Efficient Coding in the Economics of Human Brain Connectomics

Dale Zhou[1], Christopher W. Lynn[2,3], Zaixu Cui[4], Rastko Ciric[5], Graham L. Baum[6], Tyler M. Moore[4,7], David R. Roalf[4,7], John A. Detre[8], Ruben C. Gur[4,7], Raquel E. Gur[4,7], Theodore D. Satterthwaite[4,7 †], and Dani S. Bassett[4,8,9,10,11,12,13 †]

[1]Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[2]Initiative for the Theoretical Sciences, Graduate Center, City University of New York, New York, NY 10016, USA
[3]Joseph Henry Laboratories of Physics, Princeton University, Princeton, NJ 08544, USA
[4]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[5]Department of Bioengineering, Schools of Engineering and Medicine, Stanford University, Stanford, CA 94305 USA
[6]Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA 02138, USA
[7]Penn-Children's Hospital of Philadelphia Lifespan Brain Institute, Philadelphia, PA 19104, USA
[8]Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[9]Department of Physics & Astronomy, College of Arts and Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA
[10]Department of Bioengineering, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA
[11]Department of Electrical & Systems Engineering, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA
[12]Santa Fe Institute, Santa Fe, NM 87501, USA
[13]To whom correspondence should be addressed: dsb@seas.upenn.edu
[†]Co-senior authors

November 2, 2021

Abstract

In systems neuroscience, most models posit that brain regions communicate information under constraints of efficiency. Yet, evidence for efficient communication in structural brain networks characterized by hierarchical organization and highly connected hubs remains sparse. The principle of efficient coding proposes that the brain transmits maximal information in a metabolically economical or compressed form to improve future behavior. To determine how structural connectivity supports efficient coding, we develop a theory specifying minimum rates of message transmission between brain regions to achieve an expected fidelity, and we test five predictions from the theory based on random walk communication dynamics. In doing so, we introduce the metric of compression efficiency, which quantifies the trade-off between lossy compression and transmission fidelity in structural networks. In a large sample of youth (n = 1,042; age 8-23 years), we analyze structural networks derived from diffusion weighted imaging and metabolic expenditure operationalized using cerebral blood flow. We show that structural networks strike compression efficiency trade-offs consistent with theoretical predictions. We find that compression efficiency prioritizes fidelity with development, heightens when metabolic resources and myelination guide communication, explains advantages of hierarchical organization, links higher input fidelity to disproportionate areal expansion, and shows that hubs integrate information by lossy compression. Lastly, compression efficiency is predictive of behavior—beyond the conventional network efficiency metric—for cognitive domains including executive function, memory, complex reasoning, and social cognition. Our findings elucidate how macroscale connectivity supports efficient coding, and serve to foreground communication processes that utilize random walk dynamics constrained by network connectivity.

Author Summary. Macroscale communication between interconnected brain regions underpins most aspects of brain function and incurs substantial metabolic cost. Understanding efficient and behaviorally meaningful information transmission dependent on structural connectivity has remained challenging. We validate a model of communication dynamics atop the macroscale human structural connectome, finding that structural networks support dynamics that strike a balance between information transmission fidelity and lossy compression. Notably, this balance is predictive of behavior and explanatory of biology. In addition to challenging and reformulating the currently held view that communication occurs by routing dynamics along metabolically efficient direct anatomical pathways, our results suggest that connectome architecture and behavioral demands yield communication dynamics that accord to neurobiological and information theoretical principles of efficient coding and lossy compression.

Keywords. network communication dynamics, metabolic resources, lossy compression, rate-distortion theory, integration, hierarchical organization, network hubs

Introduction

The principle of compensation states that "to spend on one side, nature is forced to economise on the other side" [1]. In the economics of brain connectomics, natural selection optimizes network architecture for versatility, resilience, and efficiency under constraints of metabolism, materials, space, and time [2, 1, 3]. Brain networks—composed of nodes representing cortical regions and edges representing white matter tracts—strike evolutionary compromises between costs and adaptations [2, 1, 4, 5, 6, 7]. Further, network disruptions may contribute to the development of neuropsychiatric disorders [8, 9, 10, 11]. Because the limits of computations are intertwined with the limits of communication between brain regions [12], to understand how the brain efficiently balances resource constraints with pressures of information processing, one must begin with models of information transmission in brain networks.

Principles of neurotransmission established at the cellular level suggest that biophysical constraints on information processing may apply to the macroscopic levels of brain regions and networks [13, 14, 2]. The principle of efficient coding proposes that the brain transmits maximal information in a metabolically economical or *compressed* form to improve future behavior [15]. Efficient coding at the macroscale offers a parsimonious principle of compression characterizing the dimensionality of neural representations [16, 17, 18, 19], as well as a parsimonious principle of transmission characterizing a spectrum of network communication mechanisms [3, 20, 21, 5, 22, 23, 24]. However, it remains incompletely understood how this principle generalizes from cells and sensory systems to the macroscale connectome [13, 15].

An unexplored link between the efficient coding of compressed transmissions and macroscale brain network communication dynamics is rate-distortion theory, a major branch of information theory that establishes the mathematical foundations of *lossy data compression* for any communication channel [25]. Rate-distortion theory formalizes the link between compression and communication by determining the minimum amount of information that a source should transmit (the rate) for a target to approximately receive the input signal without exceeding an expected amount of noise (the distortion) [25]. Lossy data compression is reducing the amount of information transmitted (rate), accepting some loss of data fidelity (distortion). Using a rate-distortion model, we sought to explain how the macroscale connectome supports efficient coding from minimal assumptions.

We modeled information transmission as the passing of stochastic messages in parallel along the wiring of the

human connectome (**Figure 1A**). More precisely, we modeled information transmission using a repetition code [26, 12]. A repetition code uses redundancy—here by sending multiple copies of a message—to overcome errors in communication arising from the stochasticity of neural processes [13, 12]. A natural trade-off emerges between the redundancy and efficiency of a message: while redundant messages are more robust to errors in transmission, they also incur greater cost [26]. Thus, given an allowed error rate (or, equivalently, an expected fidelity), maximizing the efficiency of information transmission requires minimizing the redundancy of messages [12]. By quantifying the minimal number of repeated messages needed to achieve a given fidelity, questions of connectome computation and communication can be formulated as a tractable mathematical problem using stochastic processes and redundancy reduction [13, 24, 27].

A parsimonious model of information transmission in connectomes emerges naturally from two key assumptions. The first assumption is that stochastic transmission entails an economy of discrete impulses where the immediate future state only depends on the current state [26, 13]. Mathematically, this assumption casts transmission as a linear process, or a random walk, wherein message copies (identical random walkers) propagate along structural connections with probabilities proportional to the connection's microstructural integrity [32, 24]. Biologically, macroscale random walk models are supported by their ability to predict the transsynaptic spread of pathogens [33, 34, 35], as well as the directionality and spatial distribution of neural dynamics from structural connectivity [21, 36, 37, 38]. The second assumption is that the impulse can lose information but never generate additional information over successive steps of propagation [39]. Mathematically, this assumption represents the data processing inequality, which states that a random walker can only lose (and never gain) information about an information source [12]. Biologically, the assumption is supported by increasing temporal delay, signal mixing, and signal decay introduced by longer paths [40, 13].

Combining these two assumptions, if packets of information propagate along structural pathways and information can only be lost with each step, then the shortest pathway between two brain regions yields an upper bound on the fidelity with which they can communicate. This key conclusion allows one to formulate the probability that a message propagates along the shortest path as an effective fidelity for the communication between two regions, thereby operationalizing the notion of *distortion*. Moreover, by modeling messages as random walkers, one can operationalize the notion of *rate* by computing the number of messages that must be sent to ensure that at least one transmits along the shortest path; that is, to ensure that at least one message reaches a specified receiver with maximum fidelity [20]. We applied our model to 1,042 youth (aged

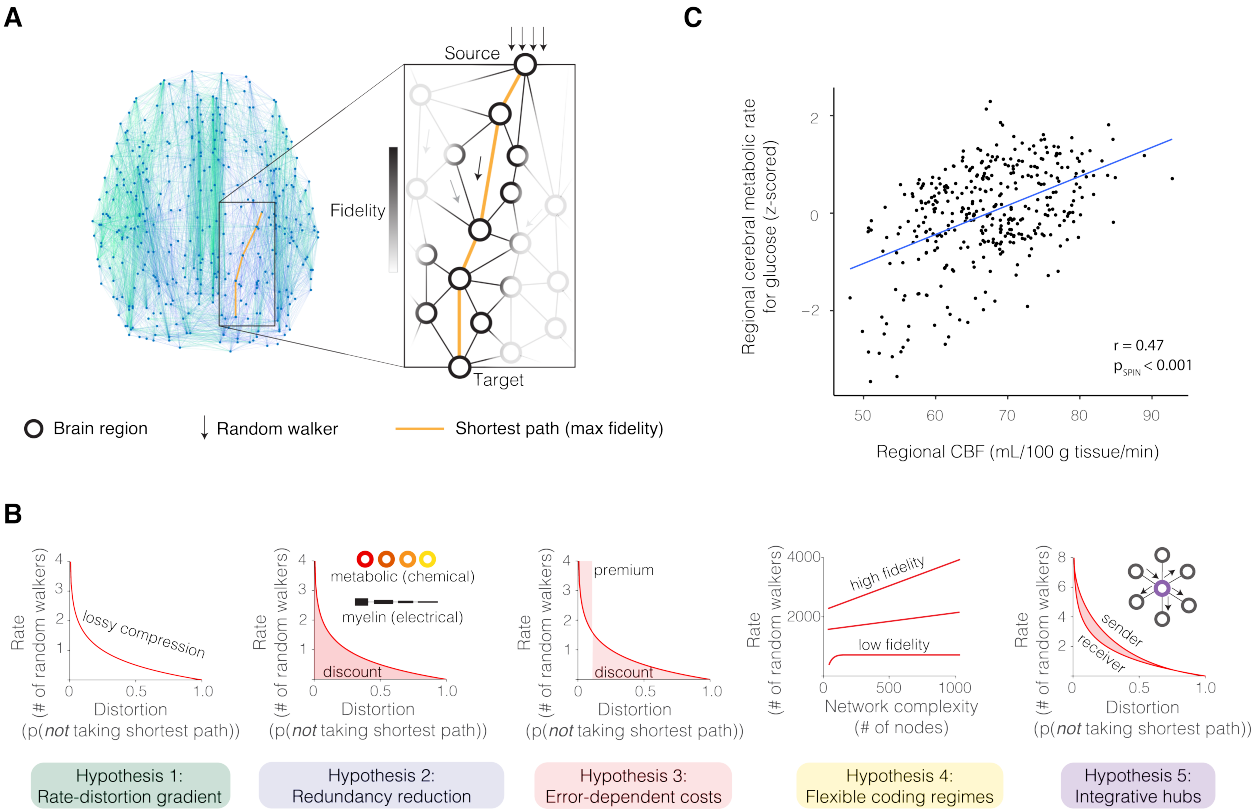

Figure 1: **The efficient coding principle and associated hypotheses. (A)** To model efficient coding, we applied rate-distortion theory, a branch of information theory that provides the foundations of lossy compression [25], to a previously proposed network measure of random walk dynamics [20]. This metric models random walks atop the structural connectome to solve for the number of random walkers required for at least one to propagate by the shortest path given an expected probability. After determining that the minimal amount of noise is achieved by signals that randomly walk along shortest paths, we calculated the optimal information rate to communicate between brain regions with an expected transmission fidelity in the rate-limited (or capacity-limited) structural network. **(B)** Across the brain, regional CBF measured from 1,042 participants in our study correlated with the regional cerebral metabolic rate for glucose acquired from published maps of 33 healthy adults (Pearson's correlation coefficient $r = 0.47, df = 358, p_{\text{SPIN}} < 0.001$), replicating prior findings and supporting the operationalization of metabolic expenditure using CBF [28, 29]. We used CBF to investigate the relationship between metabolic demands and network organization supporting either shortest path routing (a previously posited communication mechanism) or random walks (our posited communication mechanism). **(C)** Importantly, our model generates five predictions from the literatures of rate-distortion theory and neural information processing [27, 30, 31].

8-23 years) from the Philadelphia Neurodevelopmental Cohort who underwent diffusion weighted imaging (DWI; see **Supplementary Figure 1**) [41]. To operationalize metabolic expenditure, we used arterial-spin labeling (ASL) MRI, which measures cerebral blood flow (CBF) and is correlated with glucose expenditure [28, 29].

To evaluate the validity of the efficient coding model, we assessed five published predictions of any communication system adhering to rate-distortion theory, which we adapted to connectomes and distinguished from alternative explanations of brain network communication dynamics (**Figure 1B**) [31, 20, 21, 30, 27]. First, information transmission should produce a characteristic rate-distortion gradient in biological and artificial networks, where exponentially increasing information rates are required to minimize signal distortion. Second, transmission efficiency should improve with manipulations of the communication system designed to facilitate signal propagation, where information costs decrease when randomly walking messages are biased with regional differences in metabolic rates and intracortical myelin. Third, the information rate should vary as a function of the costs of error, with discounts when costs are low and premiums when costs are high. Fourth, brain network complexity should flexibly support communication regimes of varying fidelity, where a high-fidelity regime predicts information rates that monotonically increase as the network grows more complex, and a low-fidelity regime predicts asymptotic information rates indicative of lossy compression. Fifth and finally, structural hubs should integrate incoming signals to efficiently broadcast information, where hubs (compared to other brain regions) have more compressed input rates and higher transmission rates for equivalent input-output fidelity. As described below, this model advances the current understanding of how information processing is associated with behaviors in a range of cognitive domains, subject to constraints on metabolic resources and network architecture.

Results

Macroscale efficient coding can be understood by communication processes of random walks but not the alternative model of shortest path routing

To understand how the brain balances the transmission rate of stochastic messages and signal distortion across different network architectures, we developed a model positing random walk communication dynamics atop the structural connectome. We tested this model by comparing it to the alternative hypothesis of shortest-path routing (see **Supplementary Modeling/Math Notes**). The random walk model and

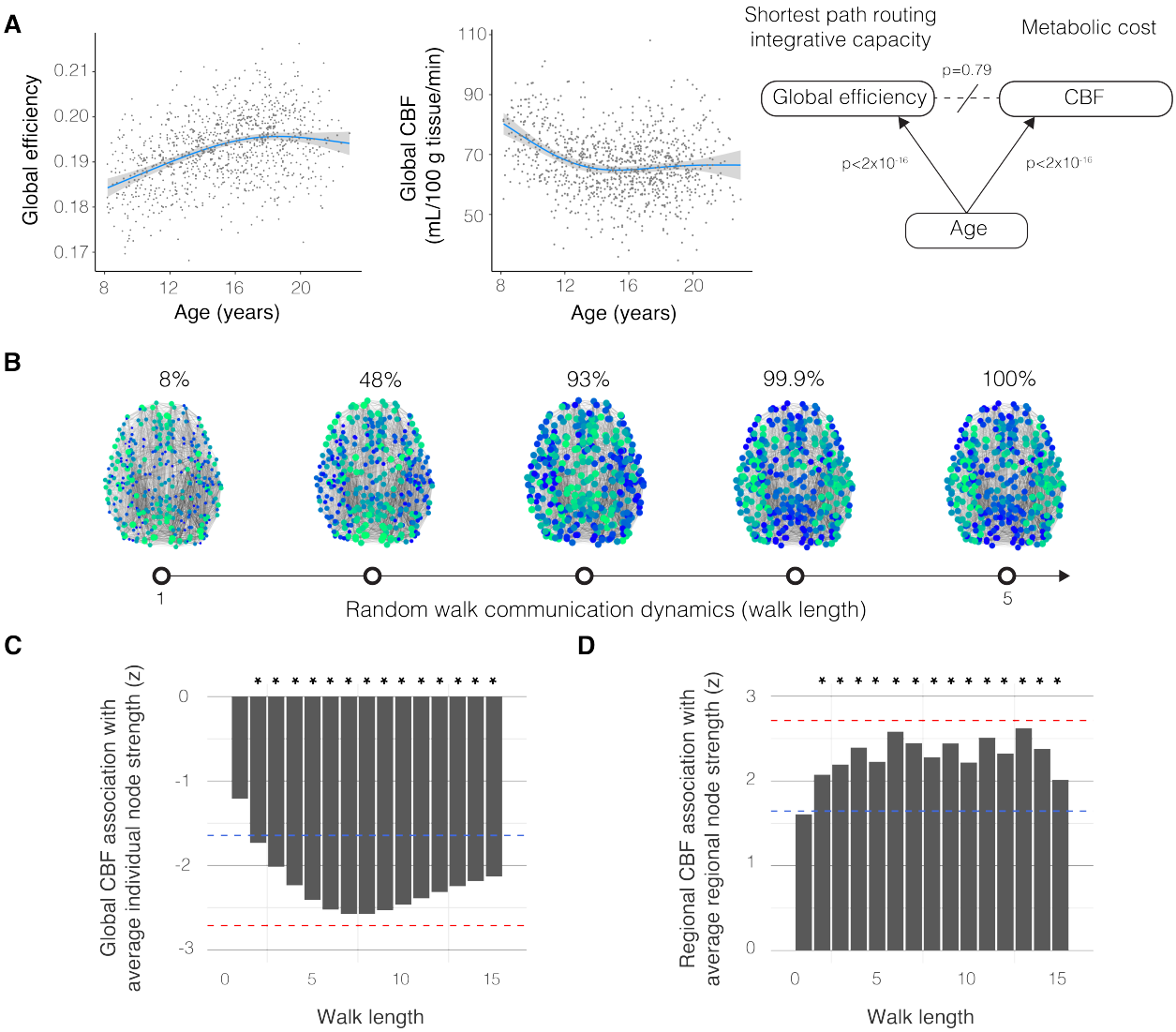

Figure 2: **Metabolic running costs support brain network architectures for random walk dynamics, not shortest path routing.** **(A)** Global efficiency increased with development while cerebral perfusion declined ($F_{\text{global efficiency, age}} = 50$, estimated $df = 3.46$, $p < 2 \times 10^{-16}$; $F_{\text{CBF,age}} = 69.22$, estimated $df = 3.74$, $p < 2 \times 10^{-16}$; confidence intervals displayed as gray bands). Contrary to prior reports, the relationship between CBF and global efficiency was confounded by age ($r = 0.01, df = 1039, p = 0.79$). Therefore, the claim of reduced metabolic cost associated with shortest path routing is weakened. **(B)** All brain regions are accessible to signals randomly walking across 5 edges or more. Bluer nodes represent greater normalized node strengths, reflecting the accessibility of the brain region to a random walker propagating along the structural connectome with differing walk lengths. Whereas the shortest path routing model predicts that only shortest path walks of length 5 or less are associated with CBF, the random walk model predicts that longer random walk paths are also associated with CBF. **(C)** Global CBF was negatively correlated with the strengths of structural paths with length 2 or more when controlling for age, sex, age-by-sex interactions, degree, density, and in-scanner motion ($t = -1.59$ to -2.81, estimated model $df = 11.45$). Asterisks denote statistical significance following correction for multiple comparisons by FDR; blue dashed line: FDR-corrected $p = 0.05$, red dashed line: Bonferroni-corrected $p = 0.05$ for 15 tests. Metabolic expenditure was related to structure supporting random walks rather than shortest path routing. **(D)** Regional CBF was positively correlated with high-integrity connections comprising the structural scaffolds supporting differing walk lengths ($\rho = 0.12$ to 0.14, $df = 358$, FDR-corrected $p < 0.05$), controlling for age, sex, age-by-sex interaction, degree, density, and in-scanner motion. Blue dashed line: FDR-corrected $p = 0.05$, red dashed line: Bonferroni-corrected $p = 0.05$ for 15 tests. Together, these data provided some evidence for metabolic cost associated with a regional profile of white matter path strengths that support random walk dynamics. Individuals with greater path strengths tended to have lower global metabolic expenditure, while brain regions with greater path strengths tended to have greater metabolic expenditure.

shortest path model are currently viewed as opposing extremes of a spectrum of communication processes [20, 5, 24]. In network neuroscience, shortest path routing anchors metrics of communication dynamics and information integration [42, 43, 24]. In cognitive neuroscience, the neural circuit related to behavior is commonly depicted as a subset of brain regions communicating their specialized information to each other across shortest and direct anatomical connections [44]. Although shortest path routing has acknowledged shortcomings as a model of communication dynamics (see **Supplementary Modeling/Math Notes**), a key extenuating hypothesis of the model is reduced metabolic cost [3, 24]. Yet, existing macroscale evidence for this model remains sparse [45].

