Dynamic neural reconfiguration for strategy switching during competitive social interactions
Supplementary Materials

Ruihan Yang, et.al.

Contents

List of Tables

List of Figures

1. Method S1. Time-varying Granger causality with signal-dependent noise

Assuming a constant effective connectivity between brain regions, the classic uses the time-series models (both GC and GCSDN) with constant model coefficients. However, this assumption may be an oversimplification of the information processing in the brain, especially during some intensive cognitive computation (e.g., the two-party bargaining game). To model the dynamic behavior, a more complicated model which can better describe the dynamic characteristics is needed. Here, we first briefly discussed the consquence when classic time-invariant model was applied to a time-varying system, and then proposed a new approach of time-varying Granger causality with signal-dependent noise (time-varying GCSDN) to measure the dynamic causality.

Consider the following time series model

$$x_t = ax_{t-1} + b(t)y_{t-1} + \varepsilon_t,$$

where ε_t is a Gaussian white noise, a is a constant coefficient, and $b(t)$ is a time-varying coefficient. The classic Granger causality can be defined as

$$\mathrm{F}_{y\to x} = \log \frac{\mathrm{var}\left(b(t)y_{t-1} + \varepsilon_t\right)}{\mathrm{var}\left(\varepsilon_t\right)},$$

and estimated by

$$\widehat{\mathrm{F}}_{y\to x} = \log \frac{\sum_{s=1}^{T-1} \hat{b}^2(s)y_s^2 + \sum_{s=1}^{T-1} \hat{\varepsilon}_s^2}{\sum_{s=1}^{T-1} \hat{\varepsilon}_s^2},$$

where $\hat{b}(t)$ is the local classic estimation at each time step t and $\hat{\varepsilon}_s$ is the residual process. If we ignore the time-varying property of the model and do the conventional least square estimation for the parameters, we can have an estimation of $\bar{b}$ and the corresponding model residual $\bar{\varepsilon}_s$, and the causality can be established as

$$\overline{\mathrm{F}}_{y\to x} = \log \frac{\sum_{s=1}^{T-1} \bar{b}^2 y_s^2 + \sum_{s=1}^{T-1} \bar{\varepsilon}_s^2}{\sum_{s=1}^{T-1} \bar{\varepsilon}_s^2}.$$

For the AR model, we showed that the causality established by assuming the constant coefficient in the model is upper bounded by the mean causality among time windows, and the condition on which the equality holds was discussed as follows. Define

$$\boldsymbol{o}_t = \begin{pmatrix} x_t \\ y_t \end{pmatrix}', \boldsymbol{\beta} = \begin{pmatrix} \bar{a} \\ \bar{b} \end{pmatrix}, \boldsymbol{\beta}_t = \begin{pmatrix} \hat{a} \\ \hat{b}(t) \end{pmatrix},$$

$$\boldsymbol{O} = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_{T-1} & x_{T-1} \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} x_2 \\ x_3 \\ \vdots \\ x_T \end{pmatrix}.$$

For the whole time series we have

$$\boldsymbol{O}'\boldsymbol{D} = \boldsymbol{O}'\boldsymbol{O}\boldsymbol{\beta},$$

and for each time step t, the following equation holds,

$$\boldsymbol{o}_t' x_{t+1} = \boldsymbol{o}_t'\boldsymbol{o}_t\boldsymbol{\beta}.$$

Since $\sum_{t=1}^{T-1} \boldsymbol{o}_t' x_{t+1} = \boldsymbol{O}'\boldsymbol{D}$, we get

$$\sum_{t=1}^{T-1} \boldsymbol{o}_t'\boldsymbol{o}_t\boldsymbol{\beta}_t = \boldsymbol{O}'\boldsymbol{O}\boldsymbol{\beta}.$$

Therefore, $\boldsymbol{\beta}$ is a weighted average of $\boldsymbol{\beta}_t$ as

$$\boldsymbol{\beta} = (\boldsymbol{O}'\boldsymbol{O})^{-1} \sum_{t=1}^{T-1} \boldsymbol{o}_t'\boldsymbol{o}_t\boldsymbol{\beta}_t.$$

Since what we consider here is the effect of the time-varying parameter $b(t)$, we further suppose that the constant parameter $a = 0$, and the above formula can be simplified as

$$\bar{b} = (\sum_{s=1}^{T-1} y_s^2)^{-1} \sum_{t=1}^{T-1} y_t^2 \hat{b}(t).$$

Since

$$\begin{aligned}
\sum_{s=1}^{T-1} \bar{b}^2 y_s^2 &= \left(\sum_{s=1}^{T-1} y_s^2\right)^{-2} \left(\sum_{t=1}^{T-1} y_t^2 \hat{b}(t)\right)^2 \left(\sum_{s=1}^{T-1} y_s^2\right) \\
&= \left(\sum_{s=1}^{T-1} y_s^2\right)^{-1} \left(\sum_{t=1}^{T-1} y_t^2 \hat{b}(t)\right)^2 \\
&\leq \left(\sum_{s=1}^{T-1} y_s^2\right)^{-1} \left(\sum_{t=1}^{T-1} \hat{b}^2(t) y_t^2\right) \left(\sum_{t=1}^{T-1} y_t^2\right) \\
&= \sum_{t=1}^{T-1} \hat{b}^2(t) y_t^2,
\end{aligned}$$

we have that

$$\overline{\mathrm{F}}_{y \to x} \leq \widehat{\mathrm{F}}_{y \to x},$$

and the equality holds only when $\sum_{t=1}^{T-1} \hat{b}^2(t) y_t^2 = c \sum_{t=1}^{T-1} y_t^2$ with a constant c. Assuming that the local estimation gives the exact value of the parameter, we can see that the causality established by ignoring the time-varying property of the parameters is upper bounded by the averaged causality among local causalities (i.e. the causality detected by each sliding window). Therefore, caution must be made when we detect the causality for a whole time series, since the causality may exist in some time windows. Therefore, if the whole time series can be divided into N time windows according to the task paradigm of the fMRI experiment, we want to estimate the effective connectivity at each time window instead of the whole time series.

In our case, to deal with the signal-dependent noise (SDN) observed in the BOLD signal[1], the proposed time-varying method is based on the Granger causality with signal-dependent noise (GCSDN) model, which has been proved to be useful in detecting the effective connectivity previously ([2] and [3]). Suppose we have two time series, $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$. Let p and q be the model orders, $\{\boldsymbol{A}_i(i = 1, \cdots, p), \boldsymbol{B}_{xy,j}, \boldsymbol{B}_{yx,j}(j = 1, \cdots, q), \boldsymbol{C}_{xy}, \boldsymbol{C}_{yx}\}$ be the model coefficient matrices, and $\boldsymbol{u}_{xy,t}, \boldsymbol{v}_{yx,t}$ be the Gaussian white noise. The model can be defined as follows:

$$\begin{pmatrix} \boldsymbol{x}_t \\ \boldsymbol{y}_t \end{pmatrix} = \sum_{i=1}^{p} \begin{pmatrix} \boldsymbol{A}_{xx,i} & \boldsymbol{A}_{xy,i} \\ \boldsymbol{A}_{yx,i} & \boldsymbol{A}_{yy,i} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_{t-i} \\ \boldsymbol{y}_{t-i} \end{pmatrix} + \begin{pmatrix} \boldsymbol{r}_{xy,t} \\ \boldsymbol{r}_{yx,t} \end{pmatrix}, \tag{S1}$$

where

$$\begin{aligned} &\boldsymbol{r}_{xy,t} = \boldsymbol{H}_{xy,t}^{1/2} \boldsymbol{u}_{xy,t}, \boldsymbol{H}_{xy,t} = \boldsymbol{C}_{xy}^{'} \boldsymbol{C}_{xy} + \sum_{j=1}^{q} \boldsymbol{B}_{xy,j}^{'} \begin{pmatrix} \boldsymbol{x}_{t-j} \\ \boldsymbol{y}_{t-j} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_{t-j}' & \boldsymbol{y}_{t-j}' \end{pmatrix} \boldsymbol{B}_{xy,j}, \\ &\boldsymbol{H}_{yx,t} = \boldsymbol{C}_{yx}^{'} \boldsymbol{C}_{yx} + \sum_{j=1}^{q} \boldsymbol{B}_{yx,j}^{'} \begin{pmatrix} \boldsymbol{x}_{t-j} \\ \boldsymbol{y}_{t-j} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_{t-j}' & \boldsymbol{y}_{t-j}' \end{pmatrix} \boldsymbol{B}_{yx,j}, \boldsymbol{r}_{yx,t} = \boldsymbol{H}_{yx,t}^{1/2} \boldsymbol{v}_{yx,t}. \end{aligned} \tag{S2}$$

For the time-varying causality, we first divide the whole time series into N time windows. At each window, the model (Eq. S1-S2) can be fit based on the direct observations and the indirect observations. Assuming the system evolved smoothly from one time window to another, the observed time series in the current window are considered as the direct observations of the model for this window, while the observed time-series data in the other windows are the indirection observations. We need the indirect observations here, since the number of data points in a given window is often too small to make reliable estimation of the model. This is especially true for the self-paced task-fMRI experiments. In a self-paced paradigm, the number of scans collected in one round can be as small as one or two. Here, we propose to make use of the indirect observations by calculating the likelihood function at the i_0^{th} time window as a weighted average among all time windows. The weight for the observations in the i^{th} window is defined as follows:

$$w_{i,i_0} = \frac{K\left(\frac{i-i_0}{h}\right)}{\sum_{j=1}^{N} K\left(\frac{j-i_0}{h}\right)}. \tag{S3}$$

