

Patterns of dynamic brain network reconfiguration shared across subjects during the learning of value

Azeez Adebimpe^a, Ankit N. Khambhati^b, Marcelo G. Mattar^c, Sharon L. Thompson-Schill^c, Daniel Romer^a,
Danielle S. Bassett^{b,d,*}

^a*Annenberg Public Policy Center, University of Pennsylvania, Philadelphia, PA, USA*

^b*Department of Bioengineering, University of Pennsylvania, PA, USA*

^c*Department of Psychology, University of Pennsylvania, PA, USA*

^d*Department of Electrical and Systems Engineering, University of Pennsylvania, PA, USA*

Abstract

Human learning is a complex process whereby future behavior is altered via the modulation of neural activity. In many cases, such activity displays markers of collective dynamics, leading to non-trivial fluctuations in patterns of functional connectivity. Such fluctuations can display characteristic structure both across time and across subjects. Yet, a fundamental understanding of how neural activity and connectivity track learning processes that are shared across subjects *versus* those that distinguish one subject from another has remained elusive. Here, we seek to address this challenge in a longitudinal experiment in which healthy adult human participants learned the values of novel objects over the course of 12 training sessions. To assess the degree to which patterns of *functional activity* were subject-general *versus* subject-specific, we calculated the intersubject correlation of fMRI BOLD time series. To perform a complementary assessment of the degree to which patterns of *statistical relations* between those time series were subject-general *versus* subject specific, we introduced a measure of intersubject functional connectivity: the Pearson correlation between the functional connectivity matrices of each subject and the functional connectivity matrices of all other subjects. Intersubject correlations in both activity and connectivity were greater than expected in non-parametric permutation tests in the lateral occipital cortex, fusiform gyrus, lingual gyrus, supramarginal area, and sensorimotor cortex. In addition, intersubject correlations in activity were greater than expected in the anterior cingulate gyrus while intersubject correlations in connectivity were greater than expected in the superior parietal cortex, cingulate gyrus, and frontal pole. Interestingly, over the whole brain intersubject correlations in activity peaked in the early stages of learning, while intersubject correlations in connectivity peaked in the later stages of learning. Finally, individual differences in performance accuracy tracked intersubject correlations in connectivity but not activity. Taken together, our results point to both a conserved and variable brain network substrate for value learning, and begin to distinguish the time scales over which these substrates vary with task performance.

Keywords: Value learning Intersubject correlation Intersubject functional connectivity Learning rate brain network dynamics

*Corresponding author

Email address: dsb@seas.upenn.edu (Danielle S. Bassett)

1. Introduction

Humans have the remarkable ability to adapt to environmental pressures through learning – a process that is accompanied by heterogeneous changes in brain structure and function. Parsing the neurophysiological mechanisms of learning could provide a better understanding of how the brain is able to fine-tune its organization and dynamics to accurately and efficiently accomplish day-to-day tasks. Importantly, such alterations continue to impact future behaviors well beyond the critical period over which learning occurs, and in a manner that is unique to each individual [? ?]. These and related empirical observations have resulted in a shift in investigational paradigms that seek to understand general mechanisms of learning and the ensuing variability in these processes that drive inter-individual differences in behavior [?].

Many factors may contribute to inter-individual variability in learning, including the mechanics of the particular skill or information that is being learned, and the strategies by which the same phenomenon, concept, or principle can be apprehended [? ?]. It is intuitively plausible that different individuals could be predisposed to engage in different learning strategies based on the nature of their central nervous system [?]. Unique genetic and environment factors present in the early stages of development can give rise to different wiring patterns in the brain [?], in turn learning to different patterns of neural activity in response to stimuli [? ? ?]. A particularly parsimonious language in which to describe and characterize such patterns is that of network science [? ?], where regions of the brain are represented as network nodes whose activity can vary over time [?], and connections between regions are represented as network edges whose strength can also vary over time [?]. While some organizational principles of brain network organization and dynamics appear to be conserved across individuals [?], others – including measures of activity [? ?] and connectivity [?] – vary appreciably [?]. It is as yet unknown to what degree these shared *versus* unique features of brain network structure and dynamics explain the processes of learning and the resultant behavior.

From a computational perspective, identifying tools that can be used to compare shared *versus* unique features of brain network structure and dynamics in a statistically principled fashion are relatively sparse. Arguably one of the most elegant potential approaches is *intersubject correlation* (ISC), which has historically been used to quantify the extent to which neuronal processes are shared across individuals [?]. ISC estimates the similarity of brain activity across subjects by using the temporal signature of neuronal activity in a particular brain region as a predictor of the neuronal activity in the corresponding brain region of another individual. Thus, ISC is a “model-free” approach that is thought to reflect the degree of inter-subject synchrony in emotional or mental states, which is not simply explained by common evoked responses to the tasks [? ?]. While ISC can effectively capture the shared neuronal response of individual brain regions, it does not measure the degree to which the functional connections between pairs of brain regions are commonly modulated across individuals. Recent work demonstrating that topological features of functional brain networks can predict the capacity of an individual to learn new skills [? ? ?], motivates the development of an approach that can quantify the inter-subject similarity of functional connectivity. Such techniques to estimate the similarity of functional brain organization of human cortex across subjects during learning in the context of both *functional activity* and *functional connectivity* constitute important tools to probe the neural basis of individual differences during learning.

In this study, we utilize ISC to measure the response of brain activity shared across individuals learning the value of novel objects over the course of 12 training sessions taking place on four days. In addition to the ISC, we introduce the intersubject functional connectivity (ISFC), which measures inter-regional functional correlations across subjects. The ISFC is estimated by calculating the Pearson correlation coefficient between the pattern of functional connectivity of a single subject and the average pattern of functional connectivity across all remaining subjects. We hypothesized that subject-general patterns of activity and connectivity would be located in motor cortex (consistent with the shared finger movement to press the response box) [? ?], visual cortex (consistent with the shared processing of the visual stimuli of the novel objects), and other areas previously associated with the learning of value, including lateral occipital cortex [?]. Based on prior work showing that learning modulates functional connectivity on an individual basis [? ? ?], we hypothesized that increase in intrinsic functional connectivity shared across the subjects over a period of learning modulated by the coherence of stimulus induced activity. Such findings would demonstrate the utility of intersubject analyses (ISC and ISFC) in the context of *functional activity* and *functional connectivity* could provide clear neurophysiological correlates and the dynamic of intrinsic functional connectivity and task-dependent activity in the human brain shared across the subjects over a period of time during learning.

2. Results

Intersubject analyses of functional brain activity and functional connectivity are powerful techniques for discriminating brain regions that adhere to generalized constraints of functional architecture from brain regions that might break free from such generalized rules and contribute to subject-specific behaviors. Simply, intersubject analyses quantify the extent to which a functional measurement (brain activity or connectivity) in one subject statistically differs from the expected distribution of the measurement in other subjects. To infer generalized constraints on the functional activity of a single brain region, we compute the inter-subject correlation (ISC) – a measure of reliability of the stimulus-driven response across subjects without any *a priori* knowledge of the temporal composition of the exact cortical response [?]. Similarly, constraints on the functional connectivity between two brain regions for a given subject can be inferred from inter-subject functional connectivity (ISFC). Together, measures of ISC and ISFC are a boon for research avenues that seek to tease apart and interpret relationships between activity and connectivity that accompany the evolution of cognitive processes over time, such as learning.