We sought to determine how brain metabolism is associated with structural signatures of shortest path versus random walk models. To quantify the extent to which a person's brain is structured to support shortest path routing, we used a statistical quantity known as the *global efficiency* [46], a commonly used measure of the average shortest path strength between all pairs of brain regions. Intuitively, global efficiency represents the ease of routing information by shortest paths and is proportional to the strength of shortest paths in a network [42]. As an operationalization of metabolic running cost, we considered CBF, which is correlated with glucose consumption [28, 29]. To test the spatial correlation between CBF and glucose consumption, we used a spatial permutation test that generates a null distribution of randomly rotated brain maps that preserves the spatial covariance structure of the original data; we denote the p-value that reflects significance as p_{SPIN} (Method 7.7.6). We observed a linear association between CBF and glucose consumption (**Figure 1C**; Pearson's correlation coefficient $r = 0.47, df = 358$, $p_{\text{SPIN}} < 0.001$).

Next, we tested if shortest path routing is linked to a decrease in metabolic expenditure, operationalized as a negative correlation between global efficiency and CBF [45, 3]. Controlling for mean gray matter density, sex, mean degree, network density, and in-scanner motion, we found that the global efficiency was negatively correlated with CBF ($r = -0.20, df = 1039, p < 0.001$), consistent with prior reports [45]. Notably, we did not regress out age in the previous analysis in order to align with the prior analysis that we aimed to replicate [45]. We were also interested in determining whether development had any effect on the relationship between global efficiency and CBF. Our interest was justified by the positive correlation between age and global efficiency (**Figure 2A**; $F = 50$, estimated $df = 3.46$, $p < 2 \times 10^{-16}$) and the negative correlation between age and CBF ($F = 69.22$, estimated $df = 3.74$, $p < 2 \times 10^{-16}$). After controlling for age we did not find a significant relationship between global efficiency and CBF ($r = 0.01, df = 1039, p = 0.79$), suggesting

that co-linearity with age drove the initial observed association between CBF and global efficiency. This null result undermines the claim that shortest path routing is associated with reduced metabolic expenditure.

Rather than being driven by shortest path routing, metabolic expenditure could instead be associated with communication by random walks. Each brain region can reach every other brain region via random walks along paths of 5 connections (**Figure 2B**). A random walker will likely not take the most efficient paths and must instead rely on the structural strengths of longer paths. Hence, if brain metabolism is associated with communication by random walks, then CBF should correlate with the strength of the white matter paths greater than length 5. To evaluate this prediction, we computed the strength of connections across different path distances using the matrix exponent of the structural network (see Method 1.5.2 and **Supplementary Figure 1B**). We then tested the association between longer paths and metabolic expenditure across individuals (**Figure 2C**) and across regions (**Figure 2D**). In first considering variation across *individuals*, we found that the average node strengths for walks of length 2 to 15 were negatively correlated with CBF ($t = -1.59$ to -2.81, estimated model $df = 11.45$, FDR-corrected $p < 0.05$), after controlling for age, sex, age-by-sex interaction, average node degree, network density, and in-scanner motion (**Figure 2C**). The negative correlations between CBF and the average connection strengths suggest that the greater the connection integrity, the lower the metabolic expenditure. In next considering variation across *brain regions*, we found that the average node strengths for walks of length 2 to 15 were positively correlated with CBF (Spearman's rank correlation coefficient $\rho = 0.12$ to 0.14, $df = 358$, FDR-corrected $p < 0.05$), after controlling for age, sex, age-by-sex interaction, average node degree, network density, and in-scanner motion (**Figure 2D**). The positive correlations between CBF and the average connection strengths suggest that brain regions with greater path strengths tended to have higher metabolic expenditure. See **Supplementary Figures 2-4** for metabolic costs associated with other connectivity metrics supporting random walks. Together, we found no evidence of metabolic expenditure associated with shortest path routing, whereas the convergent findings of an association between CBF and random walk path strengths across individuals and regions provided some evidence that metabolic running costs were linked to random walk communication dynamics.

Using the random walk model, we formalized a rate-distortion model of efficient coding by assuming that the minimal amount of noise is achieved by messages that randomly walk along shortest paths (**Figure 3A-D**). We defined rate as the number of random walkers per transmission, and distortion as the probability of the random walkers *not* taking the shortest path. We evaluated the validity of the redundancy reduction

implementation of efficient coding [26]. Reducing redundancy in repetition coding is equivalent to minimizing the number of random walkers [26, 25, 12]. To understand how the brain balances information rate and distortion, we measured the number of random walkers that are required for at least one to randomly walk along the shortest path to a target cortical region, with an expected probability (Method 1.5.5). This measure of random walk dynamics is based on a prior metric [20]. The number of random walkers can be used to calculate the transmission length of a neural message (in units of bits) or the information rate (in units of bits per second; see Method 16 and **Supplementary Figure 3**). To evaluate the roles of random walk dynamics and rate-distortion theory in the brain, we assessed five previously published predictions of rate-distortion theory and information processing (**Figure 1C**) [30, 27, 31].

Hypothesis 1: Rate-distortion gradient

The first prediction of rate-distortion theory is that communication systems including both brain networks and artificial random networks should produce an information rate that is an exponential function of distortion because biological and engineered systems are governed by the same information-theoretic trade-offs [27]. To test this prediction, we computed the average number of random walkers over all nodes in the network for a given individual, with the probability of randomly walking along the shortest path ranging from 10% to 99.9% (**Figure 3D**). We compared the number of random walkers required of structural connections in the brain with that of random walkers required of connections in random Erdős-Renyí networks, which have larger probabilities of shortest path communication compared amongst canonical random networks [46, 20]. Hence, Erdős-Renyí networks serve as an optimal benchmark for the efficiency of random walk communication and shortest path routing (Method 1.6). For each individual network, the information-theoretic trade-off between information rate and signal distortion was defined by a rate-distortion gradient (**Figure 3E**). The gradient shows that distortion increases as the information rate decreases, which is the hallmark feature of lossy compression. Next, we considered the extent to which the brain's structural connectome prioritizes compression versus fidelity. We refer to this trade-off as the *compression efficiency* (**Figure 3E**), and define it as the slope of the rate-distortion gradient (Method 1.5.6). With random walk communication dynamics, increased compression efficiency prioritizes lossy compression, while decreased compression efficiency prioritizes transmission fidelity.

Consistent with the first prediction of rate-distortion theory, we observed an exponential gradient in every individual brain network and the Erdős-Renyí random networks. Furthermore, the random networks, which

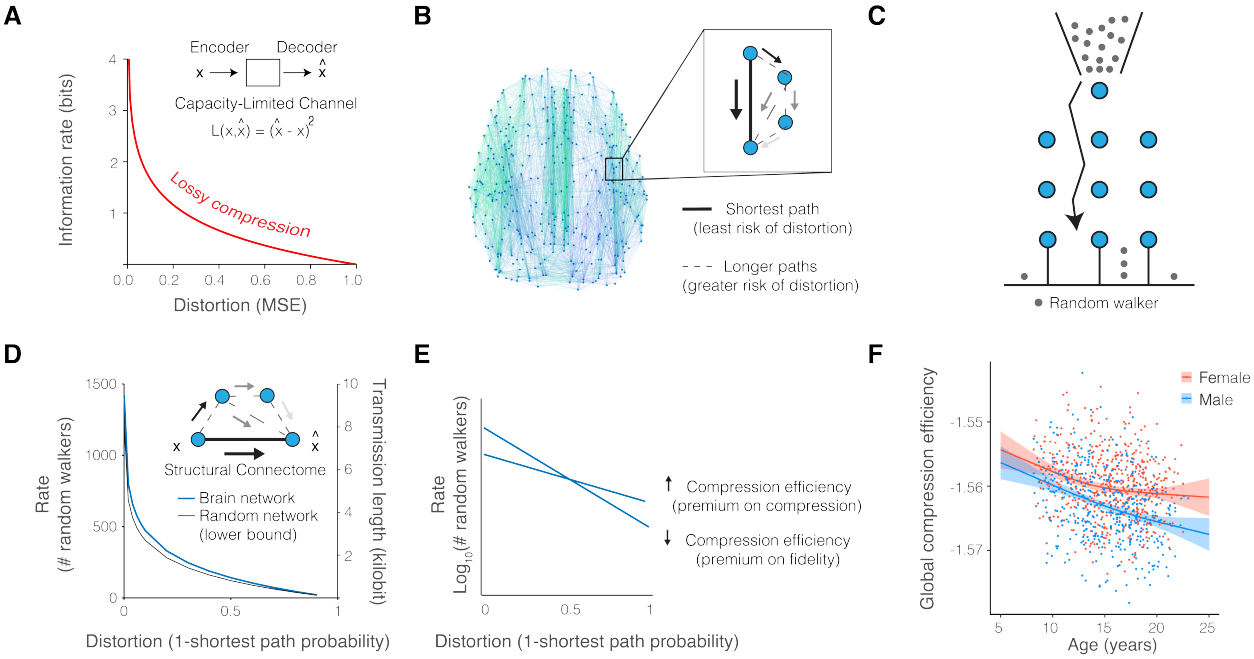

Figure 3: **Lossy compression supports efficient coding in brain networks. (A)** Rate-distortion theory is a mathematical framework that defines the required amount of information for an expected level of signal distortion during communication through capacity-limited channels. Loss functions, such as the mean squared error, are defined to map encoded and decoded signals. In a given channel, compressed messages demand lower information rates at the cost of fidelity. **(B)** In brain networks, we defined the distortion function as the probability of a randomly walking signal *not* taking the shortest path. This definition is reasonable because the temporal delay, signal mixing, and decay introduced by longer paths collectively increase distortion. **(C)** The Galton board depicts the problem of determining the minimum number of (random walkers) that the starting node should prepare to transmit in order for one resource (random walker) to propagate by the specified path given an expected probability. Formalizing the shortest path distortion function allows analytical solutions to determine the corresponding minimum rate of information using an established metric [20]. This metric provided us with the number of random walkers (or amount of information) required for at least one walker to propagate by the shortest path with an expected probability. **(D)** In applying our theory to the structural connectome, we observed a characteristic curve that was consistent with that predicted by rate-distortion theory. The artificial random networks exhibited the rate-distortion gradient also observed in brain networks, supporting the first prediction of rate-distortion theory that both biological and artifical communication systems are governed by the same information-theoretic trade-offs. The blue curve depicts the locally estimated scatterplot smoothing fit of the mean rate-distortion gradient across all individuals. The black curve depicts the mean rate-distortion gradient for Erdős-Renyí random networks whose edges maintain the weights from empirical measurements. The number of random walkers required of brain connectomes was greater than the number required of random networks, across levels of distortion ($F = 9.6 \times 10^5, df = 29120$, bootstrap 95% CI $[9.4 \times 10^5, 10.4 \times 10^5]$, $p < 2 \times 10^{-16}$, adjusted R^2=99.9%, bootstrap 95% CI [0.9989, 0.999]), highlighting the established graph theoretical notion that random networks increase shortest path accessibility but demand a biologically unrealistic cost of materials for structural connections [20]. The information content of a random walker can be calculated in order to interchange the number of random walkers for the transmission length of a neural message in units of bits (see Method 16 and **Supplementary Figure 3**). **(E)** When plotted on a logarithmic axis, the exponential rate-distortion function appears as a straight line. We measured the individually different slopes as differing *compression efficiency*. Then, we interpreted variation across individuals in the language of rate-distortion theory. For example, consider two brain networks functioning at the same low level of distortion. The brain network with the flatter slope between random walkers and distortion has greater compression efficiency because the network architecture confers information rate discounts. In comparison, the brain network with a steeper slope has reduced compression efficiency because the network architecture pays a premium for the same expected fidelity. **(F)** When we considered variation across individuals, we found that compression efficiency decreased with age and differed by sex (**Supplementary Figure 5**; $F = 27.54$, bootstrap 95% CI [1.78, 97.52], estimated $df = 2.17$, $p < 0.001$; adjusted R^2=17.2%, bootstrap 95% CI [0.13, 0.23]; confidence intervals displayed as the colored bands), suggesting that neurodevelopment places a premium on high-fidelity network communication.

are composed of more short connections than empirical brain networks, required significantly fewer random walkers than the empirical brain networks (**Figure 3D** and **Supplementary Figure 3**; $F = 10 \times 10^5, df = 29120, p < 2 \times 10^{-16}$), consistent with the intuition that a greater prevalence of short connections in the random network translates to greater likelihood of shortest path propagation [46, 20]. Rate-distortion trade-offs varied as a function of age and sex, where compression efficiency (**Figure 3E**) was negatively correlated with age ($F = 27.54$, estimated $df = 2.17$, $p < 0.001$), suggesting that neurodevelopment places a premium on fidelity (**Figure 3F**). Compression efficiency was greater on average in females compared to males ($t = 9.53$, $df = 996.82$, $p < 0.001$). The data, therefore, indicate that random walk communication dynamics on biological brain networks differ from random walks on artificial networks, yet each accords well with the prediction of rate-distortion trade-offs governing all communication systems.

Hypothesis 2: Redundancy reduction

The second prediction of rate-distortion theory is that manipulations to the physical communication system (the connectome) that are designed to facilitate information transmission will improve communication efficiency. The prediction stems from two key observations. First, between the rate-distortion gradients for structural connectomes and random networks (depicted in **Figure 3D**) exists a range of possible rate-distortion gradients produced by some other construction of communication networks [25]. Second, it is well known that brain signaling relies extensively on metabolic diffusion and devotes much of its metabolic resources to maintaining a chemical balance that supports neuron firing [13, 48]. At the longer distances of the connectome, myelin in the white matter and cerebral cortex supports the speed and efficiency of electrical signaling in subcortical fiber tracts and in cortico-cortical communication [49, 2, 50]. Hence, modifying the connectome to bias random walk dynamics according to metabolic resources and myelin mimics biological investments in communication efficiency.

The second observation above leads to the hypothesis that including biological biases in random walk probabilities based on the strength of structural connections will improve the efficiency of information transmission. We hypothesized that biasing random walkers with metabolic resources and myelin would reduce the information rate required to communicate a message with a given fidelity compared to random walkers that propagate only by connectome topology. To test this hypothesis, we biased edge weights (Method 1.5.7) representing structural connection strength by multiplying the edge weight by a bias term. Across pairs of connected brain regions, the bias term was either defined as the average, normalized metabolic rate using

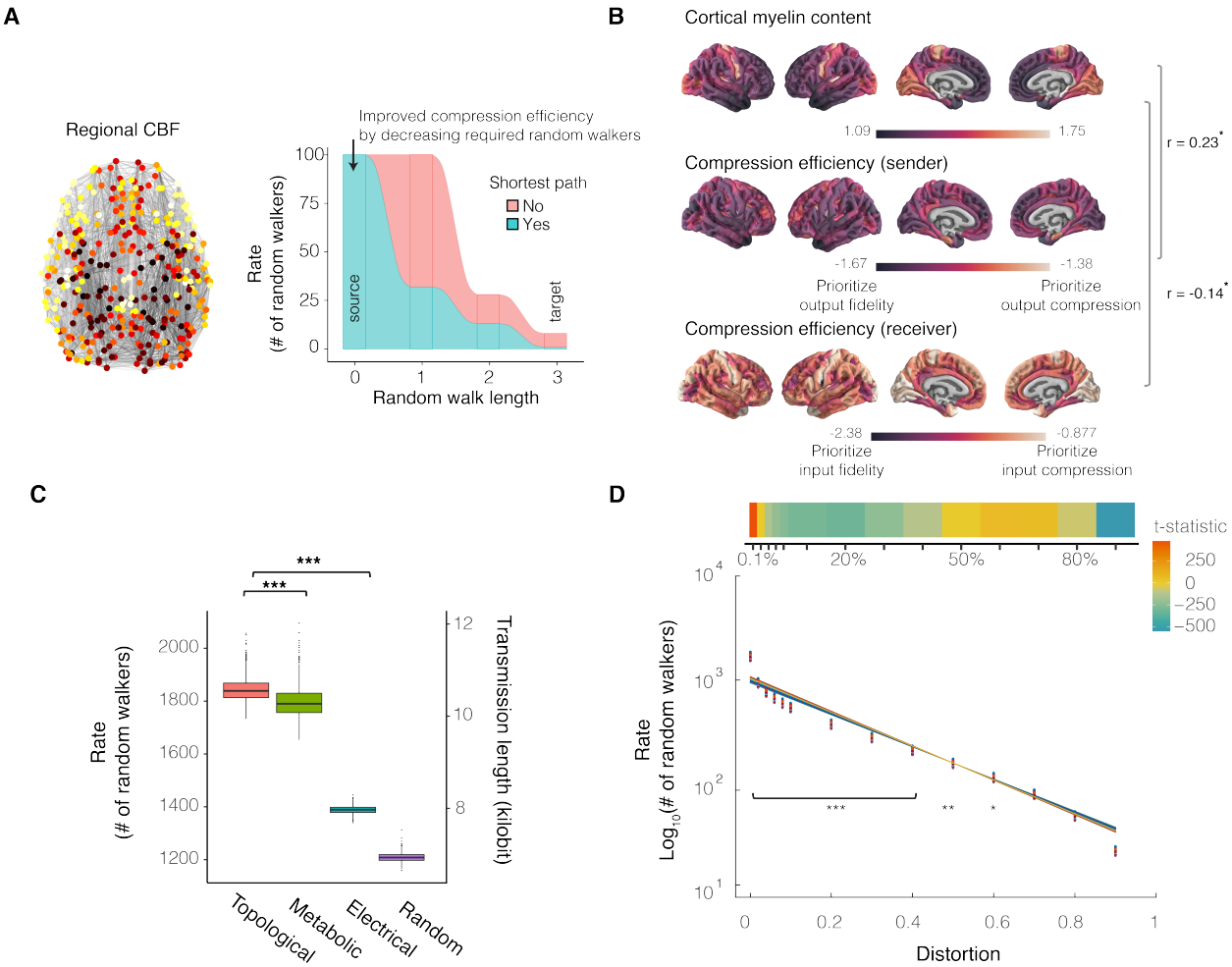

Figure 4: **Connectome architecture and biology support efficient coding. (A)** Biased random walks with transition probabilities modified by regional CBF and myelin content are biologically motivated by two fundamental modes of neural communication: biochemical diffusion and electrical signaling. Left: A biased random walk attracting walkers to regions of high CBF was modeled by constructing a biased structural connectivity matrix. The weight of an edge between region i and region j was given by the average, normalized CBF between pairs of brain regions multiplied by the weight of the structural connection. Right: The alluvial diagram depicts our expected results: if CBF acts as a substrate supporting efficient coding, then we should observe reduced minimum random walkers in random walks biased by CBF compared to random walks propagating only based on connectome topology. **(B)** Sender and receiver compression efficiency differ regionally across the cortex and describe the number of randomly walking messages required to leave from or arrive to a brain region with specified signal fidelity (**Supplementary Figure 6**). Regional values were averaged across individuals. Brain regions differ in myelination according to published maps [47]. Cortical regions with greater levels of myelin tended to have greater sender compression efficiency ($r = 0.23$, $p_{\text{SPIN, Holm-Bonferroni}} = 0.02$) and receiver compression efficiency ($r = 0.14$, $p_{\text{SPIN, Holm-Bonferroni}} = 0.046$), consistent with myelin reducing conduction delay and promoting the efficient trade-off between signal rate and fidelity to reduce the transmission rate while preserving fidelity (see **Supplementary Figure 7**). **(C)** This plot depicts the number of random walkers required to communicate with 0.01% distortion (see **Supplementary Figure 6C** for distortion ranging from 2% to 90%). In agreement with our second hypothesis that connectome structure and biology supports efficient coding, the number of required random walkers decreased in random walks biased by CBF (metabolic) compared to unbiased random walks only based on connectome topology (topological) when the level of distortion was fixed at 0.01% ($t_{\text{topology,metabolic}} = 20.87$, bootstrap 95% CI$[18.72, 23.14]$, $df = 1993.9$). Similarly, fewer random walkers were required to communicate in the intracortical myelin-biased random walk (electrical) compared to the unbiased random walk $t_{\text{topology,electrical}} = 295.93$, bootstrap 95% CI$[281.19, 312.76]$, $df = 1225.9$). Random walks on random networks required the least random walkers consistent with the notion that random networks increase shortest path

