



Can brain state be manipulated to emphasize individual differences in functional connectivity?[☆]



Emily S. Finn^{a,*}, Dustin Scheinost^b, Daniel M. Finn^b, Xilin Shen^b, Xenophon Papademetris^{b,c}, R. Todd Constable^{a,b,d}

^a Interdepartmental Neuroscience Program, Yale University, New Haven, CT, USA

^b Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA

^c Department of Biomedical Engineering, Yale University, New Haven, CT, USA

^d Department of Neurosurgery, Yale School of Medicine, New Haven, CT, USA

ARTICLE INFO

Keywords:

fMRI
Functional connectivity
Individual differences
Brain state
Scan condition
Resting state
Task
Human Connectome Project

ABSTRACT

While neuroimaging studies typically collapse data from many subjects, brain functional organization varies between individuals, and characterizing this variability is crucial for relating brain activity to behavioral phenotypes. Rest has become the default state for probing individual differences, chiefly because it is easy to acquire and a supposed neutral backdrop. However, the assumption that rest is the optimal condition for individual differences research is largely untested. In fact, other brain states may afford a better ratio of within- to between-subject variability, facilitating biomarker discovery. Depending on the trait or behavior under study, certain tasks may bring out meaningful idiosyncrasies across subjects, essentially enhancing the individual signal in networks of interest beyond what can be measured at rest. Here, we review theoretical considerations and existing work on how brain state influences individual differences in functional connectivity, present some preliminary analyses of within- and between-subject variability across conditions using data from the Human Connectome Project, and outline questions for future study.

Introduction

While more than two decades of neuroimaging have established a general blueprint for brain functional organization, less is known about the quality and quantity of the individual variation that occurs atop this blueprint (Van Horn et al., 2008). There are two main reasons to study individual differences in the course of neuroimaging research. First, of basic scientific concern, more precise descriptions of brain activity in single subjects moves us closer to a mechanistic understanding of how neural events give rise to cognition and action. Second, of practical concern, mapping from individual brains to individual behaviors is a crucial step in developing imaging-based biomarkers with real-world utility.

Functional connectivity, a family of methods investigating correlated activity between two or more brain regions, has exploded in popularity in recent years (Power et al., 2014; van den Heuvel and Hulshoff Pol, 2010). Even if task-evoked activity patterns in individual brain regions are grossly similar (which they may well not be, at least for certain tasks), functional connectivity analyses increase the features

under study by a power of two—from univariate activity in N brain regions to interactions between $N \times N$ pairs of brain regions. This vastly expands the space for both within- and between-subject variation. Because functional connectivity is thought to reveal intrinsic brain organization, many have high hopes that if biomarkers are to be found in the fMRI signal, it will be through functional connectivity analysis (Biswal et al., 2010; Castellanos et al., 2013; Kelly et al., 2012).

“Resting state,” in which subjects lie quietly in the scanner and do nothing in particular, is the default condition for measuring functional connectivity. Individual differences in resting-state connectivity have been linked to a variety of traits and behaviors, including intelligence (Finn et al., 2015; Hearne et al., 2016; van den Heuvel et al., 2009), sustained attention (Rosenberg et al., 2015), impulsivity (Li et al., 2013), and a wide variety of demographic and lifestyle factors (Smith et al., 2015). However, because functional connectivity is not static—connection strengths vary during a single resting-state session (Betzel et al., 2016; Chang and Glover, 2010; Gonzalez-Castillo et al., 2014; Hutchison et al., 2013) and between sessions (Laumann et al., 2015; Shine et al., 2016), as well as with various cognitive tasks (Cole et al.,

[☆] Part of the NeuroImage Special Issue on Functional Architecture of the Brain.

* Corresponding author.

E-mail address: emily.finn@yale.edu (E.S. Finn).

2013; Gonzalez-Castillo et al., 2015)—individual differences in functional connectivity are likely not static either. In other words, both within- and between-subject variability in connectivity probably depend on the condition in which connectivity is measured. Understanding how brain state affects within- and between-subject variability in functional connectivity is of basic interest in cognitive neuroscience, and will also be crucial for developing robust connectivity-based biomarkers. Yet, to date, this question has received relatively little attention from the field.

Our goals here are threefold. First, we offer some theoretical considerations and review existing work on how brain state (as manipulated by scan condition) influences individual differences in functional connectivity. Second, we present some observations of within- and between-subject variability as a function of state using data from the Human Connectome Project. Finally, we summarize several outstanding questions for future study. While this review may pose more questions than it answers, by highlighting these issues, we hope to inspire other researchers to pursue these lines of investigation.

Theoretical considerations and review

A thought experiment

Hypothetically, what would be the ideal condition for measuring individual differences? Simply maximizing between-subject variability—i.e., making subjects look as different as possible from one another—is not necessarily the answer. Rather, the optimal condition should make subjects look as different as possible while also retaining

the most important features of each individual. Consider the toy example in Fig. 1. Imagine that the ground-truth similarity between these three individuals is $r=0.50$ (middle box), but this is a latent value and not directly observable. Instead, we must measure individual differences in one of various experimental conditions, each of which is associated with different levels of between-subject similarity (measured with correlation) and individual-subject identifiability (a subjective metric reflecting ease of recognition). Randomly scrambling the images makes them less similar to one another, but renders the individuals all but unidentifiable (lower left). Conversely, blurring the images makes them more similar, but likewise impedes identifiability (lower right). Neither manipulation is ideal for studying individual differences; thus it seems between-subject variability is not the most important factor.

Dissociable from similarity is identifiability, a metric that is, in this case, subjective, but is a proxy for trait-level variance associated with some phenotype of interest. There are two ways to boost identifiability: by exaggerating the most prominent features of each individual (the “caricature” condition, upper left), which incidentally makes subjects look more different from one another, or by blurring irrelevant features while retaining and enriching relevant ones (“spotlight,” upper right). The second approach actually makes subjects look overall more similar to one another while also making it easy to recognize individuals. Both caricature and spotlight conditions may prove useful for studying individual differences; which conditions are best for which subject populations and phenotypes of interest is an empirical question that we encourage investigators to explore.

We note that our primary focus here is on conditions that minimize

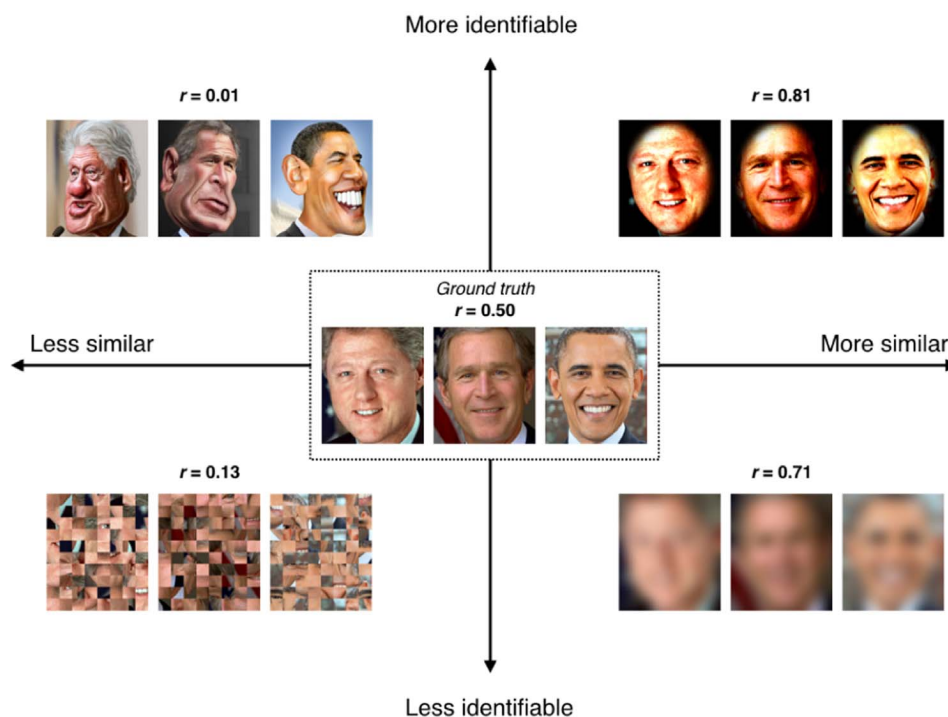


Fig. 1. What is the optimal brain state for measuring individual differences? A thought experiment to illustrate how between-subject variability can be dissociated from single-subject identifiability. Consider images of three individuals with a ground-truth similarity of $r=0.50$ (average of the correlation coefficients between all three pairs of image RGB values; middle box), but this is a latent value and not directly observable. Instead, individual differences are measured in various experimental conditions, each of which is associated with different levels of between-subject similarity (measured with r) and individual-subject identifiability (a subjective metric reflecting ease of recognition). A condition that makes subjects look maximally different (lower left) is not ideal for studying individual differences, since it may destroy key features of each individual. On the other hand, increasing similarity across subjects at the expense of obscuring any individual features (lower right) is also not ideal. Thus, the optimal brain state for measuring individual differences is likely one that evokes some level of divergence across subjects while stabilizing the most important features of each subject. Such a state may evoke either an overall increase in between-subject variability (upper left, “caricature” condition) or, somewhat paradoxically, a decrease in overall between-subject variability (upper right, “selective enhancement” condition), as long as the key individual features are preserved or even enhanced. *Image credits: Presidential photographs, whitehouse.gov; Clinton caricature, www.flickr.com/photos/donkeyhotey/10964745624/; Bush caricature, www.flickr.com/photos/donkeyhotey/29513525751/; Obama caricature, www.flickr.com/photos/donkeyhotey/5601868538/. All caricatures provided under Creative Commons Attribution 2.0 Generic license.*

within-subject variance, with an eye toward developing connectivity-based biomarkers for largely static, trait-related variables, such as intrinsic cognitive abilities, personality traits, or clinically relevant risk factors indicating present or future diagnostic status. However, far from being all noise, many within-subject changes carry biologically meaningful information. Conditions that minimize within-subject variance might obscure information relevant to behavioral changes with development, disease progression or treatment response. Thus, conditions that serve as the best biomarkers for static traits may not be ideal for capturing meaningful within-subject variance, and vice versa.