In this study, we investigated functional mechanisms that facilitate network reorganization associated with value learning by assessing inter-subject correlation (ISC) (Fig. 1A) and inter-subject functional connectivity (ISFC) (Fig. 1B) based on functional MRI (fMRI) measured in 20 healthy subjects (9 females; ages 19-53 years; mean age = 26.7 years) at rest and during a task-based learning experiment over four consecutive days. To compute ISC and ISFC, we first extracted the BOLD time series of 112 brain regions defined by the Harvard-Oxford atlas [? ?] – including cortical and sub-cortical regions. We next measured ISC by separately correlating the BOLD signal for each of N brain region between every pair of S subjects – resulting in an $S \times S$ correlation matrix for each brain regions (see Materials and Methods and Supplementary Information). Intuitively, ISC represents a correlation of brain activity between the same region across subjects. Unlike the generalized linear model, which is a generalization of

ordinary linear regression that estimates the relationship between responses and measurements, ISC is a model-free, non-parametric approach. We then measured ISFC by computing the Pearson correlation between each subject's $N \times N$ functional connectivity matrix (constructed using a wavelet coherence measure) and the average of all other, non-held-out subjects' functional connectivity matrices – resulting in an $N \times N$ correlation matrix for each subject (see Materials and Methods and Supplementary Information). A critical distinction between ISC and ISFC is that ISFC includes the information about the inter-regional, stimulus-induced functional connectivity patterns that might not be revealed by ISC. Specifically, ISFC characterizes the correlation between one subject brain network and brain network of all other subjects.

2.1. Dynamic constraints on functional architecture during learning

First, we asked whether inter-subject correlation of functional activity and functional connectivity is homogenous across brain areas, or whether some brain areas are more constrained in their activity and connectivity to the group-level than other brain areas. We expected significantly greater inter-subject correlation of functional activity among brain regions associated with basic sensory functions than among brain regions associated with abstract, cognitive processing. To test our hypothesis, we first assessed brain regions for which the ISC was statistically significant (t-test, $p < 0.05$, corrected for multiple comparisons) across all scan sessions (Fig. 2A). We found significantly high ISC among brain regions in the occipital lobe, particularly lateral occipital cortex, lingual gyrus, pericalcarine, and fusiform gyrus. We also observed significantly high ISC among sensorimotor regions, including supramarginal gyrus, right superior frontal region and superior motor areas. Separately, we verified that these brain regions also exhibited significantly high ISC during each individual experimental session (Fig. S3A-B). Next, we tested the heterogeneity of ISFC across brain regions (Fig. 2B). Across all scan sessions, we observed significantly high average ISFC (t-test, $p = 0.05$, corrected for multiple comparison) in the lingual gyrus, fusiform gyrus, cuneus, and precuneus – brain regions between occipital and temporal lobes – and in inferior parietal region, superior parietal region, cingulate gyrus, and frontal pole. We separately verified that these brain regions exhibited significantly high ISFC during each individual experimental session (Fig. S3D-E). Importantly, our results demonstrate a clear distinction in the brain regions that have high ISC and high ISFC. Namely, ISC uniquely captured anterior cingulate and ISFC uniquely captured cingulate gyrus and frontal pole. These differences may be explained by the fact that ISFC estimated inter-regional correlation based on the *functional connectivity* while ISC quantified the reliability of common response of each brain region across all the subjects which based on the *functional activity*. Overall, we observed distinct regional variation in measured ISC and ISFC across the brain, and that these brain regions exhibit robustly high inter-subject correlations across experimental sessions.

Next, we asked whether inter-subject correlation of functional activity and functional connectivity distinguish different phases of value learning. We expected to find an interaction between ISC and ISFC during learning, such that functional activity would be highly aligned across subjects early during learning, indicated by high values of ISC, and that functional connectivity would be highly aligned later during learning, indicated by high values of ISFC. Our hypothesis is driven by the notion that initial phases of learning would evoke common patterns of functional activity across individuals, which might in turn reinforce common patterns of functional connections across subjects near the end of learning. We first examined the dynamics of ISC during value learning (Fig. 2C) and

found that the average ISC of all regions increased from day 1 to day 2 and, subsequently, continued decreasing through day 4 (One-way ANOVA, $F(3, 448) = 38.3121, p = 2.7e - 21$). Overall, the average ISC was significantly greater during all four days of value learning compare to at rest ($t = 11.49, p = 1.28 \times 10^{-20}$; Fig. S2A). The resulting trends in ISC dynamics during value learning suggest that ISC might support two phases of learning: (i) increased ISC between day 1 and day 2 may be associated with increasing constraints on functional activity, perhaps as a result of common neurophysiologic mechanisms across subjects that facilitate early stage learning of the task mechanics, and (ii) decreased ISC from day 2 through day 4 may be associated with less constrained dynamics, perhaps as subjects explore cognitive strategies to increase their performance on the task. Interestingly, the average ISC of subjects at rest was near zero (Figure S2A-B, see supplementary results), suggesting that functional activity is minimally constrained in the absence of a common stimulus. Across the four days, the average ISC of the region during value learning across the region were different (One-way ANOVA, $F(3, 448) = 38.31, p = 2.7 * 10^{-21}$) and there was no significant difference between day 1 and day 3 ($F(1, 112) = 2.73, p = 0.1010$). We next examined the dynamics of ISFC during value learning (Fig. 2D) and found a significant difference in average ISFC between brain regions across the four days ($F(3, 448) = 22.77, p = 9.95 * 10^{-14}$). Specifically, ISFC significantly increased from day 1 to day 2 and peaked during day 3. The increasing trend in ISFC during value learning suggests that functional connectivity becomes more constrained to a common topological pattern across subjects, perhaps reinforcing the functional network associated with performing the value learning task.

We then explicitly tested for potential interactions between ISC and ISFC over time. Overall, we found that ISC and ISFC were significantly positively correlated on each of the four days (day 1: $r = 0.2684, p = 0.0430$, day 2: $r = 0.3185, p = 0.0000$, day3: $r = 0.1970, p = 0.0375$, day 4: $r = 0.2982, p = 0.0015$), suggesting that increased group-level constraints on functional activity are related to increased group-level constraints on functional connectivity. However, despite the correlation between ISC and ISFC on any given day, we observed that their rate of change through the duration of the task was indeed significantly different (Fig. 2E and Table 2 for statistical differences). During early stage learning on days 1 and 2, ISC significantly exceeded ISFC – suggesting that functional activity is more rigidity constrained to common organizational rules across the cohort than functional connectivity. During later stage learning on day 3 and 4, ISFC significantly exceeded ISC – suggesting that functional activity becomes more autonomous than functional connectivity and functional activity becomes more constrained than functional activity. These nuanced dynamics point to a potential driver-follower mechanism of constrained functional activity preceding constrained functional connectivity over four days of value learning.

