brain circuits. Using synthetic
87 data, we generate canonical circuit motifs to simulate circuits

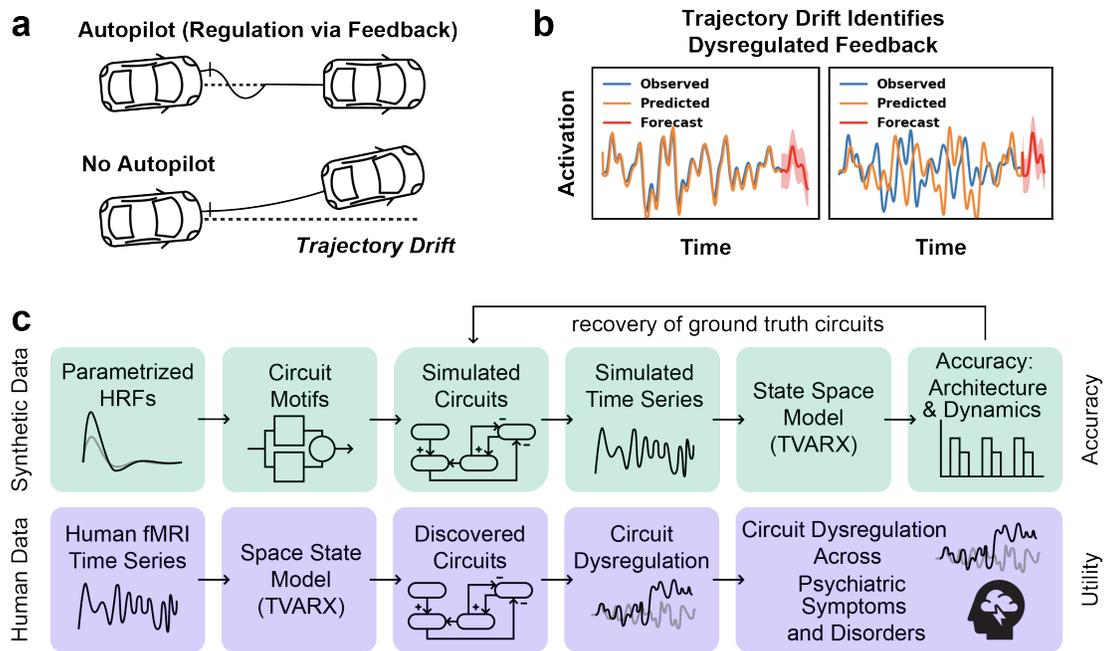


Fig. 1. Trajectory Drift as a Measure of Feedback Control Error, and Thus Circuit Dysregulation. (a) In a classic engineering control application, such as autopilot, a vehicle corrects for deviations from its desired trajectory through negative feedback. As such, the difference between the autopilot's actual versus desired trajectories provides a measure of its *control error*, or dysregulation. (b) As per the autopilot example, we use trajectory drift as control error to calculate dysregulation across fMRI-derived control circuits, and demonstrate its application for three psychiatric use cases: *neuroticism*, *depression*, and *bipolar disorder*. (c) Schematic of the pipeline for the discovery of circuit architecture and dynamics from human fMRI and simulated time series, using Time-Varying Autoregressive Model with Exogenous Inputs (TVARX) and other state space models. We use trajectory drift between predicted and actual trajectories to quantify circuit dysregulation across subjects with varying degrees of severity for psychiatric symptoms.

with varying architectures and dynamics to test our framework's ability to recover both (Figure 1c top row). Second, having validated the framework on synthetic data, we then apply the framework to UK Biobank fMRI data. Using tasks designed to dissociate processing of affective versus non-affective stimuli (14, 15), we extract the control circuit selectively engaged by each. Third, from each individual's circuits we calculate the circuit's trajectory control error, which quantifies its degree of dysregulation. From these control errors, we statistically test the relationship between circuit-wide dysregulation and psychiatric symptoms (Figure 1c bottom row).

Results

Recovering circuit motifs from dynamic outputs

We first evaluate our framework using circuit motifs. Synthetic circuits are constructed by connecting nodes, each with its own transfer function, according to three basic motifs: *series*, *parallel* and *feedback*. These motifs are then combined in a modular fashion, to create larger circuits of varying levels of complexity (See Methods, Figure 4a).

The transfer functions used in our experiments were designed to resemble the hemodynamic response function (HRF) (16) extensively used to model blood oxygen level dependent (BOLD) (17) signals measured using functional magnetic resonance imaging (fMRI) (Figure 4b). The HRF function is parameterized by response height, time-to-peak

and full-width at half-max. In our simulations, each node had different parameter values for the HRF, as previously shown for human data (9, 12).

The transfer function for each motif is an algebraic combination of node transfer functions (Figure 4c) Each motif also had an inverse variant. Serial and parallel connections each had both excitatory and inhibitory variants, while feedback loops had both positive and negative variants. The inverse variants are obtained by inverting the sign of their corresponding algebraic expressions (See Methods, Figure 4c). Note that although the node transfer functions are parameterized HRFs, successive connections and their corresponding algebraic expressions applied to HRFs can result in complex transfer functions.

We evaluate the ability to recover the canonical circuits both in terms of architecture and trajectory dynamics using a range of models with varying levels of complexity (Figure 2a). All of these models, with the exception of DCM, can be generalized by our state space system equations (See Methods, equation 3, Table 2). A detailed discussion regarding comparison with DCM is presented in a subsequent subsection.

The architecture is estimated through classifying relationship between each pair of nodes. For each node, parameters of its state space model are learned against time series of all other nodes (details in Methods). Based on the learned feed-forward matrix, a time-varying causality graph is established