Candidate brain states for measuring individual differences

Rest, the default condition for functional connectivity research, has undeniable benefits (Greicius, 2008; Smith et al., 2013). Resting-state data is convenient to acquire and standardize across studies and sites; it requires little or no subject participation and is thus free of attention, performance or motivation confounds; it is flexibly analyzed, lending itself to a variety of data-driven techniques; and it is robust to practice effects, making it well suited to longitudinal designs.

However, many of rest's "advantages," such as the lack of experimenter-imposed stimuli or behavior, come at a price. Rest is a task in and of itself, just an unconstrained one (Morcom and Fletcher, 2007). Careful investigations have found that resting-state connectivity can vary with factors such as whether subjects' eyes are open or closed (Patriat et al., 2013), how awake they are (Chang et al., 2016; Tagliazucchi and Laufs, 2014), the content of their spontaneous thoughts (Christoff et al., 2009; Gorgolewski et al., 2014; Shirer et al., 2012), mood (Harrison et al., 2008), after effects of any tasks they recently completed (Barnes et al., 2009; Tung et al., 2013), and several other factors (Duncan and Northoff, 2013). Because there is no explicitly measured input and output, it is difficult to separate the individual resting-state signal into meaningful trait-related variance and less interesting state-related components.

What if there were a brain state better than rest for measuring individual differences? Think of a stress test, in which patients exercise in a controlled setting while a physician monitors their cardiac activity; such tests help diagnose heart disorders before resting heart rhythms become abnormal. Analogously, certain in-scanner tasks could act as neuropsychiatric "stress tests" to enhance individual differences in the general population, or, in at-risk individuals, to reveal abnormal patterns of brain activity before they show up at rest.

While a variety of tasks might serve this role, naturalistic paradigms—e.g., having subjects watch a movie or listen to a story in the scanner—are especially intriguing candidates. By imposing a standardized yet engaging stimulus on all subjects, naturalistic tasks evoke rich patterns of brain activity. These patterns lend themselves to functional connectivity analysis as well as other data-driven techniques such as inter-subject correlation (ISC) (Hasson et al., 2004) and inter-subject functional connectivity (ISFC) (Simony et al., 2016), which are model-free ways to isolate stimulus-dependent brain activity from spontaneous activity and noise. Because these techniques rely on activity that is time-locked across individuals, they cannot be applied to resting-state data. Functional connectivity and ISC/ISFC analysis applied to naturalistic tasks also have several advantages over traditional event-related task approaches: these analyses do not require *a priori* modeling of specific events and/or assumptions about the functional specificity of individual brain regions; there is no need to assume a fixed hemodynamic response function; and they allow for the characterization of the full spatiotemporal richness of both evoked and intrinsic brain activity.

The techniques of ISC and ISFC crucially depend on some baseline degree of between-subject similarity in activity patterns (Hasson et al., 2010), but they may nevertheless be extended to study variation occurring over and above this baseline. Studies have reported that the brains of those with autism look less similar to one another than the

brains of neurotypical individuals while watching movies of social interactions (Byrge et al., 2015; Salmi et al., 2013); the degree of asynchrony may scale with autism-like phenotypes in both the patient and control groups (Salmi et al., 2013). This latter result suggests that this approach could prove fruitful for studying individual differences even in subclinical populations. Might healthy subjects be clustered into distinct groups according to ISC or ISFC during stimulus exposure, and might these groups relate to personality or cognitive traits?

Intuitively, it seems that naturalistic stimuli may be a "happy medium" between rest, which is entirely unconstrained, and traditional cognitive tasks, which may be overly constrained and lack ecological validity: naturalistic stimuli impose a meaningful timecourse across subjects while still allowing for individual variation in brain activity and behavioral responses, and lend themselves to a broader set of analyses than either pure rest or pure event-related task designs. However, this remains to be tested empirically, as few studies have directly compared insights gleaned from traditional tasks versus those gleaned from naturalistic tasks in the same subject group. In one encouraging example, Cantlon and Li (Cantlon and Li, 2013) found that the "neural maturity" of children's responses in certain brain regions—as measured by ISC between child and adult activity patterns—during natural viewing of educational Sesame Street videos predicted their performance on a standardized math test. Predictions of test performance could only be generated from data acquired during natural viewing and analyzed using ISC, and not simply from activation magnitudes in the same regions evoked by a classic event-related matching task. This result supports the hypothesis that certain naturalistic tasks may evince behaviorally relevant individual differences above and beyond what can be measured in other, more traditional conditions.

In another recent study using rest and two different natural viewing conditions, we found that movies made subjects look more similar to one another as well as to themselves, and boosted accuracy of individual-subject identification compared to the rest scans (Vanderwal et al., 2016). This suggests that movies may fall into the "spotlight" category of conditions (cf. Fig. 1, upper right corner), making them desirable for individual differences research.

While naturalistic tasks are promising, any task that elicits variable brain activity and/or behavior across subjects is a worthwhile candidate. For example, highly interactive paradigms—such as gambling and other decision making tasks (Helfinstein et al., 2014; Pushkarskaya et al., 2015), learning tasks (e.g., having subjects acquire novel vocabulary (Breitenstein et al., 2005) or motor skills (Bassett et al., 2011)), or virtual reality games (Spiers and Maguire, 2006)—can also draw out interesting individual differences.

On a practical note, subjects tend to move less in the scanner if they are engaged in a task rather than simply resting (Vanderwal et al., 2015). Because head motion is a notorious confound for functional connectivity analyses (Power et al., 2015; Satterthwaite et al., 2012), and also because head motion is a "trait" that tends to be consistent within subjects across scans, it is especially important to minimize head motion in data sets designed to study individual differences. Drowsiness is also a potential confound (Chang et al., 2016; Tagliazucchi and Laufs, 2014), yet tasks can promote more consistent arousal both within and across subjects, since people have an easier time staying awake when they are engaged in a task (Vanderwal et al., 2015). To the extent that tasks can help alleviate these two major confounds, investigators may choose to use them over rest for purely practical reasons even as we continue to explore their effects on individual differences.

Of course, the question of which condition is best for measuring individual differences may not have a one-size-fits-all answer; it likely depends on both the population and specific traits under study. However, certain categories of conditions—for example, naturalistic tasks—may prove more sensitive than others for drawing out meaningful idiosyncrasies (Dubois, 2016). Within such categories, stimuli

may be customized to probe specific traits of interest. For example, one might use a stimulus with rich social scenarios to study theory of mind and how it is disrupted in autism spectrum disorders (following [Byrge et al., 2015](#)), or one with suspicious content to probe threat perception in the general population as well as those suffering from paranoid delusions.

Previous work on brain state and individual differences

Much previous work has been concerned with separating the functional connectivity signal into state versus trait components: for example, by investigating the test-retest reliability of resting-state connectivity within individuals (depending on the choices of connectivity analysis and reliability metric, it's generally low to moderate ([Birn et al., 2013](#); [Braun et al., 2012](#); [Shehzad et al., 2009](#); [Zuo and Xing, 2014](#)), or comparing group-averaged network organization across brain states (it's grossly similar between rest and any of several tasks ([Cole et al., 2014](#); [Smith et al., 2009](#))). However, there is little work at the intersection of these lines of research, investigating how brain state affects test-retest reliability of single subjects, or measurements of individual differences across subjects.

Using a large lifespan sample, [Geerligs et al. \(2015\)](#) reported between-subject variability in whole-brain functional connectivity during three different conditions: resting, performing a sensorimotor task, and watching a movie. Based on correlations between pairs of individual-subject connectomes, they found that subjects looked most similar during the sensorimotor task, intermediately similar at rest and least similar during movie watching. While significant, differences between conditions were subtle (mean between-subject correlations were 0.43, 0.42 and 0.4 in the sensorimotor, rest and movie conditions, respectively). The authors also found that the relative similarity of any given pair of subjects was not necessarily preserved across conditions; rather, different conditions had distinct effects on different pairs of subjects.