2.2. Regional variability of functional constraints during learning

Next, we asked “If the dynamics of whole-brain inter-subject correlations map unto different phases of value learning, which brain systems might be most complicit in these phases?”. To investigate this question, we first partitioned brain regions into objectively defined functional modules using the popular GenLouvain community detection algorithm [?] (see Materials and Methods and SI). Briefly, community detection is applied to the functional network constructed from data of each task session and parses brain regions into functional modules such that brain regions within the same module exhibit strong functional connections and brain regions between different modules have weak functional connections. In order to obtain a single representative partitioning of brain

regions across subjects and scans, we computed the module allegiance matrix [?] – capturing the probability that two regions belong to the same functional module (Fig. 3A). By applying a final round of community detection to the module allegiance matrix, we identified seven modules that were associated with different putative brain systems, including fronto-temporal (FT) which covered most of the limbic lobe, sensorimotor network (SM), auditory network (AUD) including hippocampus and amygdala, the common default mode network (DMN), Language (LAN) network, Visual (VIS) network and three subcortical regions - putamen, caudate and thalamus (PCT) (Fig. 3B; see Table 1 for list of regions in each community).

We next examined learning-based, regional ISC and ISFC dynamics within these brain systems. First, we observed greater ISC (Fig.3C) in sensorimotor (SM), default mode network (DMN) and visual (VIS) systems (Figure S5 for statistical results) than the remaining brain systems during all four sessions of the learning task suggesting that these brain systems are more constrained to patterns of functional activity common to all participants (Fig. 3C). The SM, DMN, and VIS brain systems are generally involved in sensory processing accompanying the value task, and thus it is intuitive that functional activity in their underlying brain regions would be most constrained across subjects. Second, regional ISC dynamics followed a similar time course to the global ISC dynamics reported earlier (Fig. 2C), in that ISC increased between day 1 and day 2 and decreased thereafter. There were significant differences (one way anova) of ISC across all days for each community excluding the language network ($F(3, 40) = 2.10, p = 0.120$). We next assessed regional variability in ISFC, and observed that DMN, SM, LAN, and VIS systems (Fig.3D) exhibited consistently greater ISFC than other brain systems (t-test, $p < 0.05$, see Figure S5 for statistical results). The lower of global ISFC in the first day was evidence from the regional distribution of ISFC (Fig. S1B) and increased successfully in the following days. The lingual and fusiform gyri and sensorimotor brain regions were significantly correlated from the second to fourth day, and other regions including cingulate gyrus, supramarginal and frontal regions most especially in day 3 and day 4 were found correlated (Fig. S1B). These results suggest a steady integration of the functional brain network across subjects during value learning task.

We next explicitly examined the regional interactions between ISC and ISFC during value learning. Similar to global ISC and ISFC, we compared differences in ISC and ISFC in each of the seven brain systems through the four learning sessions (Fig. 4; see Table 3 for the statistical differences for each brain system). In all brain systems, we found a greater ISC than ISFC during days 1 and 2 and a greater ISFC than ISC during days 3 and 4, suggesting that high constraint on functional activity preceding high constraint on functional connectivity may be a generalized rule that brain systems follow during value learning. We did, however, find that the effect of the dynamical interaction between ISC and ISFC varied between brain systems. In particular, we measured the largest difference between ISC and ISFC in SM and VIS systems, especially during days 3 and 4 of late stage learning. Given the potential functional role of these brain systems in stimulus processing, our result suggest that functional connectivity becomes comparatively more constrained than functional activity during late stage learning when subjects are presumably better acquainted with the task. The DMN also produced similar differences between ISC and ISFC like SM and VIS but with smaller differences during day 3 and 4. However, AUD and LAN with lower ISC and ISFC compare to other communities (Fig 3C and D) show smaller gap between ISC and ISFC (Fig. 4) curves. Also, the decreased in ISC from day 2 to day 3 in LAN and AUD were not that large compared to DMN, VIS and SM network communities. There were lesser differences between ISC and ISFC for the FT and PCT

network communities compare to other network communities most especially during day 3 and 4.

2.3. ISFC and value learning task accuracy

Lastly, we examined whether inter-individual differences in the constraints of functional activity and connectivity explain measured performance on the learning task. We hypothesized some of the regions might be induced with the accuracy of the task and learning as function of days. First, we tested whether participants demonstrated a clear learning effect and found that average task accuracy increased with each learning session (Fig. 5A) and that there were obvious changes in accuracy across the days (Two-way ANOVA, $F(3, 80) = 3.69, p = 0.0169$). Next, we computed the Pearson correlation coefficient between each participant’s task accuracy and the ISFC. However, we observed a significant relationship between average task accuracy and average ISFC over the four days ($r = 0.7686, p = 0.0702$). We therefore explored the brain regions’ ISFC that are mostly correlated value learning accuracy in order to gauge which regions specifically drive inter-individual differences in value learning task accuracy. In each day, ISFC of the right occipital fusiform gyrus ($r = 0.4892, p = 0.0286$), left caudate ($r = 0.5192, p = 0.0192$, Figure 6C) and right caudate ($r = 0.4601, p = 0.0412$, Figure 5B) were significantly correlated with the learning task accuracy across the subjects. However, the resistivity which measure the flow the network interaction were significantly correlated with the task performance accuracy.

3. Discussion

In this study, we used intersubject analyses to understand the respective roles of functional activity and functional connectivity in value learning. Importantly, intersubject analyses allowed us to tease apart large-scale neuronal processes that drive the learning of object value from processes that help subjects strategize their use of the learned information to improve task performance. Critically, we found a dual role for functional activity and functional connectivity during value learning, in which: (i) commonly expressed patterns of functional activity during the early stages of learning become more individualized during later stages of learning, and (ii) individualized patterns of functional connectivity during the early stages of learning become more commonly expressed during later stages of learning. These dynamic processes suggest that the early stages of learning recruits common patterns of functional activity that help consolidate the learned information within common patterns of functional connectivity and that the later stages of learning recruits individualized patterns of functional activity that represent subject-specific processes that help optimize behavior. Generally, our findings demonstrated that subjects undergoing value learning express common stimulus-induced activity and intrinsic connectivity that evolves based on the stage of the learning process.

3.1. Intersubject brain network pattern during value learning

Are intersubject relationships homogenous across brain regions, or are some cognitive systems more likely to be correlated in their activity or connectivity than others? Regionally, we identified several cognitive brain systems – visual, sensorimotor and default mode – that demonstrated significantly greater intersubject correlation and intersubject functional connectivity than other brain systems. In particular, we witnessed significantly high ISC and ISFC within lateral occipital cortex, lingual gyrus, fusiform gyrus and pericalcarine, and all of the occipital

lobe, which suggests that value learning commonly recruits brain regions responsible for visual perception, visual memory and decision making based on chosen value [?]. We also witnessed significantly high ISC and ISFC in sensorimotor and temporoparietal regions, suggesting that brain regions involved in attention, mentalization, and the attribution of self-belief are also commonly involved in value learning [? ?].