Figure 4: **(continued.)** accessibility but demand a greater materials cost ($t_{\text{topology,random}} = 405.39$, bootstrap 95% CI[$386.3, 427.70$], $df = 1307.3$; ***: $p < 0.001$). **(D)** Here, we depict rate-distortion gradients per individual structural connectome for communicating across a range of transmission fidelity (1-distortion). When plotted on a logarithmic axis, the exponential functions appear linear. The level of distortion and type of random walk fully explained variance in the number of random walkers ($F = 8.1 \times 10^5$, bootstrap 95% CI [$7.8 \times 10^5, 8.5 \times 10^5$], df=58240, adjusted R^2=99.9%, bootstrap 95% CI [0.9986, 0.9987]; all p-values corrected using the Holm-Bonferroni method for family-wise error rate; $* : p < 0.05$, $** : p < 0.01$, $*** : p < 0.001$). According to the third prediction of rate-distortion theory, we should expect asymmetries about the rate-distortion gradient where there are asymmetric costs of error. Specifically, the cost of error should be greater during neural dynamics requiring high-fidelity communication compared to low-fidelity communication. Across individual brain networks, the random walkers required for 0.1% distortion were greater than predicted by the rate-distortion gradient, reflecting a premium placed on very high fidelity signaling. At levels of distortion greater than 2% and less than 60%, the random walkers required were less than that predicted by the rate-distortion gradient. The premium cost of high-fidelity communication compared to discounts of low-fidelity communication supports the third prediction of rate-distortion theory.

CBF or the average, normalized cortical myelin content using published maps of T2/T1w MRI measures with histological validation (**Figure 4A**) [47]. By modeling network communication dynamics, one can calculate directed patterns of transmission as inputs into (receiver) and outputs from (sender) brain regions [37]. We separately computed the send and receive compression efficiency of brain regions to better understand the biological relevance of transmitted information sent or received across the connectome (**Figure 4B**; Method 1.5.6).

We found that brain regions that prioritized input fidelity and output compression tended to have greater myelin content (**Figure 4B**; sender $r = 0.23$, $df = 358$, $p_{\text{SPIN, Holm-Bonferroni}} = 0.02$, receiver $r = -0.14$, $df = 358$, $p_{\text{SPIN, Holm-Bonferroni}} = 0.046$), consistent with myelin's function in neurotransmission efficiency and speed. For a channel communicating at 0.1% distortion (and across all distortions; **Supplementary Figure 6C**), biasing structural edge weights by the biological properties of metabolic and electrical signaling resulted in more efficient communication by reducing the number of redundant random walkers required (**Figure 4C**; $t_{metabolic} = 20.87$, bootstrap 95% CI [$18.72, 23.14$], $df = 1993.9$; $t_{electrical} = 295.93$, bootstrap 95% CI [$281.19, 312.76$], $df = 1225.9$, $p < 0.001$). While both metabolic and electrical signaling supported more efficient communication, electrical signaling was more efficient than metabolic signaling [13, 50]. Compared to rewired null networks preserving the degree sequence (**Supplementary Figure 6D**), structural topology and metabolic resources support communication that prioritizes fidelity ($t_{\text{topological,degree-preserving}}(2074.4) = 121.02, p < 0.001$; $t_{\text{metabolic,degree-preserving}}(2025.8) = 87.78, p < 0.001$), while myelination supports communication that prioritizes compression efficiency ($t_{\text{electrical, degree-preserving}}$ (1208.3) = -122.62, $p < 0.001$). The minimum number of random walkers required for distortion levels less than 60% was explained by the interaction of the distortion level with the type of biased random walk (see

Figure 4C, $F = 6 \times 10^5, df = 29120$, $p_{\text{Holm-Bonferroni}} < 0.05$). Together, these results support the prediction that biological investments can augment connectivity to support efficient communication, especially when transmission prioritizes fidelity.

Hypothesis 3: Error-dependent costs

The third prediction of rate-distortion theory is that the information rate should vary as a function of the costs of errors in communication systems that interact with their environment. If errors are more costly for networks operating at high fidelity, then we should observe an information rate surpassing the minimum predicted by rate-distortion theory. In contrast, if errors are less costly for networks operating at low fidelity, then we should observe no more than the minimum predicted information rate. In testing this prediction, we observed that brain networks commit more random walkers than required for very low levels of distortion, such as 0.1%, but allocate the predicted number of random walkers or fewer to guarantee levels of distortion between 2% and 60% (**Figure 4D**). Hence, the third prediction of rate-distortion theory was consistent with our observation of a premium placed on very low signal distortion and a discounted cost of greater distortion.

Hypothesis 4: Flexible coding regimes

The fourth prediction of rate-distortion theory proposes that communication systems, including the connectome, have distinct network properties that support information transmission in a flexibly high- or low-fidelity regime. With increasing information processing demands, a high-fidelity regime will continue to place a premium on accuracy, whereas a low-fidelity regime will tolerate noise in support of lossy compression. The operating regime depends on the behavioral demands of the environment, indicating the need for a flexible regime that can simultaneously support both high- and low-fidelity communication. In this section, we assess the fourth prediction of rate-distortion theory by testing the more precise hypotheses that large brain networks support communication in a high-fidelity regime, indirect pathways supports a low-fidelity regime, and hierarchical organization supports a flexible regime. We explain and test each hypothesis in turn.

Large networks support a high-fidelity regime and indirect pathways support a low-fidelity regime

Rate-distortion theory predicts that, in a high-fidelity regime, the information rate will monotonically increase with the complexity of the communication system in order to continue to place a premium on accuracy (**Figure 5A**). To evaluate this prediction, we operationalized complexity as network size because size

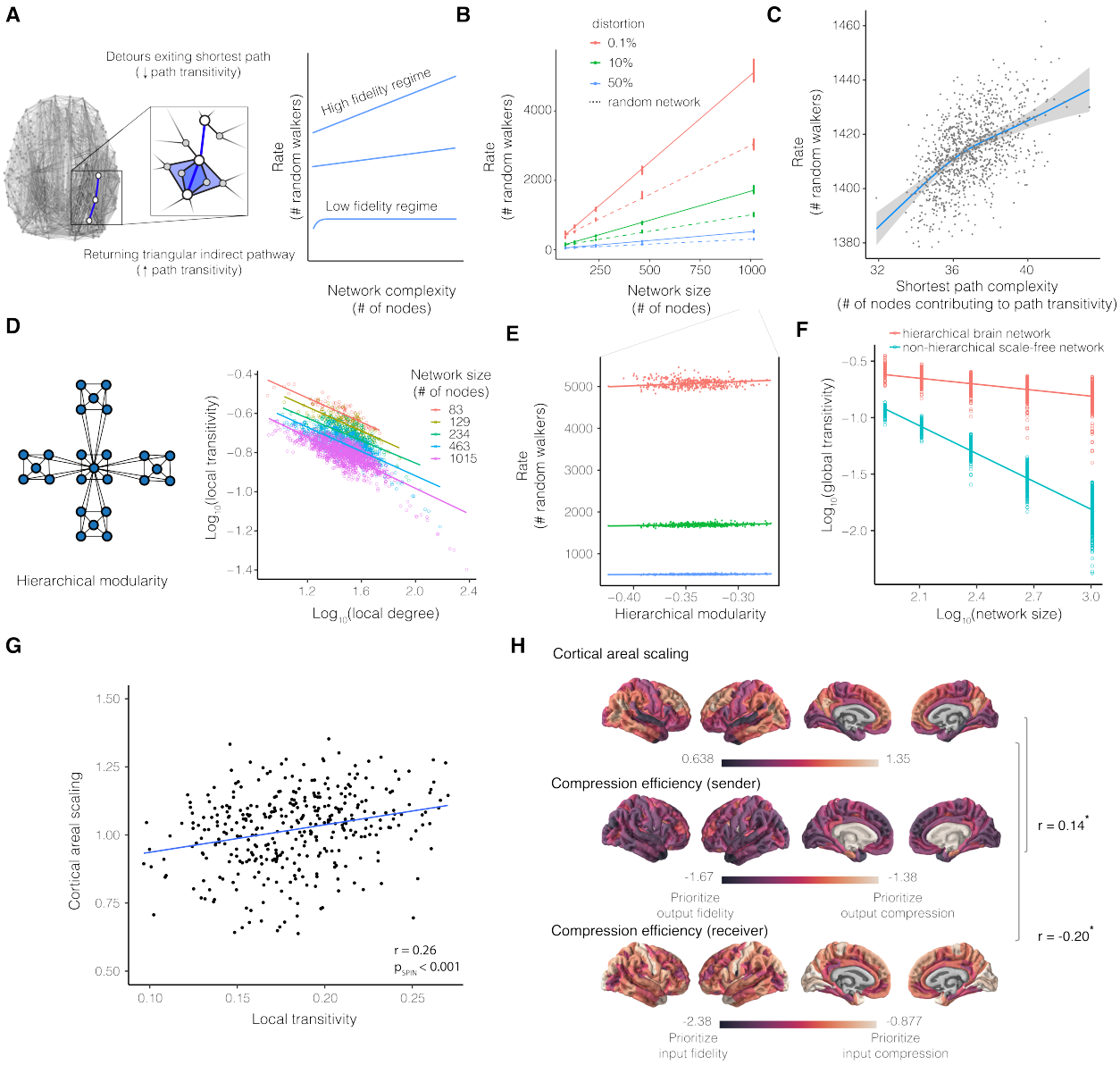

Figure 5: **Constraints of compression efficiency and brain network complexity. (A)** High versus low fidelity communication regimes make different predictions regarding the minimum number of random walkers required for an expected distortion and system complexity. We define system complexity as the network size, or the number of functionally and cytoarchitectonically distinct brain regions. In the high fidelity regime, the brain should place a premium on high fidelity signaling by allocating a monotonically increasing information rate to achieve low distortion despite increasing system complexity. In the low fidelity regime, the brain should place a premium on lossy compression by asymptotically capping the information rate despite increasing system complexity. We expected the number of triangular pathways along the shortest path (i.e., the path transitivity) to approximate the shortest paths with a one-connection longer detour. The fewer the pathways approximating the shortest and highest fidelity pathway, the lower the fidelity of communication. **(B)** Number of random walkers scaled monotonically with network size, consistent with the prediction of a high fidelity regime. Individual brain networks were reparcellated using the Lausanne brain atlases with 83, 129, 234, 463, and 1015 parcels, respectively. Together, the monotonically increasing resource demand highlights a trade-off between high fidelity communication and network complexity, indicating an additional evolutionary constraint on brain network size and complexity. **(C)** Shortest path complexity is the average number of nodes comprising path transitivity. Number of random walkers scaled asymptotically with inter-individual differences in path transitivity, consistent with a low-fidelity regime for lossy compression and storage savings. The non-linear model outperformed a linear model (non-linear $AIC = 7902$, linear $AIC = 7915$; non-linear $BIC = 7964$, linear $BIC = 7968$; confidence intervals displayed as gray band). Lossy compression of a signal

Figure 5: **continued**. by differing densities of path transitivity is akin to image compression by pixel resolution. Reduced path transitivity, or pixel resolution, tends to result in more compression, while greater path transitivity supports greater fidelity. **(D)** Hierarchical modularity is a form of network complexity that is characterized by small groups of nodes organized successively into increasingly larger groups (left). A hallmark of hierarchical modularity is a strict scaling of the transitivity and degree, representing the clustering and scale-free property of nodes (right). Hierarchical organization was evident across the average reparcellated Lausanne brain networks. **(E)** Hierarchical organization was monotonically related to increasing resource demands, but to a lesser cost compared to network size. Hence, hierarchical modularity may support both a high- and low-fidelity regime by increasingly balancing global and local network complexity. **(F)** Hierarchically modular brain networks allowed for greater transitivity (clustering) despite increasing network sizes compared to scale-free networks that do not have a hierarchical architecture. **(G)** Cortical areal scaling in neurodevelopment reflects patterns of evolutionary remodeling. Brain regions with greater transitivity, which supports greater fidelity, tended to disproportionately expand in relation to total brain size during neurodevelopment. **(H)** Brain regions that have higher sender compression efficiency tended to disproportionately expand in relation to total brain size during neurodevelopment ($r = 0.14$, $df = 358$, $p_{\text{SPIN, Holm-Bonferroni}} = 0.046$). Cortical regions with low receiver compression efficiency, placing a premium on information processing fidelity, tended to disproportionately expand in relation to whole brain growth ($r = -0.20$, $df = 358$, $p_{\text{SPIN, Holm-Bonferroni}} = 0.03$). Positively scaling regions that prioritize compression-efficient broadcasting of messages arriving with high fidelity may reflect evolutionary expansion of brain regions with high information processing capacity, whereas negatively scaling regions that prioritize high-fidelity broadcasting of compressed messages may permit other modes of material, spatial, and metabolic cost efficiency.

determines the number of possible states (or nodes) available to each random walker [30]. We re-parcellated each individual brain network at different spatial resolutions to generate brain and random null networks of 5 different sizes representing the original network. We compared the rate for brain networks to the rate for random networks of matched size. In testing this prediction, we observed that the minimum number of random walkers increased monotonically with network size, consistent with a high-fidelity regime, and at a rate different to random networks with matched sizes (**Figure 5B**). Larger brain networks support high fidelity communication by placing a greater premium on accuracy than do larger random networks.

In addition to high-fidelity communication, in a low-fidelity regime, rate-distortion theory predicts that the information rate should plateau as a function of network complexity in order to tolerate noise in support of lossy compression. To evaluate this prediction, we operationalized network complexity supporting low fidelity communication by using path transitivity. Path transitivity quantifies the number of indirect pathways along the shortest path which are a one-connection longer detour (**Figure 5A**, Method 1.5.3). Along the shortest paths, the transitivity is the number of triangles formed by one edge in the shortest path and two connected edges composing an indirect pathway which exits and immediately returns to the shortest path. Operationalizing low fidelity with path transitivity stemmed from interpreting path transitivity using our model assumptions of random walk dynamics.

Applying our random walk model assumptions to path transitivity, if the shortest path represents the struc-

ture supporting highest fidelity because longer paths introduce information loss, then path transitivity's indirect pathways which are just one connection longer than the shortest path are the next-best paths for fidelity. Thus, path transitivity quantifies indirect pathways which offer a random walker the best *approximations* of the highest fidelity path, or the best lossy compression. To operationalize the complexity of the communication system contributing to low-fidelity transmission (better able to tolerate noise in support of lossy compression), we measured the number of nodes in the indirect pathways of path transitivity, which we termed shortest path complexity. In testing the prediction that information rate should plateau as a function of network complexity in low-fidelity regimes, we found that the number of random walkers plateaued non-linearly as a function of shortest path complexity, consistent with low-fidelity communication (**Figure 5C**). Model selection criteria support the non-linear form compared to a linear version of the same model (non-linear $AIC = 7902$, linear $AIC = 7915$; non-linear $BIC = 7964$, linear $BIC = 7968$). The non-linear fit of these data suggest that path transitivity supports low-fidelity communication that is tolerant to noise.

Hierarchical organization flexibly and efficiently supports both a high-fidelity and a low-fidelity regime

Our findings suggest that high fidelity depends on network *size* while low fidelity depends on network *transitivity*. Prior work has shown that hierarchical organization supports the simultaneous presence of networks of *large size and high transitivity* [51]. Therefore, we hypothesized that if the brain network has hierarchical organization, then such organization may enable flexible switching between high- and low-fidelity regimes. Hierarchical organization, wherein submodules are nested into successively larger but less densely interconnected modules, is thought to support efficient spatial embedding as well as specialized information transfer (**Figure 5D**, left) [52]. We first sought to assess if brain networks exhibit hierarchical organization. Hierarchical organization imposes a strict scaling law between transitivity and degree; the slope of this relationship can be used to identify the presence of hierarchically modular organization in real networks. This strict scaling is distinctive of hierarchical networks because increasing the size of a generic non-hierarchical network containing highly connected hubs (degree) will tend to diminish transitivity (clustering); however, hierarchical organization is known to decouple the size and transitivity of a network, which allows each property to vary individually [51]. The characteristic scaling of reduced transitivity of brain regions with higher degree was present across the 5 brain network sizes (83 nodes: $t = -15.31$, bootstrap 95% CI $[-19.37, -12.29]$, $p < 0.001$, $R^2 = 0.74$; 129 nodes: $t = -14.82$, bootstrap 95% CI $[-18.48, -11.93]$, $p < 0.001, R^2 = 0.63$; 234 nodes: $t = -16.96$, bootstrap 95% CI $[-20.43, -13.75], p < 0.001, R^2 = 0.55$; 463 nodes: $t = -26.47$,

bootstrap 95% CI $[-30.02, -23.16]$, $p < 0.001, R^2 = 0.60$; 1015 nodes: $t = -41.88$, bootstrap 95% CI $[-46.10, -38.11], p < 0.001, R^2 = 0.63$; $F = 917.3, df = 1913, p < 0.001$; **Figure 5D**, right). Hence, the connectome exhibits hierarchical organization.

We next assessed the hypothesis that hierarchical organization permits networks to have simultaneously large size and transitivity: two core network properties hypothesized to underlie high- and low-fidelity communication regimes [51]. Hence, we sought to test whether hierarchical organization exhibits the hallmarks of both high- and low-fidelity regimes. If hierarchical organization supports a high-fidelity regime, then the information rate will monotonically increase with the complexity of the communication system. In testing this hypothesis, we found that the scaling characteristic of hierarchical organization in brain networks was associated with monotonically greater information rate ($F = 2002$, bootstrap 95% CI [1946.6, 2061.4], $df = 13389, p < 0.001, R^2 = 0.62,$ bootstrap 95% CI$[0.62, 0.63]$, **Figure 5E**). Considering the slope of the increasing rate, networks with greater hierarchical organization supported high fidelity communication more efficiently ($t = 32.4$, bootstrap 95% CI $[29.96, 34.79]$, $p < 0.001$) than larger networks ($t = 2640.32,$ bootstrap 95% CI $[2556.3, 2725.07]$, $p < 0.001$). Taken together, our findings suggest that hierarchical organization, which is already known to be spatially efficient [52], also supports high-fidelity network communication and does so more efficiently than large networks without hierarchical organization.