In a recent paper using data from the HCP 500-subject release, [Shah et al.](#) investigated reproducibility of both group-level and individual functional connectivity across rest and the HCP task suite ([Shah et al., 2016](#)). Condition induced systematic differences in group-averaged connectivity, but individual differences in connectivity were relatively maintained across conditions. While these results are informative, as the authors themselves acknowledge, any effects of condition were likely confounded by substantial differences in scan duration between conditions.

Single-subject identification

Brain state may affect how similar subjects look to one another, but can it go so far as to make an individual look like someone else? In other words, what is the ratio of within- to between-subject variability in the face of changing scan conditions? Two theoretical scenarios are depicted in [Fig. 2](#). In the first ([Fig. 2](#), left panel), different states (labeled *a*, *b* and *c*) can push subjects around their personal “connectome space,” but these spaces are non-overlapping. Thus, subject accounts for more of the variance than state, and subjects should always be perfectly identifiable. In the second scenario ([Fig. 2](#), right panel), the variance associated with state expands such that individual connectome spaces do overlap, and in some states a given subject might be mistaken for someone else (for example, subject 1 might be mistaken for subject 3 in state *b*, and any of the three subjects might be mistaken for a different subject in state *c*). Quantifying the variance associated with subject versus brain state is important for understanding state- versus trait-related components of functional connectivity.

We addressed this question in a previous study ([Finn et al., 2015](#)) using HCP data from 126 healthy subjects each scanned during two rest sessions and four different tasks (working memory, motor,

language and emotion). A whole-brain functional connectivity profile calculated from one scan condition could be used to identify the same individual from a different scan condition with very high accuracy: as high as 94% between the two rest sessions, and between 54% and 88% between a rest-task pair or two different tasks (note that chance in all cases is less than 1%, or approximately 1/126). This indicates that much of the variance in functional connectivity is intrinsic to an individual, and not associated with how the brain is engaged during scanning. Similar results have now been reported using other data sets ([Airan et al., 2016](#)).

However, just because subject accounts for more variance than brain state does not mean that tasks do not induce important changes in connectivity ([Gonzalez-Castillo et al., 2015](#); [Richiardi et al., 2011](#)). The drop in accuracy we observed when identifying subjects across rest and task or two different tasks is likely due to some combination of two factors: certain states make people look less like themselves (increasing within-subject variability), and/or certain states make people look more similar to one another (decreasing between-subject variability, thus increasing the chance that one individual will be confused for another). Here we empirically investigate these possibilities.

Observations from HCP data

Building on previous work ([Geerligs et al., 2015](#); [Shah et al., 2016](#)), in this section we present results of exploratory analyses to indicate that brain state, as manipulated by scan condition, impacts both within- and between-subject variability in functional connectivity.

Methods

Data were obtained from the Human Connectome Project (HCP; [Van Essen et al., 2013](#)), 900 subjects release. Analyses were limited to 716 subjects that had complete data for each of nine functional scans: EMOTION, GAMBLING, LANGUAGE, MOTOR, RELATIONAL, REST1, REST2, SOCIAL and WORKING MEMORY (WM). For details of scan parameters, see [Uğurbil et al. \(2013\)](#). Brief descriptions of each task are provided in [Table 1](#); see [Barch et al. \(2013\)](#) and [Smith et al. \(2013\)](#) for further detail on the task and rest scans, respectively. In contrast to previous work ([Shah et al., 2016](#)), we control for differences in the length of scans by truncating timecourses for all conditions to an equal length, allowing us to make stronger inferences about how specific brain states affect both within- and between-subject variability.

We note that while the HCP contains a diverse set of tasks, they are all traditional paradigms chosen because they evoke well-characterized activity in specific neural systems—in other words, they act as functional localizers ([Barch et al., 2013](#)). One limitation of the current study, then, is because the HCP does not include any naturalistic tasks, we are not able to directly compare the performance of traditional cognitive tasks with naturalistic stimuli for drawing out individual differences. However, given the size and quality of the HCP data set—containing by far the most subjects scanned under the most conditions—it can still provide important insights and a proof-of-principle that brain state affects the measurement of individual differences.

Connectivity matrices were calculated as described in [Finn et al. \(2015\)](#). Starting with the minimally preprocessed HCP data ([Glasser et al., 2013](#)), further preprocessing steps were performed using BioImage Suite ([Joshi et al., 2011](#)) and included regressing 12 motion parameters, regressing the mean time courses of the white matter and cerebro-spinal fluid as well as the global signal, removing the linear trend, and low-pass filtering. Task connectivity was calculated based on the “raw” task timecourses, with no regression of task-evoked activity. Following preprocessing, we applied a functional brain parcellation defined on resting-state data from a separate group of healthy adults with 268 nodes covering the cortex, subcortex and cerebellum ([Shen et al., 2013](#)). After averaging the signal in each node, we calculated the Pearson correlation between the timecourses of each pair of nodes and

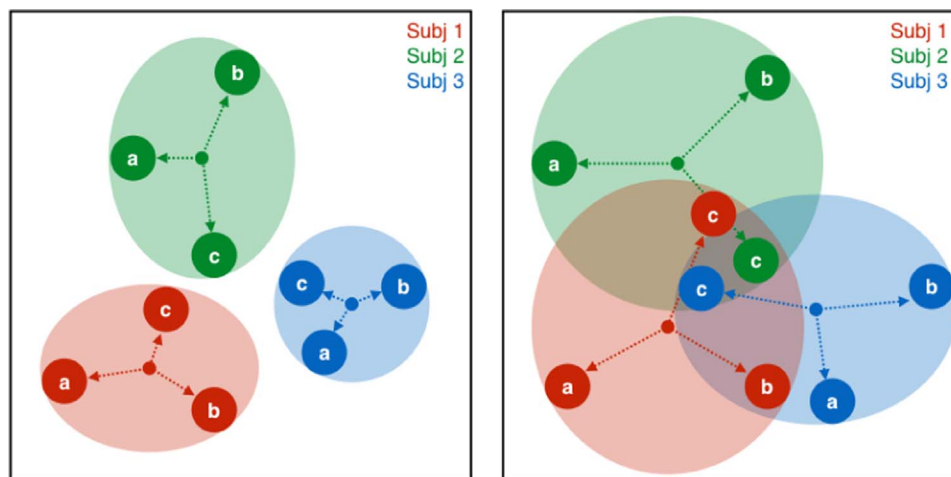


Fig. 2. Hypothetical depictions of within- versus between-subject variability as a function of brain state. Each color (red, green, blue) is a subject, each of whom is measured in three different brain states (*a*, *b*, *c*). In the left panel, while states can push subjects around their personal “connectome space,” these spaces are non-overlapping. Thus, subject accounts for more of the variance than state, and subjects should always be perfectly identifiable. In the second scenario (right panel), the variance associated with state expands such that individual connectome spaces do overlap, and in some states a given subject might be mistaken for someone else (for example, subject 1 for subject 2 in state *c*).

normalized the resulting *r*-values to *z*-scores using the Fisher transform, resulting in a 268×268 symmetrical connectivity matrix of edge strengths. Note that each condition comprised two functional runs with opposite phase-encoding directions (left-right, “LR”, and right-left, “RL”). We calculated connectivity matrices for each run separately and then averaged these two into a single connectivity matrix per subject per condition. Connectivity matrices included both positive and negative correlations, and were not binarized or otherwise thresholded in any way.

Several studies have shown that longer scan durations substantially increase stability of functional connectivity estimates both within and

across subjects (Airan et al., 2016; Birn et al., 2013; Laumann et al., 2015; Mueller et al., 2015) (although results in (Airan et al., 2016) suggest that 3–4 min of acquisition time is sufficient to differentiate subjects). Functional scans in the HCP protocol differed considerably in duration: the two resting-state acquisitions were the longest with 1,200 volumes each (14:24 min:s), while the emotion scan was shortest, at only 176 volumes (2:06). To ensure that scan duration did not confound any effects of condition, unless otherwise noted, results presented here are based on connectivity matrices calculated using only the first 176 volumes from each run.

Because connectivity matrices are symmetric, we extracted the

Table 1

Brief description of HCP functional scan conditions. Note that there are two runs per condition (one each with right-left and left-right phase encoding), thus total amount of data is equal to twice the value in the third column. For more detail, see Smith et al. (2013) for rest and Barch et al. (2013) for tasks.

