Optimal Control Costs of Brain State Transitions in Linear Stochastic Systems

Shunsuke Kamiya[1], Genji Kawakita[1], Shuntaro Sasai[2], Jun Kitazono[1], and Masafumi Oizumi*[1]

[1]Graduate School of Arts and Sciences, The University of Tokyo
[2]Araya Inc.

Abstract

The brain is a system that performs numerous functions by controlling its states. Quantifying the cost of this control is essential as it reveals how the brain can be controlled based on the minimization of the control cost, and which brain regions are most important to the optimal control of transitions. Despite its great potential, the current control paradigm in neuroscience uses a deterministic framework and is therefore unable to consider stochasticity, severely limiting its application to neural data. Here, to resolve this limitation, we propose a novel framework for the evaluation of control costs based on a linear stochastic model. Following our previous work, we quantified the optimal control cost as the minimal Kullback-Leibler divergence between the uncontrolled and controlled processes. In the linear model, we established an analytical expression for minimal cost and showed that we can decompose it into the cost for controlling the mean and covariance of brain activity. To evaluate the utility of our novel framework, we examined the significant brain regions in the optimal control of transitions from the resting state to seven cognitive task states in whole-brain imaging data. We found that, in realizing the different transitions, the lower visual areas played a commonly significant role in controlling the means, while the posterior cingulate cortex played a commonly significant role in controlling the covariances.

1 Introduction

The brain is a highly complex dynamical network that flexibly transitions to various states to execute a myriad of functions [1, 2, 3]. In this regard, the brain can be considered a system that modulates its internal states to desired states, in accordance with the function the individual needs to perform [4, 5]. Among the many transitions that bring the system into the various states it requires, some state transitions are more difficult to control than others, depending on the dynamical properties of the neuronal systems. In other words, controlling transitions to some states incurs greater "costs" than controlling transitions to others. Providing a theoretical framework for quantifying transition costs, or control costs, is important for evaluating the difficulty of the shifts between various brain states, and possibly in explaining cognitive loads [6, 7], sleep-awake differences [8], habituation of cognitive tasks [9], effects of hallucinogenic drugs [10], and psychiatric [7] and neurological disorders [11] with a quantifiable measure. Therefore, the development of such a framework for quantifying control cost in the brain is a vital topic in neuroscience.

A rigorous and promising framework to assess control cost was provided by an approach using control theory, which was first introduced in neuroscience in a pioneering work by Gu et al. [5]. Originally rooted in engineering [12, 13], control theory provides theoretical tools for investigating the dynamical properties of complex systems [14, 15]. Its application to neuroscience is now opening new doors to mechanistically explaining neural behaviors from brain

structures [7, 9, 16, 17, 18, 19]. Applying the control theory to neuroscientific use requires that the dynamics of a neural system be modeled. Previous studies have used a linear deterministic model as an approximation of real neural dynamics. This model offers two main advantages. First, the optimal control problem, which looks for the controlled path that minimizes the cost —i.e. the most efficient way to control transitions —is analytically tractable. The optimal cost is quantified as the minimal value of the total input necessary to realize a transition [20], which can be easily solved under the classical linear systems theory [13, 14]. Second, the framework enables the contribution of each brain area in controlling state transitions to be quantitatively evaluated. Since the model explicitly incorporates a control input term in its formulation, the control framework provides insights into which brain areas are important for controlling state transitions [5, 9, 16].

Nevertheless, despite these strengths, the deterministic framework does not take account of an important property of neural activity: it neglects noise or stochasticity in neural systems. Since neural noises are known to be ubiquitous in the brain and to play critical roles in information processing [21, 22], overlooking the stochasticity of neural systems may result in an inaccurate estimation of the control costs.

In our recent work [6], we introduced a novel stochastic framework to deal with the issue of stochasticity in neural systems. We defined control costs as the Kullback-Leibler (KL) divergence between an uncontrolled and a controlled processes, which provides a more general framework for considering control costs in state transitions. We call the cost defined as the KL divergence the *stochastic control cost* hereafter. To quantify the stochastic control cost, we need a probabilistic model which approximates the brain dynamics. In the previous work [6], we employed a discrete-state stochastic process for modeling the dynamics of functional magnetic resonance imaging (fMRI) data. Although this discrete model has been widely used in other studies [16, 23], it has two major limitations. First, the cost defined as the KL divergence is not analytically tractable, which severely limits the size of systems that can be analyzed. Second, because the control inputs are not explicitly modeled, we cannot examine which regions significantly contribute to controlling state transitions.

In this study, to overcome these problems and limitations in the previous studies, we propose a novel framework to quantify control costs in linear stochastic neural systems. This framework takes advantages of both the linear control theoretic framework [5, 9, 17] and the stochastic control cost proposed in our previous work [6]. We modeled brain dynamics using linear stochastic differential equations. Thanks to this linearity, we can obtain an analytical expression of the stochastic control cost, defined as the KL divergence between the uncontrolled and controlled processes (Fig. 1). This analytical expression significantly reduces the computational burden and enables us to compute the control cost in a high dimensional system. Furthermore, as we include a control input term in our model, we can identify those brain regions playing a significant role in the control of state transitions.

In addition to determining the analytical expression of the stochastic control cost, we also showed that we can decompose this expression into the cost for controlling the mean (referred to as *mean control cost*) and that for controlling the covariance (*covariance control cost*). We proved that the mean control cost is identical to the control cost in the deterministic setting, which has been quantified in previous neuroscientific applications [5, 9, 17]. The covariance control cost, on the other hand, is the cost of controlling the co-variation among the system, which has not been quantified in previous studies in neuroscience. Thus, decomposition into the mean and covariance control costs clarifies the mathematical relationship between the conventional deterministic control costs and the stochastic control cost.

The decomposition of the stochastic control cost is significant from a neuroscientific viewpoint, as it provides a unified view for examining the control of brain state transition. In the neuroscientific context, mean corresponds to the magnitude of brain activity, while covariance corresponds to co-variation in brain activity in different brain areas and is often referred to as

the functional connectivity [24, 25]. Although change in these two characteristics —magnitude of brain activity and functional connectivity —have conventionally been considered separately, the control theoretical framework enables us to quantitatively compare their significance in a unified manner in terms of control cost. For example, for some transitions, we can tell whether the change in the functional connectivity is negligible —or, equally significant —compared to changes in the magnitude of brain activity from the viewpoint of the control cost.

After formulating the theoretical framework, we then posed the following two questions. First, how important is it to take account of covariance (functional connectivity) in estimating control cost? And second, what brain areas are significant in controlling brain state transitions? To address these questions, we applied our new method to real neural data. For this we used whole-brain fMRI data of 352 healthy adults, recorded as part of the Human Connectome Project (HCP) [26]. These data consist of blood-oxygen-level-dependent (BOLD) signals during various cognitive tasks, including resting states.

To address the first question, we examined the influence of the covariance control cost, since it accounts for the impact of considering covariance. We found that the influence of the covariance control cost did not appear negligible, indicating that it is important to incorporate covariance in estimating control costs in state transitions. To address the second question, we computed the amount of control inputs at each brain region. We discovered that the lower visual areas and the posterior cingulate cortex (PCC) play important roles in controlling state transitions, but in different ways: the PCC acts in controlling the covariance and the lower visual areas act in controlling the mean in addition to the covariance.

2 Stochastic Model and Stochastic Control Cost

2.1 Control Cost in Deterministic Systems

Before formulating the control costs in stochastic systems, we start with a conventional deterministic framework and a method for quantifying control cost under this deterministic setting. The dynamics of the brain is modeled with an n-dimensional state space model, where a brain state $x \in \mathbb{R}^n$ at time $t \in [0, T]$ $(T > 0)$ consists of n scalar values that represent the magnitudes of brain activity. Then for each $t \in [0, T]$, $x(t)$ is assumed as a point in an n-dimensional Euclidean space. A state transition of the brain is described as a trajectory on which a point travels from one state to another. An uncontrolled transition is described using linear dynamics, such as

$$\dot{x}(t) = Ax(t). \tag{1}$$

Here, $x : [0, T] \to \mathbb{R}^n$ is a vector that represents the magnitudes of activity of all nodes. $A \in \mathbb{R}^{n \times n}$ is a connectivity matrix whose elements represent connectivity weights for each pair of nodes.

We consider situations where a brain state switches from an initial state $x(0) = x_0 \in \mathbb{R}^n$ to a target state $x(T) = x_T \in \mathbb{R}^n$ that is different from the state when following its uncontrolled dynamics (equation (3)). To realize such a transition, we assume that a control input $u(t)$ is given. We can incorporate the control input into the dynamics as follows:

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{2}$$

where $u : [0, T] \to \mathbb{R}^p$ is a control input and $B \in \mathbb{R}^{n \times m}$ is an input matrix that determines the nodes assigned with control inputs $(n, m \in \mathbb{N})$. We assume that this system is controllable (i.e., there exists $u(t)$ that enables the transition from $\forall x_0$ to $\forall x_T$). Here, we limit ourselves to cases where $B = I_n$ (the $n \times n$ identity matrix), which implies that the input is given independently to all nodes. In other words, we consider the system with an input $v : [0, T] \to \mathbb{R}^n$,

$$\dot{x}(t) = Ax(t) + v(t). \tag{3}$$

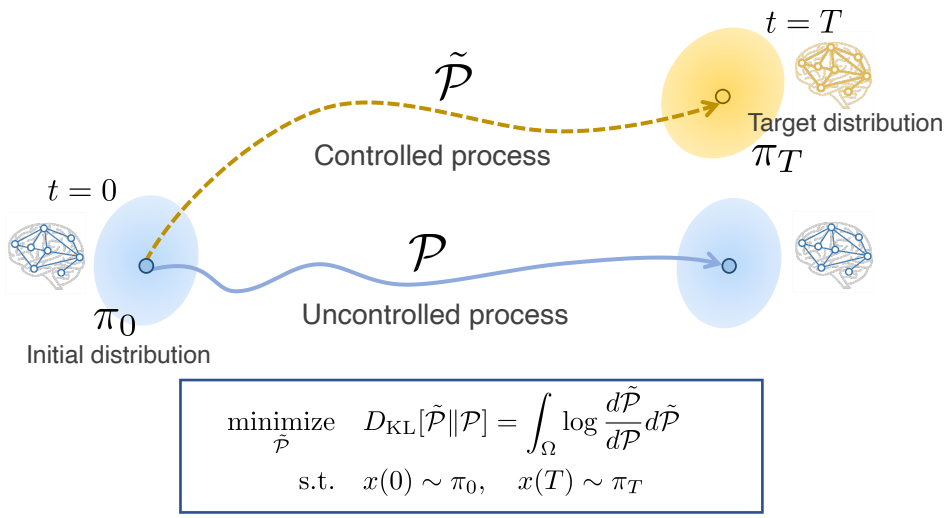


Figure 1: Schematic of our framework for quantification of control costs in the brain in linear stochastic systems. We model a brain state to follow a certain probability distribution π_0 at time $t = 0$ (blue ellipse on the left). In uncontrolled dynamics, the brain state stays in the same distribution (blue trajectory & blue ellipse on the right). However, in a state transition, the brain dynamics changes so that it reaches a target distribution π_T at time $t = T$ (the gold ellipse). We call this altered trajectory the controlled process (gold trajectory). To evaluate how close the controlled process is to the uncontrolled one, we employ the KL divergence between the two processes as the cost function, marginalized with the initial and the target distributions (blue square). The distributions of the processes are defined on a path space, the space composed of $\mathbb{R}^n$-valued continuous functions defined on $[0, T]$. This type of KL optimization problem on a path space is referred to as the Schrödinger's bridge problem.