Interestingly, we found select brain regions whose ISFC exceeded their ISC – including frontal pole, insular cortex, precuneus, cuneus and cingulate gyrus. These areas are commonly associated with higher cognitive functions such as flexibility of thinking, problem solving, cognitive inhibition and attentional control[?]. One explanation for the high ISFC exhibited by these brain regions is that they represent important cognitive control areas that are likely to act as functional hubs, which may be commonly integrated with other brain regions across subjects. Specifically, frontal pole frequently interacts with the anterior cingulate gyrus during reward-guided learning [?] and precuneus – although widely known to play an important hub in the default mode network (DMN) [?] – is often activated during episodic memory retrieval and self-processing operations [?]. The higher at the precuneus might be related higher reaction time in chosen correct value during value learning task[?] which relate to subject memory of previous similar task. In summary, greater similarity of functional activity appears to occur within predominantly sensory-specific brain regions and greater similarity of functional connectivity appears to occur in regions commonly associated with higher cognitive function.

3.2. *Dynamic intersubject learning curve*

An advantage of our experimental approach is the ability to examine changes in neuronal processes throughout different phases of value learning – that is, we were able to track intersubject relationships dynamically. During early stages of learning, we observed similar patterns of brain activity across subjects, suggesting that participants generally exhibit similar temporal course of BOLD activation in response to value learning task stimuli. During later stages of learning, we observed a decrease in the ISC across subjects, suggesting that participants may adapt to the procedures of the task. We speculate that this dynamical shift in ISC also supports a mechanism in which individualized pattern of activity supports inter-individual variability in task performance. Complimentary to ISC, the ISFC gradually increased between the first and second day and subsequently increased exponentially through the fourth day. This trajectory is reminiscent of the typical shape of a traditional learning curve [?]. Indeed, the empirically observed trajectory of ISFC suggests that the functional connectivity within and between brain networks of all the participants points to its role in an adaptive mechanism for value learning [? ? ?]. The changes in the intersubject large-scale functional brain networks may also provide important neural markers of goal-directed training performance irrespective of training platform and may give more information about the specific changes in brain activity and connectivity that affect behavioral outcomes.

Importantly, we found that cognitive systems were heterogeneously influential in the learning-induced increase of ISFC and decrease of ISC. The significant change in ISC and ISFC of the visual system could be attributed to common inputs to the sensory system for processing of stimulus information. The complimentary increase in the ISFC of the sensorimotor system may result from motor coordination of hand and finger movement to perform the button-pushing during task response [? ?]. We observed the largest difference between ISC and ISFC during the later stages of learning. The resulting changes suggest a role in visual and sensorimotor systems in supporting visual

identification of objects, interpretation of value, and motor coordination of hand-finger movement to guide more accurate and efficient decision making after the critical learning period. The lower ISC in the later stage of learning could not be ascribed to reduced functional neural activity at the visual and sensorimotor regions rather than individual differences in temporal time courses of BOLD signals after adaptation to the task and once the procedure is over learned and mastered in the early stage [? ?]. Interestingly, our findings also demonstrated involvement of DMN, such that over the course of value learning connectivity within DMN becomes more generalized across subjects and activity within DMN become more individualized across subjects. Although DMN is often viewed as a task-negative system that is typically more active during resting state processes, here we report evidence that the architecture of this system evolves as individuals learn the value of different objects. We speculate that changes in DMN architecture might support an improved physiological capability of more rapidly disengaging task-negative processes in support of improved task efficiency.

3.3. Functional network drivers of value learning

Once the learning is achieved, what mechanisms optimize the functionalization of learned information for task performance? We show evidence that visual and sensorimotor systems are involved in encoding task mechanics and value learning. First, ISFC was significantly correlated with task accuracy, suggesting that behavior might be improved when the functional network reorganizes according to a generalized rule common across subjects. Furthermore, individual variation in stimulus-induced functional activity might be responsible for modulating subject-specific behavior during the value learning tasks [? ? ?]. Our observation that the relationship between ISFC and task accuracy was driven by the bilateral caudate nucleus agrees with previous findings that the region is critical to various forms of reward-based learning, such as value learning [?]. Caudate nucleus, an sub-region of dorsal striatum, is thought to play an important role in decision making – a cognitive capability that is important in choice-driven behaviors related to determining the value of an object [?]. In particular, we found a positive relationship between ISFC and caudate that offers a role for dorsal striatum is one of the intrinsic neural drivers that modulate the subject behavior to increase the performance accuracy by interacting with other brain systems especially the feedback-related learning[?]. Our results findings suggest that interregional coupling of brain activity with the dorsal striatum may be promising focal point of future work in the identification of target brain areas that contribute to individual deficits in general learning.

4. Materials and Methods

4.1. Experimental setup and procedure

Participants learned the monetary value of 12 novel visual stimuli in a reinforcement learning paradigm. Learning occurred over the course of four MRI scan sessions conducted on four consecutive days. The novel stimuli were 3-dimensional shapes generated with a custom built MATLAB toolbox (code available at <http://github.com/saarela/ShapeToolbox>). ShapeToolbox allows the generation of three-dimensional radial frequency patterns by modulating basis shapes, such as spheres, with an arbitrary combination of sinusoidal modulations in different frequencies, phases, amplitudes, and orientations. A large number of shapes were generated

by selecting combinations of parameters at random. From this set, we selected twelve that were considered to be sufficiently distinct from one another. A different monetary value, varying from \$1.00 to \$12.00 in integer steps, was assigned to each shape. These values were uncorrelated with any parameter of the sinusoidal modulations, so that visual features were not informative of value. Participants completed 20 minutes of the main task protocol on each scan session, learning the values of the 12 shapes through feedback. The sessions comprised of three scans of 6.6 minutes each, starting with 16.5 seconds of a blank gray screen, followed by 132 experimental trials (2.75 seconds each), and ending with another period of 16.5 seconds of a blank gray screen. Stimuli were back-projected onto a screen viewed by the participant through a mirror mounted on the head coil and subtended 4 degrees of visual angle, with 10 degrees separating the center of the two shapes. Each presentation lasted 2.5 seconds and, at any point within a trial, participants entered their responses on a 4-button response pad indicating their shape selection with a leftmost or rightmost button press. Stimuli were presented in a pseudorandom sequence with every pair of shapes presented once per scan. Feedback was provided as soon as a response was entered and lasted until the end of the stimulus presentation period. Participants were randomly assigned to two groups depending on the type of feedback received. In the RELATIVE feedback case, the selected shape was highlighted with a green or red square, indicating whether the selected shape was the most valuable of the pair or not, respectively. In the ABSOLUTE feedback case, the actual value of the selected shape (with variation) was displayed in white font. Between each run, both groups received feedback about the total amount of money accumulated up to that point.