In the current paper, we used the Gaussian kernel, among many other choices [4]. A larger h is recommended, if the time window given by the task paradigm is short, and a smaller h is preferred, otherwise. Here, we set $h = 0.17$ for the numerical simulation and $h = 4$ for the fMRI data analysis. Define

$$\boldsymbol{\theta}^{i_0} = \left\{\boldsymbol{A}_k^{i_0}, k = 1, \cdots, p, \boldsymbol{B}_{xy,j}^{i_0}, \boldsymbol{B}_{yx,j}^{i_0}, j = 1, \cdots, q, \boldsymbol{C}_{xy}^{i_0}, \boldsymbol{C}_{yx}^{i_0}\right\}, \tag{S4}$$

the objective function (i.e.,the log-likelihood function) can be rewritten for the time window as follows

$$\mathrm{LLF}_{i_0}(\boldsymbol{\theta}) = \sum_{t=p\vee q}^{T} \sum_{i=1}^{N} \chi_{i^{th} \text{ window}}(t) w_{i,i_0} l_{\boldsymbol{\theta}}(t), \tag{S5}$$

where

$$l_{\boldsymbol{\theta}}(t) = -\frac{1}{2}\ln|\boldsymbol{H}_t| - \frac{1}{2}\boldsymbol{r}_t'\boldsymbol{H}_t^{-1}\boldsymbol{r}_t, t = p \vee q, \cdots, T. \tag{S6}$$

Consider the stationary conditions of the model (Eq. S1)[1], the model parameters at the i_0^{th} window can be estimated by solving the following constrained optimization problem.

$$\widehat{\boldsymbol{\theta}}^{i_0} = \arg\max_{\boldsymbol{\theta}} \text{LLF}_{i_0}(\boldsymbol{\theta}), \text{ s.t. stability conditions hold.} \tag{S7}$$

Taking the causality $\boldsymbol{y} \to \boldsymbol{x}$ at the i_0^{th} time window as example, we consider the following two models for $\boldsymbol{x}_t$,

$$\begin{aligned} \boldsymbol{x}_t &= \sum_{k=1}^{p} \boldsymbol{A}_{x,k}^{i_0}\boldsymbol{x}_{t-i} + \left(\boldsymbol{H}_{xx,t}^{i_0}\right)^{1/2}\boldsymbol{u}_{xx,t}, \\ \boldsymbol{H}_{xx,t}^{i_0} &= \boldsymbol{C}_{xx}^{i_0\,\prime}\boldsymbol{C}_{xx}^{i_0} + \sum_{j=1}^{q} \boldsymbol{B}_{xx,j}^{i_0\,\prime}\boldsymbol{x}_{t-j}\boldsymbol{x}_{t-j}'\boldsymbol{B}_{xx,j}^{i_0}, \end{aligned} \tag{S8}$$

and

$$\begin{aligned} \boldsymbol{x}_t &= \sum_{k=1}^{p} \boldsymbol{A}_{xy,k}^{i_0}\boldsymbol{x}_{t-k} + \sum_{k=1}^{p} \boldsymbol{A}_{yx,k}^{i_0}\boldsymbol{y}_{t-k} + \left(\boldsymbol{H}_{xy,t}^{i_0}\right)^{1/2}\boldsymbol{u}_{xy,t}, \\ \boldsymbol{H}_{xy,t}^{i_0} &= \boldsymbol{C}_{xy}^{i_0\,\prime}\boldsymbol{C}_{xy}^{i_0} + \sum_{j=1}^{q} \boldsymbol{B}_{xy,j}^{i_0\,\prime}\begin{pmatrix}\boldsymbol{x}_{t-j}\\ \boldsymbol{y}_{t-j}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_{t-j}' & \boldsymbol{y}_{t-j}'\end{pmatrix}\boldsymbol{B}_{xy,j}^{i_0}. \end{aligned} \tag{S9}$$

At each time window, the model coefficients are assumed to be constants. Define

$$\begin{aligned} \boldsymbol{\theta}_{\text{restricted}}^{i_0} &= \left\{\boldsymbol{A}_{x,k}^{i_0}, k = 1, \cdots, p, \boldsymbol{B}_{xx,j}^{i_0}, j = 1, \cdots, q, \boldsymbol{C}_{xx}^{i_0}\right\}, \\ \boldsymbol{\theta}_{\text{full}}^{i_0} &= \left\{\boldsymbol{A}_{xy,k}^{i_0}, \boldsymbol{A}_{yx,k}^{i_0}, k = 1, \cdots, p, \boldsymbol{B}_{xy,j}^{i_0}, j = 1, \cdots, q, \boldsymbol{C}_{xy}^{i_0}\right\}, \end{aligned}$$

We can obtained the estimated $\widehat{\boldsymbol{\theta}}_{\text{restricted}}^{i_0}$ and $\widehat{\boldsymbol{\theta}}_{\text{full}}^{i_0}$ by solving the constrained optimization problem defined in (S7). Now, the causal influence from $\boldsymbol{y}$ to $\boldsymbol{x}$ can be measured by the likelihood ratio given by the above two prediction models for $\boldsymbol{x}_t$,

$$\text{IF}_{\boldsymbol{y}\to\boldsymbol{x}}^{i_0} = \frac{\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_{\text{restricted}}^{i_0} \mid \{\boldsymbol{x}_t\}_{t=1}^{T}\right)}{\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_{\text{full}}^{i_0} \mid \{\boldsymbol{x}_t\}_{t=1}^{T}, \{\boldsymbol{y}_t\}_{t=1}^{T}\right)},$$

where the $\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_{\text{restricted}}^{i_0} \mid \{\boldsymbol{x}_t\}_{t=1}^{T}\right)$ is the likelihood function for the restricted model (Eq. S8), and the $\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_{\text{full}}^{i_0} \mid \{\boldsymbol{x}_t\}_{t=1}^{T}, \{\boldsymbol{y}_t\}_{t=1}^{T}\right)$ is the likelihood function for the full model (Eq. S9). Therefore, the likelihood ratio test can be used for causal inference,

$$lr_{\boldsymbol{y}\to\boldsymbol{x}} = -2\left[\log\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_{\text{restricted}} \mid \{\boldsymbol{x}_t\}_{t=1}^{T}\right) - \log\mathcal{L}\left(\widehat{\boldsymbol{\theta}}_{\text{full}} \mid \{\boldsymbol{x}_t\}_{t=1}^{T}, \{\boldsymbol{y}_t\}_{t=1}^{T}\right)\right]. \tag{S10}$$

The test statistic $lr_{\boldsymbol{y}\to\boldsymbol{x}}$ is approximately chi-squared distributed with the degrees of freedom $df_{\text{full}} - df_{\text{restricted}}$, where df_{full} and $df_{\text{restricted}}$ are the numbers of free parameters of the full model (Eq. S9) and the restricted model (Eq. S8), respectively. Given the TR was 2 s, we used $p = q = 1$ in this analysis

following the literature [2], [5], and [6]. A Matlab toolbox of this algorithm is also available at https://github.com/qluo2018/GCSDN.

2. Method S2. Numerical Simulations of time-varying systems

To illustrate the performance of the time-varying GCSDN model, we conducted two simulation studies. Suppose we had two brain regions, X and Y. The effective connectivity between them could be described by the time series model, either with the assumption of the constant noise level or the signal-dependent noise level. The observed time-series data of the activities for these two regions could be simulated by the models, and the information flow ($\overline{\mathrm{IF}}$) could be detected by the classic GC, GCSDN or time-varying GCSDN model.

2.1. Time series with a Gaussian white noise

The time-series data with 1000 time steps were generated by the following model, where x_t, y_t were the time series of one-dimensional, representing the activities of brain region X and Y, respectively. $u_{xy,t}$, $v_{xy,t}$ were the Gaussian white noise with variance 0.5,

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} A_{11}(t) & A_{12}(t) \\ A_{21}(t) & A_{22}(t) \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} u_{xy,t} \\ u_{yx,t} \end{pmatrix}. \tag{S1}$$

The time-varying causality was modeled by the corresponding coefficients

$$A_{11}(t) = 0.1, \quad A_{12}(t) = 0.4\left(\frac{t - t_1}{1000 - t_1}\right),$$
$$A_{21}(t) = 0.4\left(1 - \frac{t}{t_2}\right), \quad A_{12}(t) = 0.1\sqrt{2}.$$

Setting $t_1 = 500$ in the model (Eq. S1), we had a strong effective connectivity from Y to X through the coefficient A_{12} at the beginning, and this excitatory influence weakened to be undetectable as the positive coefficient A_{12} decreased linearly to zero. As the coefficient A_{12} became negative, an inhibitory effect strengthened as the absolute value of the coefficient A_{12} increased. Similarly, the effective connectivity from X to Y evolved in an opposite pattern from inhibitory to excitatory by setting $t_2 = 500$.

2.2. Time-varying model with a signal-dependent noise

The time-series data with 1000 time steps were generated by the following model for two time series, representing the activities of brain region X and Y, respectively.

$$\begin{aligned}
\begin{pmatrix} x_t \\ y_t \end{pmatrix} &= \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1\sqrt{2} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} r_{xy,t} \\ r_{yx,t} \end{pmatrix}, \\
\begin{pmatrix} r_{xy,t} \\ r_{yx,t} \end{pmatrix} &= \begin{pmatrix} H_{xy,t}^{\frac{1}{2}} & 0 \\ 0 & H_{yx,t}^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} u_{xy,t} \\ u_{yx,t} \end{pmatrix}, \\
H_{xy,t} &= 1 + \begin{pmatrix} B_{xx}(t) & B_{xy}(t) \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} \begin{pmatrix} x'_{t-1} & y'_{t-1} \end{pmatrix} \begin{pmatrix} B_{xx}(t) \\ B_{xy}(t) \end{pmatrix}, \\
H_{yx,t} &= 1 + \begin{pmatrix} B_{yx}(t) & B_{yy}(t) \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} \begin{pmatrix} x'_{t-1} & y'_{t-1} \end{pmatrix} \begin{pmatrix} B_{yx}(t) \\ B_{yy}(t) \end{pmatrix}.
\end{aligned} \tag{S2}$$

The time-varying causality was modeled by the corresponding coefficients

$$B_{xx}(t) = \sqrt{0.5}, \quad B_{xy,t}(t) = \begin{cases} \sqrt{0.6}\left(1 - \frac{t}{t_1}\right) & t \le t_1 \\ 0 & t > t_1 \end{cases},$$

$$B_{yx,t}(t) = \begin{cases} 0 & t < t_2 \\ \sqrt{0.6}\left(\frac{t-t_2}{1000-t_2}\right) & t \ge t_2 \end{cases}, \quad B_{yy}(t) = \sqrt{0.5}.$$

A significant non-zero value of the coefficient B_{xy} lead to an effective connectivity from Y to X, while the coefficient B_{yx} indicated an effective connectivity from X to Y. Setting $t_1 = 600$ and $t_2 = 400$ in model (Eq. S2), the strong influence from Y to X at the beginning of the simulation weakened linearly to zero at 600s. At 400s, an influence from X to Y began to increase linearly till the end of the simulation. With the proposed time-varying algorithm, we expected to detect an effective connectivity from Y to X, but not from X to Y at the beginning, while a strong effective connectivity from X to Y, but not from Y to X at the end of the simulation.