The control cost (also called control energy) $\mathcal{J}_{\mathrm{cont}}$ is defined as the total amount of control input required to steer the system from x_0 to the target x_T and is expressed as

$$\mathcal{J}_{\mathrm{cont}} = \int_0^T \|v\|_2^2 \, dt, \tag{4}$$

under

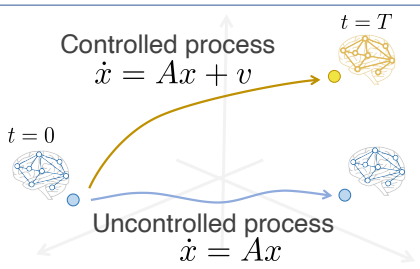
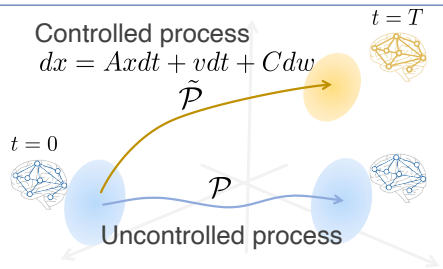
$$x(0) = x_0, \quad x(T) = x_T. \tag{5}$$

The minimum of this integral represents the input minimally needed to realize the state transition that satisfies the marginal conditions (equation (5)). It is written as

$$\mathcal{J}^*_{\mathrm{cont}} = \min_v \int_0^T \|v\|_2^2 \, dt, \tag{6}$$

and referred to as the minimal control cost. The problem of minimizing the control cost is called the *optimal control problem* in control theory. Hereafter, we call the optimal control cost simply the "control cost" because we are only concerned with the minimal value in this study. We also refer to this cost as the deterministic cost, since this cost is defined on the deterministic model (equation (3)). This minimum is known to exist, and the optimal value $\mathcal{J}^*_{\mathrm{cont}}$ is readily solved (see Supplementary Note 1.). This metric has been used as the control cost in previous studies in neuroscience [7, 16, 17, 18, 19, 27].

Table 1: Comparison of deterministic and stochastic state transitions in a brain from the perspective of control theory. The **Previous framework** column describes the deterministic model of state transitions. The uncontrolled process (blue arrow) is expressed as a linear differential equation $\dot{x}(t) = Ax(t)$ (A is a connectivity matrix of the brain network). A brain state transition is modeled as a transition from a point to another in an n-dimensional state space. Along with a control input v, the brain shifts its state to x_T, another point in the state space, at time T (dynamics drawn in yellow). The cost function is often set as the time integral of the squared input. The **Proposed framework** column explains the stochastic model of brain state transitions. Here, the uncontrolled process (blue arrow) and a controlled process (gold arrow) are given as stochastic processes. A brain state transition is viewed as a shift from an n-dimensional probability distribution to another. As a control cost, the KL divergence between the two processes is examined.

	Previous framework	Proposed framework
		
Dynamics	Deterministic	**Stochastic process**
Transition	Point → point	**Prob. dist. → Prob. dist.**
Control cost	Squared integral $\min_u \int_0^T \lVert u \rVert_2^2 \, dt.$	**KL divergence** $\int \log \frac{d\tilde{\mathcal{P}}}{d\mathcal{P}} d\tilde{\mathcal{P}}$

2.2 Formulation of a Stochastic Linear Model

To take account of noise and fluctuation in brain activity [28], we make an extension for three characteristics in the control model of the brain: dynamics, transitions, and the control cost (Tab.1). In this section, we explain these characteristics in detail.

First, we model the brain dynamics through a stochastic process instead of conventional deterministic processes as described in equation (3) and equation (2). We then model the dynamics without control, or the *uncontrolled process* described as an $\mathbb{R}^n$-valued Ornstein Uhlenbeck (OU) process $x(t)$ ($t \in [0, T]$),

$$dx(t) = Ax(t)dt + Cdw(t), \tag{7}$$

where each element of $x(t)$ corresponds to the magnitudes of activities in a brain region. Here, $w(t)$ is a standard p-dimensional normal Brownian motion, and $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{n \times p}$ ($p \leq n$, $n, p \in \mathbb{N}$). We assume $C\,{}^{\mathrm{t}}C$ to be nonsingular (${}^{\mathrm{t}}C$ denotes the matrix transpose of C), which is equivalent to $\mathrm{rank}(C) = n$ (so p must be equal to n).

Second, to model the control of transitions, we consider these as from a probability distribution to another probability distribution, instead of as the point-to-point transitions that have been considered in previous studies [5, 9]. In the present study, we consider the control of the

system between Gaussian distributions from time $t = 0$ to time $t = T$. We assume that the distribution of x a t $t = 0$ follows a Gaussian distribution with mean $\mu_0 \in \mathbb{R}^n$ and covariance $\Sigma_0 \in \mathrm{PSD}(n)$ (set of n-by-n positive semi-definite matrices), denoted by $x(0) \sim \mathcal{N}(\mu_0, \Sigma_0)$ (we call this the *initial distribution*). When the system follows equation (7), the distribution of $x(T)$ is uniquely determined:

$$x(T) \sim \mathcal{N}\left(e^{AT}x(0),\ e^{AT}\Sigma_0 e^{{}^{\mathrm{t}}AT} + \int_0^T e^{A(T-\tau)}C\,{}^{\mathrm{t}}Ce^{{}^{\mathrm{t}}A(T-\tau)}d\tau\right) =: \mathcal{N}(m_T, S_T). \tag{8}$$

Specifically, if

$$\mu_0 = 0, \quad \Sigma_0 = \int_0^\infty e^{A\tau}C\,{}^{\mathrm{t}}Ce^{{}^{\mathrm{t}}A\tau}d\tau, \tag{9}$$

the process (equation (7)) stays in the same probability distribution in $\forall t \geq 0$ (steady state distribution), and thus $\mu_0 = m_T = 0$ and $\Sigma_0 = S_T$. The existence of the steady state distribution is guaranteed if the real parts of all the eigenvalues of A are negative. To describe the control of state transitions, we consider altering the dynamics and the probability distribution at time $t = T$ by giving a certain input. We consider steering the system from the initial distribution $\mathcal{N}(\mu_0, \Sigma_0)$ to a given distribution $\mathcal{N}(\mu_T, \Sigma_T)$ ($\mu_T \in \mathbb{R}^n, \Sigma_T \in \mathbb{R}^{n\times n}$) at $t = T$, which is different from the final distribution $\mathcal{N}(m_T, S_T)$ in the uncontrolled process. We call $\mathcal{N}(\mu_T, \Sigma_T)$ a *target distribution*, and a process that reaches $\mathcal{N}(\mu_T, \Sigma_T)$ at $t = T$ a *controlled process*.

Third, for the control cost, we adopt the Kullback-Leibler (KL) divergence between the uncontrolled and controlled processes [6, 29, 30, 31]. A KL divergence is a metric that measures the closeness between two probability distributions. Thus, the control cost, defined by the KL divergence, measures the closeness between the uncontrolled and controlled processes. Note that this KL divergence is not the one between the two probability distributions of brain states at a particular time point, such as the KL divergence between the initial distribution $\mathcal{N}(\mu_0, \Sigma_0)$ and the target distribution $\mathcal{N}(\mu_T, \Sigma_T)$, but rather the divergence between the probability distributions of the entire paths in the uncontrolled and controlled process from time 0 to T.

In this study, we consider the optimal control problem based on the KL divergence. There are many possible controlled processes that take the system to a given target distribution. Among them, we consider the optimal controlled process, whose control cost is minimized. This process is the closest controlled process to the uncontrolled process in the KL sense. To express this mathematically, let us denote the probability distribution induced by the uncontrolled process by $\mathcal{P}$ and that by the controlled process ($\tilde{\mathcal{P}}$) which are defined on an abstract space composed of a set of $\mathbb{R}^n$-valued continuous functions on $[0, T]$ (denoted by $S = C([0, T],\ \mathbb{R}^n)$). The KL divergence between these paths is written as

$$D_{\mathrm{KL}}\left(\tilde{\mathcal{P}}, \mathcal{P}\right) = \int_S \log\frac{d\tilde{\mathcal{P}}}{d\mathcal{P}}d\tilde{\mathcal{P}}. \tag{10}$$

where $\frac{d\tilde{\mathcal{P}}}{d\mathcal{P}}$ is a Radon-Nikodym derivative (we consider when $\mathcal{P}$ and $\tilde{\mathcal{P}}$ are equivalent: a mathematical condition that guarantees the existence of $\frac{d\tilde{\mathcal{P}}}{d\mathcal{P}}$). Then the probability distribution of the optimal controlled process we seek is the following $\mathcal{Q}$:

$$\mathcal{Q} = \arg\min_{\tilde{\mathcal{P}}} D_{\mathrm{KL}}\left(\tilde{\mathcal{P}}, \mathcal{P}\right). \tag{11}$$

with the boundary conditions;

$$x(0) \sim \mathcal{N}(\mu_0, \Sigma_0), \quad x(T) \sim \mathcal{N}(\mu_T, \Sigma_T), \tag{12}$$

if the minimum exists. We call the minimum the *stochastic control cost*,

$$\min_{\tilde{\mathcal{P}}} D_{\mathrm{KL}}\left(\tilde{\mathcal{P}}, \mathcal{P}\right) = D_{\mathrm{KL}}\left(\mathcal{Q}, \mathcal{P}\right) = \int_S \log\frac{d\mathcal{Q}}{d\mathcal{P}}d\mathcal{Q}. \tag{13}$$

Again, the proposed framework is compared to the conventional deterministic framework in Table 1.