4.2. MRI Data collection and preprocessing

Magnetic resonance images were obtained at the Hospital of the University of Pennsylvania using a 3.0 T Siemens Trio MRI scanner equipped with a 32-channel head coil. T1-weighted structural images of the whole brain were acquired on the first scan session using a three-dimensional magnetization-prepared rapid acquisition gradient echo pulse sequence (repetition time (TR) 1620 ms; echo time (TE) 3.09 ms; inversion time 950 ms; voxel size 1 mm by 1 mm by 1 mm; matrix size 190 by 263 by 165). A field map was also acquired at each scan session (TR 1200 ms; TE1 4.06 ms; TE2 6.52 ms; flip angle 60° ; voxel size 3.4 mm by 3.4 mm by 4.0 mm; field of view 220 mm; matrix size 64 by 64 by 52) to correct geometric distortion caused by magnetic field inhomogeneity. In all experimental runs with a behavioral task, T2*-weighted images sensitive to blood oxygenation level-dependent contrasts were acquired using a slice accelerated multi-band echo planar pulse sequence (TR 2,000 ms; TE 25 ms; flip angle 60° ; voxel size 1.5 mm by 1.5 mm by 1.5 mm; field of view 192 mm; matrix size 128 by 128 by 80). In all resting state runs, T2*-weighted images sensitive to blood oxygenation level-dependent contrasts were acquired using a slice accelerated multi-band echo planar pulse sequence (TR 500 ms; TE 30 ms; flip angle 30° ; voxel size 3.0 mm by 3.0 mm by 3.0 mm; field of view 192 mm; matrix size 64 by 64 by 48).

Cortical reconstruction and volumetric segmentation of the structural data was performed with the Freesurfer image analysis suite[?]. Boundary-Based Registration between structural and mean functional image was performed with Freesurfer bbgregister [?]. Preprocessing of the fMRI data was carried out using FEAT (fMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). The following pre-statistics processing was applied: EPI distortion correction using FUGUE [?]; motion correction using MCFLIRT [?]; slice-timing correction using Fourier-space time series phase-shifting; non-brain removal using

330 BET [?]; grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; highpass temporal filtering (Gaussian-weighted least-squares straight line fitting, with $\sigma = 50.0s$). Nuisance time series were voxelwise regressed from the preprocessed data. Nuisance regressors included (i) three translation (X, Y, Z) and three rotation (pitch, yaw, roll) time series derived by retrospective head motion correction ($R = [X, Y, Z, \text{pitch}, \text{yaw}, \text{roll}]$), together with first derivative and square expansion terms, for a total of 24 motion regressors [?];
 335 (ii) the five first principal components of non-neural sources of noise, estimated by averaging signals within white matter and cerebrospinal fluid masks, obtained with Freesurfer segmentation tools and removed using the anatomical CompCor method (aCompCor) [?]; and (iii) an 24 estimate of a local source of noise, estimated by averaging signals derived from the white matter region located within a 15 mm radius from each voxel, using the ANATICOR method [?]. Global signal was not regressed out of voxel time series due to its controversial application to resting
 340 state fMRI data[? ?].

We parcellated the brain into 112 cortical and subcortical regions, separated by hemisphere using the structural Harvard-Oxford atlas of the FMRIB (Oxford Centre for Functional Magnetic Resonance Imaging of the Brain) Software Library (FSL; Version 5.0.4) [? ?]. We warped the MNI152 regions into subject-specific native space using FSL FNIRT and nearest-neighbor interpolation and calculated the average BOLD signal across all gray matter
 345 voxels within each region. The participant’s gray matter voxels were defined using the anatomical segmentation provided by Freesurfer, projected into subject’s EPI space with bbregister. For each individual fMRI, we extracted regional mean BOLD time series by averaging voxel time series in each of the 112 regions.

4.3. Inter-subject correlation (ISC) and inter-subject functional connectivity (ISFC)

We used the Pearson correlation coefficient to compute inter-subject correlation (ISC) of each brain region for
 350 each subject. First, we calculated region-wise temporal correlation between every pair of subjects as:

$$r_{ij} = \frac{\sum_{t=1}^N [(x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)]}{\sqrt{\sum_{t=1}^N (x_i(t) - \bar{x}_i)^2 \sum_{t=1}^N (x_j(t) - \bar{x}_j)^2}}, \quad (1)$$

where N is the number of time points of time series data, r_{ij} is the correlation coefficient of a region between the times series x_i and x_j of the i th and j th subjects, respectively. To test the significance of each pair of correlations between one region fMRI BOLD signals of two subjects, we performed a fully non-parametric regional permutation test with 10,000 randomization [?]. This non-parametric test accounts for slow-scale autocorrelative structure in
 355 the BOLD time series [?], by removing removing phase information from each BOLD signal, through Fourier phase randomization. To correct for multiple comparisons, we selected the maximum value of correlation value in a given permutation procedure and defined the threshold for a pair correlation as $q \cdot 100$ th percentile of null distribution of maximum values. The threshold for both rest and task condition was set at $q = 0.01$ and the following thresholds were obtained: rest: $r_{th} = 0.0475$; task: $r_{th} = 0.1172$. . Next, we averaged only significant correlation values out
 360 of 190 correlation values, r_{ij} from all subject pairs to one ISC for each region:

$$ISC = \frac{2}{M(M-1)} \sum_{i=1}^M \sum_{j=2, j>1}^M r_{ij}, \quad (2)$$

where M is the number of subject and in our study with 20 subjects, the ISC was averaged over significant 190 subject pairs. The procedure (see SI for more information) was followed for each scan, day, and for both rest and task conditions. Average ISC for each region was obtained for each day during learning task for three scans per day.

We constructed functional brain networks for each experimental session of each subject, by computing a wavelet transform coherence (WTC) measure between every possible pair of regional BOLD time series [?]. To calculate the WTC, we first applied a wavelet decomposition to the BOLD time series of each brain regions by successively convolving the time series with the Morlet wavelet in a frequency range of 0.05 – 0.11Hz (see SI). The wavelet transformed signal is a complex number expressed as a function of time and frequency. Next, we computed the correlation between wavelet transformed signals associated with each possible pair of brain regions. The wavelet transform of each time series were smoothed in both time and frequency (or scale) before computing the correlation measure [? ?](see SI). We repeated this procedure for all pairs of regions and tabulated the resulting WTC values in a weighted adjacency matrix \mathbf{A} for each experimental session and subject.

Inter-subject functional connectivity (ISFC) was obtained from the functional connectivity (FC) matrices of all the subjects. ISFC estimated correlation between one subject and all other subjects. We computed the ISFC of each subject as the correlation between single subject FC matrix and average of all other subjects FC matrices as

$$ISFC_i = \frac{1}{N} A_i \left[\frac{1}{n-1} \sum_{j \neq i} A_j \right] \quad (3)$$

where A is the adjacency functional connectivity matrix of a subject and n and N is the total number of subjects and total number of regions respectively.