2.3. Simulation results for the time-varying Granger causality with Gaussian white noise

Assuming the effective connectivity between two brain regions X and Y by the simulation model (Eq. S1; Fig. S1A), we employed the classic GC model to infer the information flow between X and Y by using the whole time series data. In 1000 repeats, only 0.5% and 0.8% significant information flow was detected for $X \to Y$ and $Y \to X$, respectively (Fig. S1B). We also applied the GCSDN model to detect the information flow on the whole time series data, and found few significant information flow on both direction, only 13.2% significant information flow was detected for $X \to Y$ and 16.1% for $Y \to X$ (Fig. S1B). When we divided

the simulated time series data into five time windows, both methods identified significant information flows as specified by the simulation model (Eq. S1). At the first and the last two time windows, both the classic GC model and the GCSDN model could detect more than 94% of the significant information flows in the 1000 repeated simulations. In the 3rd time window, where no significant interaction was specified by the model (Eq. S1), the false positive detection was less than 1% (Fig. S1B). Applying the proposed the time-varying GCSDN method to these five time windows, similar performance was achieved (Fig. S1D). With the time-varying GCSDN method, we found strong information flow between X and Y in both directions at the beginning, and this information flow weakened as the causal coefficients in the model (Eq. S1) decreased. At the 3rd time window, these two time series stopped to exchange any information. As the absolute value of the causal coefficients in the model (Eq. S1) increased from the third time window, stronger information flow was detected again (Fig. S1D). We also estimated the coefficients of the AR-BEKK model for each window, and found that the estimated coefficients were equal to the average of the time-varying parameters at each window (Fig. S1C). These results suggested that the time invariant models were no longer applicable when time-varying property was significant.

2.4. Simulation results for the time-varying Granger causality with signal-dependent noise

We repeated the simulation of the model (Eq. S2) for 1000 times (Fig. S2A), and found that the classic GC method failed to detect any significant information flow in the presence of the signal-dependent noise, no matter we used the whole time series data or divided it into five time windows (Fig. S2B). Applying the GCSDN method to the whole time series, we detected the significant information flows between X and Y in both directions (Fig. S2B). In five time windows, the time-varying GCSDN method accurately identified the significant information flows from Y to X at the first two windows and from X to Y at the last two windows, but no significant information flow at the 3rd time window (Fig. S2D). Compared with the GCSDN method, the proposed time-varying GCSDN revealed more details of the time-varying system, and estimated the time-varying coefficient accurately at each time window (Fig. S2C), which may be particularly useful when we want to investigate the evolving pattern of interaction between the key brain regions the underlying time-varying behavior.

3. Method S3. Expectation-Maximum algorithm to identify the behavioral window

To learn the parameters for the HMM, we applied the Expectation-Maximum algorithm. With the same notations used as described in the main text, the iterative estimation of parameters is as follows:

E-steps: for all time-state pairs (t, k),

1) Recursively compute the forward probabilities

$$\alpha_{t,k}(i) = \mathbb{P}\left(\boldsymbol{O}_{1:t,k}, q_t = q_i | \boldsymbol{\lambda}_t\right),$$

and the backward probabilities

$$\beta_{t,k}(i) = \mathbb{P}\left(\boldsymbol{O}_{T:-1:t+1,k} | q_t = q_i, \boldsymbol{\lambda}_t\right).$$

2) Compute the probabilities of the state occupation

$$\mathbb{P}\left(q_t = q_i, \boldsymbol{O}_{1:T,k} | \boldsymbol{\lambda}_t\right) = \alpha_{t,k}(i)\beta_{t,k}(i),$$

$$\mathbb{P}\left(q_t = q_i, q_{t+1} = q_j, \boldsymbol{O}_{1:T,k} | \boldsymbol{\lambda}_t\right) = \alpha_{t,k}(i)\beta_{t,k}(i)a_{ij}^t \mathbb{P}\left(\boldsymbol{O}_{t+1,k} | q_{t+1} = q_j, \boldsymbol{\lambda}_t\right).$$

M-steps:

Based on the estimated probabilities of the state occupation, we can re-estimate the HMM parameters

$$a_{ij,t+1} = \frac{\sum_{k=1}^{N}\sum_{t=1}^{T-1}\mathbb{P}\left(q_t = q_i, q_{t+1} = q_j, \boldsymbol{O}_{1:T,k} | \boldsymbol{\lambda}_t\right)}{\sum_{k=1}^{N}\sum_{t=1}^{T-1}\mathbb{P}\left(q_t = q_i, \boldsymbol{O}_{1:T,k} | \boldsymbol{\lambda}_t\right)},$$

$$\pi_{i,t+1} = \frac{\sum_{k=1}^{N}\mathbb{P}\left(q_1 = q_i, \boldsymbol{O}_{1:T,k} | \boldsymbol{\lambda}_t\right)}{\sum_{k=1}^{N}\mathbb{P}\left(\boldsymbol{O}_{1:T,k} | \boldsymbol{\lambda}_t\right)},$$

$$\boldsymbol{\mu}_{i,t+1} = \frac{\sum_{k=1}^{N}\sum_{t=1}^{T}\mathbb{P}\left(q_t = q_i, \boldsymbol{O}_{1:T,k} | \boldsymbol{\lambda}_t\right) \cdot \boldsymbol{o}_{t,k}}{\sum_{k=1}^{N}\sum_{t=1}^{T}\mathbb{P}\left(q_t = q_i, \boldsymbol{O}_{1:T,k} | \boldsymbol{\lambda}_t\right)},$$

$$\boldsymbol{\Sigma}_{i,t+1} = \frac{\sum_{k=1}^{N}\sum_{t=1}^{T}\mathbb{P}\left(q_t = q_i, \boldsymbol{O}_{1:T,k} | \boldsymbol{\lambda}_t\right) \cdot (\boldsymbol{o}_{t,k} - \boldsymbol{\mu}_i^{t+1})(\boldsymbol{o}_{t,k} - \boldsymbol{\mu}_i^{t+1})'}{\sum_{k=1}^{N}\sum_{t=1}^{T}\mathbb{P}\left(q_t = q_i, \boldsymbol{O}_{1:T,k} | \boldsymbol{\lambda}_t\right)}.$$

where N is the number of subjects and T is the length of the time-series observation. Here, we had 76 subjects with 59 observations for each.

We calculated the information revelation (IR) and R^2 at the identified behavioral windows and plotted the clustering results in the two-dimensional feature space (the x-axis represents information revelation (IR) and the y-axis represents R^2). However, the R^2 calculated at the incremental window $[25, 35]$ of subject 29 was too small, making the incremental window class and the conservative window class overlapped (Fig. S4D). Therefore, we did not consider this behavioral window as an incremental window in the main analysis, and the main findings remained to be the same when retaining this behavioral window (Fig. S4A, B, C).

4. Method S4. Estimate the hemodynamic response function with the GLM model

As the BOLD signal could be driven by both the task design and the information exchange between regions, it was necessary to regress out the activation of the whole brain from the BOLD signal before calculating the effective connectivity. Therefore, we estimated the hemodynamic response function of the brain regions of interest by the means of the GLM model[7]. Given the matrix formed by the stacking of k basis elements $\boldsymbol{B} = [\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k]$ (here we used the 3-HRF basis, i.e., hrf (with time and dispersion derivatives) in the SPM8), we first convolved the basis elements with the event-train of each condition (i.e., trail onset, thinking, choice making), then down-sampled the signal to the same sampling rate as the BOLD signal. Next, the designed matrix $\boldsymbol{X}_B$ was formed by the above regressors and a matrix of nuisance parameters, such as the trends and the motion parameters (i.e., three translations and three rotations). The estimated vector was then given by $\widehat{\boldsymbol{\beta}}_{\boldsymbol{B}} = \boldsymbol{X}_B^{\dagger}\boldsymbol{y}$. We used the estimated vector $\widehat{\boldsymbol{\beta}}_B$ to calculate the hrf for each condition. To ensure that the residuals were white noise, the BOLD signal and the design matrix were pre-whitened before the estimation[8].

5. Method S5. Accounting for the possible confounding effect of the hemodynamic response function on the findings

Instead of the classic Granger causality, the model applied to the BOLD signal was the Granger causality with signal-dependent noise (GCSDN) in our case. Therefore, it was necessary to examine whether the information flow detected by the GCSDN model could still reflect the correct underlying effective connectivity with the possible confounding effect of the hemodynamic response function. We examined the effect of the hemodynamic response function in the following aspects.

5.1. Compare the HRF delay between two brain regions

The delay between the hemodynamic response and the neuronal activity was estimated by the time required for the HRF to reach its peak value from the onset. Comparing such delays among three ROIs with the pairwise t-test, we found no significant difference in the condition of thinking (Fig. S10A-C). Therefore, the regional variation of HRF would not be a significant problem in the current study.

5.2. Numerical simulations of the HRF delay

Since the sensitivity of the analyses for effective connectivity depends on the neuronal transmission delay between the source and the effect regions and their relative HRF delay (i.e., the time required for the HRF to reach its peak value from the onset.)[9], we assessed the performance of the tvGCSDN model at different levels of HRF delay and neuronal transmission delay.