The optimization problem of minimizing the KL divergence between two stochastic processes (equation (11)) has been referred to as *Schrödinger's bridge problem* [29, 30]. This problem was originally proposed by Erwin Schrödinger for finding the most probable path that moving particles take [32]. The Schrödinger bridge problem was subsequently found to be equivalent to an optimal control problem, and has been studied in the field of control theory [30, 31, 33]. To our knowledge, our recent study [6] is the first work in neuroscience to use the KL minimization problem in examining control cost in brain dynamics.

2.3 Equivalence between KL cost and Quadratic Cost

Next, we show that KL minimization boils down to an optimal control problem where the cost function is the expectation of the quadratic input with respect to controlled process [34, 31]. First, we observe the following two propositions.

Proposition 2.1 (Chen [33])**.** The next two probability laws are equal:

- $\mathcal{Q} = \arg\min_{\tilde{\mathcal{P}}} D_{\mathrm{KL}}\left(\tilde{\mathcal{P}}, \mathcal{P}\right)$.
- The probability law induced by the next process:

$$dx(t) = Ax(t)dt + v^*\left(x(t), t\right)dt + Cdw(t) \tag{14}$$

where $v^* : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$ is the process that minimizes the next functional of $v : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$,

$$\mathcal{J}(v) = \mathbb{E}_{\mathcal{P}'}\left[\int_0^T \|v\|^2_{(C\,{}^{\mathrm{t}}C)^{-1}}\, dt\right]. \tag{15}$$

Here, $\|v\|^2_{(C\,{}^{\mathrm{t}}C)^{-1}} = {}^{\mathrm{t}}v\left(C\,{}^{\mathrm{t}}C\right)^{-1} v$, and v moves over the set such that the process

$$dx(t) = Ax(t)dt + v\left(x(t), t\right)dt + Cdw(t) \tag{16}$$

satisfies the Gaussian boundary conditions

$$x(0) \sim \mathcal{N}(\mu_0, \Sigma_0), \quad x(T) \sim \mathcal{N}(\mu_T, \Sigma_T). \tag{17}$$

and

$$\mathbb{E}_{\mathcal{P}'}\left[\int_0^T \|v\|^2_{(C\,{}^{\mathrm{t}}C)^{-1}}\, dt\right] < \infty. \tag{18}$$

Proof. See [33]. □

Proposition 2.2. Let the uncontrolled process (equation (7)) and the controlled process (equation (16)) admit solutions, and $\mathcal{P}$ and $\mathcal{P}'$ probability distributions induced by the uncontrolled and controlled process, respectively. We assume that the two process have the same initial distribution π_0 (i.e., the probability distribution of $x(0)$) that is absolutely continuous with respect to the Lebesgue measure. If equation (18) stands, then

$$D_{\mathrm{KL}}\left(\mathcal{P}', \mathcal{P}\right) = \frac{1}{2}\mathbb{E}_{\mathcal{P}'}\left[\int_0^T \|v\|^2_{(C\,{}^{\mathrm{t}}C)^{-1}}\, dt\right], \tag{19}$$

where $\mathbb{E}_{\mathcal{P}'}[\cdot]$ represents the expectation on a probability law $\mathcal{P}'$.

Proof. We make use of the Girsanov theorem, which is a rule for transforming measures of processes expressed with different Brownian motions. See Supplementary Note 3.. □

To summarize the two propositions,

- The KL minimization problem is equivalent to the optimization problem where the cost function is a quadratic function (equation (15)).
- Among numerous dynamics that satisfy equation (17), we only need to consider those that can be described in the form (equation (16)).
- The value of the KL divergence itself is equal to the quadratic cost function (equation (19)).

3 Theoretical results

3.1 Analytical solution of the stochastic control cost

Having proposed our new framework and some key features previously acknowledged, we move on to presenting our own results. We start by showing the analytical solution of the optimal value of equation (15), which is described as follows:

$$\begin{aligned}\mathcal{J}^* =& {}^{\mathrm{t}}\mu_d G(T)^{-1}\mu_d - {}^{\mathrm{t}}\mu_T \Pi(T)\mu_t + {}^{\mathrm{t}}\mu_0 \Pi(0)\mu_0 \\ &+ \int_0^T \mathrm{tr}\left(C\,{}^{\mathrm{t}}C\Pi(t)\right) dt - \mathrm{tr}\left(\Pi(T)\Sigma_T - \Pi(0)\Sigma_0\right).\end{aligned} \tag{20}$$

In the equation, $G : [0,T] \to \mathbb{R}^{n\times n}$, $\Pi : [0,T] \to \mathbb{R}^{n\times n}$, $\Psi : [0,T]\times[0,T] \to \mathbb{R}^{n\times n}$, and $\mu_d := \mu_T - \Psi(T,0)\mu_0$. See Supplementary Note 2. for the definitions and detailed derivations. Note that $\Pi(t)$, $G(t)$, and $\Psi(t,s)$ depend on A, C, Σ_0 and Σ_T. This gives the analytical expression of the stochastic control cost.

3.2 Decomposition of the Stochastic Cost

We next demonstrate that equation (20) can be decomposed into two parts: the cost needed to steer the mean (called *mean control cost*) and the cost needed to steer the covariance (called *covariance control cost*). More specifically, when we decompose this as

$$\mathcal{J}^* = \mathcal{J}^*_\mu + \mathcal{J}^*_\Sigma, \tag{21}$$

where

$$\mathcal{J}^*_\mu = {}^{\mathrm{t}}\mu_d G(T)^{-1}\mu_d - {}^{\mathrm{t}}\mu_T \Pi(T)\mu_t + {}^{\mathrm{t}}\mu_0 \Pi(0)\mu_0, \tag{22}$$

$$\mathcal{J}^*_\Sigma = \int_0^T \mathrm{tr}\left(C\,{}^{\mathrm{t}}C\Pi(t)\right) dt - \mathrm{tr}\left(\Pi(T)\Sigma_T - \Pi(0)\Sigma_0\right), \tag{23}$$

we show that we can consider $\mathcal{J}^*_\mu$ as the cost needed steer the mean (mean control cost), and $\mathcal{J}^*_\Sigma$ as the cost needed to steer the covariance (covariance control cost). This decomposition is shown schematically in Fig. 2.

We found that the mean control cost $\mathcal{J}^*_\mu$ depends only on the marginal means (μ_0 and μ_T), while the covariance control cost $\mathcal{J}^*_\Sigma$ depends only on the marginal covariances (Σ_0 and Σ_T). To explain these dependencies, we start by examining $\mathcal{J}^*_\Sigma$. We can see that equation (23) depends only on the marginal covariances and is independent of the marginal means, since $\Pi(t)$ relies only on Σ_0 and Σ_T (in addition to A and C). We can then interpret these terms to represent the cost of navigating the covariance from Σ_0 to Σ_T. On the other hand, $\mathcal{J}^*_\mu$ (equation (22)) seems to depend on both the marginal means and marginal covariances, since G and Π are dependent

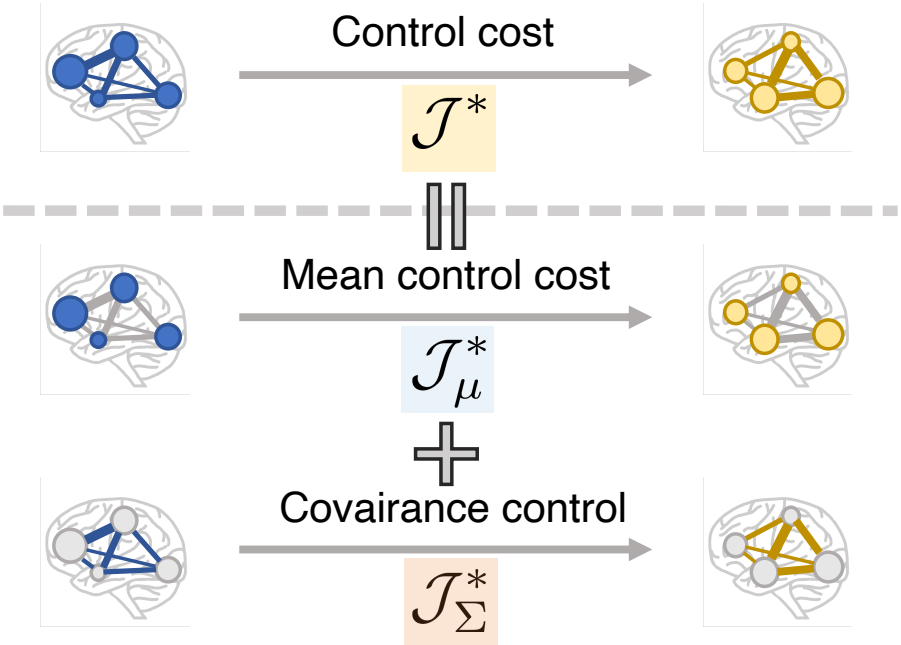


Figure 2: Schematic illustration of the decomposition of stochastic control cost into the mean and covariance control. The stochastic control cost from one state to another is described as a sum of the cost to control the mean and covariance from the initial state to the target state. The radius of the circle represents the magnitude of the mean, while the thickness of an edge represents the covariance value between two nodes.

on Σ_0 and Σ_T. Surprisingly, however, we found that equation (22) actually does *not* depend on the marginal covariances, but rather only on the means. In fact, $\mathcal{J}^*_\mu$ is further rewritten as

$$\mathcal{J}^*_\mu = {}^{\mathrm{t}}(\mu_T - \Phi(T,0)\mu_0)M(T)^{-1}(\mu_T - \Phi(T,0)\mu_0). \tag{24}$$

See Supplementary Note 4. for the details. Thus, equation (22) can be seen as the cost for controlling the mean from μ_0 to μ_T.

Furthermore, under a specific setting, we found that the mean control cost (Eq .(24)) is equal to the deterministic control cost discussed in 2.1, and has been used to examine control properties in previous studies [5, 7, 9, 10, 17, 19]. This correspondence between the mean control cost and the deterministic control cost indicates that the stochastic control cost, defined as the KL divergence, can be considered an extension of the conventional deterministic control cost. This correspondence is explained in detail and mathematically proved in Supplementary Note 4..