HCP name	Abbrev.	Frames per run (duration)	Description
REST1 (first day) REST2 (second day)	R1, R2	1,200 (14:24)	Participants are instructed to keep their eyes open, maintain “relaxed” fixation on a white cross on a dark background, think of nothing in particular, and not fall asleep.
EMOTION	EMO	176 (2:06)	Participants decide which of two faces presented at the bottom of the screen match the face at the top of the screen. Faces have either angry or fearful expressions. In a control condition, participants perform the same task with shapes instead of faces.
GAMBLING	GAM	253 (3:02)	Participants guess whether the number on a mystery card (possible range 1–9) is more or less than 5. There are three possible outcomes: \$1 gain for a correct guess, \$0.50 loss for an incorrect guess, and no change for a neutral trial (card=5).
LANGUAGE	LAN	316 (3:47)	Participants engage in alternating blocks of a story task and a math task, both aurally presented. In the story task, they hear brief fables and complete a forced-choice task between two options for the topic of the fable (e.g., revenge versus reciprocity). In the math task, they solve addition and subtraction problems, also with two response options.
MOTOR	MOT	284 (3:24)	Participants tap their left or right fingers, squeeze their left or right toes, or move their tongue according to visual cues.
RELATIONAL	REL	232 (2:46)	Participants are presented with two pairs of objects, and must determine if the second pair of objects differs along the same dimension as the first pair (e.g., shape or texture). In a control condition, they simply determine whether a single object matches either of the objects in a separate pair on a given dimension.
SOCIAL	SOC	274 (3:17)	Participants view 20-s video clips of shapes either interacting (theory-of-mind condition) or moving randomly, and choose one of three potential responses (‘interaction’, ‘no interaction’, or ‘not sure’).
WM (working memory)	WM	405 (4:51)	Participants perform a visual <i>n</i> -back task, with blocked 0-back and 2-back conditions using four stimulus categories, also blocked (faces, place, tools, body parts).

unique elements by taking the upper triangle of the matrix; this results in a $1 \times 35,778$ vector of edge values for each subject for each condition. These vectors can then be compared using Pearson correlation either between different subjects in the same condition (yielding a 716×716 between-subject correlation matrix for each condition, with 255,970 unique values representing similarity between all possible subject pairs), or within the same subject across conditions (yielding a single 9×9 within-subject correlation matrix for each subject).

Of note, subjects moved more during some conditions than others. To conservatively estimate differences in head motion across conditions, we conducted paired t-tests between individual mean framewise displacements (HCP: *Movement_RelativeRMS.txt*; first 176 volumes) in all possible pairs of nine conditions (36 total); several of these reached (uncorrected) significance. Unfortunately, removing outlier subjects with high head motion across all nine conditions did not alleviate this effect.

Differences in head motion are a potential confound because as we might expect, head motion makes measurements of functional connectivity more variable. Using a median split of subjects based on mean framewise displacement (FD) in the REST1 condition (first 176 volumes), mean between-subject correlations in functional connectivity were 0.39 ± 0.07 within the low-motion group but dropped to 0.32 ± 0.07 in the high-motion group; this difference is highly significant (two-tailed t-test: $t(127,804)=151$, $p < 0.0001$). To assess the extent to which head motion explains differences in between-subject variability across conditions, we correlated each condition's grand mean FD (average of mean FD values from all subjects; one value per condition) with mean between-subject correlation in connectivity (average of the upper triangle of the between-subject correlation matrix; one value per condition). These two measures were not strongly related linearly (Pearson $r(7)=0.17$, $p=0.67$) or nonlinearly (Spearman $r(7)=0.03$, $p=0.95$). We also note that Geerligts et al. carefully investigated the impact of head motion on individual differences and concluded that “in most cases the state effects are larger in the motion-matched sample than in the full sample, suggesting that motion might be associated with an underestimation of true state effects” (Geerligts et al., 2015). Thus, while motion is an ever-present confound, we believe that it does not account for the majority of the condition effects observed here.

We also checked for differences in head motion between males and females, since we investigate sex-condition interactions below. We conducted nine two-tailed t-tests, one per condition, comparing the distributions of mean framewise displacements between female subjects ($n=392$) and male subjects ($n=324$). The only conditions that showed trend-level differences were MOTOR ($t(714)=1.95$, $p=0.052$; females > males) and REST1 ($t(714)=1.65$, $p=0.098$; females > males). All other p-values were 0.10 or greater.

Subjects look more similar to one another in certain brain states

Our exploratory analysis showed that between-subject variability does, in fact, change with brain state. Strikingly, the RELATIONAL task was associated with the highest between-subject similarity, while the two REST sessions, along with the MOTOR session, showed the lowest between-subject similarity; the other six task sessions showed intermediate similarity (Fig. 3a).

Overall, individuals tended to look more similar during tasks than during rest. This is perhaps not surprising: while rest lacks any temporal structure, tasks impose a common external stimulus on all subjects, which should induce stereotyped activity across subjects. However, the rank order of the tasks is also of interest. Here, we have chosen to calculate connectivity on the “raw” task data without regressing any task-evoked activity; this was done to be consistent with our earlier work (Finn et al., 2015) as well as to be more conservative in the identification experiments described below (any task-evoked activity patterns should be more similar across subjects than intrinsic fluctuations, so leaving them in

should make the identification problem harder). Thus, it is possible that differences among tasks could be due in part to basic features such as the modality, rate and timing of stimulus presentation; future investigations should compare between-subject similarity calculated with and without task-based regressors. But other, higher-level factors could also be at play, including nature of the task, difficulty level, and differences in strategy, attention or motivation levels across subjects.

While these features are challenging to quantify, the behavioral data for a subset of conditions can be informative. We focused on the four tasks that measured accuracy on a continuous scale from 0 to 1. Fig. 3b shows distributions of performance on the RELATIONAL, EMOTION, WM and LANGUAGE tasks. Performance was most variable for RELATIONAL, followed by LANGUAGE; in contrast, WM and especially EMOTION performance were heavily left-skewed, indicating a ceiling effect. Thus we might infer that the RELATIONAL and LANGUAGE tasks were most difficult.

Of note, a single task, RELATIONAL, simultaneously produced the most variability in behavior and the least variability in functional connectivity. Why might this be? One interpretation is resource allocation: if we assume that subjects are more engaged during more difficult tasks, then we might expect that more of the brain is devoted to task-relevant processing, essentially propagating the task effect through more of the connectivity matrix in a way that is similar across subjects. Contrast the RELATIONAL task with the MOTOR task, during which subjects simply moved their fingers and toes on cue: here cognitive demands are low, leaving many resources left over for task-irrelevant mind-wandering that is more like rest, and thus less likely to be similar across subjects.

Future studies could test this hypothesis by parametrically manipulating difficulty within a single task paradigm. For example, one could use an n -back task to determine if across-subject similarity increases as n (and thus difficulty) increases. One caution is that engagement may follow an inverted U such that subjects give up as task demands become unreasonable. Overly difficult tasks may then paradoxically begin to resemble rest, especially in less motivated populations or those that struggle to maintain attention.

While difficulty as a proxy for engagement may account for some of the differences between tasks, it is clearly not the only factor at play: as a counterexample to the cognitive load interpretation, LANGUAGE was associated with highly variable performance—suggesting difficulty—but relatively low across-subject similarity. Because the LANGUAGE session comprised alternating blocks of two quite dissimilar tasks (narrative listening and arithmetic; see Table 1), it is difficult to interpret this result in the context of features of the task itself. Diverging strategies, especially for more open-ended tasks, may drive down between-subject similarity. Future studies might test this by manipulating strategy with explicit instruction or via post-hoc interview.

Similarity is generally higher among females than males, but effects vary with brain state

We hypothesized that sex would affect measures of between-subject variability, and it might interact with brain state such that sex effects would be stronger in some states than others. To test this, for each condition, we compared correlations of female subjects with other female subjects to those of male subjects with other male subjects. (Note that female and male brains differ on several dimensions such as overall volume, gray-to-white matter ratio, regional morphometry measures, and structural connectivity (Allen et al., 2003; Goldstein et al., 2001; Gong et al., 2011; Good et al., 2001) that were not controlled for here. However, these anatomical features should be static with respect to scan condition, and thus in this case we can interpret relative differences between scan conditions as true functional effects.)