Consider two brain regions X and Y, the HRF delay of X was longer than that of Y and the relative HRF delay varied in 0-1s. The underlying effective connectivity between them could be described by the time series model. Suppose X was the source region and Y was the effect region, the model was defined as follows,

$$\begin{aligned} &\begin{cases} x_t = A_{xx}x_{t-l} + H_{xy,t}^{\frac{1}{2}}u_{xy,t} \\ y_t = A_{yx}x_{t-l} + A_{yy}y_{t-l} + H_{yx,t}^{\frac{1}{2}}u_{yx,t}, \end{cases} \\ &H_{xy,t} = 0.1 + 0.01x_{t-l}^2, \\ &H_{yx,t} = 0.1 + (B_{yx}x_{t-l} + 0.01y_{t-l})^2, \end{aligned} \tag{S1}$$

where l represented the neuronal transmission delay. Single-cell recordings in monkeys had shown that the median latencies increased by approximately 20 ms from one brain region in the visual hierarchy to the next [10]. Considering that the size of the human brain increased compared with that of the monkeys, the neuronal delay could be longer. Therefore, the neuronal delay l was set to vary from 40ms to 140ms. Following the settings of simulation in the literature[9], the autocorrelation parameter A_{xx} and A_{yy} was set to 0.9, and

the influence from X to Y was set to 0.5 for A_{xy} in the autoregressive model and $\sqrt{0.005}$ for B_{yx} in the signal-dependent noise respectively.

We first generated two time series x_t and y_t by the Granger causality model with signal-dependent noise (Eq. S1), both signals were simulated for 300,000 time steps of 1 ms (300 s). Next, the signal was convolved with the hemodynamic response functions of the corresponding region, which was generated by the SPM8. Gaussian white noise was then added to the signal, representing the physiological noise in the BOLD response. Subsequently, the signal was down-sampled to the same rate as the BOLD signal (i.e., 2 HZ), and the Gaussian noise was again added to represent the acquisition noise. The signals were normalized to zero mean and unit-variance after each step above and the total amount of added noise was 20%.

We repeated the above experiment 100 times. The GCSDN model was applied to calculate the information flow ($\overline{\mathrm{IF}}$) of the simulated time series in both directions. We detected the dominant direction of the information flow by the difference in the information flow $r_{diff} = -\frac{1}{2}\left(\overline{\mathrm{IF}}_{X\to Y} - \overline{\mathrm{IF}}_{Y\to X}\right)$, where $\overline{\mathrm{IF}}_{X\to Y}$ and $\overline{\mathrm{IF}}_{Y\to X}$ were two chi-squared distribution statistics with the same degrees of freedom. Therefore, the distribution function of r_{diff} is

$$T_m(x) = \frac{1}{2^m\sqrt{\pi}\Gamma(m+\frac{1}{2})}x^m K_m(x),$$

where $K_m(x)$ is a modified Bessel function, $\Gamma(\cdot)$ is a Gamma function, and $m = d_x d_y - \frac{1}{2}$ [11]. A table for the two-sided one and five percent quantiles of this distribution can be found in [11]. In the current paper, we took $d_x = d_y = 1$. Therefore, 4.61 represents a significant level of 0.01, if $r_{diff} < -4.61$, we detected a effective connectivity from X to Y, if $r_{diff} > 4.61$, we detected a effective connectivity from Y to X. Otherwise, no significant effective connectivity could be detected. We counted the times of $r_{diff} < -4.61, \left(\overline{\mathrm{IF}}_{Y\to X} < \overline{\mathrm{IF}}_{X\to Y}\right)$ and $r_{diff} > 4.61, \left(\overline{\mathrm{IF}}_{X\to Y} < \overline{\mathrm{IF}}_{Y\to X}\right)$, then calculated the proportion of inverted, correct and non-significant results.

It could be found that when the underlying effective connectivity was from X to Y, where the source region had longer HRF delay than the target region, though the number of detected non-significant results increased, there were few inverted results(Fig. S11A). The proportion of inverted results decreased when neuronal delay increased. Assuming no difference in HRF delay, the proportion of detected inverted results was less than 1% in all conditions of neuronal delays. Together with the findings of no significant regional variation in the HRF delays of 3 ROIs (Method S4), we believe that the dynamic information flow estimated by the tvGCSDN was a reliable measurement of the strength of the effective connectivity.

We also simulated the time series with Y-to-X underlying effective connectivity 100 times, the effect region had a longer HRF delay than the source region in this case. GCSDN was applied to the time series to detect the information flow, but no inverted result was found (Fig. S11B).

Table S1: Characteristics of the behavioral windows defined by the hidden Markov model.

Incr. is short for incremental window, Cons. for conservative window and Strat. for strategic window. The characteristics were reported by mean ± standard deviation.

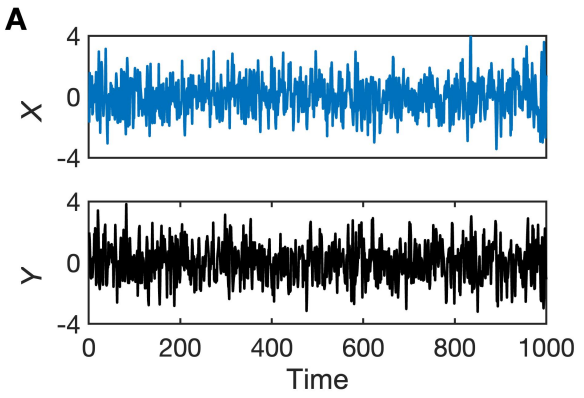
	Incr.	Cons.	Strat	p-value
Length	27.14 ± 18.36	20.5 ± 18.27	29.89 ± 19.75	0.1453
Starting time	15.05 ± 16.14	18.18 ± 16.96	22.64 ± 16.68	0.107
The values of the virtual items	5.54 ± 0.71	5.76 ± 1.05	5.64 ± 0.78	0.45

Table S2: The correlation between demographics of the buyers and the number of different types of behavioral windows. SES - social economic status; IQ - intelligence quotient. Each measurement was reported by Pearson correlation coefficient/p-value.

Correlation / p-value	Incr.	Cons.	Strat.
Sex	**-0.39/0.0004**	-0.073/0.53	0.19/0.094
Age	**0.36/0.0015**	-0.21/0.074	-0.27/0.019
SES	-0.15/0.21	0.23/0.051	0.03/0.80
Earning	-0.26/0.023	0.16/0.16	0.06/0.60
IQ	-0.48/0.0089	0.22/0.25	0.26/0.18

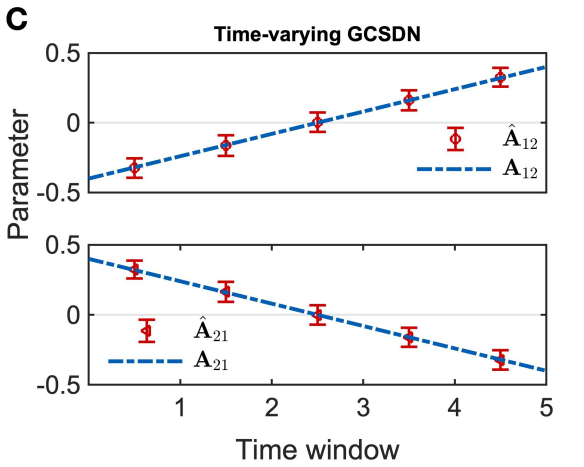
Table S3: The identification of strategies when different intervals were chosen to calculate the observations. The information revelation(IR) was reported by mean ± standard deviation.

length of interval	DBI	information revelation (IR)		
		Incr.	Cons.	Strat.
5	0.5131	0.49±0.19	0.07±0.15	-0.61±0.22
7	0.5824	0.48±0.19	0.13±0.11	-0.59±0.26
9	0.7748	0.48±0.2	0.13±0.16	-0.62±0.23
7(focus on past)	0.7986	0.5±0.2	0.17±0.16	-0.58±0.28



B

Granger	Classic GC		GCSDN	
Causality	X → Y	Y → X	X → Y	Y → X
Whole TS	0.5%	0.8%	13.2%	16.1%
Window 1	97.8%	96.8%	95%	94.2%
Window 2	30.5%	32.2%	22.7%	22.7%
Window 3	0.7%	0.7%	0.6%	0.5%
Window 4	30%	31.1%	20.6%	21.9%
Window 5	96.4%	98%	93.6%	95.7%



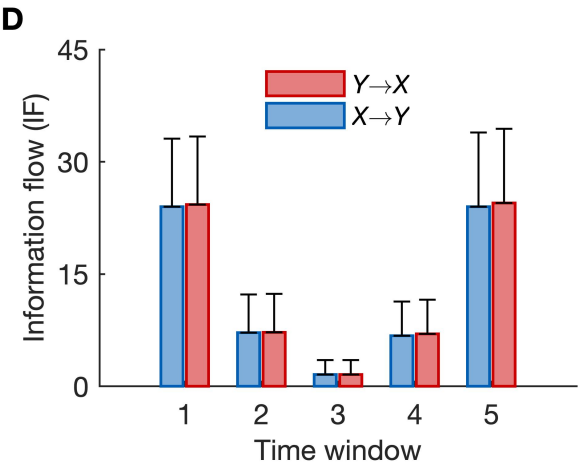
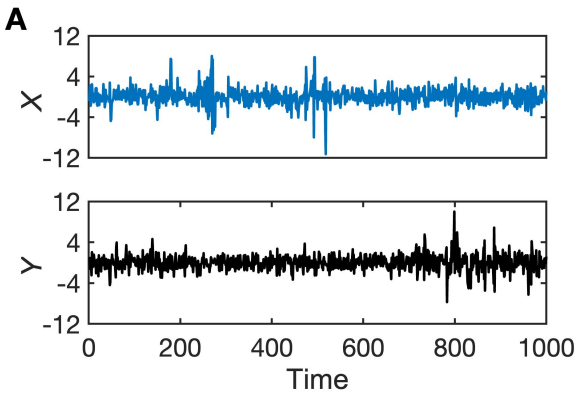
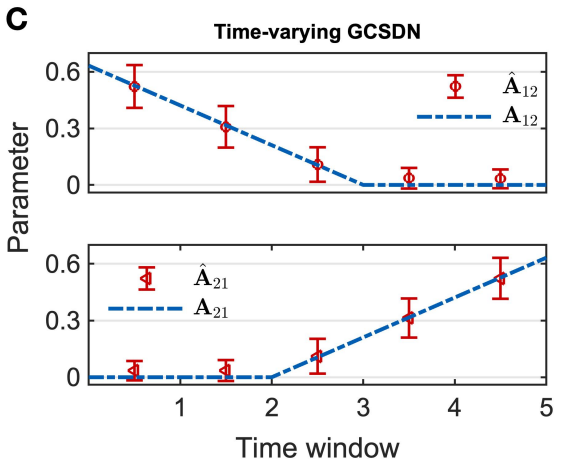


Figure S1: Detection of information flow in a system with time-varying coefficient and Gaussian white noise.
(A) The example of the simulated time series data by model (Eq. S1). (B) Results of the information flow inferred by both the classic GC and the GCSDN methods on the simulation data. These models were applied to both the whole time series and the 5 time windows, the rates of the significant detections in the 1000 repeats of the simulation were reported. (C) The time-varying parameter estimated by Time-varying GCSDN at each time window. Red points and error bars represent the mean and standard deviation, respectively. (D) The information flow between X and Y that measured by the likelihood ratio of Time-varying GCSDN method at each time window, error bars represent the standard deviation.