The covariance control cost, on the other hand, is the cost of shifting the covariance of the probability distribution. This cost has been ignored in previous neuroscience studies [5, 7, 9, 10, 17, 19]. We will cover the neuroscientific implication of the covariance control cost, which can be rephrased as the cost for controlling functional connectivity, in the Discussion.

The significance of this decomposition is that it enables us to separately quantify the influence of taking into account the covariance of brain activities in the control cost (Fig. 2). As we saw, the mean control cost depends only on the marginal means and can be shown to be equivalent to the deterministic control cost under a certain setting. On the other hand, the convariance control cost depends only on the marginal covariances. Thus, the effect of taking account of the covariance of brain activities is reflected only in the covariance control cost.

3.3 Input Given at Each Node

Lastly, we evaluate the contribution of each node to control transitions by the total amount of inputs given at each node in the brain network.

The amount of inputs provided at each node can be calculated as follows. Again, the optimal dynamics is described as

$$dx = Axdt + v^* dt + Cdw. \tag{25}$$

We let $v^* = {}^{\mathrm{t}}(v_1^*, \cdots, v_{100}^*)$ where $v_k^* : [0, T] \to \mathbb{R}$ $(k = 1, \cdots, 100)$. The expectation of the total input given to the k'th node (denoted by $\mathcal{I}(k),\ k = 1, \cdots, 100$) is then given as

$$\mathcal{I}(k) = \mathbb{E}\left[\int_0^T \|v_k^*\|^2 dt\right], \tag{26}$$

which is readily transformed into

$$\mathcal{I}(k) = \left(S_C \int_0^T \left(\Pi\left(\Sigma(t) + \mu(t)\,{}^{\mathrm{t}}\mu(t)\right)\Pi - 2m(t)\,{}^{\mathrm{t}}\mu(t)\Pi + m(t)\,{}^{\mathrm{t}}m(t)\right) dt\ S_C\right)_{kk}. \tag{27}$$

The subscript $(\cdot)_{ij}$ denotes the (i, j)-th entry of a matrix, and $m : [0, T] \to \mathbb{R}^n$ is a function for arranging the transient mean. See Supplementary Note 5. for the derivation, and $m : [0, T] \to \mathbb{R}$ is a certain function defined in Supplementary Note 2.. This value can be decomposed in a similar way to the stochastic control cost (equation (21)),

$$\mathcal{I}(k) = \mathcal{I}_\mu(k) + \mathcal{I}_\Sigma(k), \tag{28}$$

where

$$\mathcal{I}_\mu(k) = \left(S_C \int_0^T (\Pi(t)\mu(t) - m(t))\,{}^{\mathrm{t}}(\Pi(t)\mu(t) - m(t))\ dt\ S_C\right)_{kk}, \tag{29}$$

$$\mathcal{I}_\Sigma(k) = \left(S_C \int_0^T \Pi(t)\Sigma(t)\Pi(t)\ dt\ S_C\right)_{kk}. \tag{30}$$

As $m(t)$ is the function for shifting the mean of the brain dynamics, the first term $\mathcal{I}_\mu(k)$ can be interpreted as the cost needed to control the mean, which we call the *mean input* at the k'th ROI. The latter $\mathcal{I}_\Sigma(k)$ is the cost of controlling the covariance, which we refer to the *covariance input* at the k'th ROI.

3.4 Significance of the Newly Suggested Stochastic Cost

Here we summarize the significance of our novel framework and the stochastic cost (equation (20)). Previous studies [31, 34] have shown that the quadratic cost (equation (15)) is equivalent to the KL divergence between the uncontrolled and controlled processes. There are three novel findings:

1. The analytical solution of the stochastic cost (equation (20)) can be disintegrated into two portions, the energy of driving the mean (mean control) and that for the covariance (covariance control) (equation (21) and Fig. 2).

2. The mean control cost turns out to be identical to the conventional deterministic control energy equation (3) in specific occasions (equation (24)). This clarifies the correspondence between the deterministic and the stochastic costs. The covariance control cost has not been quantified in previous studies

3. By investigating the control input assigned to each node, we can compute the amount of total input given at one node. This node-level input can also be decomposed into two parts, the input necessary for controlling the mean and one for the covariance.

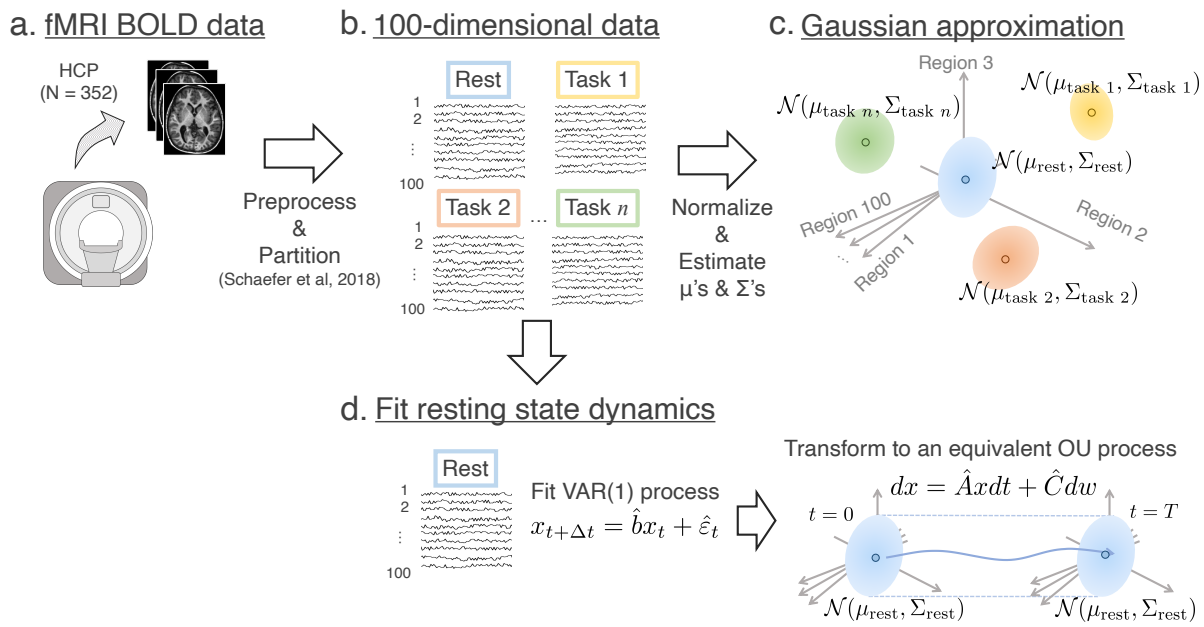


Figure 3: Data acquisition techniques. **a**. The fMRI BOLD data in HCP underwent minimal preprocessing and were then divided into 100 regions of interests (ROI). **b**. This gives us 100-dimensional time series data in our hands. **c**. The data at rest and under each task state were approximated as a Gaussian distribution in the 100-dimensional space whose individual coordinates represent the magnitude of activity of an ROI.

4 Results on Application to fMRI data

So far we have formulated a method to quantify the stochastic control cost and contribution of each brain region on state transitions. With this method, we then address the following two problems:

- How important is it to take account of the covariance of brain state probability distributions?
- What brain areas are significant in optimally controlling the brain state transitions?

To address the first question, i.e. to assess the influence of covariance, one can compare the covariance control cost $\mathcal{J}_\Sigma^*$ to the mean control cost $\mathcal{J}_\mu^*$. We have seen that the mean control cost in the stochastic model is equal to the optimal control cost in the deterministic model (Section 3.2), where we consider the control from one point to another point. So imagine, for instance, a case where the mean control cost is far greater than the covariance control cost and hence accounts for the most part of the overall stochastic cost. We might then need to consider only the point-to-point control and might not have to investigate distribution-to-distribution control as we did in the stochastic model. To evaluate the necessity of incorporating probability distributions in our model, we first need to examine to what extent the covariance control cost accounts for the stochastic cost.

To address the second question regarding important brain regions in controlling brain state transitions, we computed control input v for each region as defined in Section 3.3. By decomposing the control input of each region into the mean input ($\mathcal{I}_\mu(k)$) and covariance input ($\mathcal{I}_\Sigma(k)$), we can identify regions that are important for shifting the mean and covariance of brain activity. This decomposition enables us to identify not only the brain regions that need to be activated but also those regions that are crucial to the reconfiguration of co-variation of whole brain activity.

Table 2: Ratios of covarariance control compared to the mean control cost in the transition from rest to each task. In each of the 100 bootstrap trials, the covariance control cost is divided by the mean control cost, and the average and the standard deviations of the 100 trials are computed.

Task name	Ratio (Covariance control / mean control cost)
Emotion	2.83 ± 0.019
Gambling	0.78 ± 0.003
Language	2.56 ± 0.013
Motor	2.71 ± 0.019
Relational	0.92 ± 0.004
Social	0.53 ± 0.002
WM	0.45 ± 0.002

To address these questions, we used whole-brain functional magnetic resonance imaging (fMRI) data recorded as part of the Human Connectome Project (HCP) [35]. These data consist of recordings of 990 healthy adults. We used a subset of data composed of 352 subjects based on a criteria proposed in [36]. For each subject, the dataset contains blood-oxygen-level-dependent (BOLD) signals, including a pair of scans at rest and scans under seven different cognitive tasks: emotion, gambling, motor, language, relational, social, and working memory. After a standard preprocessing procedure including denoising with nuisance regressors, the whole voxels were partitioned into 100 regions of interests (ROIs) (Fig. 3-a & b) following the method proposed in [37].

After preprocessing and parcellation, we applied our framework to the fMRI data. Here, we assumed that the uncontrolled process is the resting state dynamics of the brain and the controlled dynamics is the transition from the probability distribution of the resting state to that of a task state. Under this assumption, we computed the control cost from the resting state probability distribution to a task distribution. To perform this computation, we need to infer two classes of probability distributions: i) the probability distributions that characterize the resting state dynamics, and ii) the task state probability distribution. For i), we employed a regression using the sparse vector autoregressive (VAR) model (Fig. 3-d) on the resting state BOLD data. We assumed the resting state dynamics to be the steady state dynamics and computed the steady state probability distribution ($\mathcal{N}(\mu_0, \Sigma_0)$) and the transition probability distribution (characterized by the matrices A and C). For ii), the mean (μ_T) and covariance (Σ_T) are inferred using the sample mean and covariance in the BOLD signals after normalization (Fig. 3-c).