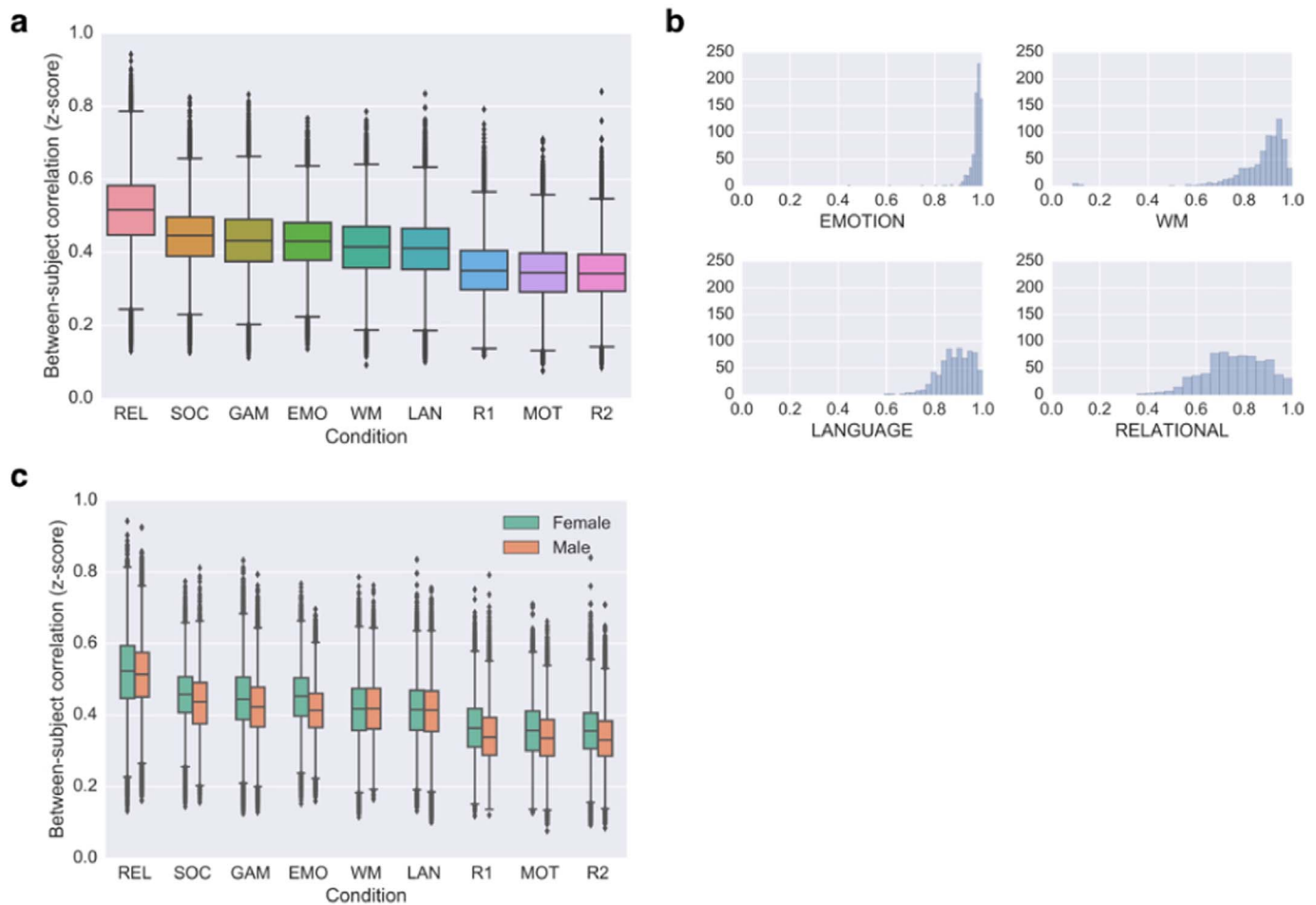


Fig. 3. Individuals look more similar during certain brain states than others. (a) Between-subject similarity as a function of scan condition. Values were derived from the upper triangle of a 716×716 subjects×subjects correlation matrix, representing the correlation (Fisher-transformed r value) between all possible pairs of subjects' whole-brain functional connectivity profiles in each respective condition. Whiskers denote 1.5×interquartile range. (b) Histograms showing the distribution of performance on the four tasks that measured accuracy on a continuous scale from 0 to 1. Horizontal axis is accuracy; vertical axis is subject count ($n=716$ for each condition). (c) Sex-condition interaction on between-subject similarity. Green denotes female-to-female correlations; orange denotes male-to-male correlations. See Table 1 for condition abbreviations.

In nearly all conditions, females were more similar to other females than males were to other males (Fig. 3c). While t -tests comparing within-female correlations to within-male correlations were statistically significant for all nine conditions (all $p < 10^{-7}$), effect sizes varied by condition: we observed the largest effect for the EMOTION task (Cohen's $d=0.49$), with intermediate effect sizes for REST1, REST2, GAMBLING, MOTOR and SOCIAL ($d=0.25$ – 0.29), and small effect sizes for LANGUAGE and RELATIONAL ($d=0.03$ and 0.05 , respectively). The only condition for which within-male correlations exceeded within-female correlations was WM; however, this effect was quite small ($d=-0.03$).

Previous work has reported sex differences in resting-state connectivity (Biswal et al., 2010; Kilpatrick et al., 2006; Satterthwaite et al., 2015; Scheinost et al., 2014), and in the location and magnitude of task-evoked activity (Domes et al., 2010; Goldstein et al., 2005; Gur et al., 2000; Stevens and Hamann, 2012). Interestingly, large sex differences in both brain activity and behavior are observed during emotion tasks (Whittle et al., 2011); our results indicate that this may be due at least in part to increased variability in males. These results suggest that investigators should take into account sex when choosing conditions to maximize individual variability: tasks that maximize variability in females may not do the same for males, and vice versa. This may be especially important when studying traits relevant to neuropsychiatric illnesses that disproportionately affect one sex or the other.

Within subjects, certain brain states increase variability more than others

Next, we investigated within-subject similarity as a function of brain state by correlating pairs of connectivity profiles from the same individual; this gives a 9×9 similarity matrix for each subject. The average matrix across individuals is shown in Fig. 4a. The lower triangle is based on connectivity calculated using only the first 176 volumes for each run, while the upper triangle was calculated using all available data for each condition (see Table 1 for scan durations).

Interestingly, given equal scan durations, two different tasks sometimes showed higher within-subject similarity than the two rest scans. For example, notice that mean similarity between the RELATIONAL condition and the EMOTION, GAMBLING and WM conditions all exceeded mean similarity between REST1 and REST2 (Fig. 4a, lower triangle). This underscores the idea that rest may not be the optimal condition for discovering biomarkers, since its unconstrained nature leads to more non-stationarities—which are essentially noise, at least until we better understand their dynamics—that reduce within-subject stability.

However, when the amount of data used to calculate connectivity matrices is increased to all available data per condition, the pattern of results changes noticeably: the mean correlation between REST1 and REST2 is extremely high ($r=0.97$), far exceeding any rest-task or task-task pair. Note that this result is based on approximately 30 minutes of

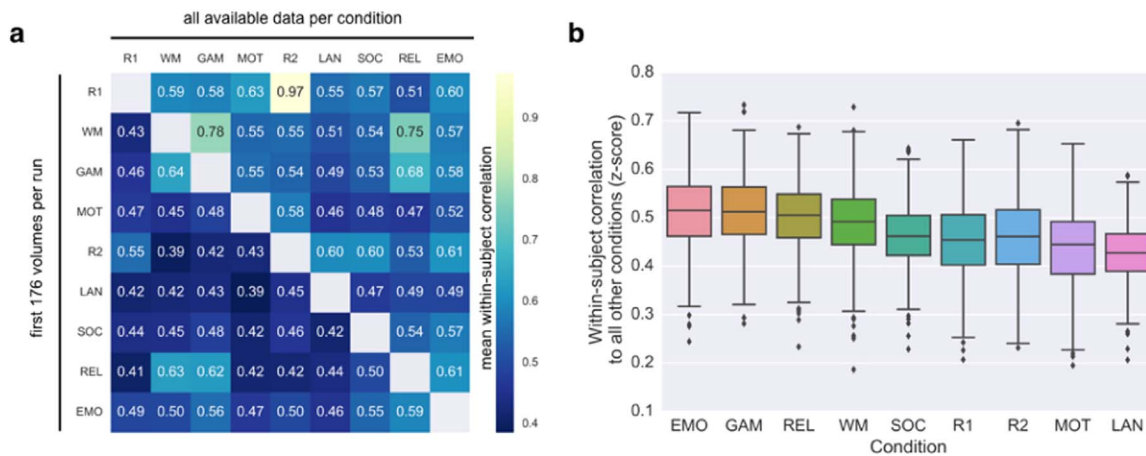


Fig. 4. Within-individual similarity varies between pairs of brain states. (a) Within-subject correlations between connectivity matrices acquired during each of 36 pairs of conditions. Each cell ij represents the mean within-subject correlation between the connectivity patterns in condition i and condition j . Lower triangle is based on connectivity matrices calculated using equal amounts of data per condition. Upper triangle is based on connectivity matrices calculated using all available data per condition (see Table 1 for scan durations). (b) Estimating how similar each state is to all other states within individuals. For each condition for each subject, correlations to all eight other conditions are averaged (thus each subject contributes one value to the boxplot per condition). See Table 1 for condition abbreviations.

data in each session (Fig. 4a, upper triangle), compared to $r=0.55$ using slightly more than four minutes of data (lower triangle).

These results recapitulate the importance of scan duration when estimating single-subject measures of functional connectivity, and also suggest a possible interaction between scan duration and condition on within-subject reliability. One possibility is that because rest is noisier, rest scans may take longer to converge on a subject's "true" profile than more constrained conditions. Shah et al. (2016) report that reproducibility of single connections for single subjects is roughly a linear function of the square root of imaging time, but this function may vary by condition (recall that scan duration and condition are confounded in the HCP data set if all available data is used); to conclusively address this question, we will need longer task acquisitions than those currently available in the HCP, or repeat scans of the same task. If it is indeed the case that rate of convergence varies between rest and various tasks, the question of which condition to use becomes a practical one: investigators could weigh subjects' tolerance for being in the scanner longer versus their ability to perform tasks while in the scanner to help decide which condition is optimal for a particular population.