B

Granger	Classic GC		GCSDN	
Causality	X → Y	Y → X	X → Y	Y → X
Whole TS	1.9%	2.6%	78.1%	78.2%
Window 1	0.2%	5.9%	0.5%	92.8%
Window 2	0.9%	2.8%	0.8%	54.7%
Window 3	0.9%	1%	4.3%	3.1%
Window 4	1.6%	0.4%	53.8%	0.7%
Window 5	4.1%	0.8%	93.7%	0.3%



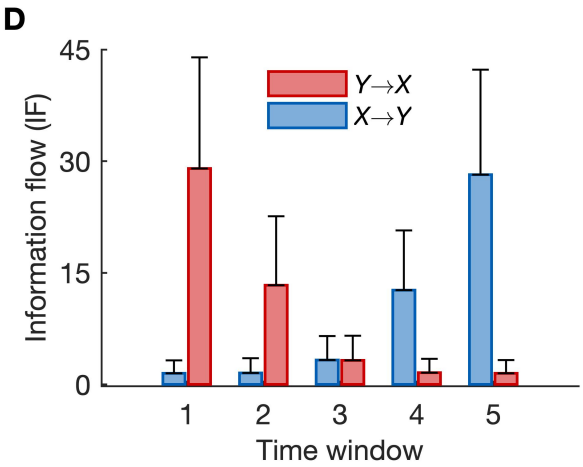


Figure S2: Detection of information flow in a system with time-varying coefficient and signal-dependent noise. (A) The example of the simulated time series data by model (Eq. S2). (B) Results of the information flow inferred by both the classic GC and the GCSDN methods on the simulation data. These models were applied to both the whole time series and the 5 time windows, the rates of the significant detections in the 1000 repeats of the simulation were reported. (C) The time-varying parameter estimated by Time-varying GCSDN at each time window. Red points and error bars represent the mean and standard deviation, respectively. (D) The information flow between X and Y that measured by the likelihood ratio of Time-varying GCSDN method at each time window, error bars represent the standard deviation.

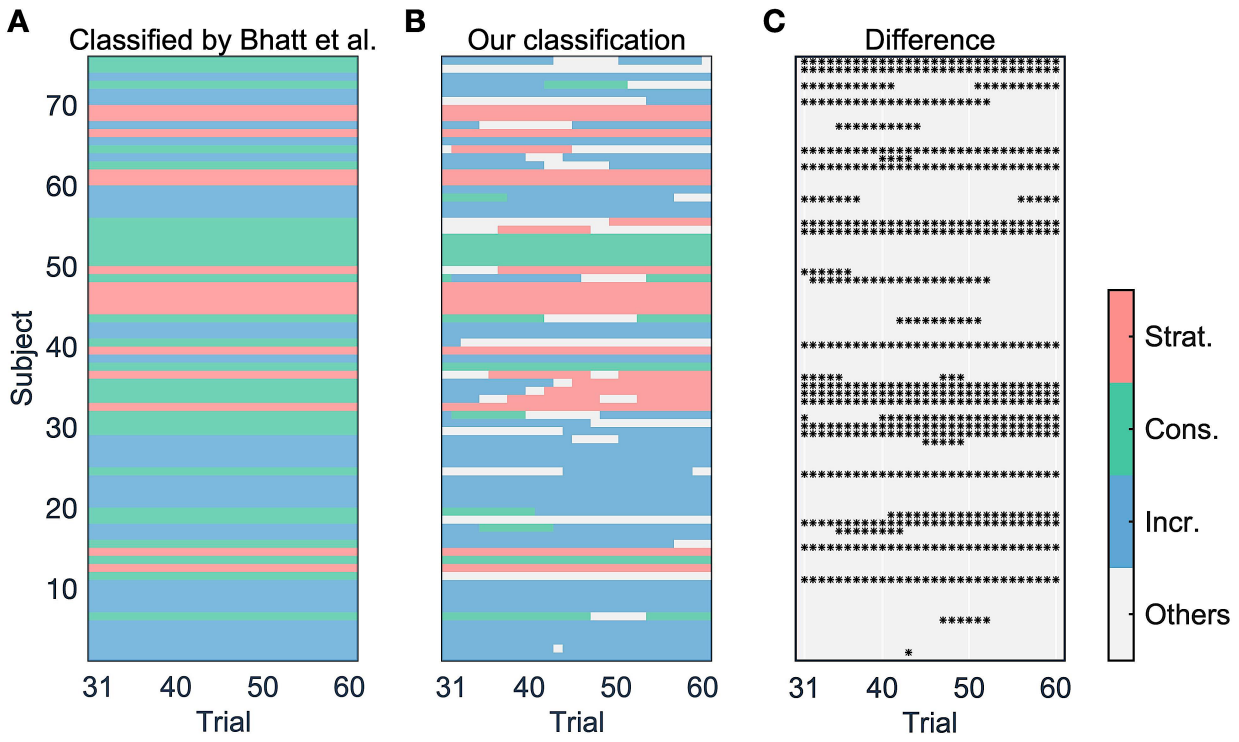


Figure S3: Comparison of behavioral classification between time-invariant and time-varying approaches.
(A) Classification of behavioral windows by the time-invariant approach in Bhatt et al. 2010. (B) Classification of behavioral windows by the time-varying approach in the current study. (C) Comparison of the grouping of the behavioral windows among three groups defined by Bhatt et al . 2010 and the current study.

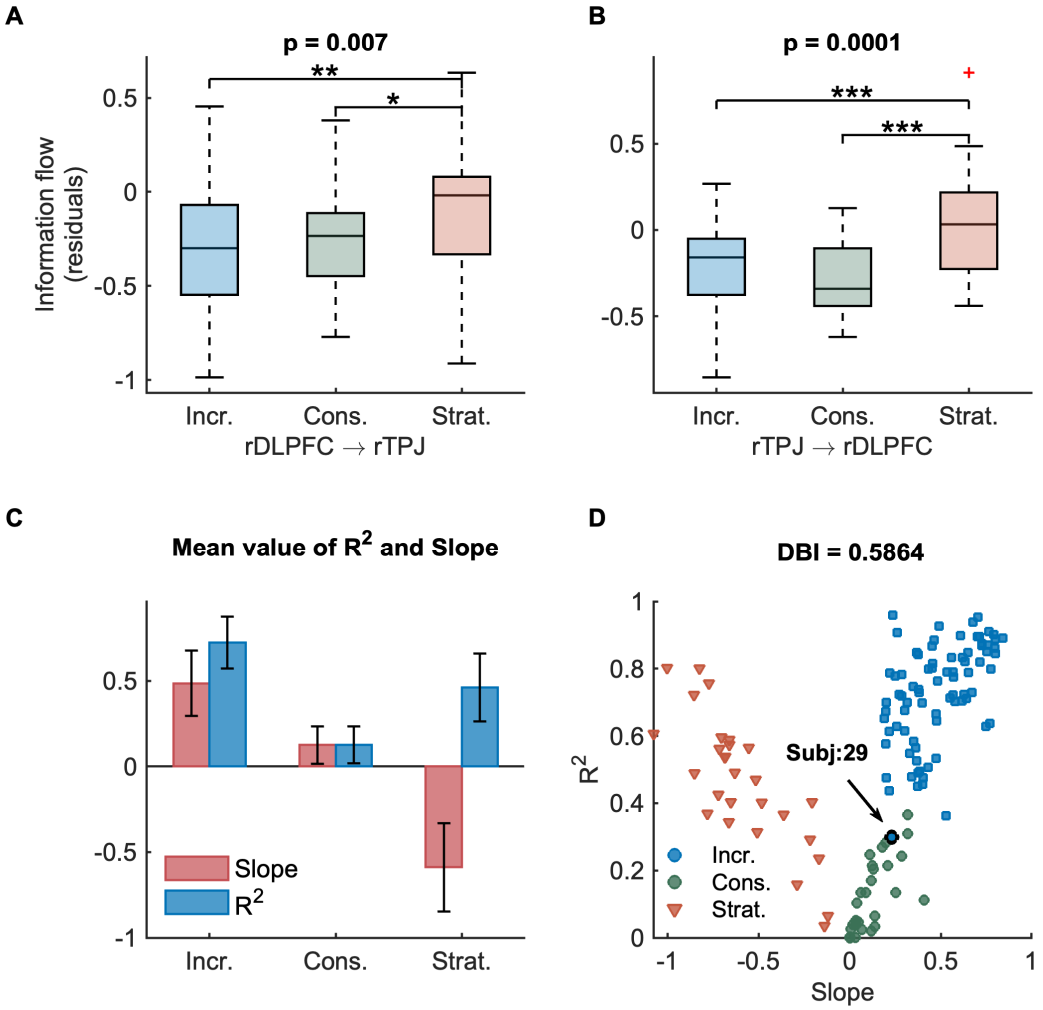


Figure S4: Main findings remained when the ambiguous behavior window of Subject 29 was retained.