To enable flexible switching, hierarchical organization should also support a low-fidelity regime. If path transitivity supports a low-fidelity regime better able to tolerate noise, then hierarchical organization may support a low-fidelity regime by allowing for greater global transitivity in hierarchical networks compared to non-hierarchical networks. In testing this prediction, we found that the hierarchical organization of the connectome exhibited greater transitivity than non-hierarchical scale-free networks (generated with Method 1.6); and this increased transitivity became more pronounced with increasing network size (size-by-hierarchical network type interaction $t = -37.22$, bootstrap 95% CI $[-40.57, -34.12]$, $p < 0.001$; size-by-scale-free network type interaction $t = -125.32,$ bootstrap 95% CI $[-131.71, -118.70]$, $p < 0.001$; **Figure 5F**). Thus, hierarchical organization could contribute to the flexibility of efficient coding by preserving high network transitivity for low fidelity communication that prioritizes noise tolerance and large network size for high fidelity communication that prioritizes accuracy.

Brain regions disproportionately scale in relation to regional prioritization of network communication fidelity or lossy compression

Having found that hierarchical organization preserves high transitivity despite large network size, we sought to understand how transitivity is associated with the areal expansion of brain regions. Evolutionarily new connections may support higher-order and flexible information processing, emerging from disproportionate expansion of the association cortex [4]. In contrast, brain regions that are disproportionately out-scaled by total brain expansion may save material, space, and metabolic resources. To explore how sender and receiver compression efficiency relates to cortical areal expansion, we used published maps of areal scaling; here, allometric scaling coefficients were defined by the non-linear ratios of surface area change to total brain size change over development. Brain regions that disproportionately expanded in relation to total brain size during neurodevelopment tended to have greater transitivity ($r = 0.26$, $df = 358$, $p_{\text{SPIN, Holm-Bonferroni}} = 0.001$), which may support noise-tolerant lossy compression with minimal loss of fidelity in association cortex regions thought to underpin higher-order and flexible information processing. Disproportionately expanding brain regions tended to have greater sender ($r = 0.14$, $df = 358$, $p_{\text{SPIN, Holm-Bonferroni}} = 0.045$; **Figure 5H**) and reduced receiver ($r = -0.20$, $df = 358$, $p_{\text{SPIN, Holm-Bonferroni}} = 0.03$; **Figure 5H**) compression efficiency. While brain regions that disproportionately expanded tend to prioritize messages received with high fidelity, brain regions that are disproportionately out-scaled by total brain expansion tend to receive compressed messages.

Hypothesis 5: Integrative hubs

The fifth and final hypothesis of our model posits that the structural hubs of the brain's highly interconnected rich club supports information integration of randomly walking signals [31]. To explain the hypothesized information integration roles of rich-club structural hubs, we investigated the compression efficiency of messages randomly walking into and out of hub regions compared to that of other regions. In order to identify the rich-club hubs, we computed the normalized rich-club coefficient and identified 43 highly interconnected structural hubs (**Figure 6A**). Next, we computed the send and receive compression efficiency of rich-club hubs compared to all other regions. In support of their hypothesized function, we found that the rich-club hubs required receiving fewer random walkers compared to other regions (Wilcox rank sum test, $W = 12829, p < 0.001$), suggesting prioritization of information compression (or integration). For the rich-club hub to transmit outgoing messages with a fidelity that is equivalent to the incoming messages, the rich-club hubs required sending more random walkers compared to other brain regions ($W = 64, p < 0.001$),

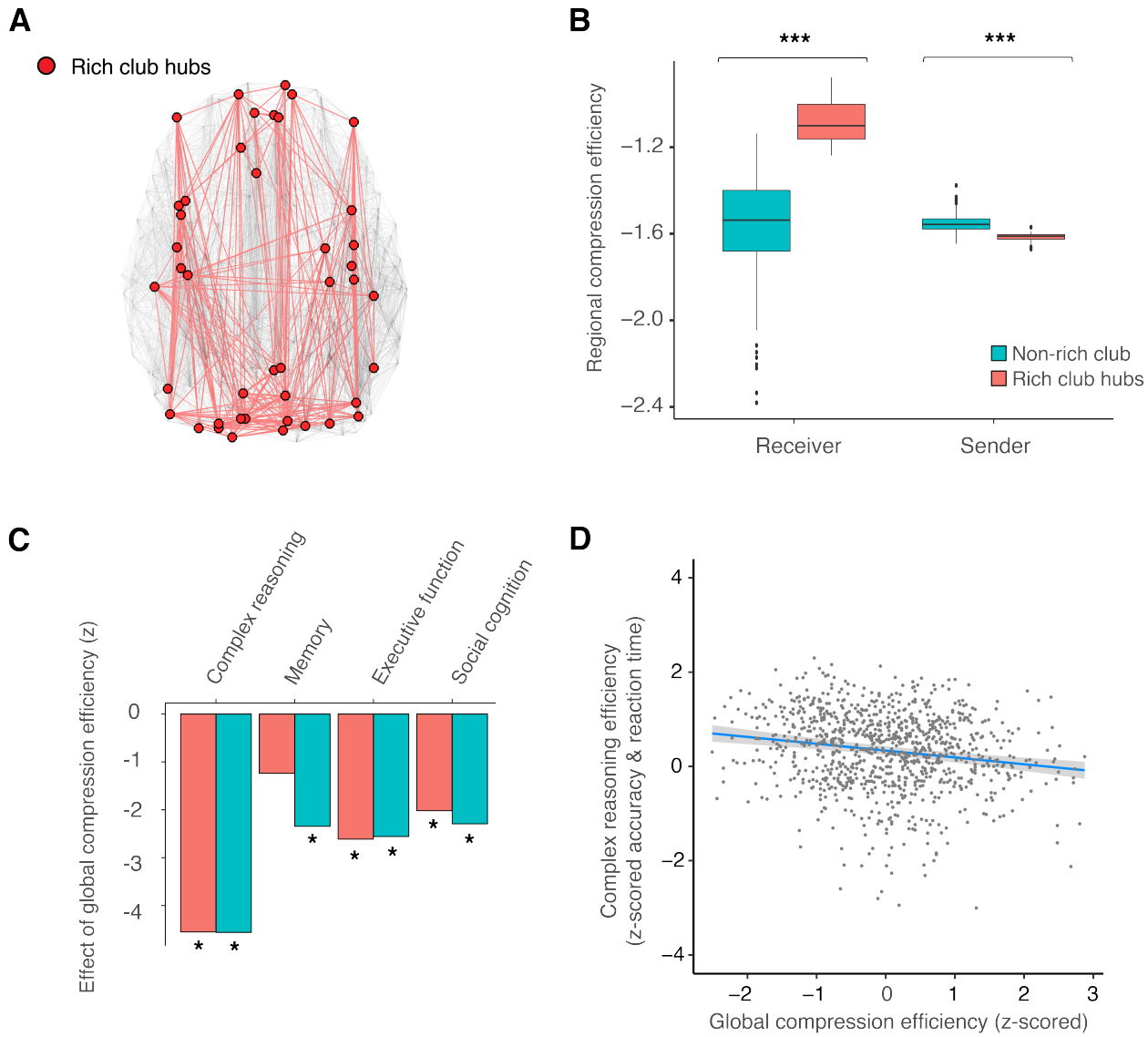

Figure 6: **Compression efficiency explains the integrative role of rich-club hubs and individual differences in cognitive efficiency. (A)** The rich club consists of highly interconnected structural hubs, thought to be the backbone of the brain connectome. Across participants, the rich club was comprised of 43 highly connected brain regions. **(B)** The receiver compression efficiency of rich clubs was greater than other brain regions ($W = 12829, \text{bootstrap } 95\% \text{ CI}[9415.6, 16635.34], p < 0.001$), suggesting that the organization of the structural network prioritizes integration and compression of information arriving to rich-club hubs. The sender compression efficiency of rich clubs was reduced compared to other brain regions, supporting the fidelity of information broadcasting ($W = 64, \text{bootstrap } 95\% \text{ CI}[13.0, 138.03], p < 0.001$). **(C)** Individuals having non-rich-club regions with decreased compression efficiency tended to exhibit greater efficiency (combined speed and accuracy) of complex reasoning (all p-values corrected using the Holm-Bonferroni method; $t = -4.72$, bootstrap 95% CI [-6.63, -2.72], model adjusted R^2=0.21, bootstrap 95% CI [0.16, 0.26], estimated model $df = 10.50$, $p = 2 \times 10^{-5}$), executive function ($t = -2.85$, bootstrap 95% CI [-4.85, -1.0], model adjusted R^2=0.50, bootstrap 95% CI [0.46, 0.56], estimated model $df = 10.64$, $p = 0.03$), and social cognition ($t = -2.30$, bootstrap 95% CI [-4.21, -0.33], model adjusted R^2=0.22, bootstrap 95% CI [0.17, 0.26], estimated model $df = 10.45$, $p = 0.04$), controlling for age, sex, age-by-sex interaction, degree, density, and in-scanner motion. Non-rich-club compression efficiency was not associated with memory efficiency ($t = -1.61$, bootstrap 95% CI [-3.29, 0.07], estimated model $df = 9.85$, $p = 0.11$). Individuals with less rich-club compression efficiency tended to exhibit greater efficiency of complex reasoning ($t = -4.73$, bootstrap 95% CI [-6.58, -2.82], model adjusted R^2=0.21, bootstrap 95% CI [0.16, 0.26], $p = 1 \times 10^{-7}$), memory efficiency ($t = -2.60$, bootstrap 95% CI [-4.41, -0.79], model adjusted R^2=0.24, bootstrap 95% CI [0.16, 0.26], $p = 0.03$), social cognition

Figure 6: **continued**. ($t = -2.55$, bootstrap 95% CI [-4.50, -0.73], model adjusted R^2=0.22, bootstrap 95% CI [0.16, 0.25], $p = 0.02$), and executive function ($t = -2.80$, bootstrap 95% CI [-4.75, -0.87], model adjusted R^2=0.50, bootstrap 95% CI [0.45, 0.56], $p = 0.03$). Overall, individuals with brain structural networks prioritizing fidelity tended to perform with greater accuracy and/or speed in various cognitive functions. Asterisks depict significance following family-wise error correction. **(D)** Individual differences in complex reasoning efficiency were negatively associated with individual differences in compression efficiency ($t = -4.95$, estimated model $df = 11.52$, $p < 0.001$; confidence intervals displayed as the gray band), controlling for global efficiency, age, sex, age-by-sex interaction, degree, density, and in-scanner motion. See **Supplementary Figure 9** for additional scatterplots of the correlation between cognitive efficiency and compression efficiency, as well as for separated metrics of accuracy and speed. For comparison to a commonly used metric of shortest-path information integration, we include global efficiency in the model, which was positively correlated with complex reasoning efficiency ($t = 2.68$, estimated model $df = 11.52$, $p < 0.01$). Global efficiency was not partially correlated with compression efficiency ($r = -0.04, df = 997, p = 0.18$), controlling for age, sex, age-by-sex interaction, degree, density, and motion. Hence, reduced compression efficiency prioritizing fidelity explains individual differences in complex reasoning efficiency.

supporting the notion that rich-club hubs serve as high-fidelity information broadcasting sources. The function of hubs apparently compensates for high metabolic cost [53]. However, we did not find evidence of greater CBF in rich-club hubs compared to non-hubs, suggesting that hubs may be more metabolically efficient than previously known (**Supplementary Figure 8**). This may potentially be due to hubs being compression efficient receivers (see **Supplementary Results** and **Discussion** for additional explanations). Though we did not find evidence of high metabolic cost, the contrasting roles of prioritizing input compression and output fidelity within rich-club hubs was consistent with our hypothesis and the current understanding of rich-club hubs as the information integration centers and broadcasters of the brain's network [31].

Compression efficiency predicts behavioral performance

Efficient coding predicts that the brain should not only balance fidelity and information compression as quantified by compression efficiency, but should do so in a way that improves future behavior [54, 15]. Thus, we next sought to evaluate the association between compression efficiency in rich-club hubs and cognitive performance in a diverse battery of tasks. In light of trade-offs between communication fidelity and information compression, we hypothesized that compression efficiency would correlate with cognitive efficiency, defined as the combined speed and accuracy of task performance. Compression efficiency prioritizing lossy compression and low-dimensionality should predict worse performance [55]. To assess the relationships between compression efficiency and cognitive efficiency, we used four independent cognitive domains that have been established by confirmatory factor analysis to assess individual variation in tasks of complex reasoning, memory, executive function, and social cognition (Method 1.2). Consistent with the hypothesis, we found that individuals having rich-club hubs with reduced compression efficiency, prioritizing transmission fidelity, tended to exhibit increased cognitive efficiency of complex reasoning (all p-values corrected using the Holm-

Bonferroni family-wise error method; $t = -4.73$, bootstrap 95% CI [-6.58, -2.82], model adjusted R^2=0.21, bootstrap 95% CI [0.16, 0.26], estimated model $df = 10.59$, $p = 2 \times 10^{-7}$), memory ($t = -2.60$, bootstrap 95% CI [-4.41, -0.79], model adjusted R^2=0.24, bootstrap 95% CI [0.16, 0.26], estimated model $df = 9.85$, $p = 0.03$), executive function ($t = -2.80$, bootstrap 95% CI [-4.75, -0.87], model adjusted R^2=0.50, bootstrap 95% CI [0.45, 0.56], estimated model $df = 10.69$, $p = 0.03$), and social cognition ($t = -2.55$, bootstrap 95% CI [-4.50, -0.73], model adjusted R^2=0.22, bootstrap 95% CI [0.16, 0.25], estimated model $df = 10.47$, $p = 0.02$). See **Supplementary Figure 9** for similar non-hub correlations, as well as speed and accuracy modeled separately. Our finding that the compression efficiency of rich-club hubs was associated with cognitive efficiency is consistent with the notion that the integration and broadcasting function of hubs contributes to cognitive performance [31]. Importantly, compression efficiency explained variation in cognitive efficiency even when controlling for the commonly used shortest-path measure of global efficiency (**Figure 6D**; compression efficiency $t = -4.95$, estimated model $df = 11.52$, $p < 0.001$; global efficiency $t = 2.68$, estimated model $df = 11.52$, $p < 0.01$). Collectively, individuals with connectomes prioritizing fidelity tended to perform with greater cognitive efficiency in a diverse range of functions, consistent with the understanding of high-dimensionality neural representations predicting flexible behavioral performance [31, 55].

Discussion

The efficient coding principle and rate-distortion theory constrain models of brain network communication, explaining how connectome biology and architecture prioritize either communication fidelity or lossy compression [14, 2, 56, 57, 3, 20, 58, 5, 59, 22, 24]. Information compression can improve prediction and generalization, separate relevant features of information, and efficiently utilize limited capacity by balancing the unreliability of stochasticity with the redundancy in messages [13, 60, 58, 61, 27]. We observed a rate-distortion gradient for every individual, consistent with the conceptualization of the brain structural network as a communication channel with limited capacity. While each individual's network adheres to rate-distortion theory, differences in structural connectivity lead to substantial variance in the compression efficiency. In addition to some of the variation being explained by age and sex, our results also indicate that variance could be driven by biological differences, such as in metabolic expenditure and myelin, as well as topological differences, such as transitivity, degree, and hierarchical organization. We found that biological investments of metabolic resources and myelin in connectome topology, hierarchically interconnected brain regions, and highly connected network hubs each supported efficient coding and compressed communica-

tion dynamics. We found evidence consistent with five predictions of rate-distortion theory adapted from prior literature, corroborating the validity of this theoretical model of macroscale efficient coding [31, 30, 27].

In addition to these findings, we reported two important null results which could be explained in part by the macroscale efficient coding model. First, we did not find evidence for the hypothesis that metabolic savings are associated with shortest path routing communication dynamics, a key redeeming feature against acknowledged theoretical shortcomings of this routing model [45, 3, 24]. Rather, we found evidence of efficient coding, implemented using random walks, as in other areas of neuroscience [26, 62, 2, 13, 60, 63, 58, 64, 15, 65]. Efficient coding generated unique predictions of communication as a stochastic process governed by trade-offs between lossy compression and transmission fidelity [66]. The evidence we presented was consistent with these predictions and would be difficult to explain with alternative models of shortest-path routing (see **Supplementary Modeling/Math Notes** and **Discussion**). This null result challenges the validity of the shortest path routing model as anchoring one end of a hypothesized spectrum of communication mechanisms and the usage of shortest path routing metrics to quantify the integrative capacity of particular brain areas or whole brain connectivity [24]. Rather, the macroscale efficient coding model implemented by random walk dynamics provides strong theoretical interpretation of information integration in hubs as lossy compression [31, 67].

Second, we did not find evidence for high metabolic cost associated with structural network hubs. Rather, we found that hubs are regions that receive inputs with greater compression efficiency; regions prioritizing input compression tend to be disproportionately out-scaled by total brain expansion during development, consistent with cortical consolidation (thinning and myelination) [6]. In contrast, regions prioritizing input fidelity tend to disproportionately expand in relation to total brain growth during development, consistent with theories positing that evolutionary expansion and new connections support the flexibility of neural activity for higher-order cognition [68, 3, 4, 47, 69, 7, 70, 71]. Structural network hubs support unique information integrative processes, apparently offsetting high metabolic, spatial, and material costs [31, 72, 9, 69, 67, 73, 74]. Hence, hub dysfunction may be particularly costly. Our null result motivates reconsideration of metabolic costs, at least for the simplest and most common definition of a structural connectivity hub using the measure of degree centrality (see **Supplementary Figure 8**, **Results**, and **Discussion**) [73]. An alternative hypothesis to high metabolic cost of hubs is that long-run metabolic savings depend on frequent usage of hubs to compensate for the expense of maintaining their size and connectivity [75, 76]. Although further in-

vestigation of the metabolic costs of rich-club hubs is warranted, our findings nevertheless reinforce a wealth of evidence emphasizing the importance of the development, resilience, and function of hubs in cognition and psychopathology [31, 9, 77, 23, 67, 6, 69, 10].

Our work admits several theoretical and methodological limitations. First, regionally aggregated brain signals are not discrete Markovian messages and do not have goals like reaching specific targets. As in recent work, our model is a deliberately simplified but useful abstraction of macroscale brain network communication [23]. Second, although we modeled random walk dynamics in light of prior methodological decisions and information theory benchmarks [20], compression efficiency can be implemented using alternative approaches. Several methodological limitations should also be considered. The accurate reconstruction of white-matter pathways using diffusion imaging and tractography remains limited [78]. While we chose to model structural connections using fractional anisotropy because it is one of the most widely used measures of white matter microstructure, future research could evaluate these same hypotheses in structural networks built from different measures and of other species [79, 80, 81]. Moreover, non-invasive measurements of CBF with high sensitivity and spatial resolution remain challenging. However, we acquired images using an ASL sequence providing greater sensitivity and approximately four times higher spatial resolution than prior developmental studies of CBF [82]. Lastly, our data was cross-sectional, limiting the inferences that we could draw about neurodevelopmental processes.