As for the target time constant, we used various T values ranging from 0.1 to 6.0 seconds, and observed that results did not qualitatively change (Supplementary Fig. 1). Thus, below, we show the results with $T = 1.0$ seconds as a representative.

To estimate the probability distributions and the stochastic costs, a bootstrapping method was adopted for statistical robustness. The estimation was carried out with 100 randomly chosen subjects out of 352 subjects and was repeated 100 times independently. See the Methods for further details.

4.1 Mean and Covariance Control Cost in the Stochastic Model

Using the estimated probability distributions and the matrices A and C, we computed the stochastic control cost from rest to the seven tasks. For each task, we separately computed the mean control cost ($\mathcal{J}^*_\mu$), the covariance control cost ($\mathcal{J}^*_\Sigma$), and their ratio ($\mathcal{J}^*_\Sigma/\mathcal{J}^*_\mu$).

First, to assess the contribution of the covariance control cost compared to the mean control

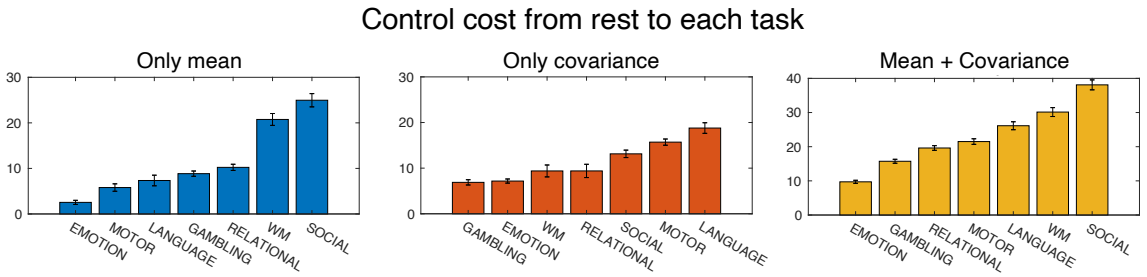


Figure 4: Stochastic control cost from the resting state to task states. Left panel, mean control cost values ($\mathcal{J}_\mu^*$); middle, covariance control values ($\mathcal{J}_\Sigma^*$); and right, total control values ($\mathcal{J}^* = \mathcal{J}_\mu^* + \mathcal{J}_\Sigma^*$) in ascending order.

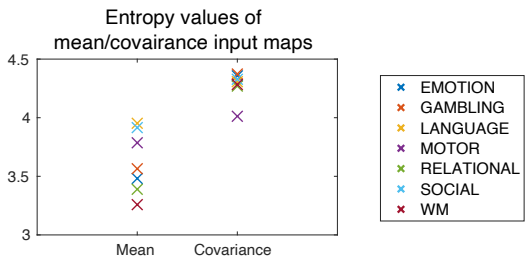


Figure 5: Entropy values of the mean and covariance input maps for each task The values were averaged over 100 bootstrap trials.

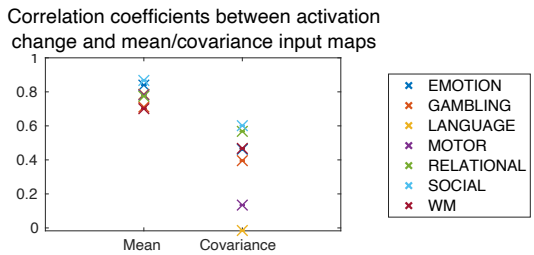


Figure 6: Correlation coefficients between the absolute values of t-values of activation and the input maps. The values were averaged over 100 bootstrap trials.

cost, we show the ratio in Tab. 2. We found that covariance control costs are as large as, or slightly larger than, the mean control costs. The ratios range from 0.4 to 2.8 depending on the task, with an average value of about 1.5. This result indicates that the contribution of covariance control cost to the overall control cost is as important as that of mean control cost. We also found that these ratios of covariance control cost to mean control cost take various values depending on the specific task. For example, the language task needs a larger covariance control costs than the mean, while the opposite is the case for WM.

To identify the influence of the covariance control cost, we then examined if incorporating the covariance control cost changes the order of magnitude of total control costs for the seven tasks in the dataset. We found that incorporating covariance control cost changes the order of magnitude of the control costs for the seven tasks. We show the mean, the covariance, and the stochastic control cost (= mean control cost + covariance control cost) of the seven tasks in ascending order (Fig. 4). The left panel (blue bars) shows the mean control cost values, the middle panel (orange bars) the covariance control cost, and the right panel (yellow bars) the total stochastic control costs. We can see that the order changes when we consider covariance control cost in addition to mean control cost. For example, the mean control cost for the gambling task is larger than that of the motor task; however, when we take account of covariance control cost, this relationship reverses. Thus, we see that the covariance control cost changes the result when estimating the magnitude relationships of control costs.

4.2 Control Inputs at Brain Regions

To identify which areas are important in optimally controlling brain state transitions, we next computed the amount of inputs given to each brain region (see 3.3). We computed the mean

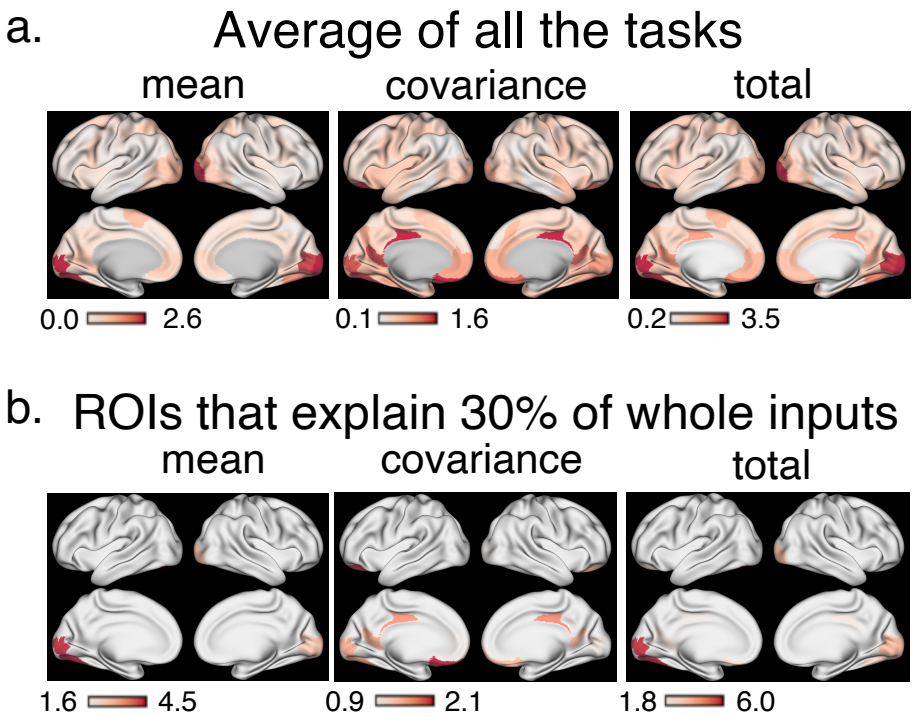


Figure 7: Mean, covariance, and total input maps averaged over the seven tasks. a. Average of the mean, covariance, and total inputs for each 100 ROIs over seven tasks in the HCP dataset. The numerical values indicate percentages compared to the sum of the total input to the whole ROIs. **b.** The ROIs that account for 30% of the whole mean, covariance, and total inputs shown in Fig. 7-a. The numerical values indicate the percentages compared to the sum of the mean, covariance, and total input to the whole ROIs, respectively.

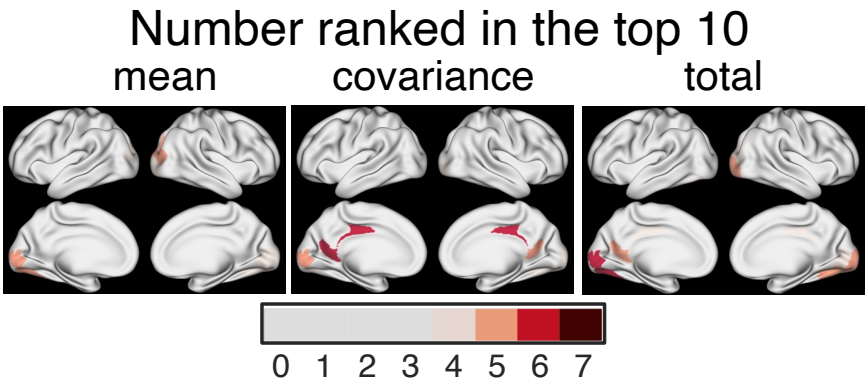


Figure 8: Number of times an ROI was ranked in the top ten brain regions that contributed most from rest to the seven tasks. From left to right, each panel shows the areas that contribute to the mean, covariance and total control, respectively.

input $\mathcal{I}_\mu(k)$ (equation (29)) and covariance input $\mathcal{I}_\Sigma(k)$ (equation (30)) at each ROI, obtaining a pair of 100-dimensional vectors ($(\mathcal{I}_\mu(1), \cdots, \mathcal{I}_\mu(100)), (\mathcal{I}_\Sigma(1), \cdots \mathcal{I}_\Sigma(100))$) and their sum $((\mathcal{I}(1), \cdots, \mathcal{I}(100)))$. We can think of these vectors as brain maps of the amount of input required to alter the mean and covariance, and the total inputs. We call these maps the *mean*

input map, *covariance input map*, and *total input map*. We computed the inputs and obtained these three maps for each of the seven tasks.

We found that the mean and covariance input maps showed quantitatively different patterns in each task. As shown in Supplementary Fig. 2, mean inputs are typically located in a small number of regions, while the covariance input maps are more widely distributed throughout the brain. We quantitatively evaluated the difference in distributions of the input maps by computing the entropy of the input maps over 100 ROIs for each task. The entropy S of a vector quantifies how dispersed the elements of the vector are; if the inputs are assigned to widely distributed regions, the entropy takes a large value. See the Methods section for a detailed description of how the entropy was calculated. As shown in Fig. 5, the entropy of the covariance input maps is larger than that of the mean input maps in all seven tasks. This result suggests that the covariance input maps are more dispersed over the whole brain regions.