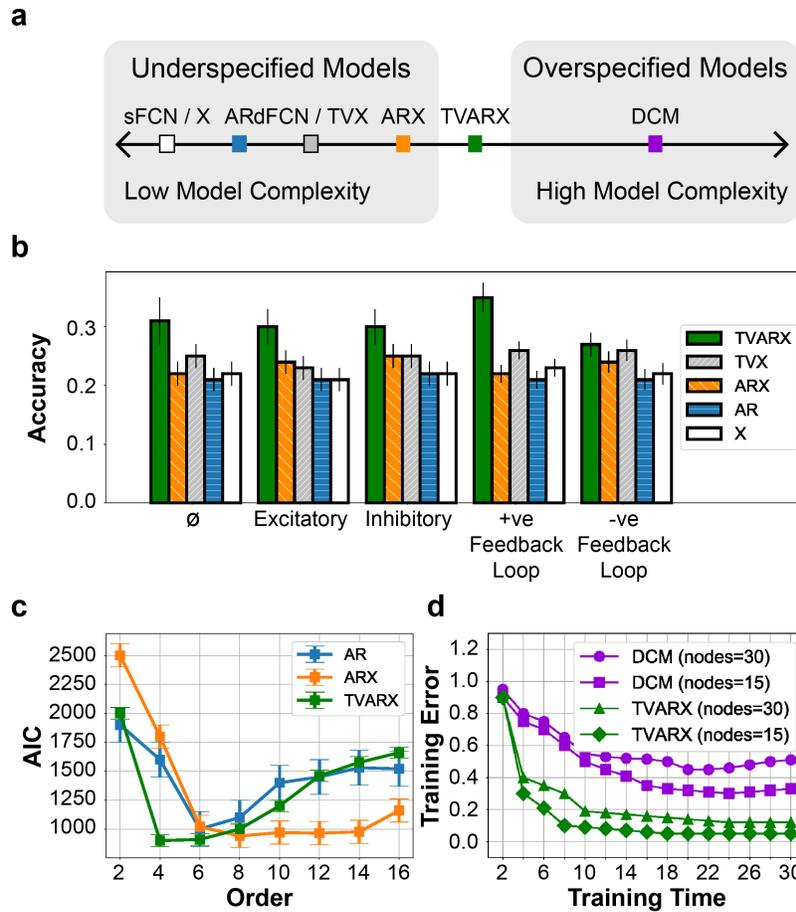


Fig. 2. Using Synthetic Data, We Compare the Performance (Accuracy and Speed) of System Identification Algorithms in Recovering Control Circuit Architecture and Dynamics (a) In this work, we fill the gap in literature between under-specified models that model networks, rather than circuits, and over-specified models that cannot scale computationally and therefore fail to extract circuits for large number of regions using shorter (5 minute) time series (b) Accuracy scores of classification of simulated connections using different state space models discussed in Table 2. TVARX performed best at recovering the original circuit topology (c) AIC scores across models with respect to order of autoregressive component, which accounts for increasing complexity. TVARX performs best even when penalized for having the larger number of parameters. (d) Comparison of TVARX with (stochastic) Dynamic Causal Modeling (DCM) on human task-based fMRI. DCM fails to converge for shorter (5 minute) fMRI time series as well as for larger number of nodes.

(Figure 4e) and Eulerian/elementary circuits are identified. The connections of these circuits are further classified into one of the four connection types.

Figure 2b provides the accuracy scores for each model with respect to classification of each connection type. The absence of DCM in Figure 2b reflects the fact that it failed to converge and thus did not yield meaningful results for our synthetic circuits. The Time-Varying autoRegressive with eXogenous inputs (TVARX) model outperforms other, simpler, models in identifying each connection type and thus recovering the overall circuit architecture.

To account for varying model complexities, in our evaluations for predicting trajectories we compute the Akaike Information Criterion (AIC) for different models as we increase the autoregressive order (Figure 2c). Only models that include past states were included in these comparisons, as majority of parameters are part of the autoregressive component and only these models are capable of generating future pre-

dicted non-linear trajectories. Here again, the TVARX model outperforms other models we evaluated in our experiments, even after accounting for the larger number of parameters.

Data-Driven Circuit Discovery Using Human fMRI

Participants from UK-Biobank (N=19,831) (13) were scanned while engaged in a task designed to elicit affective and non-affective sensorimotor processing. These were used to identify circuits selectively activated for each type of processing.

Task Design

During fMRI scans, participants were administered the Hariri faces/shapes "emotion" task (14, 15), as also implemented in the Human Connectome Project (HCP) (18) but with shorter overall duration and hence fewer total stimulus block repeats. Participants were presented with alternating blocks of trials

with visual stimuli consisting of human faces and geometric shapes (circles, ellipses), with brief periods of rest in between. For facial expressions, 12 different images were used: six of each gender and affect (angry or fearful), all from a standardized set of pictures of facial affect (19). In each trial, either three faces or three shapes were presented in a triangular configuration: one centered above the other two. The participants were asked to indicate, pressing a button, which stimulus on the bottom row matched the stimulus on the top row. The response triggered either the next trial or an eight second period of rest. The total length of the scan for each subject was 4 minutes; we obtained data for N=19,831 participants. The data were acquired on harmonized Siemens 3T Skyra scanners. The scans are 2.4mm isotropic with TR of 0.735s and 332 frames per run. Each subject had one run. The parcellation used in our experiments was provided by UK Biobank and included 139 regions of interest (ROIs). These ROIs are defined in MNI152 space, combining parcellations from several atlases: the Harvard-Oxford cortical and subcortical atlases (20, 21) and the Diedrichsen cerebellar atlas (22). Further information for the dataset is provided in Methods.

Comparison of TVARX with Dynamic Causal Modeling

In Figure 2 (d), we present a comparison between TVARX and DCM with respect to training time and corresponding training error. Even for small circuits (nodes ≤ 30), DCM fails to converge for time series of 4 minutes (332 time-points) as available in UK-Biobank. As shown, not only does TVARX converge considerably faster, but training error for DCM does not decrease monotonically, indicating failure to converge.

This is not surprising, since DCM is designed to be hypothesis-driven, testing competing models from a pre-specified set of nodes(6). Several competing hypotheses that constitute a model space are specified in the form of sub-graphs, which are then compared using Bayesian model selection. Increasing the number of nodes is challenging because the number of extrinsic (between-node) connections or edges increases with the square of the number of nodes. This can lead to models with an enormous number of free parameters and profound conditional dependencies among the parameters. Furthermore, the computational time required to invert these models grows exponentially with the number of free parameters. More recent variants have been developed to successfully address this issue, but were not compared to TVARX because of other limitations: spectral DCM is not appropriate for measuring homeostatic regulation in response to driving inputs, and regression/sparse DCM (7, 8) constrains the HRF in ways that can introduce confounds in clinical populations (9–12).

Prefrontal-Limbic and Ventral Stream Subcircuits

A striking feature of our results is that TVARX overwhelmingly implicates the same two circuits across individuals (Figure 3). These are highly consistent with those re-

ported in the basic and clinical neuroscience literature despite their inference from data alone. The *prefrontal-limbic subcircuit* (PFLC) composed of the thalamus, hippocampus, orbitofrontal cortex (OFC) and ventromedial prefrontal cortex (vmPFC), includes key regions associated with affective processing(23, 24). The non-affective condition was equally informative in identifying the *ventral stream subcircuit* (VS) composed of thalamus, inferior temporal gyrus (ITG) and inferior frontal gyrus (IFG). While the VS is often understood in terms of recognition of objects and form representation (25, 26), the VS also regulates perceptual uncertainty in the context of threat-assessment ("threat generalization"). Indeed, it is dysregulation of this subcircuit—in response to *non-affective*, rather than affective stimuli—that we have previously shown (in four independent datasets, totaling N=226) to track the spectrum of trait to clinical anxiety (27).

The two subcircuits were found to be mutually interacting with each other, with the pivot point centered at the thalamus, the only region shared by both circuits. Our system identification methods identified the thalamus to have two inputs, both inhibitory connections, one from hippocampus for the prefrontal-limbic subcircuit and one from inferior frontal gyrus for the ventral stream subcircuit; i.e.

$$\text{Hippocampus} \xrightarrow[\text{PFLC}]{-} \text{Thalamus} \xleftarrow[\text{VS}]{-} \text{IFG}^1$$

Similarly, the thalamus has two outputs, both excitatory connections, one to hippocampus in the PFLC and the other to ITG in VS i.e.

$$\text{Hippocampus} \xleftarrow[\text{PFLC}]{+} \text{Thalamus} \xrightarrow[\text{VS}]{+} \text{ITG}$$

In this way, the thalamus *completes* one of two competing negative feedback loops, one for each of the two identified subcircuits.