In summary, our study advances understanding of how brain metabolism and architecture at the macroscale supports communication dynamics in complex brain networks as efficient coding. Our model provides a simple framework for investigating efficient coding at the whole brain network level and an important basis to build towards computational network models of special cases of efficient coding, such as robust, sparse, and predictive coding [15]. In quantifying integrative communication dynamics and lossy compression of the macroscale brain network, these results are relevant to future research on low-dimensional neural representations [55, 17, 16], efficient control of functional network dynamics [83, 84], and hierarchical abstraction of behaviorally relevant cortical representations [85, 86, 87]. Future research could investigate whether other biological properties of the network support efficient coding, such as gradients in brain structure and function [88, 89, 90, 91, 92, 38, 93]. Neurodevelopmental processes vary with the naturalistic environment and socioeconomic background, suggesting that they are shaped by exposure to different experiences and expectations [94]. We hypothesize that the neurodevelopmental range of compression efficiency reflects a relationship be-

tween source coding (statistics or compressibility of the naturalistic input information) and channel coding (compression efficiency), which demarcate zones of possible brain network communication [60, 66]. Lastly, the compression efficiency metric offers a novel tool to test leading hypotheses of dysconnectivity [8], hubopathy [9, 10], disrupted information integration [95, 67], and neural noise [96] in neuropsychiatric disorders.

Author Contributions

D.Z wrote the paper. C.W.L., T.D.S., and D.S.B. edited the paper. D.Z. developed the theory with input from C.W.L, T.D.S., and D.S.B. R.C., G.L.B., and Z.C. preprocessed the data. D.Z. performed the analysis with input from Z.C. T.M.M. performed data preprocessing and interpretation of statistical models. D.R. preprocessed the data. J.D. developed imaging acquisition methods. R.E.G. acquired funding for data collection and performed data collection. R.C.G. provided expertise in cognitive phenotyping. D.Z., T.D.S, and D.S.B. designed the study. D.S.B. and T.D.S. acquired funding to support theory development and data analysis, and contributed to theory and data interpretation.

Acknowledgments

We acknowledge helpful discussions with Dr. Jennifer Stiso, Dr. Richard Betzel, Dr. David Lydon-Staley, Dr. Lorenzo Caciagli, Adon Rosen, and Dr. Bart Larsen. The work was largely supported by the John D. and Catherine T. MacArthur Foundation, the ISI Foundation, the Paul G. Allen Family Foundation, the Alfred P. Sloan Foundation, the NSF CAREER award PHY-1554488, NIH R01MH113550, NIH R01MH112847, and NIH R21MH106799. Secondary support was also provided by the Army Research Office (Bassett-W911NF-14-1-0679, Grafton-W911NF-16-1-0474) and the Army Research Laboratory (W911NF-10-2-0022). D.Z. acknowledges support from the National Institute of Mental Health F31MH126569. C.W.L. acknowledges support from the James S. McDonnell Foundation 21st Century Science Initiative Understanding Dynamic and Multi-scale Systems - Postdoctoral Fellowship Award. The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

Competing Interests

The authors declare that they have no competing interests.

1 Materials and Methods

1.1 Participants

As described in detail elsewhere [41], diffusion weighted imaging (DWI) and arterial-spin labeling (ASL) data were acquired for the Philadelphia Neurodevelopmental Cohort (PNC), a large community-based study of neurodevelopment. The subjects used in this paper are a subset of the 1,601 subjects who completed the cross-sectional imaging protocol. We excluded participants with health-related exclusionary criteria (n=154) and with scans that failed a rigorous quality assurance protocol for DWI (n=162) [97]. We further excluded subjects with incomplete or poor ASL and field map scans (n=60). Finally, participants with poor quality T1-weighted anatomical reconstructions (n=10) were removed from the sample. The final sample contained 1042 subjects (mean age=15.35, SD=3.38 years; 467 males, 575 females). Study procedures were approved by the Institutional Review Board of the Children's Hospital of Philadelphia and the University of Pennsylvania. All adult participants provided informed consent; all minors provided assent and their parent or guardian provided informed consent.

1.2 Cognitive Assessment

All participants were asked to complete the Penn Computerized Neurocognitive Battery (CNB). The battery consists of 14 tests adapted from tasks typically applied in functional neuroimaging, and which measure cognitive performance in four broad domains [41]. The domains included: (1) executive control (i.e., abstraction and flexibility, attention, and working memory), (2) episodic memory (i.e., verbal, facial, and spatial), (3) complex cognition (i.e., verbal reasoning, nonverbal reasoning, and spatial processing), (4) social cognition (i.e., emotion identification, emotion intensity differentiation, and age differentiation), and (5) sensorimotor and motor speed. Performance was operationalized as z-transformed accuracy and speed. The speed scores were multiplied by -1 so that higher indicates faster performance, and efficiency scores were calculated as the mean of these accuracy and speed z-scores. The efficiency scores were then z-transformed again, to achieve mean = 0 and SD = 1.0 for all scores. Confirmatory factor analysis supported a model of four latent factors corresponding to the cognitive efficiency of executive function, episodic memory, complex cognition, and social cognition [98]. Hence, we used these four cognitive efficiency factors in our analyses. In a factor solution separately modeling accuracy and speed, the accuracy factors correspond to (1) executive and complex cognition, (2) social cognition, and (3) memory. The speed factors correspond to (1) fast speed (e.g.,

working memory and attention tasks requiring constant vigilance), (2) episodic memory speed, and (3) slow speed (e.g., tasks requiring complex reasoning).

1.3 Image Acquisition, Preprocessing, and Network Construction

Neuroimaging acquisition and pre-processing were as previously described [41]. We depict the overall workflow of the neuroimaging and network extraction pipeline in **Supplementary Figure 1**.

1.3.1 Diffusion Weighted Imaging

As was previously described [99, 83], diffusion imaging data and all other MRI data were acquired on the same 3T Siemens Tim Trio whole-body scanner and 32-channel head coil at the Hospital of the University of Pennsylvania. DWI scans were obtained using a twice-focused spin-echo (TRSE) single-shot EPI sequence (TR = 8100 ms, TE = 82 ms, FOV = 240 mm^2/240 mm^2; Matrix = RL: 128/AP:128/Slices:70, in-plane resolution (x & y) 1.875 mm^2; slice thickness = 2 mm, gap = 0; FlipAngle = 90°/180°/180°, volumes = 71, GRAPPA factor = 3, bandwidth = 2170 Hz/pixel, PE direction = AP). The sequence employs a four-lobed diffusion encoding gradient scheme combined with a 90-180-180 spin-echo sequence designed to minimize eddy current artifacts. The complete sequence consisted of 64 diffusion-weighted directions with $b = 1000$ s/mm^2 and 7 interspersed scans where $b = 0$ s/mm^2. Scan time was about 11 min. The imaging volume was prescribed in axial orientation covering the entire cerebrum with the topmost slice just superior to the apex of the brain [97].

1.3.2 Connectome construction

Cortical gray matter was parcellated according to the Glasser atlas [100], defining 360 brain regions as nodes for each subject's structural brain network, denoted as the weighted adjacency matrix $\mathbf{A}$. To assess multiple spatial scales, cortical and subcortical gray matter was parcellated according to the Lausanne atlas [101]. Together, 89, 129, 234, 463, and 1015 dilated brain regions defined the nodes for each subject's structural brain network in the analyses of **Figure 5**.

DWI data was imported into DSI Studio software and the diffusion tensor was estimated at each voxel [102]. For deterministic tractography, whole-brain fiber tracking was implemented for each subject in DSI Studio using a modified fiber assessment by continuous tracking (FACT) algorithm with Euler interpolation, initiating 1,000,000 streamlines after removing all streamlines with length less than 10mm or greater than 400mm.

Fiber tracking was performed with an angular threshold of 45, a step size of 0.9375mm, and a fractional anisotropy (FA) threshold determined empirically by Otzu's method, which optimizes the contrast between foreground and background [102]. FA was calculated along the path of each reconstructed streamline. For each subject, edges of the structural network were defined where at least one streamline connected a pair of nodes. Edge weights were defined by the average FA along streamlines connecting any pair of nodes. The resulting structural connectivity matrices were not thresholded and contain edges weighted between 0 and 1.

1.3.3 Arterial-Spin Labeling

CBF was quantified from control-label pairs using ASLtbx [103], as was previously described [82]. We consider f as CBF, δM as the difference of the signal between the control and label acquisitions, R_{1a} as the longitudinal relaxation rate of blood, τ as the labeling time, ω as the post-labeling delay time, α as the labeling efficiency, λ as the blood/tissue water partition coefficient, and M_0 as the approximated control image intensity. Together, CBF f can be calculated according to the equation:

$$f = \frac{\Delta M \lambda R_{1a} \exp\left(\omega R_{1a}\right)}{2 M_0 \alpha} \left[1 - \exp\left(-\tau R_{1a}\right)\right]^{-1}. \tag{1}$$

Because prior work has shown that the T1 relaxation time changes substantially in development and varies by sex, this parameter was set according to previously established methods, which enhance CBF estimation accuracy and reliability in pediatric populations [104, 105]. As in prior work, the global network CBF was calculated as the average CBF across all brain regions to obtain an individual participant's CBF [82].

1.4 Brain Maps

1.4.1 Cortical Myelin

As described previously [106], cortical myelin content was calculated by dividing the T1w image signal by the T2w image signal. Specifically, we define the myelin content x^2 in the following manner:

$$\frac{\text{T1w}}{\text{T2w}} \approx \frac{x * b}{(1/x) * b} = x^2, \tag{2}$$

where x is the myelin contrast in the T1w image, $1/x$ is the myelin contrast in the T2w, and b is the receive bias field in both T1w and T2w images. We used a published atlas generated by this method [47].

1.4.2 Cortical Areal Scaling

As described previously [7], to estimate cortical areal scaling between the size of cortical regions and the total brain, regression coefficients β were estimated for $\log_{10}$(total cortical surface area) as a covariate predicting $\log_{10}$(vertex area) using spline regression models that incorporated effects of age and sex on vertex area [107]. We used the following relational form:

$$\log 10(\text{Vertex area}) \sim \text{s}(\text{age by } = \text{sex}) + \text{B}_\text{I}\left[\log 10\left(\text{total_area}\right)\right]. \tag{3}$$

When β is 1, the scaling between total brain size and brain regions is linear. When β deviates greater or less than 1, scaling is non-linearly and disproportionately expanding or contracting. We used the published atlas generated using the same data as in our study [7, 41].

1.5 Network Statistics

1.5.1 Global Efficiency

In the context of the brain structural connectome, global efficiency represents the strength of the shortest paths between brain regions supporting efficient communication. In network neuroscience, global efficiency is commonly used as a metric of a brain network's capacity for shortest path routing [3, 20, 24]. We calculated the common global efficiency statistic [46], which is defined for a graph G as:

$$\mathcal{E}_{glob}(G) = \frac{1}{\mathcal{N}(\mathcal{N}-1)} \sum_{i \neq j \epsilon G} \frac{1}{d_{ij}}, \tag{4}$$

where $\mathcal{N}$ is the number of nodes and d_{ij} is the shortest distance between node i and node j. Intuitively, a high $\mathcal{E}$ value indicates greater potential capacity for global and parallel information exchange along shortest paths, and a low $\mathcal{E}$ value indicates decreased capacity for such information exchange [46].

1.5.2 Path Strengths

Beyond shortest paths between pairs of brain regions, we also sought to measure the strength of structural connections S comprising the paths of multiple connections. As global efficiency measures the capacity of brain networks for shortest path routing, path strengths measure the capacity for stochastic communication. Path strengths are apt for assessing the network capacity for a model of stochastic transmission of impulses because paths can be represented as random walks $p = (i, j, \ldots, k)$, where p is a path and i, j, and k are

nodes in the path. As in prior work [108], the strength of the weighted connections in a path, denoted $\omega(p)$, in the graph G with adjacency matrix $\mathbf{A}$ is defined as:

$$\omega(p) = [\mathbf{A}]_{i_0 i_1} [\mathbf{A}]_{i_1 i_2} \dots [\mathbf{A}]_{i_{(l-1)} i_i}, \tag{5}$$

where the matrix products produce the strengths of all possible random walks according to the length of p, as depicted in the schematic **Figure 1B**. Then, for walks of length n, the strengths of the paths from node i to node j are defined as:

$$[\mathbf{A}^n]_{ij} = \sum_{p \in \mathcal{P}_{ij}^n} \omega(p), \tag{6}$$

where P_{ij}^n is the set of all walks from node i to node j with length n. When n=1, the matrix exponent produces a matrix with elements equal to d_{ij} from Equation 1, or the shortest distance between node i and node j. Intuitively, a high path strength represents structural paths that consist of higher integrity connections measured by DWI, whereas a low path strength indicates paths consisting of low integrity connections. To compute node strengths, the values for each node were summed. An average value was also calculated across node strengths per individual participant.

1.5.3 Path Transitivity

Shortest paths confer advantages in speed and signal fidelity when messages are transmitted by random walks. Therefore, we sought to measure a property of brain network architecture supporting random walks by shortest paths. Local detours which first leave and then re-access the shortest path serve to support such random walks, and the potential for such local detours can be estimated using a measure called path transitivity (see **Figure 5A**, left) [21]. Path transitivity was previously used to predict functional BOLD activation comparably to conventional distance or computational models of neural dynamics. To compute path transitivity, we first calculated the matching index for each pair of successive nodes i and j along the shortest path $\pi_{s \to t}$, with neighboring non-shortest path nodes k as:

$$m_{ij} = \frac{\sum_{k \neq i,j} (w_{ik} + w_{jk}) \, \Theta(w_{ik}) \, \Theta(w_{jk})}{\sum_{k \neq j} w_{ik} + \sum_{k \neq i} w_{jk}}, \tag{7}$$

where w is the connection weight, and $\Theta(w_{ik}) = 1$ if $w_{ik} > 0$, and 0 otherwise. Intuitively, the numerator is non-zero if and only if there are two locally detouring connections that make a closed triangle along the shortest path. If either of the two connections w_{ik} or w_{jk} does not exist, then the numerator is 0. With the

denominator representing the strength of all cumulative connections of the shortest path nodes, the matching index fraction then represents the density of closed triangles (i.e., transitivity) around the shortest path.

Whereas the matching index is a pairwise measure of the density of locally returning detours, path transitivity generalizes the density across the shortest path. Using the computed matching index m_{ij} for each pairwise connection Ω from source node s to target node t by the set of shortest path edges $\pi_{s\to t}$, we compute path transitivity M as:

$$M\left(\pi_{s\to t}\right) = \frac{2\sum_{i\in\Omega}\sum_{j\in\Omega} m_{ij}}{|\Omega|(|\Omega|-1)}, \tag{8}$$

where the numerator sums the matching index m_{ij} for all edges in Ω, the scale factor of 2 indicates an undirected graph, and the denominator sums over all possible edges. Intuitively, a high path transitivity M indicates that the shortest path is more densely encompassed by locally detouring triangular motifs. Low path transitivity indicates that the shortest path is surrounded by connections that deviate from the shortest path without an immediate avenue of return. An individual-level value of path transitivity was calculated as the average path transitivity across brain regions.

1.5.4 Modularity

Modularity is a common architectural feature observed in neural systems across species. A single community contains brain regions that are more highly connected to each other than to brain regions located in other communities (see **Supplementary Figure 2A**, right). Modularity of brain networks is spatially efficient, supports the development of executive function in youths, and supports flexibly adaptable functional activations according to distinct task demands [109, 52, 99, 110, 111]. To assess modularity, we apply a common community detection technique known as modularity maximization [112], in which we used a Louvain-like locally greedy algorithm [113] to maximize a modularity quality function for the adjacency matrix A. The modularity quality function is defined as:

$$Q = \frac{1}{2\mu}\sum_{ij}\left(A_{ij} - \gamma P_{ij}\right)\delta(g_i, g_j), \tag{9}$$

where $\mu = \frac{1}{2}\sum_{ij} A_{ij}$ denotes the total weight of $\mathbf{A}$, A_{ij} encodes the weight of an *edge* between node i and node j in the structural connectivity matrix, $\mathbf{P}$ represents the expected strength of connections according to a specified null model [112], γ is a structural resolution parameter that determines the size of modules, and δ is the Kronecker function which is 1 if $g_i = g_j$ and zero otherwise. As in prior work, we set γ to the default value

of 1 [110]. Intuitively, a high Q value indicates that the structural connectivity matrix contains communities, where nodes within a community are more densely connected to one another than expected under a null model. Modularity maximization is commonly used to detect community structure, and to quantitatively characterize that structure by assessing the strength and number of communities [99, 110, 112]. Regional contributions to the modularity quality function were used for analyses of brain regions.

1.5.5 Resource Efficiency

A signal that randomly walks along the shortest path between brain regions confers advantages in speed, reliability, and fidelity [3, 32, 24]. Following prior work, we sought to compute the number of random walkers beginning at node i that were required for at least one to travel along the shortest path to another node j with probability η [20, 32]. To begin, we consider the transition probability matrix by $\mathbf{U}$, defined as $\mathbf{U} = \mathbf{W}\mathbf{L}^{-1}$, where each entry W_{ij} of $\mathbf{W}$ describes the weight of the directed edge from node i to node j, and each entry L_{ii} of the diagonal matrix $\mathbf{L}$ is the strength of each node i, defined as $\sum_i W_{ij}$. Intuitively, each entry U_{ij} of $\mathbf{U}$ defines the probability of a random walker traveling from node i to node j in one step. Next, to compute the probability that a random walker travels from node i to node j along the shortest path, we define a new matrix $U'(i)$ that is equivalent to $\mathbf{U}$ but with the non-diagonal elements of row i set to zero and $U_{ii} = 1$ as an absorbant state. Then, the probability of randomly walking from i to j along the shortest path is given by:

$$1 - \sum_{n=1}^{N} [U'(i)^H]_{in}, \tag{10}$$

where H is the number of connections composing the shortest path from i to j. Similarly, the probability η of releasing r random walkers at node i and having at least one of them reach node j along the shortest path is given by:

$$\eta = 1 - \left(\sum_{n=1}^{N} [U'(i)^H]_{in} \right)^r. \tag{11}$$

Setting the above probability to some set value η, we can then solve for the number of random walkers r required to guarantee (with probability η) that at least one of them travels from i to j along the shortest path, denoted by:

$$r_{ij}(\eta) = \frac{\log(1-\eta)}{\log\left(\sum_{n=1}^{N}\left[U'(i)^H\right]_{in}\right)}. \tag{12}$$

The number of random walkers r_{ij} has been referred to as resources in prior literature [20]. In our analyses, we calculate resources r_{ij} over a range of values of η for each participant. Finally, to calculate the resource efficiency of each participant, the resource efficiency of an entire network is taken to be $1/(r_{ij}(\eta))$ averaged over all pairs of nodes i and j. With the right stochastic matrix U'_i, the resource efficiency of brain regions as message senders is $1/(r_{ij}(\eta))$ averaged over i, while brain regions as message receivers is $1/(r_{ji}(\eta))$ averaged over j.

1.5.6 Compression Efficiency

Rate-distortion theory formalizes the study of information transfer as passing signals (messages) through a capacity-limited information channel. A signal x is encoded as $\hat{x}$ with a level of distortion D that depends on the information rate R. The greater the rate, the less the distortion. The rate-distortion function $R(D)$ defines the minimum information rate required to transmit a signal corresponding to a level of signal distortion (see **Figure 3A**). Lossy compression arises from the choice of the distortion function $d(x, \hat{x})$, which implicitly determines the relevant and irrelevant features of a signal. With the true signal x mapped to the compressed signal $\hat{x}$ described by $p(\hat{x}|x)$, the rate-distortion function is defined by minimizing the mutual information of the signal and compression over the expected distortion defined as $d(x, \hat{x})_{p(x,\hat{x})} = \sum_{x \in X} \sum_{\hat{x} \in \hat{X}} p(x, \hat{x}) d(x, \hat{x})$:

$$R(D) \equiv \min_{d(x,\hat{x})} I(X, \hat{X}) = \sum_{x \in \Omega_X} \sum_{\hat{x} \in \Omega_{\hat{X}}} P(\hat{x}|x) P(x) \log_2 \left(\frac{P(\hat{x}|x)}{P(\hat{x})} \right). \tag{13}$$

By minimizing the mutual information $I(X, \hat{X})$, we arrive at a probabilistic map from the signal to the compressed representation, where the information gain between the signal and compression is as small as possible (i.e., high fidelity) to favor the most compact representations.