We then examined the relationships between the two types of control inputs and the changes in BOLD signal magnitudes when transitioning to a task. Intuitively, we may expect that larger inputs are required for brain regions that are activated or deactivated in the task. To study this relationship, we computed the correlation coefficient of the the brain activity change and the mean/covariance input maps. We subtracted BOLD signals of preparation periods from those of task periods, and then computed the t-values of the differences. We defined the absolute value of the t-values as the brain activity change. We found that the mean input maps have high correlation coefficients in all the tasks, whereas the covariance input maps have lower values (Fig. 6). This result indicates that the mean input maps are highly correlated to the changes in BOLD activation level, while the covariance input maps are less related.

To identify the brain regions that are commonly important for control of the seven cognitive tasks, we then computed the average of the mean, the covariance, and the total input maps for all seven tasks, as shown in Fig. 7-a. To calculate the contribution of each task in a fair manner, for each task, we normalized the mean($\mathcal{I}_\mu(k)$), the ($\mathcal{I}_\Sigma(k)$), and the total input map $\mathcal{I}(k),\ k = 1, \cdots, 100$ by dividing by the sum of the total input to the whole ROIs ($\sum_{k=1}^{100} \mathcal{I}(k)$), and then took the average of the seven tasks. To visually inspect particularly important ROIs, we created Fig. 7-b, which shows ROIs having the largest inputs and explaining 30% of the sum of each input to the whole ROIs ($\sum_{k=1}^{100} \mathcal{I}_\mu(k), \sum_{k=1}^{100} \mathcal{I}_\Sigma(k)$, or $\sum_{k=1}^{100} \mathcal{I}_\mu(k)$). We can see that a large amount of the mean inputs is in the visual areas (Fig. 7-a, "mean"). In contrast, the covariance inputs are slightly more distributed over the whole brain. Specifically, a large portion of inputs is concentrated in the visual area, orbitofrontal area, and the posterior cingulate cortex (PCC) (Fig. 7-a "covariance").

To validate these results on the brain regions which are important in regulating brain state transitions, we counted how many times each ROI appears in the top 10 ROIs with the largest control inputs. Although we identified commonly important regions in controlling transitions by computing the average control inputs of all tasks, the possibility exists that the control input to a certain area is extremely large for one task only and not for the others. To rule out such a possibility and to find those regions which are commonly significant for the seven tasks, we selected the top 10 significant brain regions for each task and counted how many times each region is ranked in the top 10. We performed this analysis for the mean, the covariance, and the total inputs.

This additional analysis further supported our findings that the PCC and lower visual areas are generally important in control transition to task states. The results of the additional analysis are shown in Fig. 8. The color of an ROI signifies the number of tasks for which the ROI is ranked in the top 10 (thus the numbers take integer values from zero to seven). To facilitate visualization, we only colored ROIs that appeared in the top 10 more than three times. For mean input, the lower visual areas are ranked in the top 10 in the majority of the tasks. For the covariance input, the posterior cingulate cortex (PCC) is ranked in the top 10 most frequently, followed by the lower visual areas. For the total input, the lower visual areas and the PCC

were ranked most important. Taking these findings together, we conclude that the PCC and the lower visual areas are the most significant regions for transitioning to task states. While the lower visual areas contribute to both the mean and covariance shift, the PCC contributes to shifting covariance only.

5 Discussion

In this study, we propose a novel framework based on a linear stochastic dynamics to quantify control costs between brain states. We define the minimal KL divergence between the uncontrolled and the optimally controlled paths (equation (10)) as the stochastic control cost. There are three theoretical contributions in this study. i) We give a solution for the stochastic control cost (equation (20)), using equivalence for the quadratic integral (equation (15)). ii) We show that the cost can be decomposed into two parts: one for controlling mean brain activities (mean control) and the other for controlling co-variations among various brain regions (covariance control) (Fig. 2). We also demonstrated that the mean control cost is identical to the deterministic control cost (equation (6)) that has been used in the previous studies. iii) The proposed framework allows us to measure the inputs given in each ROI, thanks to the explicit modeling of the input term (equation (26) & equation (27)). The input in each ROI can be further decomposed into inputs necessary for regulating the mean brain activity and the whole-brain functional connectivity (equation (28)).

We applied our new method to fMRI BOLD data in the HCP dataset. We first examined the impact of incorporating covariance in the model. We found that the covariance control cost was as large as the mean control cost. Furthermore, taking into account the covariance control cost changed the order of magnitude of the control costs for tasks. These two results indicate that the influence of incorporating covariance in estimating control costs is not negligible. We then investigated the brain areas that contribute to state transitions by computing mean and covariance inputs at each brain region. The results revealed that the regions of high mean input were more limited than those of high covairance input. We also found that regions with high mean inputs involve those areas whose activity changes during the tasks. In particular, we discovered that for regulating the mean, the lower visual areas turned out to be the most significant, whereas for the covariance, the PCC and the lower visual were the most significant.

5.1 Stochasticity of Control Cost

One contribution of the present framework is incorporating stochasticity in quantifying the control cost, which has not been adequately addressed in previous studies in neuroscience [5, 7, 9, 10, 17]. Our framework incorporates stochasticity in two ways, namely (1) whether the dynamics of the model is stochastic, and (2) whether the cost takes stochasticity into consideration.

Previous studies have not dealt with either of these or considered only the first point. Previous studies that utilized control theory [5, 7, 9, 17, 16, 18, 19] applied the linear state-space model with deterministic dynamics and deterministic cost (equation (6)), while another study applied a stochastic model (OU process) for brain dynamics [27] but calculated costs based on the deterministic framework.

To our knowledge, our present and recent studies [6] are the first to take account of stochasticity, not only with regard to brain dynamics but also to cost. As we saw in Fig. 4, incorporating stochasticity in quantifying the control cost significantly changes the result. Thus, in future applications of control theory to neuroscience, it will be desirable to assess the influence of stochasticity in the brain on control costs, as we did in the present work.

5.2 Advantages of the Decomposition into Mean and covariance control cost

One remarkable characteristic of the newly proposed cost is that it allows decomposition into mean control cost and covariance control cost (equation (21)). A major theoretical advantage of this decomposition is that, as seen in Section 3.2, it enables us to separately examine the influence when taking covariance into account. Here, we discuss two significant aspects of this decomposition in neuroscience.

First, thanks to the decomposition, the stochastic cost can quantitatively compare the significance of contributions of two separate phenomena in a unified manner, namely the change in magnitude and co-variation of brain activities. Conventionally, changes in the magnitude of brain activities have been assessed through the estimated coefficients (often denoted as β) after linear regression using the general linear model (GLM) [38, 39]. The change in co-variation has been evaluated through the subtraction of correlation matrices [25, 40] or through network theoretic measures [41, 42, 43]. The dynamical change in co-variation (often referred to as functional connectivity in the field of neuroimaging) is called functional reconfiguration [44, 45, 46, 47]. In this way, classically, the change in magnitudes and the co-variation of activities have been examined separately in different contexts. In contrast to these conventional studies, however, our use of decomposition has allowed us to quantitatively compare control costs for the mean and covariance from a unified perspective.

Second, the proposed framework enables us to investigate regions playing a significant role in the control of brain state transitions, as we have seen in Section 3.3. Furthermore, the contribution of a brain region can also be decomposed as in equation (28), which allows us to separately compute the cost for controlling the mean of brain activity and covariance (functional connectivity). This may provide insights into important brain regions in controlling the mean and covariance of brain activity. This topic is covered in the next section.

5.3 Significant Brain Regions in the Control of State Transitions

The present framework allows us to identify brain areas that contribute to controlling the mean and covariance separately. Here we discuss the general tendencies of the mean and covariance input maps, which are obtained by computing equation (29) and equation (30). We then examine significant areas for control, namely the PCC and the lower visual areas.

We showed the general tendencies of the mean and covariance input maps in Section 4.2. (1) Regions with high mean inputs tend to involve limited brain areas (Fig. 5) and (2) are closely related to areas whose activity magnitude changes during tasks (Fig. 6). It is intuitively understandable that to change the magnitude of activity at a region, we need to add inputs to this specific region.

In contrast to the mean inputs, (1) the covariance inputs are more widespread (Fig. 5) and (2) are less related to the regions of altered activities (Fig. 6). The first observation might reflect the fact that the alteration of covariance (termed functional reconfiguration [44, 45, 46, 47]) occurs all over the brain and thus broad input to control covariance may be necessary. The second observation means that we cannot estimate important regions for controlling covariance using only the magnitudes of activity. Examining the magnitudes of activity only may lead to the failure to identify significant regions controlling brain state transitions.

An intriguing observation is that the posterior cingulate cortex (PCC) commonly contributes to controlling the covariance in many of the tasks examined (Fig. 7-a & b, Fig. 8, the middle panels). Among previous findings related to this result, the PCC is reported to be connected with many brain areas not only structurally [48] but also functionally [49]. Although the PCC's specific functions are not yet fully understood, previous studies have revealed that it is associated with many cognitive processes, including cognitive control [50], as shown using fMRI [51], and through recordings of single-cell firing rates [52]. From these studies, it is now considered that the PCC exerts influence on various regions and might thereby play a role in altering the functional

connectivity of the brain [50, 51, 53]. Our finding that the PCC is significant in controlling covariance supports this view. Although this speculation needs further evidence, our study showed the importance of the PCC in changing functional connectivity from a control theoretic perspective.

Compared to the PCC and its contribution to controlling covariance, we found that the lower visual areas are included in the significant regions controlling both the mean and covariance (Fig. 7-a & b, Fig. 8, left panels). This might be due to the nature of the tasks recorded in the HCP dataset, where subjects were presented visual stimuli as part of the tasks. To determine whether the visual areas are generally significant in controlling state transitions even during nonvisual tasks, we need to apply our framework to cognitive tasks that involve other modalities (e.g. audition, olfaction, and tactile sensation).

5.4 Validity and Merits of the Linear Continuous Model for fMRI Data Analysis

In this study, we adopted the linear continuous-state model to describe brain dynamics in the fMRI data. It has been pointed out that a linear model may oversimplify the brain dynamics, which is complex and known to behave in nonlinear manners [54, 55, 56]. Although we acknowledge the limitation of using the linear model for the highly nonlinear brain dynamics, a recent study has shown that the neural data, including the same fMRI data as we used in this study (HCP), can be better fitted with linear models than nonlinear ones, possibly because of temporal and spatial averaging effects [57]. Accordingly, this previous study [57] provides validity for our use of a linear model for the dataset.

To account for the nonlinearity of brain dynamics, another approach to modeling brain dynamics in the fMRI data is to use the discrete-space probabilistic model [6, 16, 23]. In our recent study [6], we used the same KL divergence as control cost in brain state transition as in the present study. The difference from our study lies in that in [6] we employed a different probabilistic model, where brain states are coarse grained using k-means clustering and discrete.