We measure the relative dominance of one subcircuit versus the other at any point in time as the absolute sum of relevant entries in the time-varying feedforward matrix D_t in our state space equation (Methods, Equation 3). Note that each entry in $D_t^{(a,b)}$ represents the causal dependence of trajectory of b on trajectory of a (i.e. effective connectivity $a \rightarrow b$).

Our results show these two competing feedback loops to be alternatively dominating in strength based on the stimuli during the scan for each subject (Figure 3b). The prefrontal-limbic subcircuit was found to be the dominant loop at points in time when the subjects were tasked with matching facial stimuli of angry or fearful affect. The opposite was observed for ventral stream subcircuit, which was the dominant feedback loop when subjects were tasked with matching geometric shapes.

¹Notation: Region A $\xrightarrow[\text{circuit}]{\text{connection type}}$ Region B. Excitatory connections are denoted by + and inhibitory connections by -. Direction of the arrow head represents indicates directionality of causation.

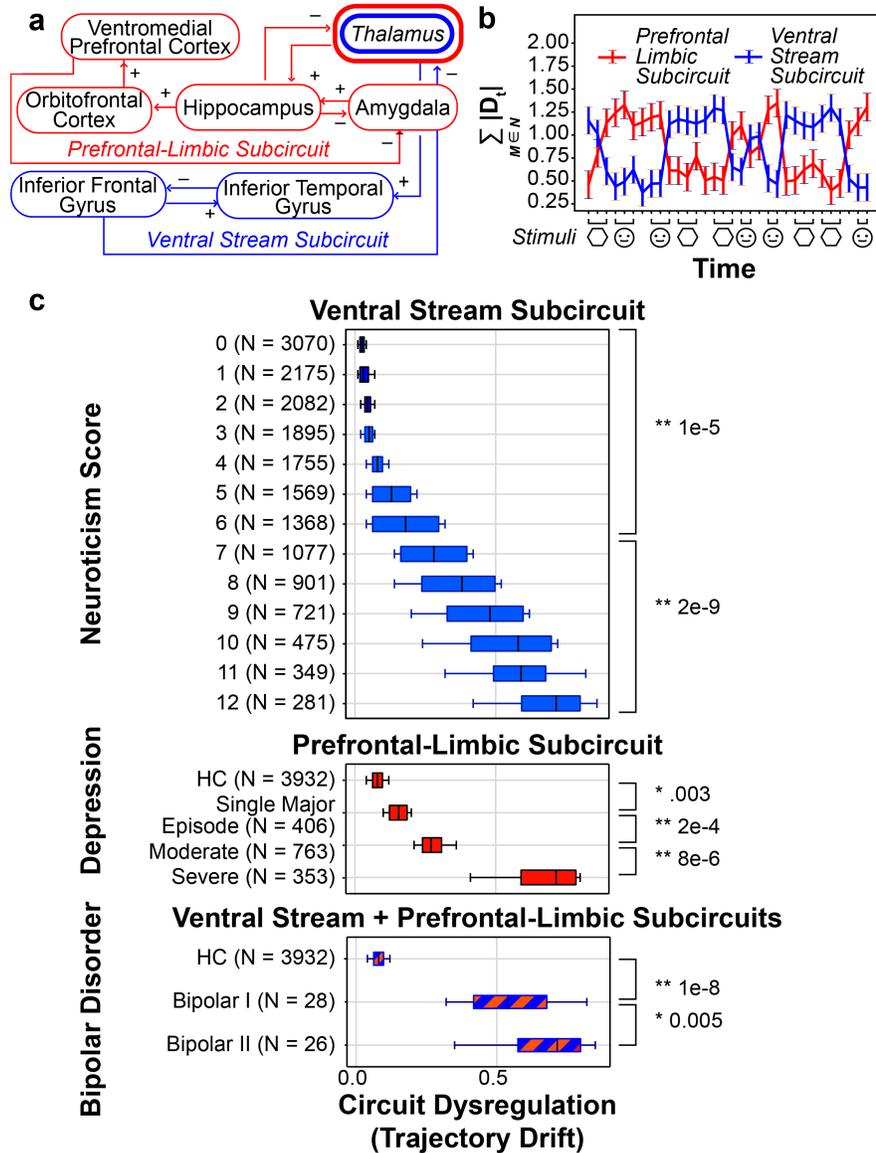


Fig. 3. Psychiatric Symptom Severity Tracks Degree of Circuit-Wide Dysregulation.(a) Two subcircuits, interacting and linked via the thalamus, were overwhelmingly implicated across subjects for each matching task; the *prefrontal-limbic* subcircuit selectively engaged during processing of affective (faces) stimuli while the *ventral stream* subcircuit selectively engaged during processing of non-affective (shapes) stimuli. (b) Results show selective dominance of each subcircuit with changing stimuli (the y-axis is the relative dominance of one circuit versus the other, defined as absolute sum of relevant entries of the feedforward matrix D in our state space equation (Eqn. 3). For prefrontal-limbic subcircuit $M = \{thalamus, hippocampus, OFC, vmPFC\}$ and for ventral stream subcircuit $M = \{thalamus, ITF, IFG\}$) (c) We use trajectory drift (measured as the mean squared error between actual and predicted trajectories) as a measure of feedback control error, or dysregulation, of each subcircuit. Dysregulation of the prefrontal-limbic subcircuit was measured as the error in the thalamus trajectory as predicted from negative feedback by the hippocampus (hippocampus \rightarrow thalamus). Dysregulation of the ventral stream was measured as error in the thalamus trajectory as predicted from negative feedback by the inferior frontal gyrus (inferior frontal gyrus \rightarrow thalamus). Trajectory drift of the ventral stream subcircuit tracks severity of neuroticism, while trajectory drift of the prefrontal-limbic circuit tracks severity of depression. For bipolar disorder, the thalamic trajectory could not be predicted from either hippocampal or IFG trajectories, and thus reflects dysregulation of the full circuit. This could be due to either more systemic problems with feedback across both circuits, or that the full circuit is receiving dysregulated inputs from another, different, circuit not identified by these tasks. Bonferroni corrected * $P < 0.05$; ** $P < 0.01$

271 Trajectory Drift as Dysregulation

272 We use trajectory drift as measure of circuit-wide control error, and therefore dysregulation. This drift is measured as
 273 the mean squared error between the actual trajectories and the predicted trajectories. We further compare these varying
 274 levels of dysregulation with the severity and type of psychiatric symptoms and diagnoses. These include scored degrees
 275
 276
 277

of neuroticism (a measure of stress vulnerability, anxiety),
 depression, and diagnosis of Type 1 and 2 bipolar disorders
 (definitions in Methods).