Similar to the mathematical framework of rate-distortion theory, we sought to specify a distortion function reflecting communication over the brain's structural network. Prior work building models of perceptual and cognitive performance have inferred distortion functions through Bayesian inference of a loss function [61, 27]. For instance, the loss function could be the squared error denoting the residual values of the true

signal minus the compression, $L = (\hat{x} - x)^2$ (**Figure 3A**). A neural rate-distortion theory has been theoretically developed [30], but remains empirically untested due in part to a lack of methodological tools at the level of brain systems. Moreover, it has been difficult to define a distortion function that incorporates both true signals x and compressed signals $\hat{x}$ in part because the measurements of these signals in human brain networks remains challenging. Here, we define an analogous framework of information transfer through capacity-limited channels in the structural network of the brain. Particularly, we build a distortion function from the simple intuition that the shortest path is the route that most reliably preserves signal fidelity, as depicted in (**Figure 3B**).

Given that a random walker propagating from node i along the shortest path to node j retains the greatest signal fidelity, we define the distortion function of any signal x from brain region i to a compressed representation $\hat{x}$ decoded in brain region j as:

$$d(x, \hat{x})_{ij} = (1 - \eta), \tag{14}$$

where η denotes the probability that a walker gets from node i to node j along the shortest path. A signal with greater probability η of propagating by the shortest path between brain region i and brain region j is at a lower risk of distortion (see **Figure 3D**). Intuitively, increased topological distance adds greater risk of signal distortion due to further transmission through capacity-limited channels (i.e. structural connections), temporal delay, and potential mixing with other signals. Given the measure of resources in Equation 12, we develop and test predictions of a novel definition of the rate $R(D)$; here, we define $R(D)$ as the resources $r_{ij}(\eta)$ required to achieve a tolerated level of distortion $d(x, \hat{x})_{ij}$:

$$R(D) \equiv r_{ij}(d(x, \hat{x})_{ij}), \tag{15}$$

as in (**Figure 3D**). When the log of resources $\log(r_{ij})$ is plotted against our metric of distortion $D = d \in 1 - \eta_{ij}$, the exponential gradient is depicted linearly (see **Figure 3E**). Because prior work focused on 50% distortion during analyses, we required the slope to intersect the mean midpoint rate at 50% distortion [20]. In addition to the precedent offered by prior work, this requirement is also reasonable given that we sought to model both high and low distortions equitably. The slope denotes the minimum number of resources required to achieve a tolerated level of distortion, which we refer to as the *compression efficiency* (3E; bottom). A steeper slope (i.e., a more negative relation) reflects reduced compression efficiency, or prioritization

of message fidelity. A flatter slope (i.e., a more positive relation) reflects increased compression efficiency, or prioritization of lossy compression. Individual variation in compression efficiency can be assessed by using the average resource efficiency across brain regions. When compression efficiency is computed for sets of brain regions by averaging across individuals, the slope can denote either messages sent from or arriving to a brain region by using the average resource efficiency over either all nodes j or all nodes i, respectively.

While prior work calculates the bit rates that arise from the stochasticity of opening and closing ion channels or releasing a synaptic vesicle [2], in this manuscript we specifically chose to focus on macroscale level neuroimaging data. Our choice stemmed in part from the fact that the surrounding background literature in neuroimaging motivated many components of our theory, and in part from the fact that the theory of efficient coding had not yet been extended to this scale. We sought to understand the role of the whole connectome in supporting efficient communication and information processing.

To calculate numerical measures of bit rates at any spatial scale, we require a probabilistic information source. In the random walk model of inter-regional communication, each brain region is an information source that activates while sending a message—that is, a random walker—to another region along the wiring of the connectome. The information content of an individual message is determined by the probability of a message being sent, which, in our setting, is equivalent to the probability of a brain region changing its activity (activating or deactivating) in statistical association with behaviors and cognitive functions. At the level of macroscale neuroimaging data, individual studies describe localized neural activation and deactivation associated with specific cognitive tasks and behavior. Meta-analyses of these individual studies can help aggregate and summarize a large quantity of data on the general probability of brain region activation across a wide range of behaviors and cognitive functions [13, 114]. We used the meta-analytic approach of NeuroSynth to obtain probability distributions across statistical maps of neural activity [114]. We then used this average probability that a brain region is active to calculate the information content of each communication event, or the transmission of one neural message.

To calculate the information content of a neural message, we used the Shannon information measure to calculate information rate in bits per second [115]. Shannon information derives from the probability of a particular random event. Intuitively, less probable events convey messages with more information content, and *vice versa*. In the context of macroscale brain networks, we sought to obtain the probability of a brain

region transmitting a neural signal measured by fMRI BOLD activation (see **Supplementary Figure 3**). The mean probability of regional activation then is used to calculate the information content per message as the Shannon information given by the following expression:

$$H = -log_2(P(Activation|Terms)), \tag{16}$$

where H denotes the information content, or surprisal, of a neural signal measured by fMRI, and $P(Activation|Terms)$ defines the average conditional probability of brain region activation given all >3,200 psychological terms encompassing an ontology that spans sensation, behavior, cognition, emotion, and disorders [114]. The average information content of a message, which we model with a random walker, is 5.74 bits. A characteristic timescale of fMRI activity is given by the repetition time (TR) of each measurement, which is 0.72 seconds in the Human Connectome Project neuroimaging sequence [116] and can vary across studies. Hence, the information rate is 5.74 bits/0.72 seconds = 7.97 bits/second per channel use. As with the estimation of information content, the main results of our paper do not depend on the choice of a characteristic timescale, because all values are converted using the same unit of bits per message. In selected early graphs, we include a secondary y-axis to indicate the conversion between units of random walkers and units of bits. To obtain the average bits per transmission for different levels of distortion, we multiplied the number of messages (random walkers) by 5.74 bits/message. We note that our theory can be applied to different neural activity measurements with connectomes reconstructed from a range of spatial scales. In future work, it will be important to analyze the predictability of bit rates measured directly from neural recordings based on theoretical estimates derived from the connectome.

1.5.7 Biased Random Walk

Given the advantages of random walkers which propagate along the shortest path, we sought to assess how brain metabolism could support the reliability and fidelity of signaling. Metabolic diffusion can be modeled as random walks over a structural connectivity matrix biased by regional CBF [117]. Electrical signal propagation could be modeled as random walks over a structural connectivity matrix biased by regional intracortical myelin [106]. To model metabolic diffusion of random walkers attracted to brain regions of high CBF, we used analytical solutions to biased random walks. First, we defined the matrix **T** of CBF-biased

transition probabilities as:

$$T_{ij}^{\alpha} = \frac{\alpha_{ij} A_{ij}}{\sum_k \alpha_{ik} A_{ik}}, \tag{17}$$

where the element of T_{ij} defines the transition probabilities of a random walker traversing edges of the structural connectivity matrix $\mathbf{A}$ which are multiplied by a bias term α. For random walkers attracted to brain regions of high CBF, the bias term α was defined as the average CBF value for each pair of brain regions. Hence, a random walker propagates over the brain's structural connections with transition probabilities of T_{ij} that reflect the integrity of structural connections and the average level of CBF between pairs of brain regions. We then substituted the $\mathbf{U}$ matrix in the resources $r_{ij}(\eta)$ of Equation 12 with T_{ij} in Equation 17 to compute the number of resources required for a biased random walker to propagate by the shortest path with a specified probability. To model electrical signal propagation, we performed the same matrix transformation as described above, using regional measurements of intracortical myelin.

1.5.8 Rich Club

Due to the importance of brain network hubs in the broadcasting of a signal [3, 118, 24], we sought to identify the set of high-degree brain regions in the rich club (see **Figure 6A**) [119]. To identify the subnetwork of rich club brain regions, we computed the weighted rich club coefficient $\Phi^z(k)$ as:

$$\phi^z(k) = \frac{Z_{>k}}{\sum_{l=1}^{E>k} z_l^{\text{ranked}}}, \tag{18}$$

where Z^{ranked} is a vector of ranked network weights, k is the degree, $Z_{>k}$ is the set of edges connecting the group of nodes with degree greater than k, and $E_{>k}$ is the number of edges connecting the group of nodes with degree greater than k. Hence, the rich-club coefficient $\Phi^z(k)$ is the ratio between the set of edge weights connected to nodes with degree greater than k and the strongest $E_{>k}$ connections. The rich-club coefficient was normalized by comparison to the rich club coefficient of random networks [119]. Random networks were created by rewiring the edges of each individual's brain network while preserving the degree distribution. The rich-club coefficient for the randomized networks $\Phi_{\text{random}}(k)$ was computed using Equation 18. Then, the normalized rich-club coefficient $\Phi_{\text{norm}}(k)$ was calculated as follows:

$$\phi_{\text{norm}}(k) = \frac{\phi(k)}{\phi_{\text{random}}(k)}, \tag{19}$$

where $\Phi_{\text{norm}}(k) > 1$ indicates the presence of a rich club organization. We tested the statistical significance of $\Phi_{\text{norm}}(k)$ using a 1-sample t-test at each level of k, with family-wise error correction for multiple tests over k. Each individual was assigned the value of their highest degree $> k$ rich club level and their nodes were ranked by rich club level. Over the group of individuals, the nodal ranks were averaged and the top 12% of nodes were selected as the rich club, following prior work [53].

1.5.9 Hierarchical Organization

Hierarchical modularity is a form of network complexity that is characterized by small groups of nodes organized successively into increasingly larger groups. Networks with hierarchical organization exhibit a distinctive scaling where transitivity (clustering coefficient) decreases exponentially as a function of degree [51]. This scaling appears linear in a log-log plot and the slope will be used to assess the hierarchical organization as the hierarchy parameter ζ. To assess hierarchical structure, we examined the relationship between the nodal transitivity (clustering coefficient) and nodal degree [51]:

$$C(k) = k^{\zeta}, \tag{20}$$

where C is the transitivity (clustering coefficient), k is the degree, and ζ represents the extent of hierarchical organization. To estimate ζ, we used similar curve-fitting methods as in estimating the compression efficiency gradient. The transitivity C was defined as:

$$C_i = \frac{2n_i}{k_i(k_i - 1)}, \tag{21}$$

where n is the number of links between the k neighbors of node i. Intuitively, a high ζ value indicates hierarchical organization with less clustering about central nodes (with higher degree), and a low ζ value indicates hierarchical organization with more clustering about central nodes (with higher degree).

1.6 Network Null Models

Random graphs are commonly used in network science to test the statistical significance of the role of some network topology against null models. We used randomly rewired graphs generated by shuffling each individual's empirical networks 20 times, as in prior work [120]. Furthermore, we generated Erdős-Renyí random networks for each individual brain network where the presence or absence of an edge was generated by a uniform probability calculated as the density of edges existing in the corresponding brain network. Edge

weights were randomly sampled from the edge weight distribution of the brain network. While the randomly rewired graphs retain empirical properties such as the degree and edge weight distributions of the individual brain networks, the Erdős-Renyí networks do not. Hence, the randomly rewired null network was used in all analyses where the degree distribution should be retained (e.g., normalized rich club coefficient), while the Erdős-Renyí network was used in analyses assessing the overall contribution of the brain network topology (e.g., compression efficiency).

We used the Erdős-Renyí networks as the null network for two reasons. First, Erdős-Renyí networks are particularly appropriate for comparing the random walk model to the alternative shortest path routing model. Prior work found that Erdős-Renyí were networks with very high global efficiency, though the networks were biologically implausible due to their connection costs [42]. Thus, the Erdős-Renyí networks serve as a benchmark for a communication architecture supporting highly efficient shortest path routing.

Second, we chose to use Erdős-Renyí networks because we sought to compare random walk communication on brain network connectivity to random walk communication on synthetic networks that optimally minimize our distortion function—that is, maximizing the shortest path probability [20]. In doing so, we can address the major criticism of the plausibility of random walk models: inefficiency [24]. Prior work compared random walks dynamics on canonical graph models, finding that the Erdős-Renyí networks had the highest probability of shortest path propagation [20]. Thus, we expected that the Erdős-Renyí networks would approximate a limit on the efficiency for random walk communication dynamics according to the rate-distortion model. Compared to this approximate limit of efficiency, we aimed to show that biological investments in brain network communication modeled using random walks could indeed be efficient. Moreover, we aimed to show that inefficiency, when viewed in the light of the biologically-established efficient coding framework, uses the redundancy of communication to overcome noise introduced by intrinsic stochasticity of neural processes to balance the transmission fidelity and lossy compression of information [26].

Our tests using the randomly rewired network evaluate the null hypothesis that an apparent rich-club property of brain networks is a trivial result of topology characteristic of random networks with some empirical properties preserved, as in prior work [119, 118]. The alternative hypothesis is that the brain network has a rich-club organization beyond the level expected in the random networks. Our tests using the Erdős-Renyí network evaluate the null hypothesis that the rates in the rate-distortion function modeling information

processing capacity in brain networks does not differ from the rates in the rate-distortion function of random networks. The alternative hypothesis is that the rate of the brain network's rate-distortion function differs from that of random networks, consistent with the notion that Erdős-Renyí networks have a greater prevalence of shortest paths compared to brain networks. We additionally used the Erdős-Renyí network to assess the hypothesis of rate-distortion theory that synthetic networks should exhibit the same information processing trade-offs (the monotonic rate-distortion gradient) as empirical brain networks [27]. We selected Erdős-Renyí networks to assess these hypotheses for two reasons. First, Erdős-Renyí networks do not retain core architectures of brain networks, such as modularity, and therefore reflect an extreme synthetic network. Second, Erdős-Renyí networks are commonly used as a benchmark for assessing shortest path prevalence due to the prominence of uniformly distributed direct pairwise connections [42, 24]. In light of the central assumption that shortest paths represent the route of highest signal fidelity in our definition of distortion, we used Erdős-Renyí networks to verify our intuition that compression efficiency should be greater in the Erdős-Renyí network than in brain networks.

Scale-free networks have a degree distribution resulting in a subset of highly connected hubs. Prior work established that global transitivity decreases exponentially with network size in a scale-free network, but that networks that are scale-free *and* have hierarchical organization can decouple transitivity from size. We created non-hierarchical scale-free networks using the Barabási-Albert algorithm (with the preferential attachment parameter set to 1 for linear preferential attachment) and re-evaluated whether the structural brain networks with hierarchical organization similarly decoupled transitivity and size [121]. Decoupling transitivity and size allows for simultaneously high levels of transitivity despite large size. In the context of efficient coding theory, we chose to compare the global transitivity of structural brain networks with scale-free networks across 5 network sizes to examine if hierarchical organization in the human structural brain networks confers a similar outcome.

1.7 Statistical Analyses

To assess the covariation of our measurements across individuals and brain regions, we used generalized additive models (GAMs) with penalized splines. GAMs allow for statistically rigorous modeling of linear and non-linear effects while minimizing over-fitting [107]. Throughout, the potential for confounding effects was addressed in our model by including covariates for age, sex, age-by-sex interaction, network degree, network density, and in-scanner motion. Due to the likelihood of inflated estimates of brain-behavior associations

despite well-powered analyses [122], we report bootstrap 95% confidence intervals for the test statistic and adjusted R^2 of each model across 1000 bootstrap samples.

1.7.1 Metabolic running costs associated with brain network architectures

We used penalized splines to estimate the nonlinear developmental patterns of global efficiency (Equation 4) and CBF, as in prior work [82, 99]. Then, we assessed the partial correlation between the residual variance (unexplained by covariates of age, sex, age-by-sex-interaction, degree, density, and motion) of global efficiency and CBF. The final models can be written as:

$$\text{Global efficiency} \sim \text{spline(age)} + \text{sex} + \text{spline(age by sex)} + \text{degree} + \text{density} + \text{motion}, \tag{22}$$

$$\text{CBF} \sim \text{spline(age)} + \text{sex} + \text{spline(age by sex)} + \text{degree} + \text{density} + \text{motion}, \tag{23}$$

and

$$\text{Residual(Global efficiency)} \sim \text{residual(CBF)}. \tag{24}$$

To evaluate the importance of age as a confound for the relationship between global efficiency and CBF, we also performed sensitivity analyses by removing selected covariates and re-assessing the model. In addition, for consistency with prior work [45], we performed the same analysis including covariates for gray matter volume and density.

To assess the relationship between CBF and the strength of structural paths supporting random walks (Equation 6), we again used penalized splines. The final model can be written as:

$$\text{CBF} \sim \text{path strength} + \text{spline(age)} + \text{sex} + \text{spline(age by sex)} + \text{degree} + \text{density} + \text{motion}. \tag{25}$$

Assessments of path strengths were corrected for false discovery rate across the statistical tests performed over the discrete path lengths.

1.7.2 Trade-offs between modularity and random walk architecture

Next, we sought to evaluate the metabolic running cost of brain network properties, in line with calls for investigation of the economic landscape of resource-constrained trade-offs between hallmark brain network architectures such as modularity (Equation 9) and new measures of brain network organization [3]. Following

our findings that CBF is associated with structural properties supporting random walks, we investigated path transitivity (Equation 8). We continued to use penalized splines to model the non-linear patterns of CBF and brain properties of interest. The final model can be written as:

$$\begin{aligned}\text{CBF} \sim \text{path transitivity} + \text{modularity} + \text{path transitivity by modularity} \\ + \text{spline(age)} + \text{sex} + \text{spline(age by sex)} + \text{degree} + \text{density} + \text{motion.}\end{aligned} \tag{26}$$

To visualize the landscape of CBF as a function of modularity and path transitivity, we plotted the GAM model response function. We described the distribution of modularity and path transitivity across individuals using frequency histograms. We calculated a map of change in global CBF with respect to both modularity and path transitivity using first-order derivatives. A saddle point suggests that adaptive compromises in network architecture are constrained by dual objectives. To quantify the location of the saddlepoint coordinate within the change map, we performed a k-nearest neighbor search of the value 0 in the gradient of first-order derivatives indicating minima and maxima.

1.7.3 Compression efficiency and development

To assess the possibility of distinct compression efficiency of brain networks compared to random networks, we calculated the resource efficiency (Equation 12) at 14 levels of distortion and performed an analysis of variance (ANOVA) test. The ANOVA model can be written as:

$$\text{Resources} \sim \text{distortion} + \text{type of network,} \tag{27}$$

where the type of the network is a categorical variable designating if the network was a brain network or a random network.

To compute compression efficiency per individual brain network, we used a polynomial regression function to find the best linear fit to the monotonic rate-distortion function according to the prediction of a linear rate-distortion gradient in semi-log space (log(resources) as a function of distortion). Next, we used a GAM model to assess the non-linear patterns of compression efficiency in development, which we can formally write as follows:

$$\text{Compression efficiency} \sim \text{spline(age)} + \text{sex} + \text{spline(age by sex)} + \text{degree} + \text{density} + \text{motion.} \tag{28}$$

To quantify the differences between the resources calculated from resource efficiency and the compression efficiency across 14 levels of distortion, we performed two-sample t-tests while controlling for family-wise error rate across multiple comparisons. For each level of distortion, we calculated the t-statistic comparing all individual resources to all individual resources predicted by the linear rate-distortion gradient.