Compared to the linear continuous model used in the present study, the discrete model applied in [6] has pros and cons. In contrast to the linear model, the discrete model can capture the nonlinearity of the brain dynamics. This capability may broaden application to various kinds of neural data other than fMRI. By taking account of nonlinearity, it can also deal with the metastable nature of brain states [1, 8].

Although the discrete model can incorporate nonlinearity, the linear model is more feasible than the discrete model in computing control costs in high-dimensional brain dynamics. In the discrete model, the computational cost of the control cost can easily explode, as the control cost is analytically intractable. We have to resort to some iterative algorithms such as Sinkhorn's algorithms to solve KL optimization [6, 58]. Unless we coarse-grain and limit the number of brain states, the computational cost overflows and the estimation of cost becomes impractical. In contrast, control cost in the linear continuous model is analytically tractable. This enables us to compute control costs in high-dimensional dynamics.

5.5 Limitations of the theoretical framework

In this study, we considered limited situations of control wherein we implemented a model in which independent inputs are assigned to all nodes. This is mathematically denoted as a stochastic system with an input term $v : [0, T] \to \mathbb{R}^n$ (equation (16)). In the standard control theoretic framework, the input term v is often described as Bu, where $B \in \mathbb{R}^{n\times m}$ is an input matrix and $u : [0, T] \to \mathbb{R}^m$ is the control input (see equation (2)). We did not incorporate the input matrix B, which means that we only consider where $B \in \mathbb{R}^{n\times n}$ is an identity matrix.

In previous studies that have utilized the deterministic control theoretical framework, the model was grounded on more general cases wherein the system is described with an input term

Bu, as in equation (2). For example, in [16, 17, 27], $B \in \mathbb{R}^{n\times m}$, $m \neq n$ was set so that it represents independent inputs in a subgroup of nodes $\{k_1, \cdots, k_m\}$,

$$(B)_{ij} = \delta_{k_j j} \quad (i = 1, \cdots, n,\ j = 1, \cdots, m), \tag{31}$$

where δ is the Kronecker delta. These studies parametrized m, the number of input nodes, to consider the best control strategy. Another study [5] modeled the input nodes to only one, and set B as an n-by-1 matrix (or an n-dimensional vector),

$$(B)_i = \begin{cases} 1 & (i = k) \\ 0 & (\text{otherwise}) \end{cases}, \tag{32}$$

for a fixed $k \in \{1, \cdots, n\}$.

Even though we can technically consider such general cases where the input matrix B satisfies $\in \mathbb{R}^{n\times m}$ $(n \neq m)$ in our stochastic framework, computing the control costs in such models is more difficult both analytically and numerically. According to a previous study [59], if B is rank deficient, we cannot find an analytical solution and hence have to resort to a numerical convex optimization with time discretization, which takes enormous computational time.

To alleviate computational time in high-dimensional data, we may need different approaches, such as dimensionality reduction. It has been reported that the major neural activity patterns are confined on a low dimensional space, named neural manifolds [60, 61, 62, 63]. Based on these findings, we may first apply the dimensionality reduction to high-dimensional neural data and then fit the model with a general input matrix B.

5.6 Future Directions

In this study, we only considered the optimal control cost where brain state transitions are controlled in an optimal manner, with minimization of stochastic control cost. However, in real neural systems, state transitions are not controlled in an optimal manner. An intriguing future direction will be to compare the optimal and actual dynamics using neural data during tasks. For example, we can expect that the control cost to an unfamiliar task may be quite large at the beginning and gradually lessen as the individual is habituated to it, i.e., becomes more experienced. The habituation process is called automaticity and has been studied extensively [9, 64, 65, 66]. Estimating the control cost from real time-series data requires the estimation of time-varying control inputs. This is generally a difficult estimation problem that requires more sophisticated techniques [67]. Quantifying the actual control cost in real neural data will provide another new insight into cognitive processes, which we plan to pursue in future research.

6 Methods

6.1 On fMRI Data

The 3T functional magnetic resonance imaging (fMRI) data of 990 subjects were obtained from the Washington University-Minnesota Consortium Human Connectome Project (HCP) [26]. To remove the effects of outliers, we picked data of 352 subjects according to criteria suggested in [36].

6.2 Data preprocessing

Minimally preprocessed fMRI data were used for the resting state and seven cognitive task states (emotion, gambling, language, motor, relational, social, and working memory). Denoising was performed by estimating nuisance regressors and subtracting them from the signal at every vertex [68]. For this, 36 nuisance regressors and spike regressors were used following a previous

study [68], consisting of (1-6) six motion parameters, (7) a white matter time series, (8) a cerebrospinal fluid time series, (9) a global signal time series, (10-18) temporal derivatives of (1-9), and (19-36) quadratic terms for (1-18). The spike regressors were computed with 1.5 mm movement as a spike identification threshold. After regressing these nuisance time courses, a band-pass filter (0.01-0.69 Hz) was applied, whose upper filter bound corresponds to the Nyquist frequency of the time series. A parcellation process proposed in [37] was used to divide the cortex into 100 brain regions, which reduced the complexity of the following analysis.

6.3 Estimation of transition & diffusion matrices

We next estimated parameters that characterize resting and task brain states. To apply the framework explained in Section 2.2, we estimated the next parameters: the mean and the covariance matrix of the initial distribution (μ_0 and Σ_0), those of the target distribution (μ_T and Σ_T), and the matrices in the uncontrolled dynamics (A and $S_C := C^{\intercal}C$). For statistical robustness, a bootstrapping method was adopted, where the estimation described below was done using data of 100 randomly chosen subjects out of 352 overall subjects, and repeated 100 times independently. The data of the 100 subjects were concatenated to obtain a single time series that is long enough for statistically reliable estimation.

Before estimating the parameters using the concatenated data across different subjects, we need to normalize the time series data. This normalization is necessary for the following reason. Basically, the absolute values of fMRI BOLD signals are meaningless by themselves, due to the bias caused by the water content or the basic blood flow in the brain tissues [69]. Because of this, one cannot directly compare values of BOLD signals in different subjects. Thus, we cannot concatenate data of different subjects using the preprocessed fMRI BOLD signals as they are. To concatenate BOLD signals across subjects, we need some kind of scaling, or normalization, of the data.

We performed the scaling based on the assumption that the sum of the time series variances in the whole ROIs are constant when a subject is not engaged in a cognitive task. First, for each subject, we computed the trace values of the empirical covariance matrices of the resting state and the task-free moments in each task. Then, we divided the whole time series data of the rest and each task by the square root of the trace value of the rest and by that of each task's task-free moment, respectively. This division makes the magnitude of the sample covariance matrix (i.e., the trace of the sample covariance matrix) of the rest and the task-free moments to be one in all subjects and tasks. After this normalization, we concatenated the time series data of the whole resting state and the task-free moments of each task state of all 100 subjects. By using the concatenated time series data, we estimated the parameters as follows.

6.3.1 Distribution of the resting state and task states

For the resting state, the empirical mean and covariance were used to represent the probability distribution. The empirical mean of the resting state was almost a 100-dimensional zero vector in each subject due to a zero-mean adjustment in preprocessing. To extract the mean of the activity during a task, one needs to know to what extent the activity in the task-performing moments is different from that in the task-free moments. We assumed that in the task-free moments, the average magnitude of activity is the same as the average of the resting state, i.e., approximately zero. Accordingly, to obtain the time series mean of the activity during the task state, the time series mean during the task-free moments was subtracted from that in the task-performing moments. The covariance matrix was calculated as the empirical covariance matrix during the task-performing moments.

6.3.2 Estimation of Parameters of the Resting State Dynamics

To estimate the matrices A and C, we need to fit the time series to the dynamics equation

$$dx = Ax(t)dt + Cdw. \tag{33}$$

When sampled at some interval of $\Delta t > 0$, this process is equivalent to the next VAR(1) process,

$$x(t + \Delta t) = ax(t) + \varepsilon_{t+\Delta t}, \tag{34}$$

where $a = e^{A\Delta t}$ (matrix exponential). We first estimated the drift term (denoted by $\hat{a}$) using the least absolute shrinkage and selection operator (LASSO) regression. A transformation

$$\hat{A} = \log \hat{a}/\Delta t \tag{35}$$

followed to obtain the estimated drift coefficient A in the dynamics.

As for $S_C = C\,{}^{\mathrm{t}}C$, one can directly infer this from the covariance of $\varepsilon_{t+\Delta t}$ ($\mathbb{E}[\varepsilon_{t+\Delta t}\,{}^{\mathrm{t}}\varepsilon_{t+\Delta t}] =: S_{\varepsilon_{t+\Delta t}}$) in equation (34) using the relationship

$$S_C = -\frac{1}{\Delta t} vec^{-1}\left[(I - a \otimes a)^{-1} \log\left(a \otimes a\right) vec\left(S_{\varepsilon_{t+\Delta t}}\right)\right]. \tag{36}$$

where vec is the vectorization operetor and vec^{-1} its inverse, and log the matrix logarithm. To obtain the estimation of S_C (denoted by $\hat{S}_C$), we substituted above the estimated $\hat{a}$ and the empirical time series covariance $\hat{S}_{\varepsilon_{t+\Delta t}}$.

6.4 Calculation of Entropy of an Input Map

The entropy S of an input map $I = {}^{\mathrm{t}}(I_1, \cdots, I_{100}) \in {\mathbb{R}_+}^{100}$ quantifies how dispersed a vector is and takes the maximum if each element I_l ($l = 1, \cdots, 100$) takes the same value. S is calculated by

$$S = -\sum_{l=1}^{100} \frac{I_l}{I_{\mathrm{sum}}} \log(\frac{I_l}{I_{\mathrm{sum}}}), \tag{37}$$

where $I_{\mathrm{sum}} = \sum_{l=1}^{100} I_l$. Note that each element I_l ($l = 1, \cdots, 100$) is positive.

7 Code Availability

The codes for computing the stochastic control cost and reproducing the figures on this paper are available at https://github.com/oizumi-lab/SB_toolbox.

8 Acknowledgement

Masafumi Oizumi and Shuntaro Sasai were supported by JST Moonshot R&D Grant Number JPMJMS2012. Masafumi Oizumi was supported by JST CREST Grant Number JPMJCR1864, and JSPS KAKENHI Grant Number 18H02713 and 20H05712, Japan.