We then summarized how similar a condition is on average to every other condition within individuals by averaging the respective row (minus the diagonal, which is always 1) from each subject's 9×9 similarity matrix and plotting the distribution of subject means for each condition (Fig. 4b). Given equal amounts of data per condition, EMOTION and GAMBLING were most similar to all other conditions (mean $r=0.51$); in other words, it seems these conditions evoke a connectivity pattern that is closest to the center of "connectome space" (cf. Fig. 2). On the other hand, the connectivity pattern associated with LANGUAGE is furthest from center (mean r with the other eight conditions $=0.43$). However, note that this result likely depends on the particular suite of tasks available in the HCP. It will be important to analyze connectivity patterns from additional tasks, to get a more complete picture of where different brain states fall within this space.

Subject identifiability varies with brain state

Thus far we have established that brain state influences both within- and between-subject variability in functional connectivity. However, recall that the optimal state for studying individual differences may not necessarily minimize or maximize either quantity on its own, but rather maximize the ratio of the two (within- to between-subject variability).

Identification experiments, in which a given connectivity profile acquired in one session is matched to the same individual's profile from a different session, can help estimate this ratio—in other words, they can help adjudicate between the possibilities depicted in Fig. 2. In the absence of test-retest sessions, identification experiments are a simple yet compelling way to estimate the ratio of within- to between-subject variability in a multi-condition data set. These experiments operate on pairs of conditions, but by pairing one condition with enough other conditions, we can get a sense of where it falls within connectome space (cf. Fig. 2); this might inform the question of which brain state is optimal for measuring individual differences.

To this end, we replicated and extended the identification experiments described in Finn et al. (2015). Briefly, for each pair of the nine scan conditions, we selected one to serve as the target and the other to serve as the database. In an iterative analysis, we compared each matrix in the target set to each of the database matrices, and assigned a predicted identity to the subject in the database whose matrix was maximally similar (as defined by Pearson correlation). After obtaining predicted identities for each matrix in the target set, the true identities were decoded and an overall accuracy was computed, expressed as number of correctly predicted identities over the total number of subjects. Finally, the roles of database and target session were reversed. There were 36 possible database-target configurations, representing various combinations of rest-rest, rest-task, and task-task pairs. Results are shown in Fig. 5a.

There are a few things to notice here. First, accuracy for all condition pairs is well above chance, which is approximately 0.001 ($1/716$). Average accuracy (0.54 ± 0.16) is lower than in (Finn et al., 2015), which is likely due to the combination of having more subjects and less data per subject (recall that to keep amount of data equal across conditions, matrices were calculated using only 4:12 min of data per condition). (Indeed, when we repeat the identification experiments using all available data for each condition, the average accuracy rises to 0.65 ± 0.14 , with 0.89 accuracy between the two REST sessions [data not shown]). Second, there is a subtle day effect, such that identification between sessions acquired on the same day tends to be higher than between those acquired on different days (compare off-diagonal rectangles to on-diagonal squares); this effect has been described previously in other data sets (Mueller et al., 2015). Finally, identification rates vary markedly between different pairs of conditions, with accuracy as high as 0.92 using GAMBLING as database and WM as target, and as low as 0.15 using REST1 as database and RELATIONAL

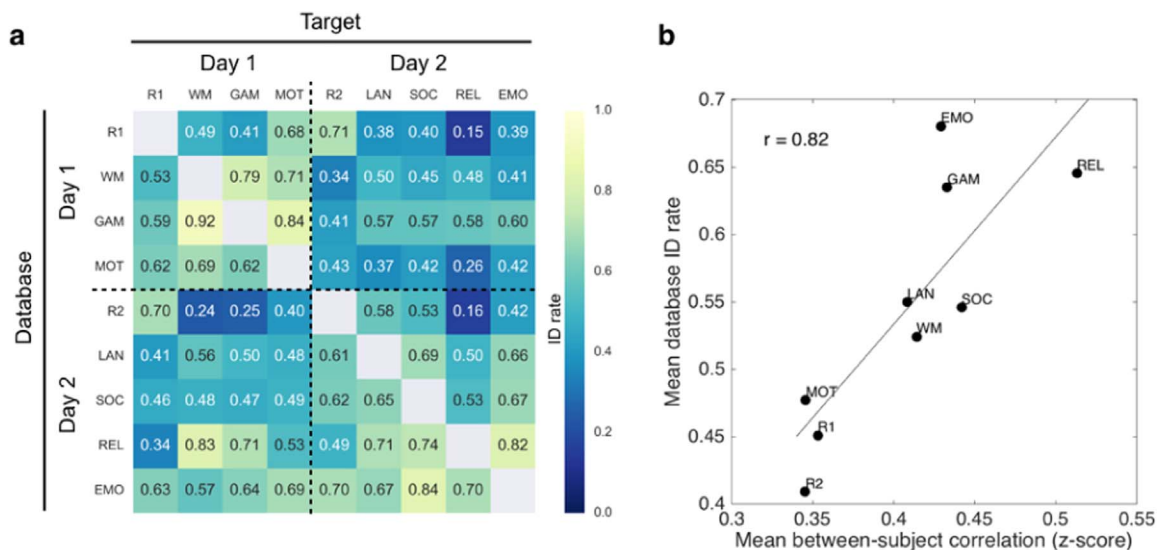


Fig. 5. Identification of individual subjects across brain states. (a) Extension of the identification experiments described in Finn et al. (2015) using $n=716$ subjects each scanned during nine conditions. Identification (ID) rate denotes the overall accuracy of an iterative algorithm that attempts to match a given subject's connectivity matrix from one condition (target) to the matrix of the same subject from a different condition (database). The roles of target and database may then be reversed, resulting in 72 possible identification experiments between rest-rest, rest-task and task-task scan pairs. Chance rate is approximately 0.001 (1/716). (b) Mean database ID rate (average across each row of the matrix in panel (a)) plotted against mean between-subject correlation (Fisher-transformed z-score; cf. Fig. 3a) for all nine conditions. Conditions that make subjects look more similar tend to make better databases for identification experiments ($r(7)=0.82$, $p=0.007$). See Table 1 for condition abbreviations.

as target. This indicates that individual differences do, in fact, change depending on the condition in which they are measured.

To further explore how identifiability changes with brain state, we averaged identification rates across the rows of the results matrix (Fig. 5a) to get an overall database score for each condition. We then plotted this score against mean between-subject similarity (cf. Fig. 3a). Interestingly, conditions that make subjects look more similar to one another also make better databases in identification experiments (Fig. 5b). Rest is the worst performing database condition, suggesting that the ways in which subjects differ at rest are not the same as the ways they differ during tasks (echoing results in (Geerligs et al., 2015)). We might speculate that in the absence of cognitive input and output, individual differences at rest default to “noisier” drivers such as arousal state, ongoing thought processes, and physiological artifacts. To reinforce the point illustrated in Fig. 1, brain states that increase between-subject similarity are not necessarily undesirable from an individual differences perspective, since any variation occurring above and beyond the common template may be more meaningful (or at least more interpretable).

Another way to interpret the high correlation depicted in Fig. 5b is that between-subject variability—the inverse of between-subject similarity—is actually a mixture of “true” (neural) inter-individual differences plus intra-individual differences that arise from both physiological (non-neural) and technical (MR-system related) sources during data acquisition (Liu, 2016). Thus, the high correlation between database score and between-subject similarity suggests that task conditions mainly serve to reduce intra-individual variability, while preserving meaningful inter-individual variance (i.e., variance related to true individual differences).

Outstanding questions

We have presented these observations as proof-of-principle that individual differences in functional connectivity do, in fact, depend on the condition in which they are measured. We hope these results provide a jumping-off point for more detailed investigations into how brain state affects both within- and between-subject variability, which

will help determine which conditions are optimal for individual differences research. In this final section, we outline several important directions for future study.

Anatomical specificity

Here we have reported within- and between-subject variability in functional connectivity across the whole brain, by considering the full connectivity matrix in our analyses. Previous work has found that between-subject variation in both structure and function is largest in high-order association regions of neocortex, including prefrontal, parietal and temporal lobes, and lowest in primary sensory cortices (Frost and Goebel, 2012; Laumann et al., 2015; Mueller et al., 2013). Also of note, within-subject test-retest reliability of connectivity profiles is not uniformly distributed across the brain; rather, the frontoparietal, posterior medial and occipital regions tend to be more reliable, while areas of high signal dropout, including the temporal lobes and ventromedial prefrontal cortex as well as subcortical regions, tend to be less reliable (Mueller et al., 2015; Noble et al., 2016). By restricting analyses to connectivity within circumscribed networks, especially those that are closely involved in task processing, and/or correcting for spatial differences in reliability (Mueller et al., 2015), we may find that individual differences are even further enhanced. Different brain states might have different effects: some states may induce a global increase in between-subject variability, with diffuse effects all over the connectivity matrix (cf. Fig. 1, upper left), while others may enhance signal-to-noise in networks of interest while leaving the rest of the brain relatively unchanged (cf. Fig. 1, upper right). Resolving the degree to which individual differences in functional connectivity are anatomically specific under various conditions will inform our understanding of how networks reconfigure across brain states, and may help increase sensitivity in the search for biomarkers in the functional connectivity signal.