1.7.4 Compression efficiency of biased random walks

To compute the compression efficiency of metabolic diffusion and electrical signaling, we modified the model of Equation 27 to instead calculate resource efficiency using the biased random walk matrices from Equation 17. The model can be written:

$$\text{Resources} \sim \text{distortion} + \text{type of random walk} + \text{distortion by type of random walk}, \tag{29}$$

where the type of random walk is a categorical variable designating unbiased random walks using the structural network, biased random walks using the structural network biased with CBF, and biased random walks using the structural network biased with intracortical myelin. To assess the hypothesis that resources differ according to the type of random walk, we performed t-tests while controlling for family-wise error rate across multiple comparisons.

1.7.5 Compression efficiency in a low or high fidelity regime

Next, we sought to test the predictions of a high or low fidelity communication regime. In a high fidelity regime, minimum resources given an expected distortion should increase monotonically as a function of network complexity. To assess whether the relationship between resources and network complexity (operationalized here as network size) is monotonic, we used a linear model written as:

$$\text{Log}_2(\text{resources}) \sim \log_{10}(\text{network size}). \tag{30}$$

In a low fidelity regime, minimum resources given an expected distortion should plateau as a function of complexity. We hypothesized that path transitivity is a property of structural networks that supports lossy compression and storage savings. The complexity of the shortest path was defined as the number of nodes contributing to path transitivity (Equation 8). To assess whether the resources non-linearly plateau as a

function of shortest path complexity, we used a GAM model written as follows:

$$\text{Log}_2(\text{resources}) \sim \text{spline}(\text{shortest path complexity}). \tag{31}$$

In light of our interest in network size and shortest path complexity (path transitivity) as measurements of high and low fidelity regimes, we next sought to measure hierarchical organization. Networks that are scale-free and have hierarchical organization can achieve simultaneously high network size and transitivity [51]. To assess whether hierarchical modularity supports a high fidelity regime, we used a linear model written as follows:

$$\text{Log}_2(\text{resources}) \sim \text{hierarchical modularity}. \tag{32}$$

We next evaluated whether brain structural networks with hierarchical modularity *and* the scale-free property achieved greater transitivity than random networks with only the scale-free property, as predicted by prior findings [51]. To evaluate the differences in transitivity according to the network properties of hierarchical organization and the scale-free property, we used the ANOVA model written as:

$$\text{Global transitivity} \sim \text{network size} + \text{type of network}, \tag{33}$$

where the type of network is a categorical variable designating either the human brain structural network or the Barabási-Albert scale-free network.

1.7.6 Compression efficiency and patterns of neurodevelopment

To explore how compression efficiency might relate to patterns of cortical myelination and areal scaling, we assessed the Spearman's correlation coefficient between myelination or scaling and send or receive compression efficiency. We also assessed the relationship between cortical areal scaling and nodal transitivity, in light of our hypothesis that greater path transitivity may support greater fidelity and the efficient coding hypothesis which states that the organization of the brain allocates neural resources according to the physical distribution of information. To further test correspondence between brain maps, we used a spatial permutation test, which generates a null distribution of randomly rotated brain maps that preserve the spatial covariance structure of the original data [123]. Using spatially-constrained null models is the state-of-the-art for comparing brain maps [123]. We used a spin test variant that reassigns parcels with no duplication. This method implements an iterative procedure that uniquely assign parcels based on Euclidean distance

of the rotated parcels, ignoring the medial wall and its location [124]. Our choice of null model introduces a trade-off between permitting slightly more liberal critical thresholds than other spatially-constrained null models and retaining the exact distribution of the original brain map to better test network-based statistics like compression efficiency and transitivity [125]. We refer to the p-value of this statistical test as p_{SPIN}. Finally, we applied the conservative Holm-Bonferroni correction for family-wise error across these tests.

1.7.7 Compression efficiency and rich-club hubs

Given the assumed integrative and broadcasting function of rich-club hubs, we sought to evaluate whether compression efficiency differed in rich-club hubs compared to other brain regions. We used the Wilcoxon rank-sum test to compare regional compression efficiency of either receiving or sending messages. Moreover, we assessed whether there was a difference between CBF in the rich-club hubs compared to other brain regions. Lastly, we tested the correlation of compression efficiency in the rich-club hubs and other brain regions with cognitive efficiency. To model non-linear patterns of cognitive efficiency, we used penalized splines controlling for potentially confounding covariates. The final model can be written as:

$$\begin{aligned} &\text{Cognitive efficiency} \sim \text{compression efficiency} + \text{spline(age)} \\ &+ \text{sex} + \text{spline(age by sex)} + \text{degree} + \text{density} + \text{motion}. \end{aligned} \tag{34}$$

Due to previous report of the relationship between cognition and global efficiency (Equation 1), we determined that compression efficiency and global efficiency were not collinear and therefore conducted a sensitivity analysis including global efficiency as a covariate. The model was written as:

$$\begin{aligned} &\text{Cognitive efficiency} \sim \text{compression efficiency} + \text{global efficiency} \\ &+ \text{spline(age)} + \text{sex} + \text{spline(age by sex)} + \text{degree} + \text{density} + \text{motion}. \end{aligned} \tag{35}$$

1.8 Code availability

Code can be found at `https://github.com/dalejn/economicsConnectomics`.

1.9 Data availability

Neuroimaging and cognitive test data were acquired from the Philadelphia Neurodevelopmental Cohort. The data reported in this paper have been deposited in the database of Genotypes and Phenotypes under accession number dbGaP: phs000607.v2.p2 (`https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?`

study_id=phs000607.v2.p2). The allometric cortical scaling maps were downloaded from a NeuroVault repository [7] and the cortical myelin maps were downloaded from a public resource [106].

1.10 Citation Diversity Statement

Recent work in several fields of science has identified a bias in citation practices such that papers from women and other minority scholars are under-cited relative to the number of such papers in the field [126, 127, 128, 129, 130, 131, 132, 133, 134]. Here we sought to proactively consider choosing references that reflect thediversity of the field in thought, form of contribution, gender, race, ethnicity, and other factors. First, we obtained the predicted gender of the first and last author of each reference by using databases that store the probability of a first name being carried by a woman [130, 135]. By this measure (and excluding self-citations to the first and last authors of our current paper), our references contain 18.07% woman(first)/woman(last), 5.75% man/woman, 23.33% woman/man, and 52.85% man/man. This method is limited in that a) names, pronouns, and social media profiles used to construct the databases may not, in every case, be indicative of gender identity and b) it cannot account for intersex, non-binary, or transgender people. Second, we obtained predicted racial/ethnic category of the first and last author of each reference by databases that store the probability of a first and last name being carried by an author of color [136, 137]. By this measure (and excluding self-citations), our references contain 7.38% author of color (first)/author of color(last), 12.18% white author/author of color, 24.82% author of color/white author, and 55.62% white author/white author. This method is limited in that a) names and Florida Voter Data to make the predictions may not be indicative of racial/ethnic identity, and b) it cannot account for Indigenous and mixed-race authors, or those who may face differential biases due to the ambiguous racialization or ethnicization of their names. We look forward to future work that could help us to better understand how to support equitable practices in science.

References

[1] M. J. West-Eberhard, *Developmental plasticity and evolution.* Oxford University Press, 2003.

[2] S. B. Laughlin, "Energy as a constraint on the coding and processing of sensory information," *Current opinion in neurobiology*, vol. 11, no. 4, pp. 475–480, 2001.

[3] E. Bullmore and O. Sporns, "The economy of brain network organization," *Nature Reviews Neuroscience*, vol. 13, no. 5, p. 336, 2012.

[4] R. L. Buckner and F. M. Krienen, "The evolution of distributed association networks in the human brain," *Trends in cognitive sciences*, vol. 17, no. 12, pp. 648–665, 2013.

[5] A. Avena-Koenigsberger, J. Goñi, R. Solé, and O. Sporns, "Network morphospace," *Journal of the Royal Society Interface*, vol. 12, no. 103, p. 20140881, 2015.

[6] K. J. Whitaker, P. E. Vértes, R. Romero-Garcia, F. Váša, M. Moutoussis, G. Prabhu, N. Weiskopf, M. F. Callaghan, K. Wagstyl, T. Rittman, *et al.*, "Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome," *Proceedings of the National Academy of Sciences*, vol. 113, no. 32, pp. 9105–9110, 2016.

[7] P. Reardon, J. Seidlitz, S. Vandekar, S. Liu, R. Patel, M. T. M. Park, A. Alexander-Bloch, L. S. Clasen, J. D. Blumenthal, F. M. Lalonde, *et al.*, "Normative brain size variation and brain shape diversity in humans," *Science*, vol. 360, no. 6394, pp. 1222–1227, 2018.

[8] A. Di Martino, D. A. Fair, C. Kelly, T. D. Satterthwaite, F. X. Castellanos, M. E. Thomason, R. C. Craddock, B. Luna, B. L. Leventhal, X.-N. Zuo, *et al.*, "Unraveling the miswired connectome: a developmental perspective," *Neuron*, vol. 83, no. 6, pp. 1335–1353, 2014.

[9] N. A. Crossley, A. Mechelli, J. Scott, F. Carletti, P. T. Fox, P. McGuire, and E. T. Bullmore, "The hubs of the human connectome are generally implicated in the anatomy of brain disorders," *Brain*, vol. 137, no. 8, pp. 2382–2395, 2014.

[10] L. L. Gollo, J. A. Roberts, V. L. Cropley, M. A. Di Biase, C. Pantelis, A. Zalesky, and M. Breakspear, "Fragility and volatility of structural hubs in the human connectome," *Nature neuroscience*, vol. 21, no. 8, p. 1107, 2018.

[11] A. N. Kaczkurkin, T. M. Moore, M. E. Calkins, R. Ciric, J. A. Detre, M. A. Elliott, E. B. Foa, A. G. de la Garza, D. R. Roalf, A. Rosen, *et al.*, "Common and dissociable regional cerebral blood flow differences associate with dimensions of psychopathology across categorical diagnoses," *Molecular psychiatry*, vol. 23, no. 10, p. 1981, 2018.

[12] T. M. Cover, *Elements of information theory.* John Wiley & Sons, 1999.

[13] P. Sterling and S. Laughlin, *Principles of Neural Design.* The MIT Press, 2015.

[14] W. B. Levy and R. A. Baxter, "Energy efficient neural codes," *Neural computation*, vol. 8, no. 3, pp. 531–543, 1996.

[15] M. Chalk, O. Marre, and G. Tkačik, "Toward a unified theory of efficient, predictive, and sparse coding," *Proceedings of the National Academy of Sciences*, vol. 115, no. 1, pp. 186–191, 2018.

[16] E. Tang, M. G. Mattar, C. Giusti, D. M. Lydon-Staley, S. L. Thompson-Schill, and D. S. Bassett, "Effective learning is accompanied by high-dimensional and efficient representations of neural activity," *Nature neuroscience*, vol. 22, no. 6, p. 1000, 2019.

[17] J. M. Shine, M. Breakspear, P. T. Bell, K. A. E. Martens, R. Shine, O. Koyejo, O. Sporns, and R. A. Poldrack, "Human cognition involves the dynamic integration of neural activity and neuromodulatory systems," *Nature neuroscience*, vol. 22, no. 2, p. 289, 2019.

[18] C. Stringer, M. Pachitariu, N. A. Steinmetz, M. Carandini, and K. D. Harris, "High-dimensional geometry of population responses in visual cortex," *Nature*, pp. 1–5, 2019.

[19] M. L. Mack, A. R. Preston, and B. C. Love, "Ventromedial prefrontal cortex compression during concept learning," *Nature Communications*, vol. 11, no. 1, pp. 1–11, 2020.

[20] J. Goni, A. Avena-Koenigsberger, N. V. de Mendizabal, M. P. van den Heuvel, R. F. Betzel, and O. Sporns, "Exploring the morphospace of communication efficiency in complex networks," *PLoS One*, vol. 8, no. 3, p. e58070, 2013.

[21] J. Goñi, M. P. van den Heuvel, A. Avena-Koenigsberger, N. V. de Mendizabal, R. F. Betzel, A. Griffa, P. Hagmann, B. Corominas-Murtra, J.-P. Thiran, and O. Sporns, "Resting-brain functional connectivity predicted by analytic measures of network communication," *Proceedings of the National Academy of Sciences*, vol. 111, no. 2, pp. 833–838, 2014.

[22] A. Avena-Koenigsberger, B. Mišić, R. X. Hawkins, A. Griffa, P. Hagmann, J. Goñi, and O. Sporns, "Path ensembles and a tradeoff between communication efficiency and resilience in the human connectome," *Brain Structure and Function*, vol. 222, no. 1, pp. 603–618, 2017.

[23] B. Mišić, R. F. Betzel, A. Nematzadeh, J. Goni, A. Griffa, P. Hagmann, A. Flammini, Y.-Y. Ahn, and O. Sporns, "Cooperative and competitive spreading dynamics on the human connectome," *Neuron*, vol. 86, no. 6, pp. 1518–1529, 2015.

[24] A. Avena-Koenigsberger, B. Misic, and O. Sporns, "Communication dynamics in complex brain networks," *Nature Reviews Neuroscience*, vol. 19, no. 1, p. 17, 2018.

[25] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec*, vol. 4, no. 142-163, p. 1, 1959.

[26] H. B. Barlow *et al.*, "Possible principles underlying the transformation of sensory messages," *Sensory communication*, vol. 1, pp. 217–234, 1961.

[27] C. R. Sims, "Efficient coding explains the universal law of generalization in human perception," *Science*, vol. 360, no. 6389, pp. 652–656, 2018.

[28] S. N. Vaishnavi, A. G. Vlassenko, M. M. Rundle, A. Z. Snyder, M. A. Mintun, and M. E. Raichle, "Regional aerobic glycolysis in the human brain," *Proceedings of the National Academy of Sciences*, vol. 107, no. 41, pp. 17757–17762, 2010.

[29] R. C. Gur, J. D. Ragland, M. Reivich, J. H. Greenberg, A. Alavi, and R. E. Gur, "Regional differences in the coupling between resting cerebral blood flow and metabolism may indicate action preparedness as a default state," *Cerebral Cortex*, vol. 19, no. 2, pp. 375–382, 2008.

[30] S. E. Marzen and S. DeDeo, "The evolution of lossy compression," *Journal of The Royal Society Interface*, vol. 14, no. 130, p. 20170166, 2017.

[31] M. P. van den Heuvel, R. S. Kahn, J. Goñi, and O. Sporns, "High-cost, high-capacity backbone for global brain communication," *Proceedings of the National Academy of Sciences*, vol. 109, no. 28, pp. 11372–11377, 2012.

[32] A. Fornito, A. Zalesky, and E. T. Bullmore, "Chapter 7 - paths, diffusion, and navigation," in *Fundamentals of Brain Network Analysis*, pp. 207 – 255, San Diego: Academic Press, 2016.

[33] A. Raj, E. LoCastro, A. Kuceyeski, D. Tosun, N. Relkin, M. Weiner, A. D. N. I. (ADNI, *et al.*, "Network diffusion model of progression predicts longitudinal patterns of atrophy and metabolism in alzheimer's disease," *Cell reports*, vol. 10, no. 3, pp. 359–369, 2015.

[34] Y.-Q. Zheng, Y. Zhang, Y. Yau, Y. Zeighami, K. Larcher, B. Misic, and A. Dagher, "Local vulnerability and global connectivity jointly shape neurodegenerative disease propagation," *PLoS biology*, vol. 17, no. 11, p. e3000495, 2019.

[35] M. X. Henderson, E. J. Cornblath, A. Darwich, B. Zhang, H. Brown, R. J. Gathagan, R. M. Sandler, D. S. Bassett, J. Q. Trojanowski, and V. M. Lee, "Spread of α-synuclein pathology through the brain connectome is modulated by selective vulnerability and predicted by network analysis," *Nature neuroscience*, vol. 22, no. 8, p. 1248, 2019.

[36] F. Abdelnour, M. Dayan, O. Devinsky, T. Thesen, and A. Raj, "Functional brain connectivity is predictable from anatomic network's laplacian eigen-structure," *NeuroImage*, vol. 172, pp. 728–739, 2018.

[37] C. Seguin, A. Razi, and A. Zalesky, "Inferring neural signalling directionality from undirected structural connectomes," *Nature communications*, vol. 10, no. 1, pp. 1–13, 2019.

[38] C. Paquola, J. Seidlitz, O. Benkarim, J. Royer, P. Klimes, R. A. Bethlehem, S. Lariviere, R. V. de Wael, J. Hall, B. Frauscher, *et al.*, "A cortical wiring space links cellular architecture, funtional dynamics and hierarchies in humans," *bioRxiv*, 2020.

[39] E. Amico, K. Abbas, D. A. Duong-Tran, U. Tipnis, M. Rajapandian, E. Chumin, M. Ventresca, J. Harezlak, and J. Goñi, "Toward an information theoretical description of communication in brain networks," *Network Neuroscience*, vol. 5, no. 3, pp. 646–665, 2021.

[40] J. D. Murray, A. Bernacchia, D. J. Freedman, R. Romo, J. D. Wallis, X. Cai, C. Padoa-Schioppa, T. Pasternak, H. Seo, D. Lee, *et al.*, "A hierarchy of intrinsic timescales across primate cortex," *Nature neuroscience*, vol. 17, no. 12, pp. 1661–1663, 2014.

[41] T. D. Satterthwaite, M. A. Elliott, K. Ruparel, J. Loughead, K. Prabhakaran, M. E. Calkins, R. Hopson, C. Jackson, J. Keefe, M. Riley, *et al.*, "Neuroimaging of the philadelphia neurodevelopmental cohort," *Neuroimage*, vol. 86, pp. 544–553, 2014.

[42] O. Sporns, "Network attributes for segregation and integration in the human brain," *Current opinion in neurobiology*, vol. 23, no. 2, pp. 162–171, 2013.

[43] C. Seguin, M. P. van den Heuvel, and A. Zalesky, "Navigation of brain networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 24, pp. 6297–6302, 2018.

[44] C. Saleeba, B. Dempsey, S. Le, A. Goodchild, and S. McMullan, "A student's guide to neural circuit tracing," *Frontiers in Neuroscience*, vol. 13, p. 897, 2019.

[45] B. Várkuti, M. Cavusoglu, A. Kullik, B. Schiffler, R. Veit, Ö. Yilmaz, W. Rosenstiel, C. Braun, K. Uludag, N. Birbaumer, *et al.*, "Quantifying the link between anatomical connectivity, gray matter volume and regional cerebral blood flow: an integrative mri study," *PLoS One*, vol. 6, no. 4, p. e14801, 2011.

[46] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical review letters*, vol. 87, no. 19, p. 198701, 2001.

[47] M. F. Glasser, M. S. Goyal, T. M. Preuss, M. E. Raichle, and D. C. Van Essen, "Trends and properties of human cerebral cortex: correlations with cortical myelin content," *Neuroimage*, vol. 93, pp. 165–175, 2014.

[48] D. Attwell and S. B. Laughlin, "An energy budget for signaling in the grey matter of the brain," *Journal of Cerebral Blood Flow & Metabolism*, vol. 21, pp. 1133–1145, 2001.

[49] H. Barbas and N. Rempel-Clower, "Cortical structure predicts the pattern of corticocortical connections.," *Cerebral cortex (New York, NY: 1991)*, vol. 7, no. 7, pp. 635–646, 1997.

[50] G. Deco, P. E. Roland, and C. C. Hilgetag, *Cortico-cortical Communication Dynamics*. Frontiers E-books, 2014.