To estimate dysregulation of the PFLC and VS subcircuits identified for N=19,831 subjects, we use the same task fMRI scans. However, unlike our identification of circuits in the previous section, here we measure dysregulation across the

		Neuroticism Score		Depression			Bipolar Disorder	
		0 (N=3070)	6 (N=1368)	HC (N=3932)	Single Major Episode (N=406)	Moderate (N=763)	HC (N=3932)	Bipolar I (N=28)
		vs.	vs.	vs.	vs.	vs.	vs.	vs.
		6 (N=1368)	12 (N=281)	Single Major Episode (N=406)	Moderate (N=763)	Severe (N=353)	Bipolar I (N=28)	Bipolar II (N=26)
DCM	<i>Failed to Converge</i>							
Trajectory Drift	Hippocampus $\xrightarrow{-}$ Thalamus			*0.003	**2e-4	**8e-6		
	Inferior Frontal Gyrus $\xrightarrow{-}$ Thalamus	**1e-5	**2e-9				**1e-8	*0.005
Functional Connectivity	Hippocampus \leftrightarrow Thalamus			0.06	0.1	*0.001	**1e-4	0.3
	Inferior Frontal Gyrus \leftrightarrow Thalamus	*0.002	**1e-4				0.12	0.26
GLM	Thalamus	0.08	*0.01	0.13	*0.012	0.16	*0.01	0.2

Table 1. Comparison of Trajectory Drift with fMRI Analytical Methods: Dynamic Causal Modeling (DCM), Functional Connectivity, and Activation Based Generalized Linear Models (GLM). We took a whole-brain purely data-driven approach in identifying circuits for the two circuit-based methods: DCM and Trajectory Drift. Of these, only Trajectory Drift was able to converge for the parcellation (139 regions of interest) and sample size (UK Biobank N=19,831). For the two subcircuits identified by Trajectory Drift: Prefrontal-Limbic and Ventral Stream, we then tested how the key regulatory components for Prefrontal-Limbic (negative feedback by the hippocampus) and Ventral Stream (negative feedback by the inferior frontal gyrus) were interpreted by non-circuit-based methods: GLM and Functional connectivity. For each comparison of psychiatric variables, we report p-values from statistical significance testing using Welch's t-test for unequal variances and sample sizes. Bonferroni corrected * $P < 0.05$; ** $P < 0.01$

entire scan, independent of the design matrix.

Our results show marked association between greater dysregulation of specific subcircuits and the severity of the psychiatric symptoms Figure 3c.

Since the thalamus was identified as a pivot point for switching between the two subcircuits, in determining which of the two competing feedback loops dominates the system, we specifically focused on regulation of the thalamus; i.e. prediction of thalamus's trajectory as a function of negative feedback by either the hippocampus (for PFLC) or the inferior frontal gyrus (for VS).

Our results show more severe neuroticism to be associated with greater trajectory drift (control error) in the ventral stream subcircuit (level 0 [N=3070] vs. 6 [N=1368] ** $p \leq 1e-5$; level 6 [N=1368] vs. 12 [N=281] ** $p \leq 2e-9$) (Figure 3c top), specific to weakened negative feedback from the inferior frontal gyrus (IFG) to the thalamus (IFG $\xrightarrow{-}$ Thalamus). This inhibitory connection is critical to stable regulation of the ventral stream and was observed to in turn result in greater dysregulation downstream with respect to thalamic outputs to the inferior temporal gyrus (ITG) (Thalamus $\xrightarrow{+}$ ITG).

In contrast, our results show more severe depression to be associated with greater trajectory drift (control error) in the prefrontal-limbic subcircuit (HC [N=3932] vs. Single Major Episode [N=406] * $p \leq 0.003$; Single Major Episode [N=406] vs. Moderate [N=763] ** $p \leq 2e-4$; Moderate [N=763] vs. Severe [N=353] ** $p \leq 8e-6$) (Figure 3c center), specific to weakened negative feedback from the hippocampus to the thalamus (Hippocampus $\xrightarrow{-}$ Thalamus). Note that this relationship is itself dependent on the excitatory inputs from the thalamus to the hippocampus (Thalamus $\xrightarrow{+}$ Hippocampus) completing the negative feedback loop.

In the case of bipolar disorder, the thalamus was observed to be dysregulated with respect to both of its inhibitory inputs (Hippocampus $\xrightarrow{-}$ Thalamus $\xleftarrow{-}$ Inferior Temporal Gyrus) with greater dysregulation observed for subjects with Bipolar I Disorder compared to subjects with Bipolar II Disorder (Figure 3 bottom) (HC [N=3932] vs. Bipolar I [N=28] * $p \leq 1e-8$; Bipolar I [N=28] vs. Bipolar II [N=26] * $p \leq 0.0005$). In the case of Bipolar Disorder I & II, the thalamic trajectory was observed to drift significantly from its predicted trajectory, but the system was not dominated by either of the two competing feedback loops. This could be due either to more systemic problems with feedback across both circuits, or that the full circuit is receiving dysregulated inputs from another, different, circuit not identified by these tasks.

Finally, we compare trajectory drifts of discovered circuits with more conventional methods currently prevalent in clinical neuroscience (Table 1). These standard methods include Stochastic DCM, correlation-based functional connectivity (28), and activation based Generalized Linear Models (GLM) (29). Note, however, that different methods answer fundamentally different (if complementary) questions: DCM and Trajectory Drift capture circuit-wide dynamics, functional connectivity provides the strength of (undirected) signaling across pairs of regions, and GLM provides activation of individual regions. To allow for a more direct comparison of our circuit-wide measure to activation and networks, in Table 1, we report results for the subcircuit regions and connections that we identified as tracking symptom severity. Our results show that modeling psychiatric disorders in terms of circuit dysregulation achieves markedly greater detection sensitivity across all three sets of psychiatric symptoms. Beyond identification of differences, however, the most important advantage of our method is that it uses data-driven methods to construct generative computational neuroscience

353 models that explicitly consider homeostatic regulation across
354 negative feedback loops. This has the potential to allow hy-
355 potheses regarding dysregulation across psychiatrically rele-
356 vant circuits to be rigorously specified and empirically tested.

357 Discussion

358 In this work, we present a scalable fMRI data-driven tech-
359 nique that allows for construction of generative circuits in the
360 human brain, and provides a quantitative measure–trajectory
361 drift–of their control error, or circuit-wide dysregulation.
362 We demonstrate the effectiveness of our technique in re-
363 covering artificially generated circuits of varied architectures
364 and transfer functions. To demonstrate the applicability of
365 the technique to computational psychiatry, we use large-
366 scale fMRI data to identify two subcircuits and demonstrate
367 that their dysregulation tracks with symptom severity with
368 markedly greater detection sensitivity than standard analytic
369 methods.