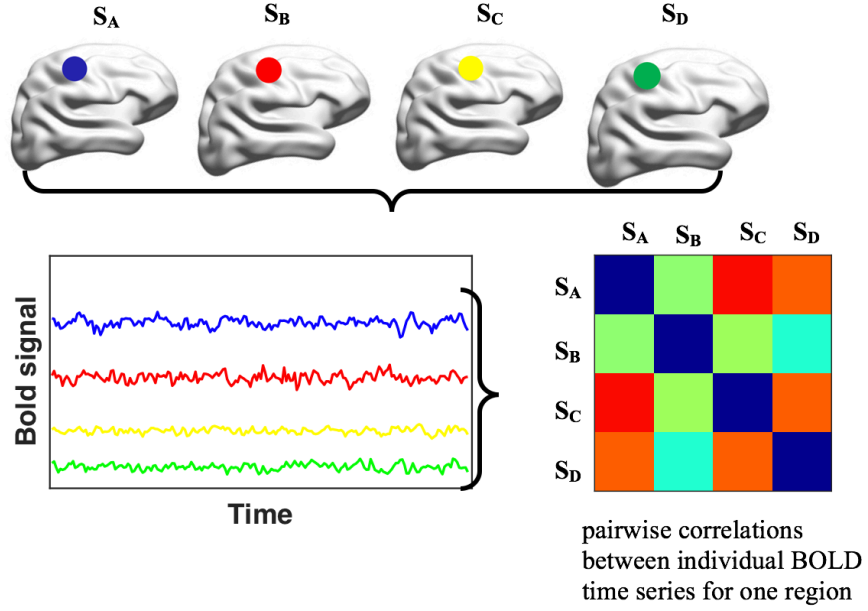
To objectively identify a common set of functional brain systems across all subjects, we partitioned the adjacency matrix into communities by optimizing the modularity index using a generalized Louvain method [? ?] developed specifically for community detection in multi-layer networks. We constructed a multi-layer network from our experimental data, by concatenating adjacency matrices from all sessions and subjects for each day and each runs. We obtained partition of brain into network communities for each scan and subject with the standard parameter of $\gamma = \omega = 1$. (See supplementary materials for details). In order to obtain single representative of partition into different communities, we computed a module allegiance matrix [?], a matrix whose entries represent the probability that corresponding regions belong to the same community across the scans and participants. We then computed the optimal community structure [?] of the network by subdivided the network into non-overlapping groups of nodes which maximizes the number of within-group edges from the module allegiance matrix.

4.4. Reliability of ISC and ISFC

For each subject, we measured the correlation between bold signals for the three scans obtained in each in order to measure the reliability of ISC. This was done for each subject and across the days on a region-by-region basis (see SI for more information). Similarly, to investigate the reliability of ISFC, we computed edge persistence (see SI) between the three scans for each day and for each subjects and was done for region-by-region basis.

References

(A) Inter-subject correlation (ISC)



(B) Inter-subject functional connectivity (ISFC)

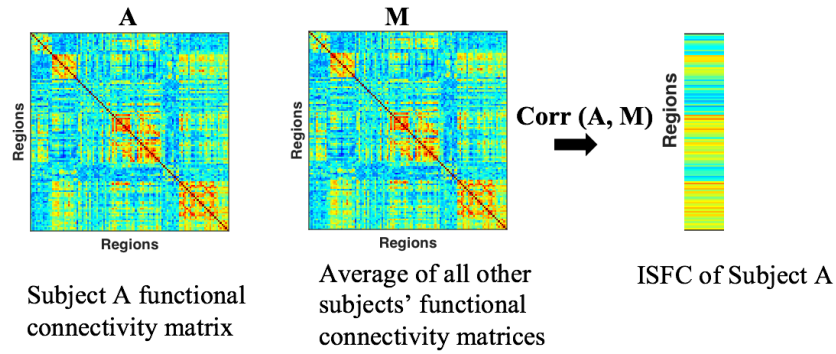


Figure 1: **Inter-subject correlation and functional connectivity.** (A) Inter-subject correlations (ISC) between times series from the same region across different subjects during task and rest predict the coherent activation of similar brain regions. This was estimated from the bold signals by the person correlation(B) Inter-subject functional connectivity (ISFC) is the correlation between connectivity of one subject and other group of subjects revealing pattern of correlations across both the regions and the subjects .

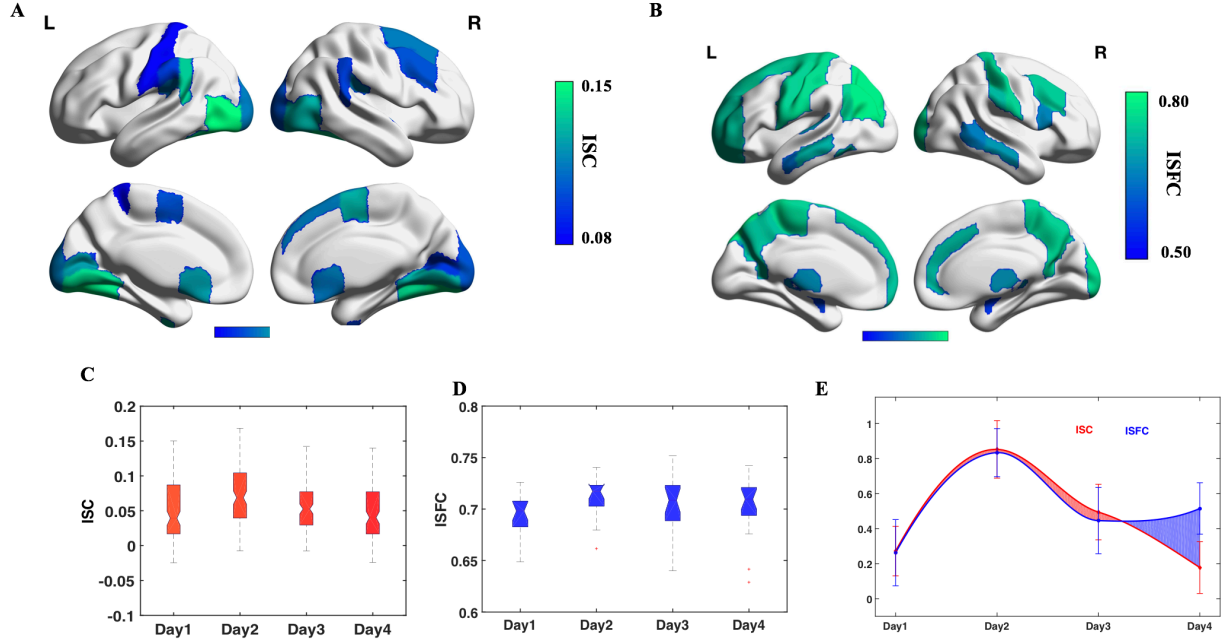


Figure 2: **Inter-subject correlation.** Dynamics ISC and ISFC (A) The consistent ISC ($p < 0.05$, corrected for multiple comparison) across the subjects for the whole four days during the value learning task, shows that lateral occipital cortex, lingual gyrus, supramarginal, sensorimotor region and anterior cingulate were more correlated across the subjects. (B) The average ISFC ($p < 0.05$, corrected for multiple comparison) across all the subjects for the whole four days revealed higher correlation at the pre-central, lingual gyrus and left suprmarginal. Other regions include precuneus, cuneus, frontal pole and cingulate gyrus. (C) Average inter-subject correlation (ISC) increased from day1 to day 2 and decreased in day 3 and day 4. There was significance difference across the whole days ($F(3, 33) = 38.31, p = 2.7 * 10^{-21}$), (D) The ISFC increased significantly from the first day, and was highest in the third day. This was evidence from large statistical difference across the days ($F(3, 33) = 935, p = 0.00$) and (E) Normalized ISC and ISFC shows their dynamics with ISC significantly higher in the first day and no significant difference was observed in day 2. However, ISFC increased in day 3 and 4 while ISC was lower (see Table 2 for statistical differences)