[51] E. Ravasz and A.-L. Barabási, "Hierarchical organization in complex networks," *Physical review E*, vol. 67, no. 2, p. 026112, 2003.

[52] D. S. Bassett, D. L. Greenfield, A. Meyer-Lindenberg, D. R. Weinberger, S. W. Moore, and E. T. Bullmore, "Efficient physical embedding of topologically complex information processing networks in brains and computer circuits," *PLoS computational biology*, vol. 6, no. 4, p. e1000748, 2010.

[53] G. Collin, O. Sporns, R. C. Mandl, and M. P. van den Heuvel, "Structural and functional aspects relating to cost and benefit of rich club organization in the human cerebral cortex," *Cerebral cortex*, vol. 24, no. 9, pp. 2258–2267, 2013.

[54] J. E. Niven, J. C. Anderson, and S. B. Laughlin, "Fly photoreceptors demonstrate energy-information trade-offs in neural coding," *PLoS Biol*, vol. 5, no. 4, p. e116, 2007.

[55] M. Rigotti, O. Barak, M. R. Warden, X.-J. Wang, N. D. Daw, E. K. Miller, and S. Fusi, "The importance of mixed selectivity in complex cognitive tasks," *Nature*, vol. 497, no. 7451, pp. 585–590, 2013.

[56] S. Achard and E. Bullmore, "Efficiency and cost of economical brain functional networks," *PLoS computational biology*, vol. 3, no. 2, p. e17, 2007.

[57] H. Johansen-Berg, "Behavioural relevance of variation in white matter microstructure," *Current opinion in neurology*, vol. 23, no. 4, pp. 351–358, 2010.

[58] S. E. Palmer, O. Marre, M. J. Berry, and W. Bialek, "Predictive information in a sensory population," *Proceedings of the National Academy of Sciences*, vol. 112, no. 22, pp. 6908–6913, 2015.

[59] M. Rubinov, "Constraints and spandrels of interareal connectomes," *Nature communications*, vol. 7, p. 13812, 2016.

[60] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," *Current opinion in neurobiology*, vol. 14, no. 4, pp. 481–487, 2004.

[61] C. R. Sims, "Rate–distortion theory and human perception," *Cognition*, vol. 152, pp. 181–198, 2016.

[62] S. B. Laughlin, R. R. d. R. van Steveninck, and J. C. Anderson, "The metabolic cost of neural information," *Nature neuroscience*, vol. 1, no. 1, p. 36, 1998.

[63] X.-X. Wei and A. A. Stocker, "A bayesian observer model constrained by efficient coding can explain'anti-bayesian'percepts," *Nature neuroscience*, vol. 18, no. 10, p. 1509, 2015.

[64] S. Denève, A. Alemi, and R. Bourdoukan, "The brain as an efficient and robust adaptive learner," *Neuron*, vol. 94, no. 5, pp. 969–977, 2017.

[65] A. I. Weber, K. Krishnamurthy, and A. L. Fairhall, "Coding principles in adaptation," *Annual review of vision science*, vol. 5, pp. 427–449, 2019.

[66] D. J. MacKay and D. J. Mac Kay, *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[67] L. Chanes and L. F. Barrett, "Redefining the role of limbic areas in cortical processing," *Trends in cognitive sciences*, vol. 20, no. 2, pp. 96–106, 2016.

[68] P. S. Goldman-Rakic, "Topography of cognition: parallel distributed networks in primate association cortex.," *Annual review of neuroscience*, 1988.

[69] L. H. Scholtens, R. Schmidt, M. A. de Reus, and M. P. van den Heuvel, "Linking macroscale graph analytical organization to microscale neuroarchitectonics in the macaque connectome," *Journal of Neuroscience*, vol. 34, no. 36, pp. 12192–12205, 2014.

[70] S. G. Vij, J. S. Nomi, D. R. Dajani, and L. Q. Uddin, "Evolution of spatial and temporal features of functional brain networks across the lifespan," *NeuroImage*, vol. 173, pp. 498–508, 2018.

[71] P. Theodoni, P. Majka, D. H. Reser, D. K. Wójcik, M. G. Rosa, and X.-J. Wang, "Structural attributes and principles of the neocortical connectome in the marmoset monkey," *bioRxiv*, 2020.

[72] P. E. Vértes, A. Alexander-Bloch, and E. T. Bullmore, "Generative models of rich clubs in hebbian neuronal networks and large-scale human brain networks," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1653, p. 20130531, 2014.

[73] S. Oldham and A. Fornito, "The development of brain network hubs," *Developmental cognitive neuroscience*, 2018.

[74] A. Avena-Koenigsberger, X. Yan, A. Kolchinsky, M. P. van den Heuvel, P. Hagmann, and O. Sporns, "A spectrum of routing strategies for brain networks," *PLOS Computational Biology*, vol. 15, pp. 1–24, 03 2019.

[75] S. S.-H. Wang, J. R. Shultz, M. J. Burish, K. H. Harrison, P. R. Hof, L. C. Towns, M. W. Wagers, and K. D. Wyatt, "Functional trade-offs in white matter axonal scaling," *Journal of Neuroscience*, vol. 28, no. 15, pp. 4047–4056, 2008.

[76] J. J. Harris and D. Attwell, "The energetics of cns white matter," *Journal of Neuroscience*, vol. 32, no. 1, pp. 356–371, 2012.

[77] X. Liang, Q. Zou, Y. He, and Y. Yang, "Coupling of functional connectivity and regional cerebral blood flow reveals a physiological basis for network hubs of the human brain," *Proceedings of the National Academy of Sciences*, vol. 110, no. 5, pp. 1929–1934, 2013.

[78] A. Zalesky, A. Fornito, L. Cocchi, L. L. Gollo, M. P. van den Heuvel, and M. Breakspear, "Connectome sensitivity or specificity: which is more important?," *Neuroimage*, vol. 142, pp. 407–420, 2016.

[79] D. K. Jones, T. R. Knösche, and R. Turner, "White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion mri," *Neuroimage*, vol. 73, pp. 239–254, 2013.

[80] E. H. Chang, M. Argyelan, M. Aggarwal, T.-S. S. Chandon, K. H. Karlsgodt, S. Mori, and A. K. Malhotra, "The role of myelination in measures of white matter integrity: Combination of diffusion tensor imaging and two-photon microscopy of clarity intact brains," *NeuroImage*, vol. 147, pp. 253–261, 2017.

[81] E. G. Teich, M. Cieslak, B. Giesbrecht, J. M. Vettel, S. T. Grafton, T. D. Satterthwaite, and D. S. Bassett, "Crystallinity characterization of white matter in the human brain," *New Journal of Physics*, vol. 23, no. 7, p. 073047, 2021.

[82] T. D. Satterthwaite, R. T. Shinohara, D. H. Wolf, R. D. Hopson, M. A. Elliott, S. N. Vandekar, K. Ruparel, M. E. Calkins, D. R. Roalf, E. D. Gennatas, *et al.*, "Impact of puberty on the evolution of cerebral perfusion during adolescence," *Proceedings of the National Academy of Sciences*, vol. 111, no. 23, pp. 8643–8648, 2014.

[83] E. Tang, C. Giusti, G. L. Baum, S. Gu, E. Pollock, A. E. Kahn, D. R. Roalf, T. M. Moore, K. Ruparel, R. C. Gur, *et al.*, "Developmental increases in white matter network controllability support a growing diversity of brain dynamics," *Nature communications*, vol. 8, no. 1, p. 1252, 2017.

[84] P. Srivastava, E. Nozari, J. Z. Kim, H. Ju, D. Zhou, C. Becker, F. Pasqualetti, G. J. Pappas, and D. S. Bassett, "Models of communication and control for brain networks: distinctions, convergence, and future outlook," *Network Neuroscience*, vol. 4, no. 4, pp. 1122–1159, 2020.

[85] A. C. Schapiro, N. B. Turk-Browne, M. M. Botvinick, and K. A. Norman, "Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1711, p. 20160049, 2017.

[86] K. L. Stachenfeld, M. M. Botvinick, and S. J. Gershman, "The hippocampus as a predictive map," *Nature neuroscience*, vol. 20, no. 11, p. 1643, 2017.

[87] I. Momennejad, "Learning structures: Predictive representations, replay, and generalization," *Current Opinion in Behavioral Sciences*, vol. 32, pp. 155–166, 2020.

[88] H. Barbas, "Pattern in the laminar origin of corticocortical connections," *Journal of Comparative Neurology*, vol. 252, no. 3, pp. 415–422, 1986.

[89] M. A. Kingsbury and B. L. Finlay, "The cortex in multidimensional space: where do cortical areas come from?," *Developmental Science*, vol. 4, no. 2, pp. 125–142, 2001.

[90] C. J. Charvet and B. L. Finlay, "Evo-devo and the primate isocortex: the central organizing role of intrinsic gradients of neurogenesis," *Brain, Behavior and Evolution*, vol. 84, no. 2, pp. 81–92, 2014.

[91] C. J. Charvet, D. J. Cahalane, and B. L. Finlay, "Systematic, cross-cortex variation in neuron numbers in rodents and primates," *Cerebral Cortex*, vol. 25, no. 1, pp. 147–160, 2015.

[92] J. M. Huntenburg, P.-L. Bazin, and D. S. Margulies, "Large-scale gradients in human cortical organization," *Trends in cognitive sciences*, vol. 22, no. 1, pp. 21–31, 2018.

[93] B. Vazquez-Rodriguez, Z.-Q. Liu, P. Hagmann, and B. Misic, "Signal propagation via cortical hierarchies," *BioRxiv*, 2020.

[94] U. A. Tooley, D. S. Bassett, and A. P. Mackey, "Environmental influences on the pace of brain development," *Nature Reviews Neuroscience*, vol. 22, no. 6, pp. 372–384, 2021.

[95] L. M. Hernandez, J. D. Rudie, S. A. Green, S. Bookheimer, and M. Dapretto, "Neural signatures of autism spectrum disorders: insights into brain network dynamics," *Neuropsychopharmacology*, vol. 40, no. 1, p. 171, 2015.

[96] I. Dinstein, D. J. Heeger, L. Lorenzi, N. J. Minshew, R. Malach, and M. Behrmann, "Unreliable evoked responses in autism," *Neuron*, vol. 75, no. 6, pp. 981–991, 2012.

[97] D. R. Roalf, M. Quarmley, M. A. Elliott, T. D. Satterthwaite, S. N. Vandekar, K. Ruparel, E. D. Gennatas, M. E. Calkins, T. M. Moore, R. Hopson, *et al.*, "The impact of quality assurance assessment on diffusion tensor imaging outcomes in a large-scale population-based cohort," *Neuroimage*, vol. 125, pp. 903–919, 2016.

[98] T. M. Moore, S. P. Reise, R. E. Gur, H. Hakonarson, and R. C. Gur, "Psychometric properties of the penn computerized neurocognitive battery.," *Neuropsychology*, vol. 29, no. 2, p. 235, 2015.

[99] G. L. Baum, R. Ciric, D. R. Roalf, R. F. Betzel, T. M. Moore, R. T. Shinohara, A. E. Kahn, S. N. Vandekar, P. E. Rupert, M. Quarmley, *et al.*, "Modular segregation of structural brain networks supports the development of executive function in youth," *Current Biology*, vol. 27, no. 11, pp. 1561–1572, 2017.

[100] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, p. 171, 2016.

[101] L. Cammoun, X. Gigandet, D. Meskaldji, J. P. Thiran, O. Sporns, K. Q. Do, P. Maeder, R. Meuli, and P. Hagmann, "Mapping the human connectome at multiple scales with diffusion spectrum mri," *Journal of neuroscience methods*, vol. 203, no. 2, pp. 386–397, 2012.

[102] F.-C. Yeh, T. D. Verstynen, Y. Wang, J. C. Fernández-Miranda, and W.-Y. I. Tseng, "Deterministic diffusion fiber tracking improved by quantitative anisotropy," *PloS one*, vol. 8, no. 11, p. e80713, 2013.

[103] Z. Wang, G. K. Aguirre, H. Rao, J. Wang, M. A. Fernández-Seara, A. R. Childress, and J. A. Detre, "Empirical optimization of asl data analysis using an asl data processing toolbox: Asltbx," *Magnetic resonance imaging*, vol. 26, no. 2, pp. 261–269, 2008.

[104] W.-C. Wu, V. Jain, C. Li, M. Giannetta, H. Hurt, F. W. Wehrli, and D. J. Wang, "In vivo venous blood t1 measurement using inversion recovery true-fisp in children and adults," *Magnetic resonance in medicine*, vol. 64, no. 4, pp. 1140–1147, 2010.

[105] V. Jain, J. Duda, B. Avants, M. Giannetta, S. X. Xie, T. Roberts, J. A. Detre, H. Hurt, F. W. Wehrli, and D. J. Wang, "Longitudinal reproducibility and accuracy of pseudo-continuous arterial spin–labeled perfusion mr imaging in typically developing children," *Radiology*, vol. 263, no. 2, pp. 527–536, 2012.

[106] M. F. Glasser and D. C. Van Essen, "Mapping human cortical areas in vivo based on myelin content as revealed by t1-and t2-weighted mri," *Journal of Neuroscience*, vol. 31, no. 32, pp. 11597–11616, 2011.

[107] S. N. Wood, "Stable and efficient multiple smoothing parameter estimation for generalized additive models," *Journal of the American Statistical Association*, vol. 99, no. 467, pp. 673–686, 2004.

[108] C. O. Becker, S. Pequito, G. J. Pappas, M. B. Miller, S. T. Grafton, D. S. Bassett, and V. M. Preciado, "Spectral mapping of brain functional connectivity from diffusion imaging," *Scientific reports*, vol. 8, no. 1, p. 1411, 2018.

[109] O. Sporns and R. F. Betzel, "Modular brain networks," *Annual review of psychology*, vol. 67, pp. 613–640, 2016.

[110] D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, and S. T. Grafton, "Dynamic reconfiguration of human brain networks during learning," *Proceedings of the National Academy of Sciences*, vol. 108, no. 18, pp. 7641–7646, 2011.

[111] M. A. Bertolero, B. T. Yeo, and M. D'Esposito, "The modular and integrative functional architecture of the human brain," *Proceedings of the National Academy of Sciences*, vol. 112, no. 49, pp. E6798–E6807, 2015.

[112] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[113] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[114] T. Yarkoni, R. Poldrack, T. Nichols, D. V. V. Essen, and T. Wager, "Large-scale automated synthesis of human functional neuroimaging data," *Nature methods*, vol. 8, pp. 665 – 670, 2011.

[115] J. V. Stone, "Principles of neural information theory a tutorial introduction," 2016.

[116] D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, *et al.*, "Function in the human connectome: task-fmri and individual differences in behavior," *Neuroimage*, vol. 80, pp. 169–189, 2013.

[117] W. Alt, "Biased random walk models for chemotaxis and related diffusion approximations," *Journal of mathematical biology*, vol. 9, no. 2, pp. 147–177, 1980.

[118] M. P. van den Heuvel and O. Sporns, "Network hubs in the human brain," *Trends in cognitive sciences*, vol. 17, no. 12, pp. 683–696, 2013.

[119] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, "Detecting rich-club ordering in complex networks," *Nature physics*, vol. 2, no. 2, p. 110, 2006.

[120] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science*, vol. 296, no. 5569, pp. 910–913, 2002.

[121] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[122] S. Marek, B. Tervo-Clemmens, F. J. Calabro, D. F. Montez, B. P. Kay, A. S. Hatoum, M. R. Donohue, W. Foran, R. L. Miller, E. Feczko, O. Miranda-Dominguez, A. M. Graham, E. A. Earl, A. J. Perrone, M. Cordova, O. Doyle, L. A. Moore, G. Conan, J. Uriarte, K. Snider, A. Tam, J. Chen, D. J. Newbold, A. Zheng, N. A. Seider, A. N. Van, T. O. Laumann, W. K. Thompson, D. J. Greene, S. E. Petersen, T. E. Nichols, B. T. Yeo, D. M. Barch, H. Garavan, B. Luna, D. A. Fair, and N. U. Dosenbach, "Towards reproducible brain-wide association studies," *bioRxiv*, 2020.

[123] A. F. Alexander-Bloch, H. Shou, S. Liu, T. D. Satterthwaite, D. C. Glahn, R. T. Shinohara, S. N. Vandekar, and A. Raznahan, "On testing for spatial correspondence between maps of human brain structure and function," *Neuroimage*, vol. 178, pp. 540–551, 2018.

[124] F. Váša, J. Seidlitz, R. Romero-Garcia, K. J. Whitaker, G. Rosenthal, P. E. Vértes, M. Shinn, A. Alexander-Bloch, P. Fonagy, R. J. Dolan, *et al.*, "Adolescent tuning of association cortex in human structural brain networks," *Cerebral Cortex*, vol. 28, no. 1, pp. 281–294, 2018.

[125] R. D. Markello and B. Misic, "Comparing spatial null models for brain maps," *NeuroImage*, vol. 236, p. 118052, 2021.

[126] S. M. Mitchell, S. Lange, and H. Brus, "Gendered citation patterns in international relations journals," *International Studies Perspectives*, vol. 14, no. 4, pp. 485–492, 2013.

[127] M. L. Dion, J. L. Sumner, and S. M. Mitchell, "Gendered citation patterns across political science and social science methodology fields," *Political Analysis*, vol. 26, no. 3, pp. 312–327, 2018.

[128] N. Caplar, S. Tacchella, and S. Birrer, "Quantitative evaluation of gender bias in astronomical publications from citation counts," *Nature Astronomy*, vol. 1, no. 6, p. 0141, 2017.

[129] D. Maliniak, R. Powers, and B. F. Walter, "The gender citation gap in international relations," *International Organization*, vol. 67, no. 4, pp. 889–922, 2013.

[130] J. D. Dworkin, K. A. Linn, E. G. Teich, P. Zurn, R. T. Shinohara, and D. S. Bassett, "The extent and drivers of gender imbalance in neuroscience reference lists," *bioRxiv*, 2020.

[131] M. A. Bertolero, J. D. Dworkin, S. U. David, C. L. Lloreda, P. Srivastava, J. Stiso, D. Zhou, K. Dzirasa, D. A. Fair, A. N. Kaczkurkin, B. J. Marlin, D. Shohamy, L. Q. Uddin, P. Zurn, and D. S. Bassett, "Racial and ethnic imbalance in neuroscience reference lists and intersections with gender," *bioRxiv*, 2020.

[132] X. Wang, J. D. Dworkin, D. Zhou, J. Stiso, E. B. Falk, D. S. Bassett, P. Zurn, and D. M. Lydon-Staley, "Gendered citation practices in the field of communication," *Annals of the International Communication Association*, 2021.

[133] P. Chatterjee and R. M. Werner, "Gender disparity in citations in high-impact journal articles," *JAMA Netw Open*, vol. 4, no. 7, p. e2114509, 2021.

[134] J. M. Fulvio, I. Akinnola, and B. R. Postle, "Gender (im)balance in citation practices in cognitive neuroscience," *J Cogn Neurosci*, vol. 33, no. 1, pp. 3–7, 2021.

[135] D. Zhou, E. J. Cornblath, J. Stiso, E. G. Teich, J. D. Dworkin, A. S. Blevins, and D. S. Bassett, "Gender diversity statement and code notebook v1.0," Feb. 2020.

[136] A. Ambekar, C. Ward, J. Mohammed, S. Male, and S. Skiena, "Name-ethnicity classification from open sources," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, pp. 49–58, 2009.

[137] G. Sood and S. Laohaprapanon, "Predicting race and ethnicity from the sequence of characters in a name," *arXiv preprint arXiv:1805.02109*, 2018.