370 fMRI has conventionally been used to either compute
371 brain activation maps, as areas of differential hemodynamic
372 response, or to quantify pairwise connectivity between brain
373 regions using Pearson correlation(30). More recent devel-
374 opments in fMRI analyses consider graph-theoretic mea-
375 sures (31) and a shift towards dynamic patterns of connectiv-
376 ity using time varying connections (32, 33). What all of these
377 methods lack, however, is a conceptual and mathematical
378 framework for considering the implications of closed feed-
379 back loops. Without these, activation maps and connectivity-
380 derived networks can suggest the presence of neural circuits,
381 but can neither define nor simulate their behavior, which in-
382 cludes their regulation. Given the assumption that psychiatric
383 disorders are grounded in the failure of circuits to maintain
384 homeostatic regulation, the ability to identify trajectories–
385 including drift from normative trajectories–is thus an impor-
386 tant step in the development of computational psychiatry and
387 its characterization of dynamical disease(34, 35).

388 ACKNOWLEDGEMENTS

389 The research described in this article was funded by the Baszucki Brain Re-
390 search Fund (L.R.M.-P.)

391 Bibliography

- 392 1. Lilliane R Mujica-Parodi and Helmut H Strey. Making Sense of Computational Psychi-
393 atry. *International Journal of Neuropsychopharmacology*, 23(5):339–347, March 2020.
394 ISSN 1461-1457. doi: 10.1093/ijnp/pyaa013. _eprint: [https://academic.oup.com/ijnp/article-](https://academic.oup.com/ijnp/article-pdf/23/5/339/33314618/pyaa013.pdf)
395 [pdf/23/5/339/33314618/pyaa013.pdf](https://academic.oup.com/ijnp/article-pdf/23/5/339/33314618/pyaa013.pdf).
- 396 2. E. J. Hinch. *Perturbation Methods*. Cambridge Texts in Applied Mathematics. Cambridge
397 University Press, 1991. doi: 10.1017/CBO9781139172189.
- 398 3. Michael CK Khoo. *Physiological control systems: analysis, simulation, and estimation*. John
399 Wiley & Sons, 2018.
- 400 4. Petar Kokotović, Hassan K. Khalil, and John O'Reilly. *Singular Perturbation Methods in*
401 *Control: Analysis and Design*. Society for Industrial and Applied Mathematics, 1999. doi:
402 [10.1137/1.9781611971118](https://doi.org/10.1137/1.9781611971118).
- 403 5. Karl J Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. *Neuroimage*, 19
404 (4):1273–1302, 2003.
- 405 6. Karl J Friston, Joshua Kahan, Bharat Biswal, and Adeel Razi. A dcm for resting state fmri.
406 *Neuroimage*, 94:396–407, 2014.
- 407 7. Stefan Frässle, Ekaterina I Lomakina, Lars Kasper, Zina M Manjaly, Alex Leff, Klaas P
408 Pruessmann, Joachim M Buhmann, and Klaas E Stephan. A generative model of whole-
409 brain effective connectivity. *Neuroimage*, 179:505–529, 2018.

8. Stefan Frässle, Ekaterina I Lomakina, Adeel Razi, Karl J Friston, Joachim M Buhmann, and
410 Klaas E Stephan. Regression dcm for fmri. *Neuroimage*, 155:406–421, 2017. 411
9. Daniel A Handwerker, John M Ollinger, and Mark D'Esposito. Variation of bold hemody-
412 namic responses across subjects and brain regions and their effects on statistical analyses.
413 *Neuroimage*, 21(4):1639–1651, 2004. 414
10. Geoffrey Karl Aguirre, Eric Zarahn, and Mark D'Esposito. The variability of human, bold
415 hemodynamic responses. *Neuroimage*, 8(4):360–369, 1998. 416
11. Olivier David, Isabelle Guillemain, Sandrine Saillet, Sebastien Rey, Colin Deransart,
417 Christoph Segebarth, and Antoine Depaulis. Identifying neural drivers with functional mri:
418 an electrophysiological validation. *PLoS biology*, 6(12):e315, 2008. 419
12. Pedro A Valdes-Sosa, Alard Roebroeck, Jean Daunizeau, and Karl Friston. Effective con-
420 nectivity: influence, causality and biophysical modeling. *Neuroimage*, 58(2):339–361, 2011. 421
13. Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T Elliott, Kevin Sharp,
422 Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, et al. The uk biobank
423 resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209, 2018. 424
14. Ahmad R Harii, Alessandro Tessitore, Venkata S Mattay, Francesco Fera, and Daniel R
425 Weinberger. The amygdala response to emotional stimuli: a comparison of faces and
426 scenes. *Neuroimage*, 17(1):317–323, 2002. 427
15. Deanna M Barch, Gregory C Burgess, Michael P Harms, Steven E Petersen, Bradley L
428 Schlaggar, Maurizio Corbetta, Matthew F Glasser, Sandra Curtiss, Sachin Dixit, Cindy Feldt,
429 et al. Function in the human connectome: task-fMRI and individual differences in behavior.
430 *Neuroimage*, 80:169–189, 2013. 431
16. Richard B Buxton, Kâmil Uludağ, David J Dubowitz, and Thomas T Liu. Modeling the
432 hemodynamic response to brain activation. *Neuroimage*, 23:S220–S233, 2004. 433
17. Seiji Ogawa, RS Menon, David W Tank, SG Kim, H Merkle, JM Ellermann, and K Ugurbil.
434 Functional brain mapping by blood oxygenation level-dependent contrast magnetic reso-
435 nance imaging. a comparison of signal characteristics with a biophysical model. *Biophysical*
436 *journal*, 64(3):803–812, 1993. 437
18. David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub,
438 Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project:
439 an overview. *Neuroimage*, 80:62–79, 2013. 440
19. Paul Ekman and Wallace V Friesen. Measuring facial movement. *Environmental psychology*
441 *and nonverbal behavior*, 1(1):56–75, 1976. 442
20. Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson,
443 Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman,
444 et al. An automated labeling system for subdividing the human cerebral cortex on mri scans
445 into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006. 446
21. Jean A Frazier, Sufen Chiu, Janis L Breeze, Nikos Makris, Nicholas Lange, David N
447 Kennedy, Martha R Herbert, Eileen K Bent, Vamsi K Koneru, Megan E Dieterich, et al.
448 Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric
449 bipolar disorder. *American Journal of Psychiatry*, 162(7):1256–1265, 2005. 450
22. Jörn Diedrichsen, Joshua H Balsters, Jonathan Flavell, Emma Cussans, and Narender
451 Ramnani. A probabilistic mr atlas of the human cerebellum. *Neuroimage*, 46(1):39–46,
452 2009. 453
23. Joseph E. LeDoux. Emotion circuits in the brain. *Annual Review of Neuroscience*, 23(1):
454 155–184, 2000. 455
24. MR Bennett. The prefrontal–limbic network in depression: Modulation by hypothalamus,
456 basal ganglia and midbrain. *Progress in neurobiology*, 93(4):468–487, 2011. 457
25. Laurent Cohen and Stanislas Dehaene. Specialization within the ventral stream: the case
458 for the visual word form area. *Neuroimage*, 22(1):466–476, 2004. 459
26. Bradford Z Mahon, Shawn C Milleville, Gioia AL Negri, Raffaella I Rumiati, Alfonso Cara-
460 mazza, and Alex Martin. Action-related properties shape object representations in the ven-
461 tral stream. *Neuron*, 55(3):507–520, 2007. 462
27. Lilliane R Mujica-Parodi, Jiok Cha, and Jonathan Gao. From anxious to reckless: a control
463 systems approach unifies prefrontal–limbic regulation across the spectrum of threat detec-
464 tion. *Frontiers in systems neuroscience*, 11:18, 2017. 465
28. Karl J Friston. Functional and effective connectivity in neuroimaging: a synthesis. *Human*
466 *brain mapping*, 2(1-2):56–78, 1994. 467
29. Karl J Friston, Andrew P Holmes, Keith J Worsley, J-P Poline, Chris D Frith, and Richard SJ
468 Frackowiak. Statistical parametric maps in functional imaging: a general linear approach.
469 *Human brain mapping*, 2(4):189–210, 1994. 470
30. Jean-Baptiste Poline and Matthew Brett. The general linear model and fmri: does love last
471 forever? *Neuroimage*, 62(2):871–880, 2012. 472
31. Danielle Smith Bassett and ED Bullmore. Small-world brain networks. *The neuroscientist*,
473 12(6):512–523, 2006. 474
32. Daniel J Lurie, Daniel Kessler, Danielle S Bassett, Richard F Betzel, Michael Breakspear,
475 Shella Kheifholz, Aaron Kucyi, Raphaël Liégeois, Martin A Lindquist, Anthony Randal McIn-
476 tosh, et al. Questions and controversies in the study of time-varying functional connectivity
477 in resting fmri. *Network Neuroscience*, 4(1):30–69, 2020. 478
33. Leonardo L Gollo and Michael Breakspear. The frustrated brain: from dynamics on motifs
479 to communities and networks. *Philosophical Transactions of the Royal Society B: Biological*
480 *Sciences*, 369(1653):20130532, 2014. 481
34. Quentin JM Huys, Michael Browning, Martin P Paulus, and Michael J Frank. Advances in
482 the computational understanding of mental illness. *Neuropsychopharmacology*, 46(1):3–19,
483 2021. 484
35. Daniel Durstewitz, Quentin JM Huys, and Georgia Koppe. Psychiatric illnesses as disorders
485 of network dynamics. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6
486 (9):865–876, 2021. 487
36. Lionel Barnett, Adam B Barrett, and Anil K Seth. Granger causality and transfer entropy are
488 equivalent for gaussian variables. *Physical review letters*, 103(23):238701, 2009. 489
37. Steen Moeller, Essa Yacoub, Cheryl A Olman, Edward Auerbach, John Strupp, Noam Harel,
490 and Kâmil Ugurbil. Multiband multislice ge-epi at 7 tesla, with 16-fold acceleration using
491 partial parallel imaging with application to high spatial and temporal whole-brain fmri. *Magnetic*
492 *resonance in medicine*, 63(5):1144–1153, 2010. 493
38. Christian F Beckmann and Stephen M Smith. Probabilistic independent component analysis
494 for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23(2):
495