Table 1: Network Communities

Community	Regions
FT	Temporal pole, Middle temporal gyrus - anterior, Frontal medial cortex, Subcallosal cortex, Orbital frontal cortex, Nucleus Accumbens
SM	Precentral gyrus, Postcentral gyrus, Superior parietal lobule, Supramarginal gyrus-anterior and posterior, Supplemental motor area, Central opercular cortex, Parietal operculum cortex
AUD	Inferior temporal gyrus-anterior and posterior, Parahippocampal gyrus-anterior and posterior, Temporal fusiform cortex-anterior and posterior, Globus pallidus, Amygdala, Hippocampus, Brainstem
DMN	Frontal pole, Insular cortex, Superior frontal gyrus, Middle frontal gyrus, Inferior frontal gyrus - pars triangularis and pars opercularis, Middle temporal gyrus-posterior, Middle temporal gyrus- temporooccipital Supramarginal gyrus-posterior Angular gyrus, Paracingulate gyrus, Cingulate gyrus, anterior Cingulate gyrus, posterior, Frontal operculum cortex,
LAN	Superior temporal gyrus-anterior and posterior, Planum polare, Heschl's gyrus, Planum temporale
VIS	Inferior temporal gyrus-temporooccipital, Lateral occipital cortex-superior and inferior, Intracalcarine cortex, Precuneus cortex, Cuneus cortex, Lingual gyrus, Temporal occipital fusiform cortex, Occipital fusiform gyrus, Supercalcarine cortex Occipital pole
PCT	Putamen, Caudate, Thalamus

Table 2: Statistical comparison of average ISC and ISFC for each day across the regions related to Fig. 2E. The statistical t-test was done with 1000 permutations

Day 1	$t = 11.0316$	$p = 1.4939 * 10^{-19}$
Day 2	$t = 0.6881$	$p = 0.9190$
Day 3	$t = -14.4227$	$p = 3.33810^{-27}$
Day 4	$t = -13.7228$	$p = 1.131 * 10^{-25}$

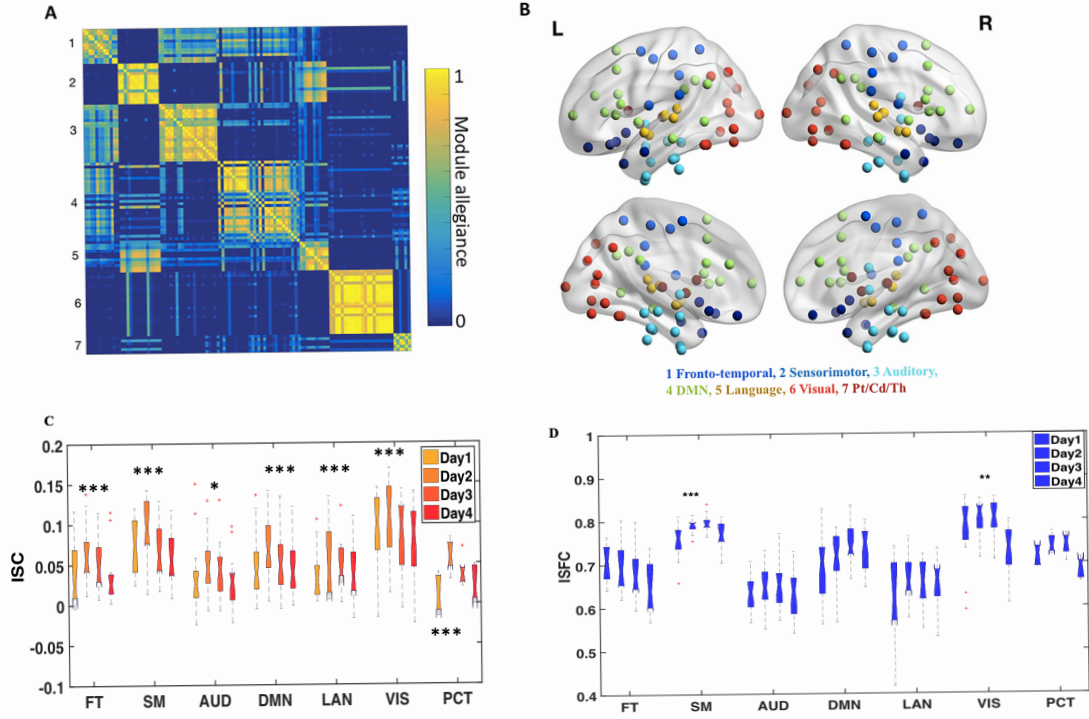


Figure 3: **Network communities, ISC and ISFC.** (A) The module allegiance matrix indicating the probability of brain region belonging to the same community based on the task. Seven communities were identified from community detections across the whole four days for all the subjects, and (B) The communities are projected on to the MRI surface and were named base on the region location and cognitive definition -frontal temporal (FT), Sensorimotor(SM), Auditory(AUD), Default mode network (DMN), Language (LAN), Visual (VIS) and Putamen-caudate- thalamus (PCT) (see Table 1 for the regions in the communities). (C) ISC for each network community the same pattern of the average ISC of the whole day. Sensorimotor, DMN and VIS systems had relatively higher ISC compared to the other regions. There was significance difference across the days in all the regions excluding language (LAN) network. (D) The network communities' ISFC revealed higher ISFC at SM, DMN, LAN and VIS networks with lower ISFC in the first day follow by increased ISFC in day 2 and day3 respectively. *Onewayanova*. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Statistical comparison of average ISC and ISFC for each network community across the days related to Fig. 4. The statistical t-test was done with 1000 permutations

Community	Day 1	Day 2	Day 3	Day 4
FT	$t = 3.1065, p = 0.0225$	$t = 0.7318, p = 0.8857$	$t = -2.7830, p = 0.0508$	$t = -4.1551, p = 0.0039$
SM	$t = 4.6775, p = 0.0030$	$t = 1.4836, p = 0.4030$	$t = -10.8110, p = 0.0000$	$t = -9.9103, p = 0.0000$
AUD	$t = 4.1756, p = 0.0010$	$t = 0.3078, p = 0.9940$	$t = -5.2241, p = 0.0000$	$t = -3.6116, p = 0.0110$
DMN	$t = 5.5727, p = 0.0000$	$t = -0.5079, p = 0.9640$	$t = -7.2368, p = 0.0000$	$t = -6.0107, p = 0.0110$
LAN	$t = 2.8748, p = 0.0840$	$t = -0.1712, p = 0.9980$	$t = -3.9685, p = 0.0156$	$t = -3.1021, p = 0.0430$
VIS	$t = 6.4058, p = 0.0000$	$t = 0.3769, p = 0.9970$	$t = -7.7063, p = 0.0000$	$t = -7.4811, p = 0.0000$
PCT	$t = 0.000, p = 1.0000$	$t = 2.2446, p = 0.1562$	$t = -3.0030, p = 0.0312$	$t = -8.8362, p = 0.0312$