Comparison of the mean information flow among three types of behavioral windows: (A) from rDLPFC to rTPJ; (B) from rTPJ to rDLPFC ((A) $^*p = 0.0369$, $^{**}p = 0.0027$. (B) $^{***}p = 0.0006$, $^{***}p = 0.0003$.).

(C) The mean and std of the information revelation (IR) and R^2 calculated at the identified behavioral windows, error bars represent the standard deviation. (D) The clustering result of behavioral windows in the two-dimensional feature space, where the x-axis represents information revelation (IR) and the y-axis represents R^2.

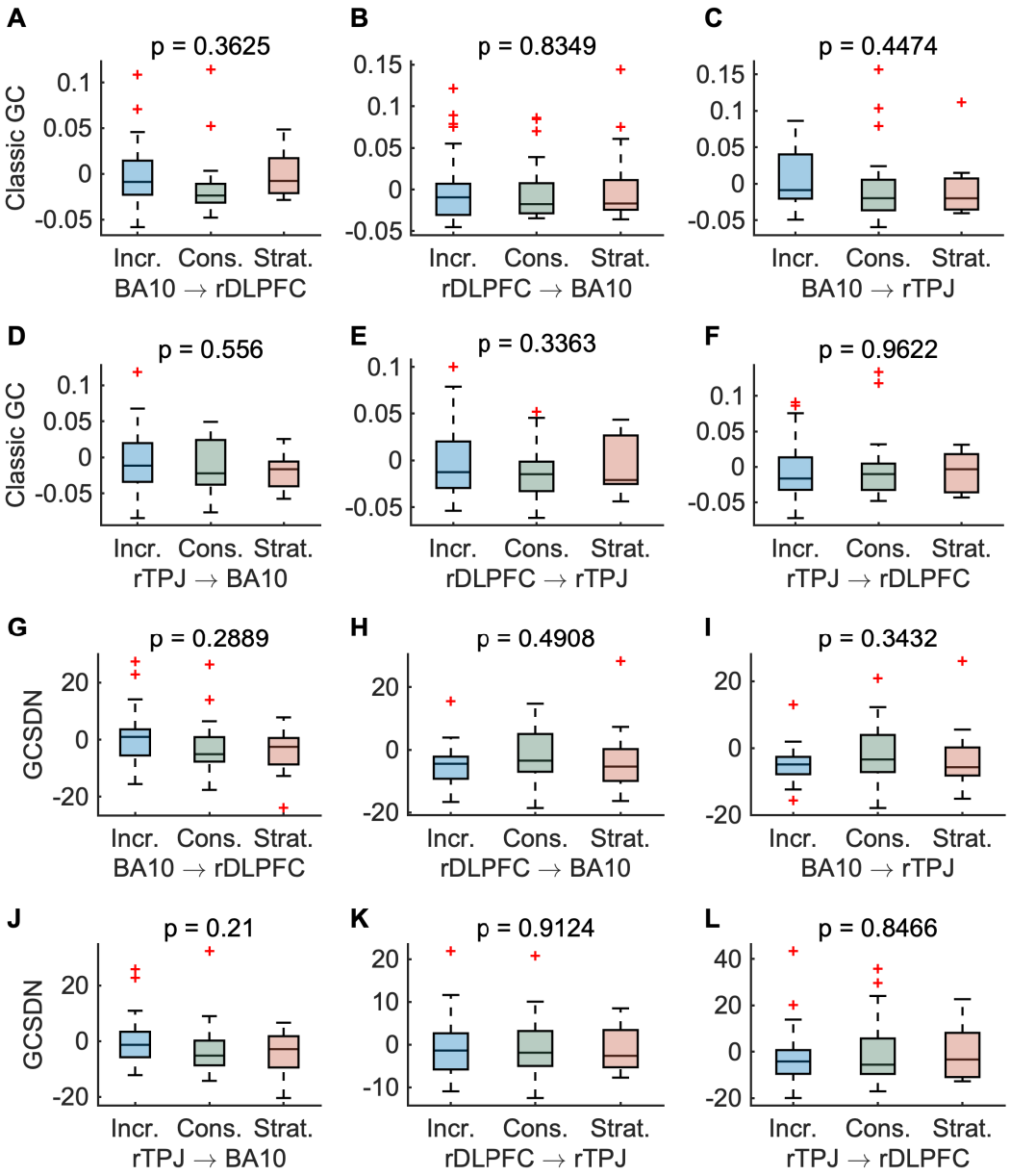


Figure S5: Comparison of the information flow estimated by the classic GC and GCSDN method.
In this time-invariant grouping of subjects by Bhatt et al. 2010, no significant group-difference could be detected from causal connectivity among three behavior groups neither by the classic GC(A-F) method nor by GCSDN method(G-L).

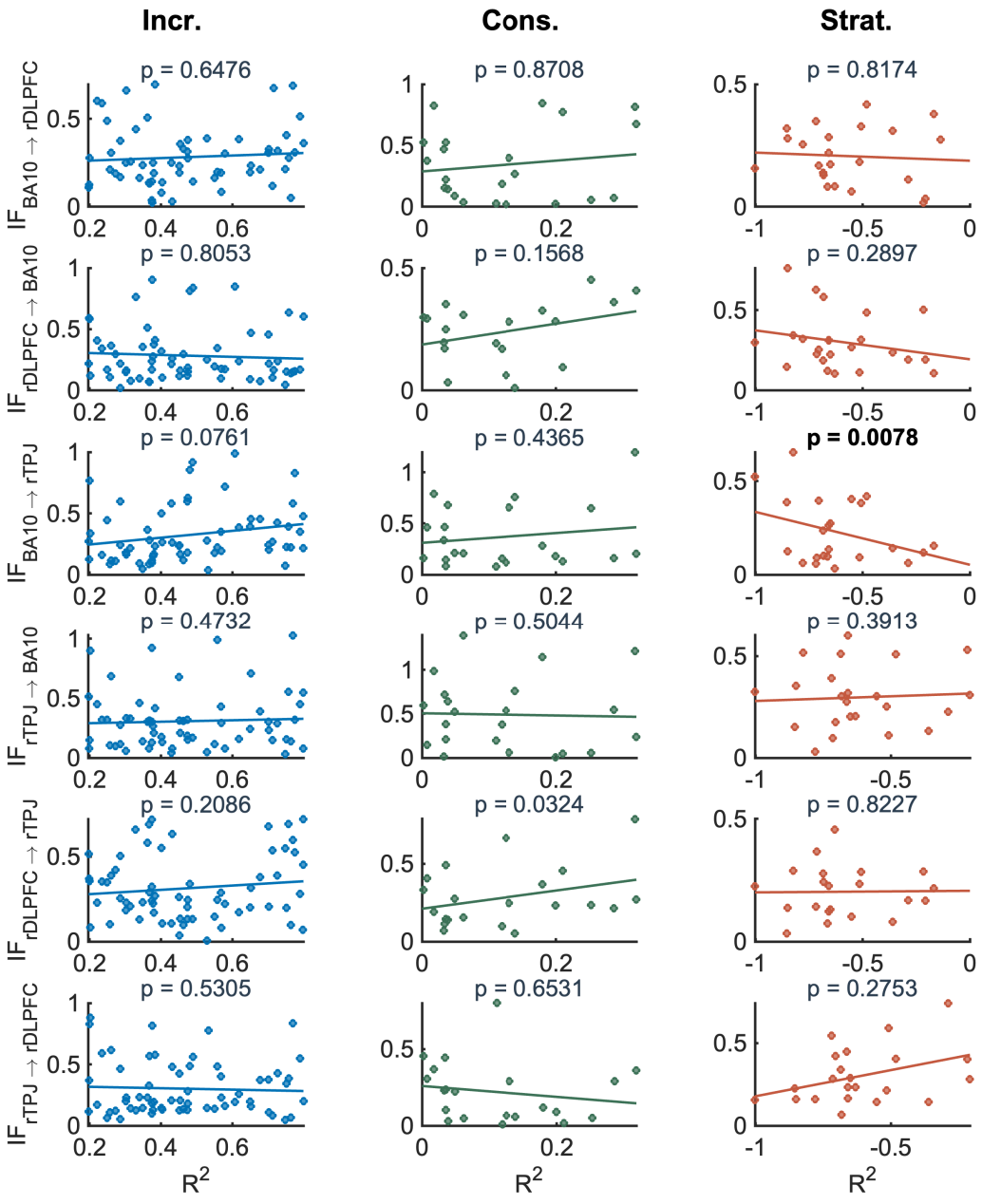


Figure S6: Behavioral correlations between the information revelation (IR) and the information flows $(\overline{\text{IF}})$ in 6 directions for 3 different types of behavioral windows.