Connectivity analysis pipeline

It is worth noting that several choices can influence measures of

functional connectivity. These include preprocessing steps, parcellation scheme, the similarity metric for BOLD timecourses (Pearson correlation versus alternatives such as partial correlation or independent components analysis) and for comparing connectivity across sessions (here we use Pearson correlation following (Geerligs et al., 2015), but intraclass correlation coefficient (Caceres et al., 2009) and distance-based metrics may also be used (Airan et al., 2016)).

One interesting question is whether the resolution of the atlas affects the detection of individual differences. In our previous work (Finn et al., 2015), we found that identification accuracy remained quite high (89%) with a coarser parcellation (the 68-node Freesurfer atlas) compared to the original result obtained with our 268-node parcellation. It is possible that increasing the resolution of the parcellation—to the order of thousands of nodes rather than hundreds, for example—would improve identification accuracy even further, and increase our sensitivity in the search for connectivity-based biomarkers. However, in its limit, we might be concerned that using smaller and smaller nodes would increase susceptibility to registration artifacts that would artificially inflate the individual signature, meaning any between-subject differences in functional connectivity might be due more to anatomical alignment error than true functional differences. On the other hand, it is also possible that moving to a higher-resolution parcellation would make the signal in each node more susceptible to motion or noise that would be different across sessions, and would thus decrease identification accuracy. The impact of these and other methodological choices cannot be exhaustively explored here but should be explored elsewhere.

Regardless of resolution, by using a predefined parcellation, we assume fixed node boundaries while allowing edge strengths to change across subjects and states. Investigating more “local” individual differences in functional alignment—i.e., the node boundaries themselves—yields valuable, complementary insights (Gordon, et al.; Laumann et al., 2015; Xu et al., 2016). New methods such as the HCP multimodal parcellation (Glasser et al., 2016) may allow for defining nodes in each subject individually without sacrificing macro-scale correspondence across subjects. Fortunately, however, any confounds related to individual differences in anatomy or intrinsic functional alignment—or, for that matter, lower-level features such as vasculature, hemodynamic response and physiology (Dubois and Adolphs, 2016)—are of minimal concern here since these features are static with respect to state.

Different populations

HCP subjects represent a tight demographic profile of healthy adults between the ages of 22 and 35, with many sets of twins and siblings. Thus, if anything, by using this data set we have likely underestimated between-subject variability in the general population. However, the fact that brain state substantially affected between-subject variability even in this relatively homogenous population underscores the importance of considering scan condition when designing experiments to probe individual differences. It will be important to see how findings from this particular set of subjects and states generalize to wider age ranges as well as other populations, including those with or at risk for neuropsychiatric illness (Koyama et al., 2016).

Behavioral prediction

The end goal of most neuroimaging research into individual differences is to relate brain measures to behavior. Ultimately, we should be striving to not simply report correlations between connectivity patterns and phenotypes, but to build predictive models that can take in neuroimaging information from a previously unseen subject and predict something about their present or future behavior, such as cognitive ability, risk for illness, or response to treatment, to name a

few (Bach et al., 2013; Gabrieli et al., 2015; Lo et al., 2015; Whelan and Garavan, 2014). With this in mind, the question is not which states maximize the ratio of within- to between-subject variability in and of itself, but rather which states maximize this ratio *while drawing out variation that is relevant to a phenotype of interest*.

To this end, investigators might try varying the input to connectivity-based predictive models (Shen et al., 2017) by testing several states and determining which yields the best behavioral prediction, using data sets such as HCP and others that contain multiple scan conditions and behavioral measures for each subject. While it is now well established that connectivity patterns are similar within individuals regardless of brain state, any effect of state that enhances individual variability in connections of interest may help pinpoint more precise relationships with behavior. For example, a model to predict attention may be more successful if it is trained on data acquired while subjects are performing a demanding attention task, even if it is ultimately applied to data acquired at rest (Rosenberg et al., 2015).

Of note, linking imaging measures to behavior crucially depends on choosing the right behavior. Advances in phenotyping will improve sensitivity in studies of brain-behavior relationships (Loeffler et al., 2015; Van Dam, et al.). To build a robust science of individual differences, we will need nuanced characterizations that move beyond coarse diagnostic categories to more closely “carve nature at its joints,” respecting the full spectrum of human variation in both healthy and clinical populations (Insel et al., 2010).

Conclusion

There is growing recognition in human neuroscience that mean-based approaches, in which data are averaged across many individuals, may obscure more than they reveal about brain-behavior relationships (Speelman and McGann, 2013). In fact, the perceived “universality” of functional brain regions and networks may be more an epiphenomenon of how we analyze our data than a reflection of how individual brains actually work. Traditionally, concerns about the signal-to-noise ratio of fMRI made it daunting to draw inferences from a single subject's scan. Furthermore, capturing enough phenotypic variation to robustly study individual differences often requires large sample sizes, which are challenging to acquire. But with recent advances in acquiring, analyzing and sharing fMRI data (Eickhoff et al., 2016; Poldrack et al., 2016), researchers are increasingly—and commendably—shifting focus from reporting group averages to discovering links between imaging measures and behavior in individual subjects (Dubois and Adolphs, 2016).

Resting-state studies presently account for the vast majority of research into individual differences in functional connectivity, and they have produced many valuable insights. However, mounting evidence, including the observations presented here, suggests that rest may not be the optimal state for studying individual differences. It is worth exploring alternative paradigms for individual differences research, as other brain states may afford an improved ratio of within- to between-subject variability, and/or enhance the individual connectivity signature in networks of interest. A better understanding of how brain state affects measurements of individual differences is important from a cognitive neuroscience perspective, and may increase our chances of finding imaging-based biomarkers with translational utility.

Acknowledgements

E.S.F. is supported by a US National Science Foundation Graduate Research Fellowship. This work was also supported by US National Institutes of Health grant EB009666 to R.T.C. Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