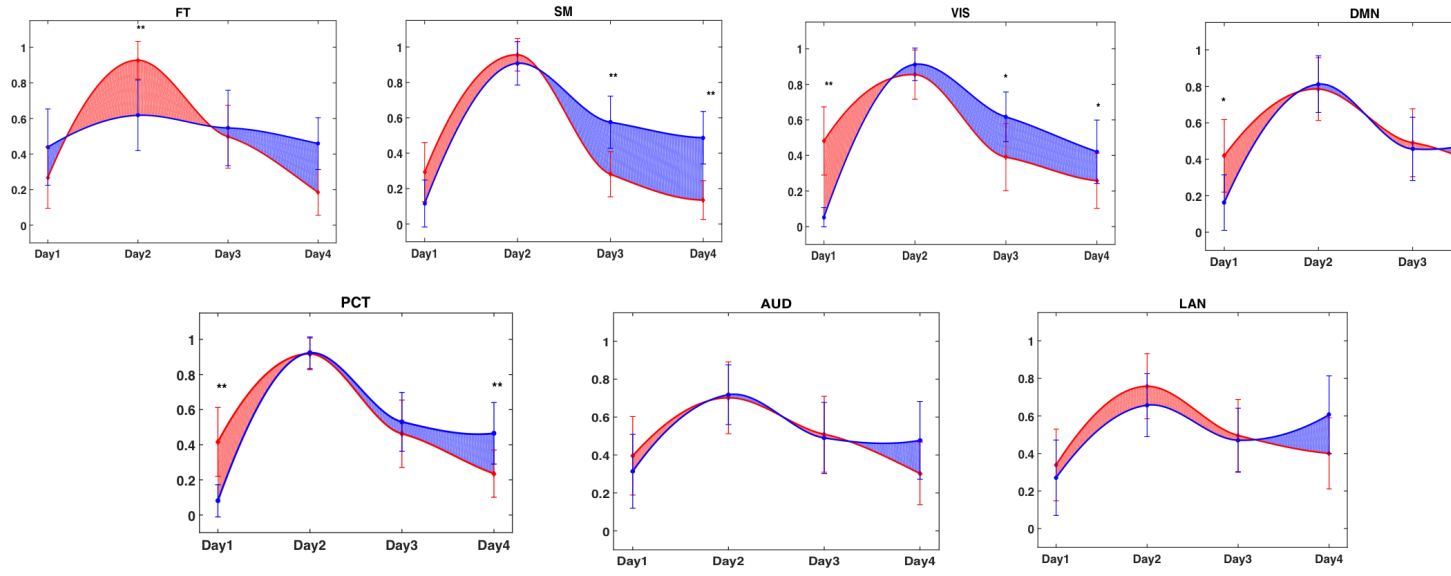


Figure 4: **Learning Curve.** The normalized ISC and ISFC curves for the seven network communities in order statistical significance (see Table 3). There was larger significant difference in day 3 and 4 especially for SM, VIS, DMN and auditory as ISC significantly decreased while ISFC increased

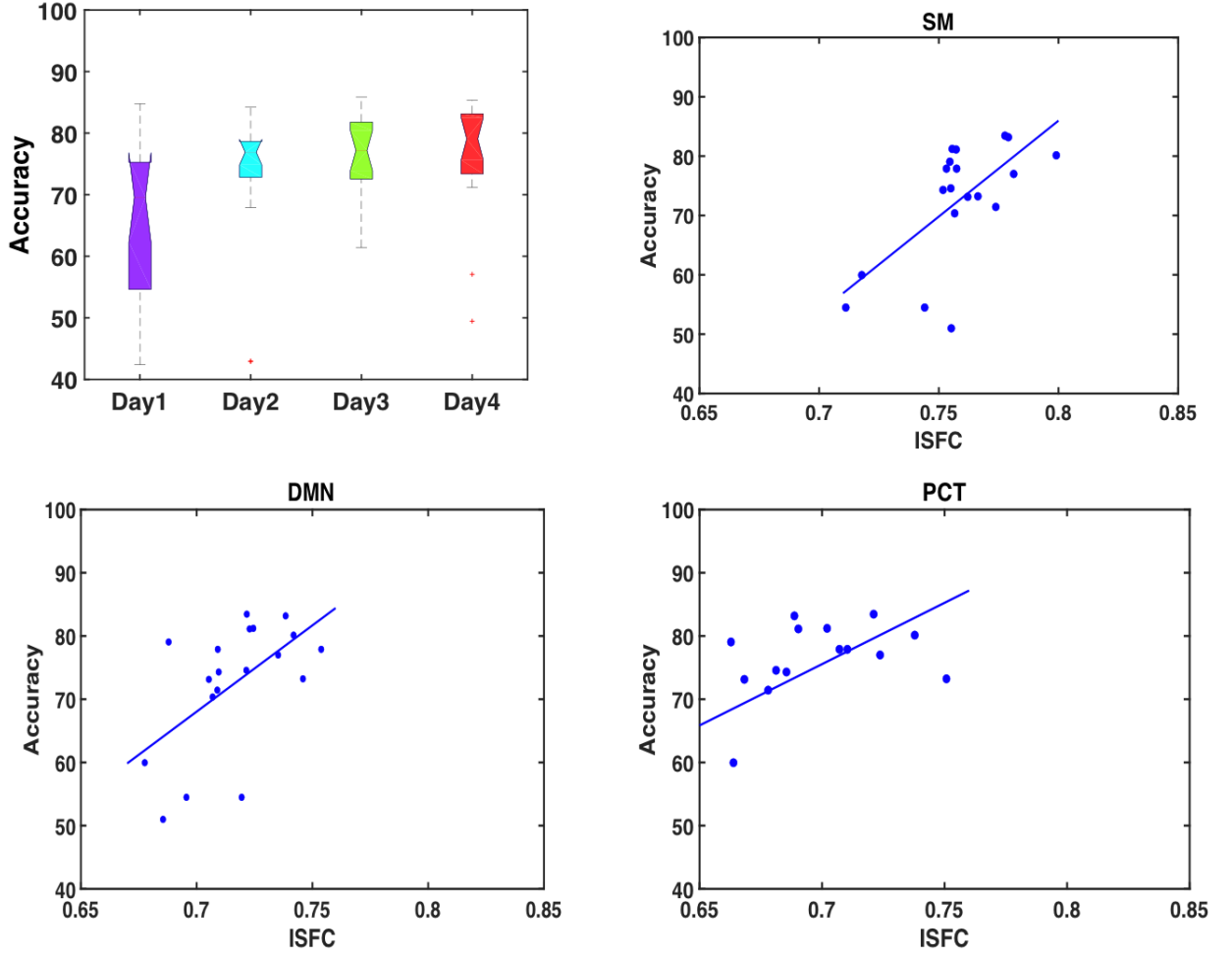


Figure 5: **Accuracy and ISFC.** (A) Average accuracy across the subject which increased with day. (B, C) Bilateral caudate's ISFC was significantly correlated with the accuracy across the subjects which indicate that ISFC is driving by the learning task performance.