References

[1] Breakspear, M. Dynamic models of large-scale brain activity. *Nat. Neurosci.* **20**, 340–352 (2017).

[2] Kringelbach, M. L. & Deco, G. Brain states and transitions: Insights from computational neuroscience. *Cell Rep.* **32**, 108128 (2020).

[3] McKenna, T. M., McMullen, T. A. & Shlesinger, M. F. The brain as a dynamic physical system. *Neuroscience* **60**, 587–605 (1994).

[4] Botvinick, M. M. & Cohen, J. D. The computational and neural basis of cognitive control: charted territory and new frontiers. *Cogn. Sci.* **38**, 1249–1285 (2014).

[5] Gu, S. *et al.* Controllability of structural brain networks. *Nat. Commun.* **6**, 8414 (2015).

[6] Kawakita, G., Kamiya, S., Sasai, S., Kitazono, J. & Oizumi, M. Quantifying brain state transition cost via schrödinger bridge. *Network Neuroscience* **6**, 118–134 (2022).

[7] Braun, U. *et al.* Brain network dynamics during working memory are modulated by dopamine and diminished in schizophrenia. *Nat. Commun.* **12**, 3478 (2021).

[8] Deco, G. *et al.* Awakening: Predicting external stimulation to force transitions between different brain states. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18088–18097 (2019).

[9] Szymula, K. P., Pasqualetti, F., Graybiel, A. M., Desrochers, T. M. & Bassett, D. S. Habit learning supported by efficiently controlled network dynamics in naive macaque monkeys (2020). `2006.14565`.

[10] Parker Singleton, S. *et al.* LSD flattens the brain's energy landscape: evidence from receptor-informed network control theory (2021).

[11] He, X. *et al.* Pathological and metabolic underpinnings of energetic inefficiency in temporal lobe epilepsy. *bioRxiv* (2021).

[12] Kalman, R. E. On the general theory of control systems. In *Proceedings First International Conference on Automatic Control, Moscow, USSR*, 481–492 (1960).

[13] Chen, C.-T. *Linear System Theory and Desig.*

[14] Liu, Y. Y., Slotine, J. J. & Barabási, A. L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).

[15] Liu, Y. Y. & Barabási, A. L. Control principles of complex systems. *Rev. Mod. Phys.* **88**, 1–61 (2016).

[16] Cornblath, E. J. *et al.* Temporal sequences of brain activity at rest are constrained by white matter structure and modulated by cognitive demands. *Commun Biol* **3**, 261 (2020).

[17] Kim, J. Z. *et al.* Role of graph architecture in controlling dynamical networks with applications to neural systems. *Nat. Phys.* **14**, 91–98 (2018).

[18] Gu, S. *et al.* Optimal trajectories of brain state transitions. *Neuroimage* **148**, 305–317 (2017).

[19] Stiso, J. *et al.* White matter network architecture guides direct electrical stimulation through optimal state transitions. *Cell Rep.* **28**, 2554–2566.e7 (2019).

[20] Kirk, D. E. *Optimal Control Theory: An Introduction* (Courier Corporation, 2004).

[21] Rieke, F. *Spikes: exploring the neural code* (MIT press, 1999).

[22] Faisal, A. A., Selen, L. P. J. & Wolpert, D. M. Noise in the nervous system. *Nat. Rev. Neurosci.* **9**, 292–303 (2008).

[23] Lynn, C. W., Cornblath, E. J., Papadopoulos, L., Bertolero, M. A. & Bassett, D. S. Broken detailed balance and entropy production in the human brain. *Proceedings of the National Academy of Sciences* **118** (2021).

[24] Honey, C. J. *et al.* Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2035–2040 (2009).

[25] Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S. & Petersen, S. E. Intrinsic and task-evoked network architectures of the human brain. *Neuron* **83**, 238–251 (2014).

[26] Van Essen, D. C. *et al.* The WU-Minn human connectome project: an overview. *Neuroimage* **80**, 62–79 (2013).

[27] Deng, S. & Gu, S. Controllability analysis of functional brain networks (2020). `2003.08278`.

[28] Rieke, F. *Spikes: exploring the neural code* (MIT press, 1999).

[29] Chen, Y., Georgiou, T. T. & Pavon, M. On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint. *J. Optim. Theory Appl.* **169**, 671–691 (2016).

[30] Léonard, C. A survey of the schrödinger problem and some of its connections with optimal transport (2013). `1308.0215`.

[31] Dai Pra, P. A stochastic control approach to reciprocal diffusion processes. *Appl. Math. Optim.* **23**, 313–329 (1991).

[32] Schrodinger, E. Uber die umkehrung der naturgesetze. sitz. ber. der preuss. *Akad. Wissen. , Berlin Phys. Math* **144** (1931).

[33] Chen, Y., Georgiou, T. T. & Pavon, M. Optimal steering of a linear stochastic system to a final probability distribution, part I. *IEEE Trans. Automat. Contr.* **61**, 1158–1169 (2016).

[34] Chen, Y., Georgiou, T. T. & Pavon, M. On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint. *J. Optim. Theory Appl.* **169**, 671–691 (2016).

[35] Van Essen, D. C. *et al.* The WU-Minn human connectome project: an overview. *Neuroimage* **80**, 62–79 (2013).

[36] Ito, T. *et al.* Task-evoked activity quenches neural correlations and variability across cortical areas. *PLoS Comput. Biol.* **16**, e1007983 (2020).

[37] Schaefer, A. *et al.* Local-Global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).

[38] Friston, K. J. *et al.* Analysis of fmri time-series revisited. *Neuroimage* **2**, 45–53 (1995).

[39] Ashby, F. G. *Statistical analysis of fMRI data* (MIT press, 2019).

[40] Cole, M. W., Ito, T., Bassett, D. S. & Schultz, D. H. Activity flow over resting-state networks shapes cognitive task activations. *Nat. Neurosci.* **19**, 1718–1726 (2016).

[41] Finc, K. *et al.* Dynamic reconfiguration of functional brain networks during working memory training. *Nature communications* **11**, 1–15 (2020).

[42] Hahn, A. *et al.* Reconfiguration of functional brain networks and metabolic cost converge during task performance. *Elife* **9**, e52443 (2020).

[43] Braun, U. *et al.* Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11678–11683 (2015).

[44] Kitzbichler, M. G., Henson, R. N. A., Smith, M. L., Nathan, P. J. & Bullmore, E. T. Cognitive effort drives workspace configuration of human brain functional networks. *J. Neurosci.* **31**, 8259–8270 (2011).

[45] Davison, E. N. *et al.* Brain network adaptability across task states. *PLoS Comput. Biol.* **11**, e1004029 (2015).

[46] Stitt, I. *et al.* Dynamic reconfiguration of cortical functional connectivity across brain states. *Sci. Rep.* **7**, 8797 (2017).

[47] Krienen, F. M., Yeo, B. T. T. & Buckner, R. L. Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369** (2014).

[48] Hagmann, P. *et al.* Mapping the structural core of human cerebral cortex. *PLoS Biol.* **6**, e159 (2008).

[49] Leech, R., Braga, R. & Sharp, D. J. Echoes of the brain within the posterior cingulate cortex. *Journal of Neuroscience* **32**, 215–222 (2012).

[50] Leech, R. & Sharp, D. J. The role of the posterior cingulate cortex in cognition and disease. *Brain* **137**, 12–32 (2014).

[51] Lin, P. *et al.* Static and dynamic posterior cingulate cortex nodal topology of default mode network predicts attention task performance. *Brain imaging and behavior* **10**, 212–225 (2016).

[52] Hayden, B., Smith, D. V. & Platt, M. Cognitive control signals in posterior cingulate cortex. *Frontiers in human neuroscience* **4**, 223 (2010).

[53] Leech, R., Kamourieh, S., Beckmann, C. F. & Sharp, D. J. Fractionating the default mode network: distinct contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. *Journal of Neuroscience* **31**, 3217–3224 (2011).

[54] Rabinovich, M. I. & Muezzinoglu, M. Nonlinear dynamics of the brain: emotion and cognition. *Physics-Uspekhi* **53**, 357 (2010).

[55] Freeman, W. J. Nonlinear gain mediating cortical stimulus-response relations. *Biol. Cybern.* **33**, 237–247 (1979).

[56] Stephan, K. E. *et al.* Nonlinear dynamic causal models for fmri. *Neuroimage* **42**, 649–662 (2008).

[57] Nozari, E. *et al.* Is the brain macroscopically linear? a system identification of resting state dynamics (2020).

[58] Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems* **26** (2013).

[59] Chen, Y., Georgiou, T. T. & Pavon, M. Optimal steering of a linear stochastic system to a final probability distribution, part II. *IEEE Trans. Automat. Contr.* **61**, 1170–1180 (2016).

[60] Sadtler, P. T. *et al.* Neural constraints on learning. *Nature* **512**, 423–426 (2014).

[61] Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural manifolds for the control of movement. *Neuron* **94**, 978–984 (2017).

[62] Gallego, J. A. *et al.* Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nat. Commun.* **9**, 4233 (2018).

[63] Gao, P. *et al.* A theory of multineuronal dimensionality, dynamics and measurement (2017).

[64] Economides, M., Kurth-Nelson, Z., Lübbert, A., Guitart-Masip, M. & Dolan, R. J. Model-Based reasoning in humans becomes automatic with training. *PLoS Comput. Biol.* **11**, e1004463 (2015).

[65] Moors, A. & De Houwer, J. Automaticity: a theoretical and conceptual analysis. *Psychol. Bull.* **132**, 297–326 (2006).

[66] Haith, A. M. & Krakauer, J. W. The multiple effects of practice: skill, habit and reduced cognitive load. *Curr Opin Behav Sci* **20**, 196–201 (2018).

[67] Milde, T. *et al.* Time-variant partial directed coherence in analysis of the cardiovascular system. a methodological study. *Physiological measurement* **32**, 1787 (2011).

[68] Satterthwaite, T. D. *et al.* An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256 (2013).

[69] Poldrack, R. A., Mumford, J. A. & Nichols, T. E. *Handbook of functional MRI data analysis* (Cambridge University Press, 2011).

[70] Föllmer, H. Time reversal on wiener space. In *Stochastic processes—mathematics and physics*, 119–129 (Springer, 1986).

[71] Pavon, M. & Wakolbinger, A. On free energy, stochastic control, and schrödinger processes (1991).

[72] Behr, M., Benner, P. & Heiland, J. Solution formulas for differential sylvester and lyapunov equations (2018). `1811.08327`.