496 137–152, 2004.
 497 39. Gholamreza Salimi-Khorshidi, Gwenaëlle Douaud, Christian F Beckmann, Matthew F
 498 Glasser, Ludovica Griffanti, and Stephen M Smith. Automatic denoising of functional mri
 499 data: combining independent component analysis and hierarchical fusion of classifiers.
 500 *Neuroimage*, 90:449–468, 2014.
 501 40. Mark W Woolrich, Brian D Ripley, Michael Brady, and Stephen M Smith. Temporal autocor-
 502 relation in univariate linear modeling of fmri data. *Neuroimage*, 14(6):1370–1386, 2001.
 503 41. Fidel Alfaro-Almagro, Mark Jenkinson, Neal K Bangerter, Jesper LR Andersson, Ludovica
 504 Griffanti, Gwenaëlle Douaud, Stamatios N Sotiropoulos, Saad Jbabdi, Moises Hernandez-
 505 Fernandez, Emmanuel Vallee, et al. Image processing and quality control for the first 10,000
 506 brain imaging datasets from uk biobank. *Neuroimage*, 166:400–424, 2018.
 507 42. Stephen B Manuck, Sarah M Brown, Erika E Forbes, and Ahmad R Hariri. Temporal stability
 508 of individual differences in amygdala reactivity. *American Journal of Psychiatry*, 164(10):
 509 1613–1614, 2007.
 510 43. Daniel J Smith, Barbara I Nicholl, Breda Cullen, Daniel Martin, Zia Ul-Haq, Jonathan Evans,
 511 Jason MR Gill, Beverly Roberts, John Gallacher, Daniel Mackay, et al. Prevalence and
 512 characteristics of probable major depression and bipolar disorder within uk biobank: cross-
 513 sectional study of 172,751 participants. *PLoS one*, 8(11):e75362, 2013.

514 Methods

515 Functional connectivity network for N regions-of-interest is
 516 traditionally defined as an $N \times N$ adjacency matrix A where

$$A_{x,y} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}, A \in \mathbb{R}^{(N \times N)}, x, y \leq N \quad (1)$$

517 When time series are normalized (zero mean and unit stan-
 518 dard deviation), a common fMRI preprocessing step, the
 519 Pearson correlation coefficient is equal to the slope of the
 520 regression. Thus resulting in a linear regression model of the
 521 form

$$y_t = A_{x,y} x_t + b_t \quad (2)$$

522 Most dynamic variants simply extend this definition by
 523 adding an additional temporal dimension resulting in a time-
 524 varying adjacency matrix $A \in \mathbb{R}^{N \times N \times T}$ where T is either
 525 the length of the time series or the number of sliding time-
 526 windows.

527 In this work, we extend this simple prevalent linear model
 528 by modeling BOLD time series observed for a brain region
 529 using a state space model of the form:

$$\begin{aligned} y_t &= Z_t \alpha_t + D_t u_t + d_t + \epsilon \\ \alpha_{t+1} &= T_t \alpha_t + B_t u_t + c_t + R_t \eta_t \end{aligned} \quad (3)$$

530 where y_t refers to the observation vector at time t , u_t refers
 531 to the input (or control) vector from other regions of the brain,
 532 α_t refers to the (unobserved) state vector at time t , and where
 533 the irregular components are defined as $\epsilon_t \sim N(0, H_t)$ and
 534 $\eta_t \sim N(0, Q_t)$.