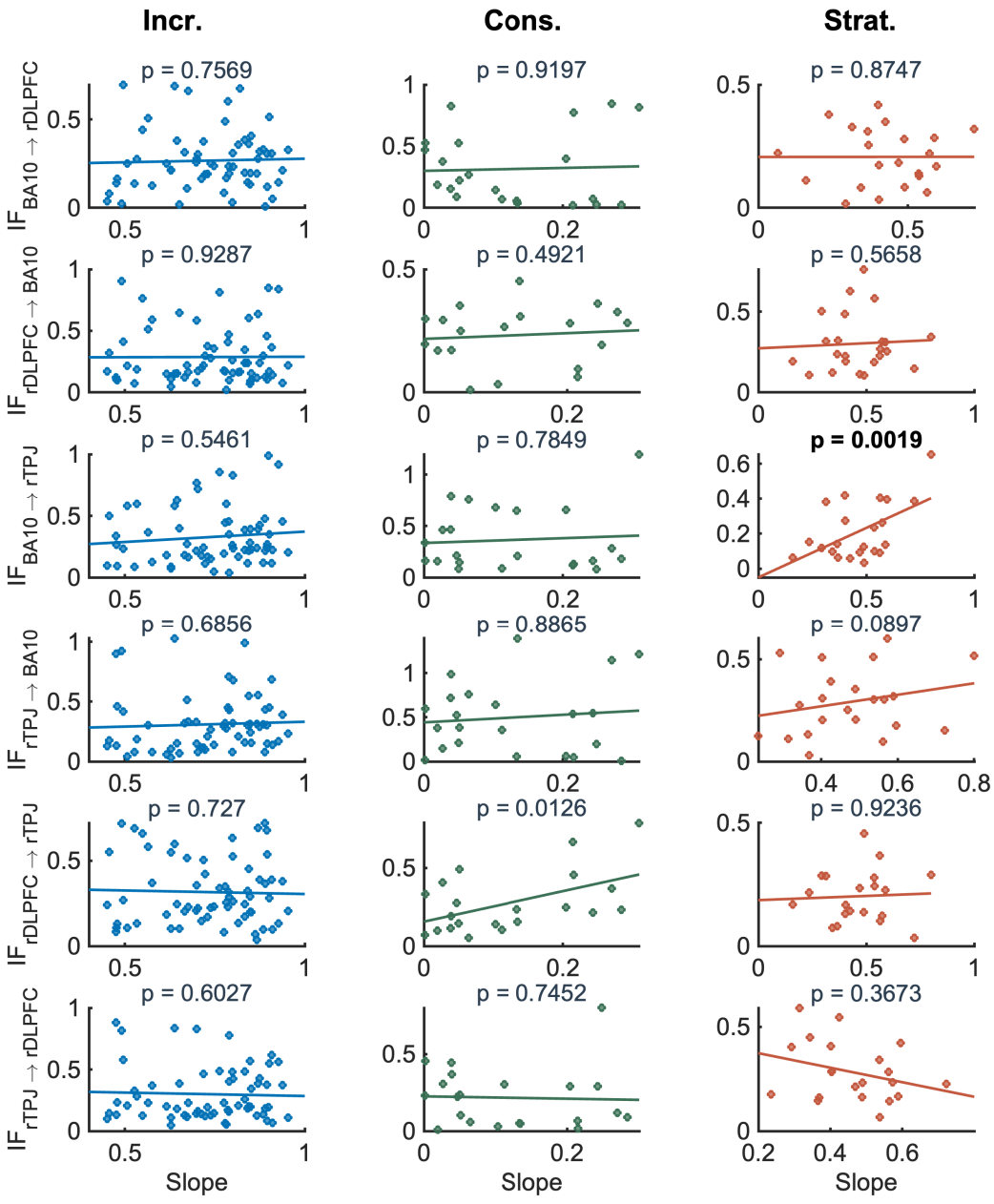


Figure S7: Behavioral correlations between the R^2 and the information flow $\left(\overline{\mathrm{IF}}\right)$ in 6 directions for 3 different types of behavioral windows.

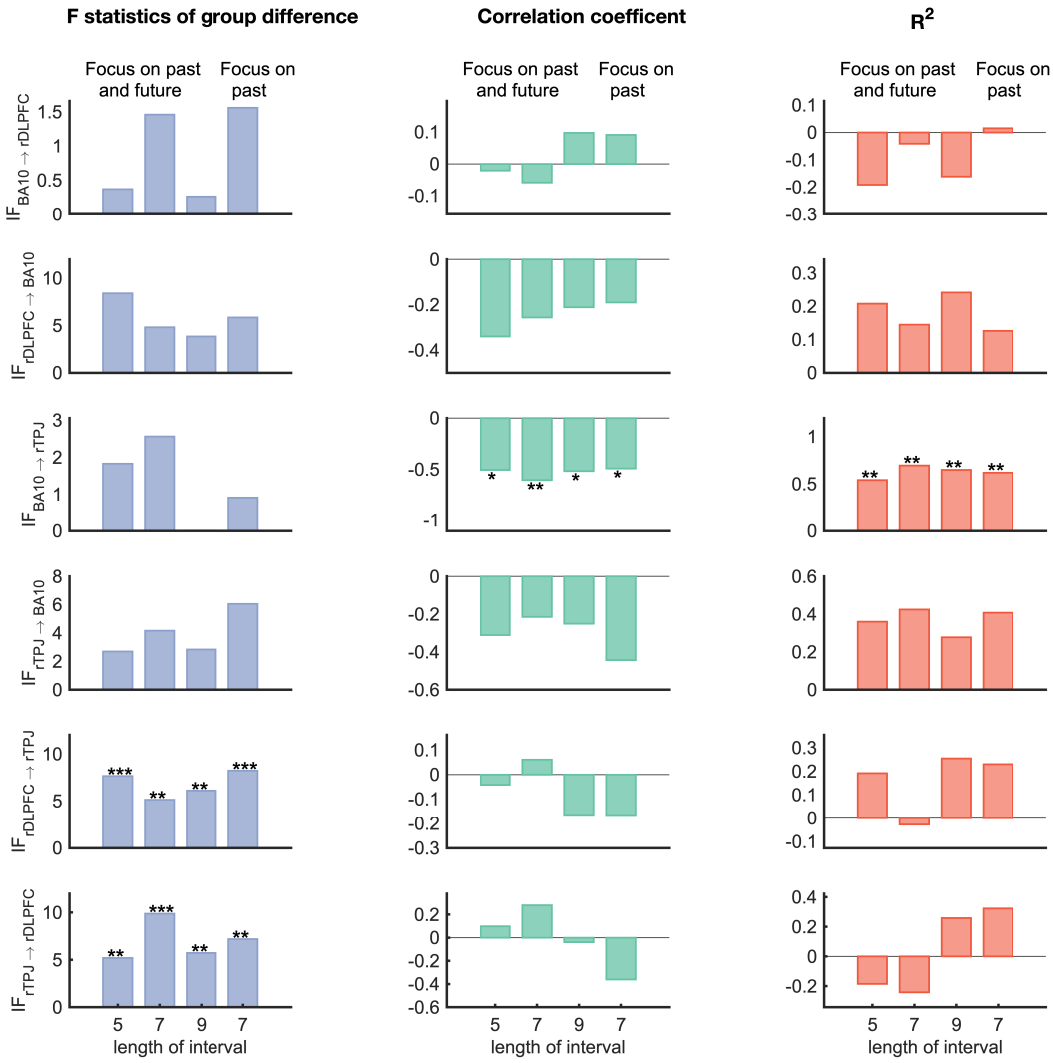


Figure S8: Results when different intervals were chosen to calculate the observations.

F statistics of the mean causal effect among 3 behavioral groups, Pearson correlation coefficient between the mean causal effect and the behavior character on behavioral windows, when different intervals were chosen to calculate the observations.

$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

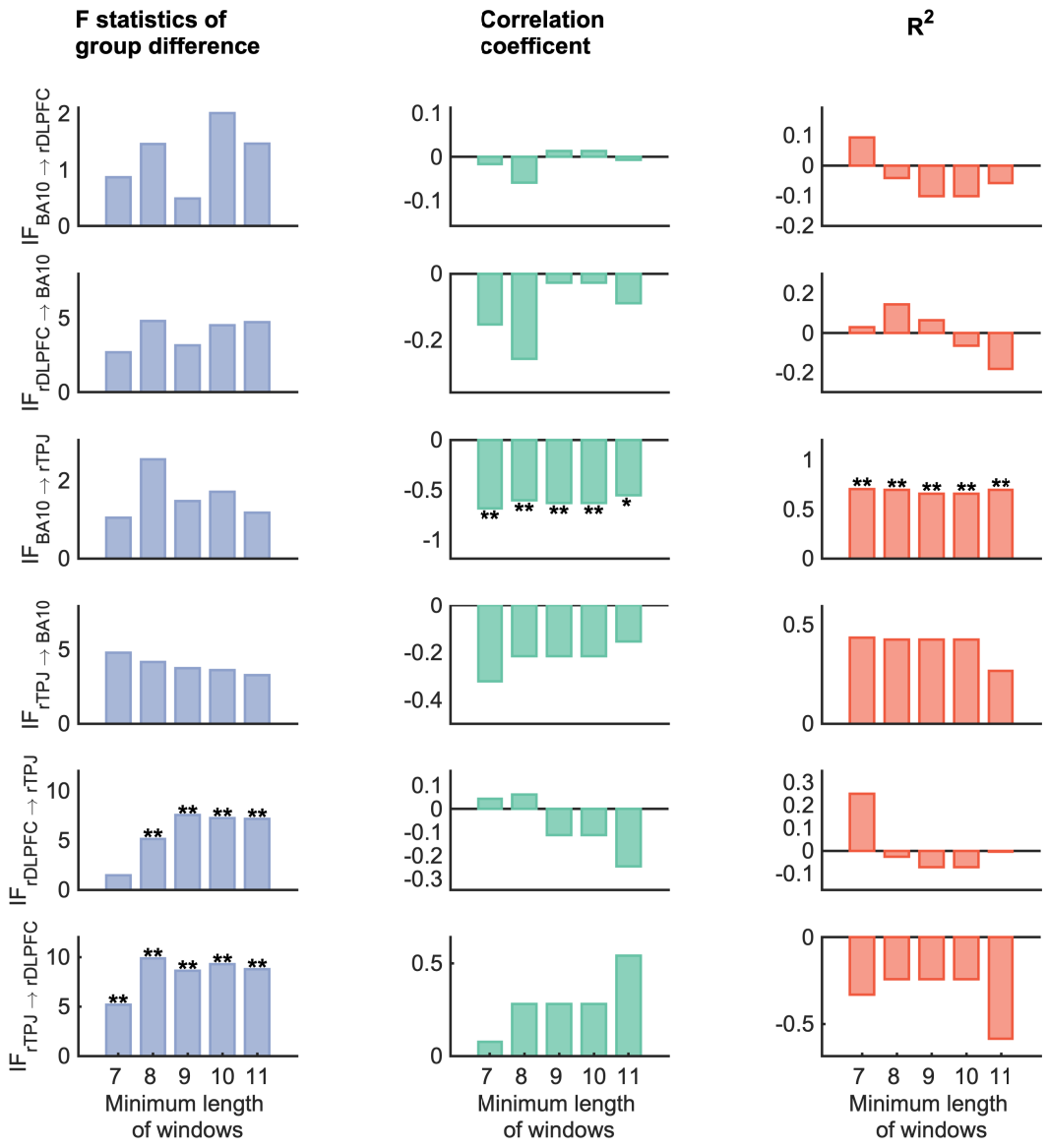
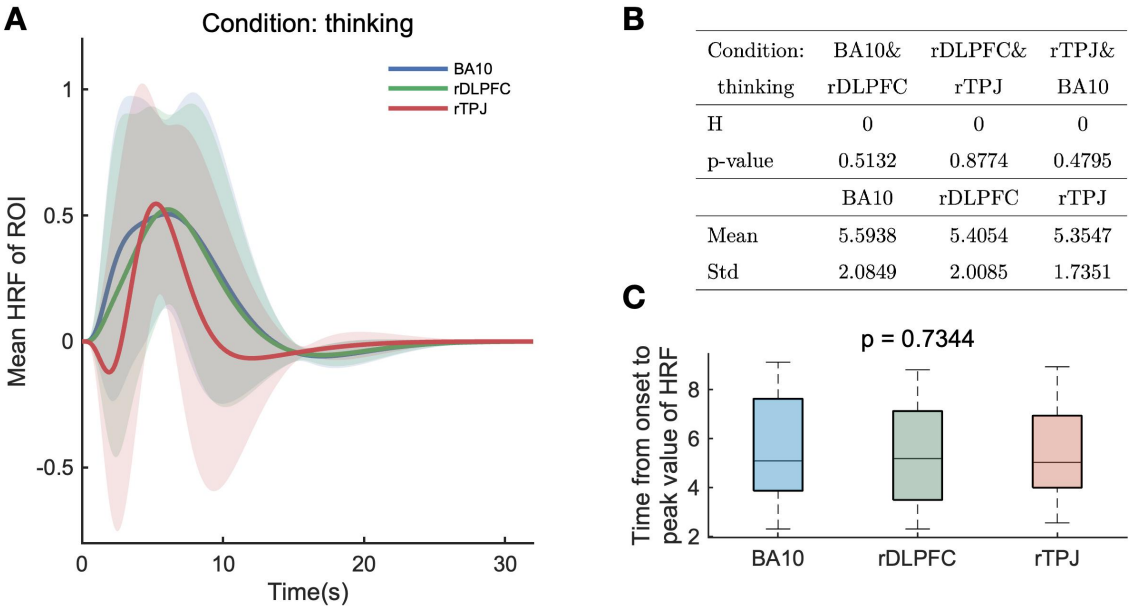