- Airani, R.D., et al., 2016. Factors affecting characterization and localization of interindividual differences in functional connectivity using MRI. *Human Brain Mapp.* 37, 1986–1997.
- Allen, J.S., et al., 2003. Sexual dimorphism and asymmetries in the gray–white composition of the human cerebrum. *Neuroimage* 18, 880–894.
- Bach, S., et al., 2013. Print-specific multimodal brain activation in kindergarten improves prediction of reading skills in second grade. *Neuroimage* 82, 605–615.
- Barch, D.M., et al., 2013. Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* 80, 169–189.
- Barnes, A., et al., 2009. Endogenous human brain dynamics recover slowly following cognitive effort. *PLoS One* 4, e6626.
- Bassett, D.S., et al., 2011. Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci. USA* 108, 7641–7646.
- Betzler, R.F., et al., 2016. Dynamic fluctuations coincide with periods of high and low modularity in resting-state functional brain networks. *Neuroimage* 127, 287–297.
- Birn, R.M., et al., 2013. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage* 83, 550–558.
- Biswal, B.B., et al., 2010. Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. USA* 107, 4734–4739.
- Braun, U., et al., 2012. Test–retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *Neuroimage* 59, 1404–1412.
- Breitenstein, C., et al., 2005. Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage* 25, 958–968.
- Byrge, L., et al., 2015. Idiosyncratic brain activation patterns are associated with poor social comprehension in autism. *J. Neurosci.* 35, 5837–5850.
- Caceres, A., et al., 2009. Measuring fMRI reliability with the intra-class correlation coefficient. *Neuroimage* 45, 758–768.
- Cantlon, J.F., Li, R., 2013. Neural activity during natural viewing of Sesame Street statistically predicts test scores in early childhood. *PLoS Biol.* 11, e1001462.
- Castellanos, F.X., et al., 2013. Clinical applications of the functional connectome. *Neuroimage* 80, 527–540.
- Chang, C., Glover, G.H., 2010. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 50, 81–98.
- Chang, C., et al., 2016. Tracking brain arousal fluctuations with fMRI. *Proc. Natl. Acad. Sci. USA*, (201520613).
- Christoff, K., et al., 2009. Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci. USA* 106, 8719–8724.
- Cole, M.W., et al., 2014. Intrinsic and task-evoked network architectures of the human brain. *Neuron* 83, 238–251.
- Cole, M.W., et al., 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* 16, 1348–1355.
- Domes, G., et al., 2010. The neural correlates of sex differences in emotional reactivity and emotion regulation. *Human Brain Mapp.* 31, 758–769.
- Dubois, J., 2016. Brain age: a state-of-mind? On the stability of functional connectivity across behavioral states. *J. Neurosci.* 36, 2325–2328.
- Dubois, J., Adolphs, R., 2016. Building a science of individual differences from fMRI. *Trends Cogn. Sci.* 20, 425–443.
- Duncan, N.W., Northoff, G., 2013. Overview of potential procedural and participant-related confounds for neuroimaging of the resting state. *J. Psychiatry Neurosci.: JPN* 38, 84–96.
- Eickhoff, S., et al., 2016. Sharing the wealth: neuroimaging data repositories. *Neuroimage* 124, 1065–1068.
- Finn, E.S., et al., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18, 1664–1671.
- Frost, M.A., Goebel, R., 2012. Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage* 59, 1369–1381.
- Gabrieli, John D.E., et al., 2015. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* 85, 11–26.
- Geerligs, L., et al., 2015. State and trait components of functional connectivity: individual differences vary with mental state. *J. Neurosci.* 35, 13949–13961.
- Glasser, M., et al., 2016. A Multi-modal parcellation of human cerebral cortex. *Nature*.
- Glasser, M.F., et al., 2013. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80, 105–124.
- Goldstein, J.M., et al., 2005. Sex differences in prefrontal cortical brain activity during fMRI of auditory verbal working memory. *Neuropsychology* 19, 509.
- Goldstein, J.M., et al., 2001. Normal sexual dimorphism of the adult human brain assessed by in vivo magnetic resonance imaging. *Cereb. Cortex* 11, 490–497.
- Gong, G., et al., 2011. Brain Connectivity: gender Makes a Difference. *Neuroscientist* 17, 575–591.
- Gonzalez-Castillo, J., et al., 2014. The spatial structure of resting state connectivity stability on the scale of minutes. *Front. Neurosci.* 8, 138.
- Gonzalez-Castillo, J., et al., 2015. Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proc. Natl. Acad. Sci. USA* 112, 8762–8767.
- Good, C.D., et al., 2001. Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *Neuroimage* 14, 685–700.
- Gordon, E.M., et al., Individual-specific features of brain systems identified with resting state functional correlations. *Neuroimage*.
- Gorgolewski, K.J., et al., 2014. A correspondence between individual differences in the brain's intrinsic functional architecture and the content and form of self-generated thoughts. *PLoS One* 9, e97176.
- Greicius, M., 2008. Resting-state functional connectivity in neuropsychiatric disorders. *Curr. Opin. Neurol.* 21, 424–430.
- Gur, R.C., et al., 2000. An fMRI study of sex differences in regional activation to a verbal and a spatial task. *Brain Lang.* 74, 157–170.
- Harrison, B.J., et al., 2008. Modulation of brain resting-state networks by sad mood induction. *PLoS One* 3, e1794.
- Hasson, U., et al., 2010. Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* 14, 40–48.
- Hasson, U., et al., 2004. Intersubject Synchronization of Cortical Activity during Natural Vision. *Science* 303, 1634–1640.
- Hearne, L.J., et al., 2016. Functional brain networks related to individual differences in human intelligence at rest. *Sci. Rep.*, 6.
- Helfinstein, S.M., et al., 2014. Predicting risky choices from brain activity patterns. *Proc. Natl. Acad. Sci. USA* 111, 2470–2475.
- Hutchison, R.M., et al., 2013. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* 80, 360–378.
- Insel, T., et al., 2010. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. psychiatry* 167, 748–751.
- Joshi, A., et al., 2011. Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics* 9, 69–84.
- Kelly, C., et al., 2012. Characterizing variation in the functional connectome: promise and pitfalls. *Trends Cogn. Sci.* 16, 181–188.
- Kilpatrick, L., et al., 2006. Sex-related differences in amygdala functional connectivity during resting conditions. *Neuroimage* 30, 452–461.
- Koyama, M.S., et al., 2016. Imaging the at-risk brain: future directions. *J. Int. Neuropsychol. Soc.* 22, 164–179.
- Laumann, T.O., et al., 2015. Functional system and areal organization of a highly sampled individual human brain. *Neuron*.
- Li, N., et al., 2013. Resting-state functional connectivity predicts impulsivity in economic decision-making. *J. Neurosci.* 33, 4886–4895.
- Liu, T.T., 2016. Noise contributions to the fMRI signal: an overview. *Neuroimage* 143, 141–151.
- Lo, A., et al., 2015. Why significant variables aren't automatically good predictors. *Proc. Natl. Acad. Sci. USA* 112, 13892–13897.
- Loeffler, M., et al., 2015. The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health* 15, 1.
- Morcom, A.M., Fletcher, P.C., 2007. Does the brain have a baseline? Why we should be resisting a rest. *Neuroimage* 37, 1073–1082.
- Mueller, S., et al., 2015. Reliability correction for functional connectivity: theory and implementation. *Human Brain Mapp.* 36, 4664–4680.
- Mueller, S., et al., 2013. Individual variability in functional connectivity architecture of the human brain. *Neuron* 77, 586–595.
- Noble, S., et al., 2016. Multisite reliability of MR-based functional connectivity. *Neuroimage*.
- Patriat, R., et al., 2013. The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated. *Neuroimage* 78, 463–473.
- Poldrack, R., et al., 2016. Scanning the Horizon: Challenges and Solutions for Neuroimaging Research. *bioRxiv*.
- Power, J.D., et al., 2014. Studying brain organization via spontaneous fMRI signal. *Neuron* 84, 681–696.
- Power, J.D., et al., 2015. Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage* 105, 536–551.
- Pushkarskaya, H., et al., 2015. Neural correlates of decision-making under ambiguity and conflict. *Front. Behav. Neurosci.* 9, 325.
- Richiardi, J., et al., 2011. Decoding brain states from fMRI connectivity graphs. *Neuroimage* 56, 616–626.
- Rosenberg, M.D., et al., 2015. A neuromarker of sustained attention from whole-brain functional connectivity. *Nat. Neurosci. Adv. online Publ.*
- Salmi, J., et al., 2013. The brains of high functioning autistic individuals do not synchronize with those of others. *NeuroImage: Clin.* 3, 489–497.
- Satterthwaite, T.D., et al., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *Neuroimage* 60, 623–632.
- Satterthwaite, T.D., et al., 2015. Linked sex differences in cognition and functional connectivity in youth. *Cereb. Cortex* 25, 2383–2394.
- Scheinost, D., et al., 2014. Sex differences in normal age trajectories of functional brain networks. *Human Brain Mapp.*
- Shah, L.M., et al., 2016. Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain Behav.* 6, (e00456-n/a).
- Shehzad, Z., et al., 2009. The resting brain: unconstrained yet reliable. *Cereb. Cortex* 19, 2209–2229.
- Shen, X., et al., 2013. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage* 82, 403–415.
- Shen, X., et al., 2017. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat. Protoc.* 12, 506–518.
- Shine, J.M., et al., 2016. Temporal metastates are associated with differential patterns of time-resolved connectivity, network topology, and attention. *Proc. Natl. Acad. Sci. USA* 113, 9888–9891.
- Shirer, W., et al., 2012. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb. Cortex* 22, 158–165.
- Simony, E., et al., 2016. Dynamic reconfiguration of the default mode network during

- narrative comprehension. *Nat. Commun.* 7, 12141.
- Smith, S.M., et al., 2013. Resting-state fMRI in the Human Connectome Project. *Neuroimage* 80, 144–168.
- Smith, S.M., et al., 2009. Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. USA* 106, 13040–13045.
- Smith, S.M., et al., 2015. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* 18, 1565–1567.
- Speelman, C., McGann, M., 2013. How mean is the mean? *Front. Psychol.*, 4.
- Spiers, H.J., Maguire, E.A., 2006. Spontaneous mentalizing during an interactive real world task: an fMRI study. *Neuropsychologia* 44, 1674–1682.
- Stevens, J.S., Hamann, S., 2012. Sex differences in brain activation to emotional stimuli: a meta-analysis of neuroimaging studies. *Neuropsychologia* 50, 1578–1593.
- Tagliazucchi, E., Laufs, H., 2014. Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* 82, 695–708.
- Tung, K.-C., et al., 2013. Alterations in resting functional connectivity due to recent motor task. *Neuroimage* 78, 316–324.
- Uğurbil, K., et al., 2013. Pushing spatial and temporal resolution for functional and diffusion MRI in the HUMAN Connectome Project. *Neuroimage* 80, 80–104.
- Van Dam, N.T., et al., Data-driven phenotypic categorization for neurobiological analyses: beyond DSM-5 Labels. *Biol. Psychiatry*.
- van den Heuvel, M.P., Hulshoff Pol, H.E., 2010. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* 20, 519–534.
- van den Heuvel, M.P., et al., 2009. Efficiency of functional brain networks and intellectual performance. *J. Neurosci.* 29, 7619–7624.
- Van Essen, D.C., et al., 2013. The WU-Minn Human Connectome Project: an overview. *Neuroimage* 80, 62–79.
- Van Horn, J.D., et al., 2008. Individual variability in brain activity: a nuisance or an opportunity? *Brain imaging Behav.* 2, 327–334.
- Vanderwal, T., et al., 2016. Individual differences in functional connectivity during naturalistic viewing conditions. *bioRxiv*.
- Vanderwal, T., et al., 2015. Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. *Neuroimage* 122, 222–232.
- Whelan, R., Garavan, H., 2014. When optimism hurts: inflated predictions in psychiatric neuroimaging. *Biol. Psychiatry* 75, 746–748.
- Whittle, S., et al., 2011. Sex differences in the neural correlates of emotion: evidence from neuroimaging. *Biol. Psychol.* 87, 319–333.
- Xu, T., et al., 2016. Assessing Variations in Areal Organization for the Intrinsic Brain: from Fingerprints to Reliability. *Cereb. Cortex*.
- Zuo, X.-N., Xing, X.-X., 2014. Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci. Biobehav. Rev.* 45, 100–118.