535 The remaining variables in the equations are matrices de-
 536 scribing the process. The total length of the time series being
 537 T , the number of ROIs being N and K being the number of
 538 states, their variable names and dimensions are as follows:
 539 design $Z \in \mathbb{R}^{N \times K \times T}$, input $B \in \mathbb{R}^{N \times K \times T}$, observation in-
 540 tercept $d \in \mathbb{R}^{N \times T}$, observation covariance $H \in \mathbb{R}^{N \times N \times T}$,

transition $T \in \mathbb{R}^{K \times K \times T}$, state intercept $c \in \mathbb{R}^{K \times T}$, selec- 541
 tion $R \in \mathbb{R}^{K \times K \times T}$, state covariance $Q \in \mathbb{R}^{K \times K \times T}$ 542

Note that this formulation is a generalized framework with 543
 prevailing definitions of static functional connectivity as cor- 544
 relations ($D_t = D_{t+1} \forall t, Z_t = O \forall t$) and dynamic functional 545
 connectivity as time-varying correlations ($Z_t = O \forall t$) as spe- 546
 cial cases. Table 2 breaks down existing models and presents 547
 a comparison with our extended model (TVARX) in terms of 548
 parameters in our state space equations. 549

The time-varying feedthrough matrix D is used as dy- 550
 namic effective connectivity between nodes. Note that ef- 551
 fective connectivity defined this way is akin to a general 552
 form of Granger causality or transfer entropy (36). This 553
 dynamic effective connectivity graph is a temporal graph 554
 as shown in 4e and can be formally defined as a set of 555
 time-stamped edges, each with its own connectivity strength 556
 $\{(a, b, t_1, D_{a,b}), (c, d, t_2, D_{c,d}), \dots, (x, y, T, D_{x,y})\}$. 557

A circuit in our framework is defined as a set of connec- 558
 tions such that there exists an elementary/Eulerian circuit 559
 (simple cycles) of length > 2 . Each connection is defined 560
 for a pair of regions and of the following four types: exci- 561
 tatory, inhibitory, negative feedback loop and positive feed- 562
 back loop defined over elements of the feedthrough matrix D . 563
 Excitatory and inhibitory connections between x and y are 564
 defined simply as effective connections where $D_{x,y} > 0$ and 565
 $D_{x,y} < 0$. Feedback loops are defined as Eulerian/elementary 566
 circuits of length = 2. Positive feedback loops are ones where 567
 both connections are excitatory. Inversely, negative feedback 568
 loops are ones with at least one inhibitory connection. 569

Parallel connections/inputs in the circuit are implicit as 570
 BOLD signal for a region y at time t is fitted against multiple 571
 inputs. All excitatory and inhibitory connections are series 572
 by default. 573

Linear time invariant systems represented in state space 574
 form can be converted into input/output transfer functions by 575
 applying Laplace transform

$$G(s) = \frac{\text{num}(s)}{\text{den}(s)} = \frac{a_0 s^m + a_1 s^{m-1} + \dots + a_m}{b_0 s^n + b_1 s^{n-1} + \dots + b_n} \quad (4)$$

where n is generally greater than or equal to m (for a proper 574
 transfer function). 575

State space systems can be manipulated using standard 576
 arithmetic operations as well as the feedback, parallel, and 577
 series. Vice versa, each of the connection types: feedback, 578
 parallel, and series represent arithmetic operations over state 579
 space systems and/or transfer functions as given in Figure 4 580
 Panel c. 581

The parameters of the TVARX model are learned by max- 582
 imizing loglikelihood via Kalman filter. The method for cal- 583
 culating the covariance matrix of parameter estimates uses 584
 outer product of gradient estimator using Broyden-Fletcher- 585
 Goldfarb-Shanno (BFGS) solver. The method by which the 586
 Hessian is numerically approximated is outer product of gra- 587
 dients. 588

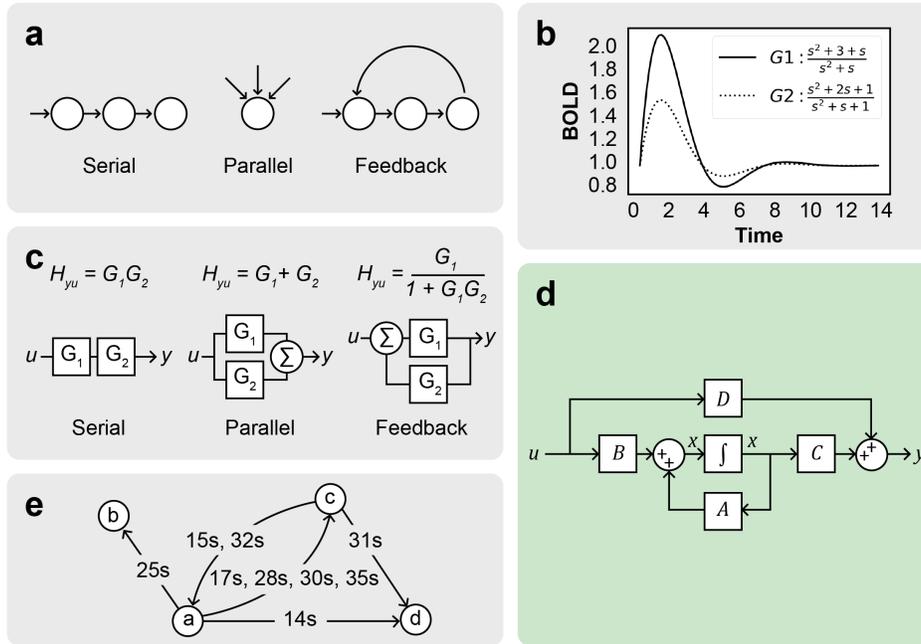


Fig. 4. Inferring Closed-Loop Circuits(a) Standard connectivity ("network") analyses depend upon linear regressions, which are only capable of modeling a very specific topology: parallel inputs. In contrast to parallel inputs A , most neurobiological circuits of relevance to psychiatry also require serial B and feedback C components, structures that could lead to an explosion of error propagation using standard statistical methods. (b) Impulse response for two hemodynamic response functions (HRFs) with different relaxation times and transfer functions. (c) To illustrate how transfer function structure changes with different circuit topologies, we show three transfer functions, each of which corresponds to a different kind of "motif," with series, parallel and feedback connections. By using pairs of inputs u and outputs y to obtain their transfer function, we systematically infer circuit topology. (d) Block diagram representation of state-space equations (e) Dynamic effective connectivity as a time-stamped temporal graph.

Model	D (feedforward)	T (transition)	B (input)	Z (design)
Correlations (sFCN / X)	Time invariant	O	O	O
Autoregressive (AR)	O	I	I	Time invariant
Time-varying Correlations (dFCN / TVX)	Time varying	O	O	O
Autoregressive w/ eXogenous inputs (ARX)	Time invariant	I	I	Time invariant
Time-varying Autoregressive w/ eXogenous inputs (TVARX)	Time varying	Time invariant	I	Time invariant

Table 2. Bridging the gap between network theory and control theory: extending existing correlation based models to our TVARX model. The breakdown and comparison in terms of state space parameters elucidates how our model is a generalized version of existing definitions of static (sFCN) and dynamic (dFCN) functional connectivity and extends networks to circuits. O : zero matrix, I : identity matrix

589 The model is fit on seventy-five percent of the time series
 590 for each subject and dysregulation is measured as the error
 591 in prediction of the remaining twenty-five percent of the time
 592 series from the actual signal. The measure of error used is
 593 mean squared error.

594 The statistical significance testing between error distribu-
 595 tions between different cohort of individuals is carried out
 596 using Welsch's t-test to account for skewed distributions be-
 597 tween healthy and diseased populations.