Figure S9: Results of different thresholds of the minimum length of a stable behavioral window.

F statistics of the mean causal effect among 3 behavioral groups, Pearson correlation coefficient between the mean causal effect and the behavior character on behavioral windows, when setting different thresholds of the minimum length of behavioral windows in six directions.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.



Condition: thinking	BA10& rDLPFC	rDLPFC& rTPJ	rTPJ& BA10
H	0	0	0
p-value	0.5132	0.8774	0.4795
	BA10	rDLPFC	rTPJ
Mean	5.5938	5.4054	5.3547
Std	2.0849	2.0085	1.7351

Figure S10: Comparison of the HRFs among three ROIs.

(A) Mean HRFs for three ROIs in the conditon of thinking. (B) Compare HRF delay (i.e., from onset to peak value of HRF) among two brain regions in the condition of thinking by pairwise t-test. (C) Group difference of the hemodynamic delay (i.e., from onset to peak value of HRF) among three groups.

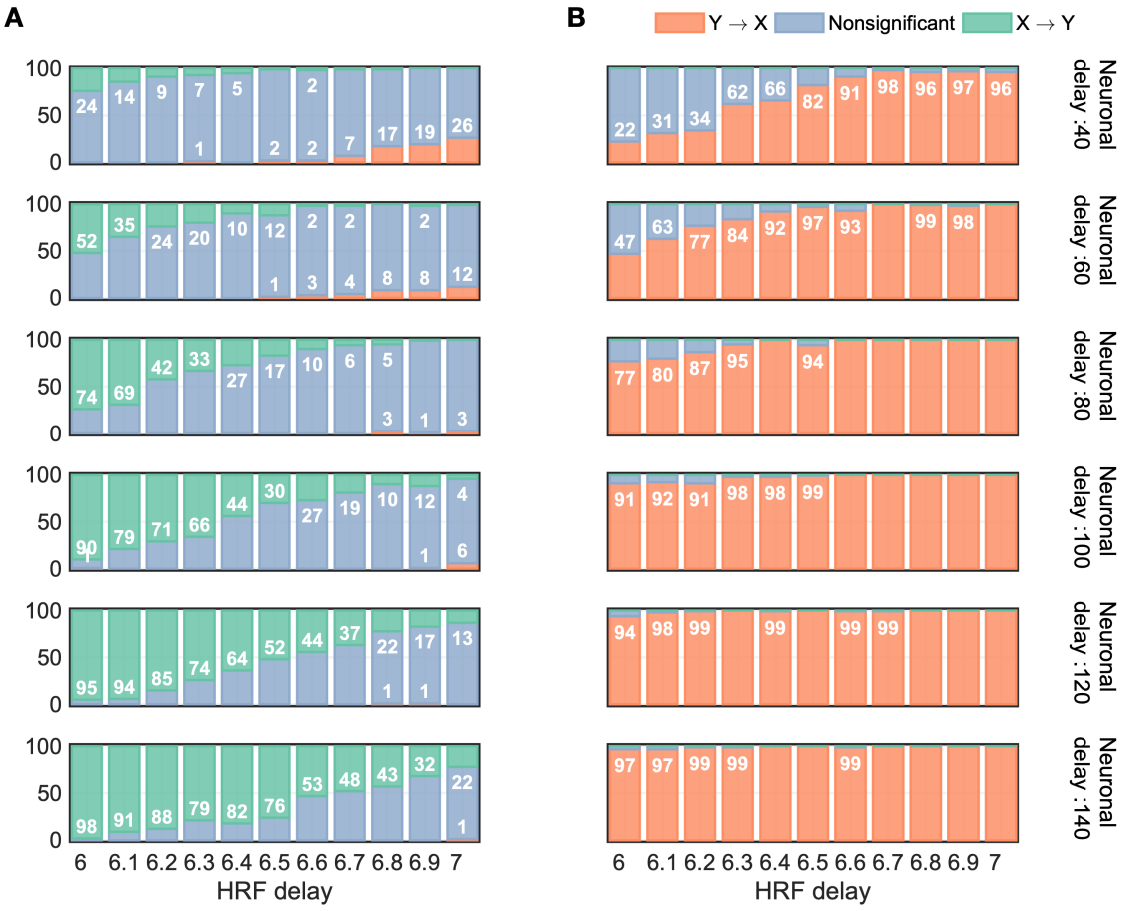


Figure S11: Model performance. Proportion of detected X-to-Y information flow (green), Y-to-X information flow (red) and non-significant results (blue) with different hrf delay and neuronal delay. (A) X-to-Y underlying effective connectivity (B) Y-to-X underlying effective connectivity.

References

[1] Q. Luo, T. Ge, J. Feng, Granger causality with signal-dependent noise, NeuroImage 57 (2011) 1422–1429. URL: http://www.sciencedirect.com/science/article/pii/S1053811911005623. doi:https://doi.org/10.1016/j.neuroimage.2011.05.054.

[2] Q. Luo, T. Ge, F. Grabenhorst, J. Feng, E. T. Rolls, Attention-Dependent Modulation of Cortical Taste Circuits Revealed by Granger Causality with Signal-Dependent Noise, PLOS Computational Biology 9 (2013) e1003265. URL: https://doi.org/10.1371/journal.pcbi.1003265. doi:10.1371/journal.pcbi.1003265.

[3] Q. Luo, Y. Ma, M. A. Bhatt, P. R. Montague, J. Feng, The Functional Architecture of the Brain Underlies Strategic Deception in Impression Management, Frontiers in Human Neuroscience 11 (2017). URL: https://www.frontiersin.org/article/10.3389/fnhum.2017.00513. doi:10.3389/fnhum.2017.00513.

[4] Q. Luo, W. Yang, D. Yi, Kernel shapes of fuzzy sets in fuzzy systems for function approximation, Information Sciences 178 (2008) 836–857. URL: http://www.sciencedirect.com/science/article/pii/S0020025507004574. doi:https://doi.org/10.1016/j.ins.2007.09.020.

[5] Ding.M., Chen.Y., Bressler.S., Granger causality: Basic theory and application to neuroscience, In: Schelter B, Winterhalder M, Timmer J, editors. Handbook of Time Series Analysis (2006) Weinheim: Wiley–VCH. doi:10.1002/9783527609970.

[6] X. Wen, L. Yao, Y. Liu, M. Ding, Causal Interactions in Attention Networks Predict Behavioral Performance, The Journal of Neuroscience 32 (2012) 1284. URL: http://www.jneurosci.org/content/32/4/1284.abstract. doi:10.1523/JNEUROSCI.2817-11.2012.

[7] K. Friston, R. Moran, A. K. Seth, Analysing connectivity with Granger causality and dynamic causal modelling, Curr Opin Neurobiol 23 (2013) 172–178. doi:10.1016/j.conb.2012.11.010.

[8] W. D. J.-B. J.Fristona, Bayesian fMRI time series analysis with spatial priors, NeuroImage 24 (2005) 350–362. URL: https://doi.org/10.1016/j.neuroimage.2004.08.034.

[9] M. B. Schippers, R. Renken, C. Keysers, The effect of intra-and inter-subject variability of hemodynamic responses on group level Granger causality analyses, NeuroImage 57 (2011) 22–36. URL: https://doi.org/10.1016/j.neuroimage.2011.02.008.

[10] Schmolesky.M.T., Wang.Y., Hanes.D.P., Thompson.K.G., Leutgeb.S., Schall.J.D., Leventhal.A.G., Signal timing across the Macaque visual system., Journal of Neurophysiology 79 (1998) 3272–3278. URL: https://doi.org/10.1152/jn.1998.79.6.3272.

[11] Knepp.D.L., Entwisle.D.R., Testing significance of differences between two chi-squares, Psychometrika 34 (1969) 331–333. URL: https://link.springer.com/content/pdf/10.1007/BF02289361.pdf.