Image Acquisition

598
 599 Task fMRI data (tfMRI) were acquired on harmonized
 600 Siemens 3T Skyra scanners at four UK Biobank imaging centres
 601 (Cheadle, Manchester, Newcastle, and Reading). The
 602 scans were 2.4mm isotropic with TR of 0.735s and 332
 603 frames per run (4 mins). Each subject had one run. The
 604 resolution of the images is 2.4x2.4x2.4 mm with a field-of-
 605 view of 88x88x64 matrix. The duration was four minutes
 606 (332 timepoints) with TR of 0.735 s and TE of 39ms, GE-
 607 EPI with x8 multislice acceleration, no iPAT, flip angle 52
 608 degrees, and fat saturation.

609 A separate "single-band reference scan" was also acquired,

610 as implemented in the Center for Magnetic Resonance Re- 664
611 search (CMRR) multiband acquisition (37). This has the 665
612 same geometry (including echo-planar imaging distortion) as
613 the timeseries data, but has higher between-tissue contrast to
614 noise, and is used as the reference scan in head motion cor-
615 rection and alignment to other modalities.

616 Spatial smoothing, using a Gaussian kernel of FWHM 5
617 mm, was applied before the intensity normalisation, and nei- 668
618 ther Independent Component Analysis (ICA) (38) nor FM- 669
619 RIB's ICA-based X-noiseifier (FIX) (39) artefact removal 670
620 was performed, both decisions being largely driven by the 671
621 shorter timeseries in the tfMRI and because of the greater 672
622 general reliance in tfMRI analysis on voxelwise timeseries 673
623 modeling. All time series signal are standardized to z-scores
624 (shifted to zero mean and scaled to unit variance) and the
625 global signal is regressed out.

626 Pre-processing and task-induced activation modeling was
627 carried out using FEAT (fMRI Expert Analysis Tool); time- 668
628 series statistical analysis was carried out using FMRIB's 669
629 Improved Linear Model (FILM) with local autocorrelation 670
630 correction (40). The timings of the blocks of the two task 671
631 conditions (shapes and faces) are defined in 2 text files. Display of 672
632 the task video and logging of participant responses is carried 673
633 out by ePrime software. The timings of the task blocks are
634 fixed and already known as well as the correctness of subject
635 responses. For more details on data collection, processing of
636 collected images and quality control, please see (41).

637 Task

638 This task was adapted from the one developed by Hariri and 668
639 colleagues which had shown evidence as a functional local- 669
640 izer (14) with moderate reliability across time (42). Partici- 670
641 pants are presented with blocks of trials that either ask them 671
642 to decide which of two faces presented on the bottom of the 672
643 screen match the face at the top of the screen, or which of 673
644 two shapes presented at the bottom of the screen match the 674
645 shape at the top of the screen. The faces have either angry or 675
646 fearful expressions. Trials are presented in blocks of 6 trials 676
647 of the same task (face or shape), with the stimulus presented 677
648 for 2 s and a 1 s inter trial interval. Each block is preceded 678
649 by a 3 s task cue ("shape" or "face"), so that each block is 21 s 679
650 including the cue. Each of the two runs include 3 face blocks 680
651 and 3 shape blocks.

652 For facial expressions, 12 different images were used, 6
653 of each gender and affect (angry or afraid), all derived from
654 a standard set of pictures of facial affect (19). Simple geo-
655 metric shapes (circles, vertical, and horizontal ellipses) were
656 used as control stimuli.

657 Subjects were asked to match one of two simultaneously
658 presented images with an identical target image. As a sen-
659 sorimotor control task, the subjects were asked to match geo-
660 metric shapes. For each face block, three images of each
661 gender and target affect (angry or fearful) were presented.
662 For each control block, six different geometric shapes were
663 presented as targets. During imaging, subjects responded by

664 pressing one of two buttons with their dominant hand, allow-
665 ing for the determination of accuracy and reaction time.

666 Clinical Variables

667 Neuroticism

668 Participants were assessed for twelve domains of neurotic
669 behaviours via the touchscreen questionnaire. Neuroticism
670 summarises the number of Yes answers across these twelve
671 questions into a single integer score for each participant. Par-
672 ticipants could answer Yes, No, Do not know or Prefer not to
673 answer. Questions included:

- 674 1. Does your mood often go up and down? 674
- 675 2. Do you ever feel 'just miserable' for no reason? 675
- 676 3. Are you an irritable person? 676
- 677 4. Are your feelings easily hurt? 677
- 678 5. Do you often feel 'fed-up'? 678
- 679 6. Would you call yourself a nervous person? 679
- 680 7. Are you a worrier? 680
- 681 8. Would you call yourself tense or 'highly strung'? 681
- 682 9. Do you worry too long after an embarrassing experi- 682
683 ence? 683
- 684 10. Do you suffer from 'nerves'? 684
- 685 11. Do you often feel lonely? 685
- 686 12. Are you often troubled by feelings of guilt? 686

687 This derived data field has come from Professor Jill Pell
688 from the Institute of Health and Wellbeing, University of
689 Glasgow (43).

690 Depression

691 Depression status of participants is defined from the touch-
692 screen questionnaire at baseline. Each of the three depression
693 states were defined based on a number of criteria:

- 694 1. Ever felt depressed for a whole week 694
- 695 2. Ever disinterested or unenthusiastic for a whole week 695
- 696 3. Only 1 episode 696
- 697 4. ≥ 2 episodes 697
- 698 5. Episode lasted ≥ 2 weeks 698
- 699 6. Ever seen a GP for nerves, anxiety, tension or depres- 699
700 sion 700
- 701 7. Ever seen a psychiatrist for nerves, anxiety, tension or 701
702 depression 702

703 Definitions Single Probable Major Depressive Episode: (1)
704 AND (3) AND (5) AND [(6) OR (7)] OR (1) AND (3) AND
705 (5) AND [(6) OR (7)]
706 Probable Recurrent Major Depression (Moderate): [(1) OR
707 (2)] AND (4) AND (5) AND (6)
708 Probable Recurrent Major Depression (Severe): [(1) OR (2)]
709 AND (4) AND (5) AND (7)

710 **Bipolar Disorder**

711 UKB data-fields from the touchscreen (which were based on
712 the Structured Clinical Interview for DSM IV Axis I Disor-
713 ders1) were classified into criteria groups to define a probable
714 case of Bipolar I or II.

715 *Bipolar I (probable mania)* was classified as (1) ever manic
716 or hyper for ≥ 2 days OR ever irritable or argumentative for
717 ≥ 2 days AND (2) manic episodes characterised by at least 3
718 of ‘more talkative’, ‘more active’, ‘needed less sleep’, ‘more
719 creative/more ideas’ AND (3) longest manic episode \geq one
720 week duration AND (4) episode needed treatment or caused
721 problems at work.

722 *Bipolar II (probable hypomania)* classified as fulfilling crite-
723 ria (1), (2) and (3) of the Bipolar I definition, NOT criteria
724 (4).