

Feature Review

Building a Science of Individual Differences from fMRI

Julien Dubois^{1,*} and Ralph Adolphs¹

To date, fMRI research has been concerned primarily with evincing generic principles of brain function through averaging data from multiple subjects. Given rapid developments in both hardware and analysis tools, the field is now poised to study fMRI-derived measures in individual subjects, and to relate these to psychological traits or genetic variations. We discuss issues of validity, reliability and statistical assessment that arise when the focus shifts to individual subjects and that are applicable also to other imaging modalities. We emphasize that individual assessment of neural function with fMRI presents specific challenges and necessitates careful consideration of anatomical and vascular between-subject variability as well as sources of within-subject variability.

From the Group to the Individual

Brain imaging with blood oxygen level-dependent functional magnetic resonance imaging (BOLD fMRI) has been used extensively since the early 1990s to understand generic aspects of brain function, typically by averaging data across individuals to improve the signal-to-noise ratio (SNR). The statistical benefits of averaging across subjects have also been leveraged in group comparisons, for example in studies of clinical populations. However, these studies have historically fallen short of a proper characterization of brain function at the level of the individual. While the importance of a fully personalized investigation of brain function has been recognized for several years [1,2], only recent technological advances now make it possible. For example, there are advances already at the acquisition level, such as higher field strength and faster acquisition, which have led to substantial SNR improvements [3]. Attempts at interpreting individual subject fMRI measurements have become a major focus in the past 5 years or so, partly driven by the rise of 'resting-state' fMRI (Box 1). There is interest in examining individual differences in relation to healthy aging [4,5], personality [6], intelligence [7,8], mood [9] and genetic polymorphism [10]. On the clinical side, there are considerable efforts to use fMRI to classify individual subjects as patient or control ([11,12]; reviewed in [13,14]), to select treatment [15], or predict future outcome ([16]; reviewed in [17]).

Several issues arise when the focus shifts from group averaging to the comparison of the statistics of individual subjects. The issues can be framed in terms of key concepts from behavioral research on individual differences, namely validity and reliability. Validity asks whether the individual differences we measure with BOLD fMRI really reflect what we intend to measure. One specific concern for validity is whether we are indeed comparing functionally homologous regions across subjects. Another is whether we are indeed comparing neural function because BOLD fMRI only provides an indirect measure of neural activity. Reliability, by contrast, asks whether a finding is stable in the face of variations that should not matter. Reliability is notably hindered by relatively well-understood noise sources such as motion and subject physiology, as well as by less well-understood ones such as neuro- or vasoactive substances that we might not

Trends

Interpretation of fMRI data at the level of individual brains is essential for characterizing brain function in health and disease.

Two core challenges are validity (do we measure what we intend to measure?) and reliability (is our measure stable in the face of variations that should not matter?) of fMRI-derived individual differences; these challenges can be partly addressed with recent tools.

Interpretation of single-subject fMRI measures relies on establishing a relationship with an independent measure in the same subjects. Out-of-sample prediction should be used over correlation analysis.

Accumulation of large samples through consortia and data sharing, as well as careful attention to statistical power issues, are crucial for reproducible research.

Whole-brain characterization in naturalistic conditions, such as while watching a movie or listening to a story, may provide an alternative to resting-state data that permits a rich link to sensory and semantic stimulus variables.

¹Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA

*Correspondence: jdubois@gmail.com (J. Dubois).

Box 1. Resting-state fMRI: The Workhorse of Individual Differences Research

Resting-state fMRI, or RS-fMRI, entails imaging subjects while they lie in the scanner doing nothing and trying not to think of anything in particular [156]. Spontaneous fluctuations in activity show reproducible correlations across brain regions; regions with correlated spontaneous activity also tend to be co-activated in task fMRI, thus establishing the relevance of spontaneous correlations (functional connectivity) to brain function [157,158]. RS-fMRI has now established itself as a leading approach to the study of brain organization [156].

A major reason for the widespread adoption of the RS-fMRI approach is its minimal requirements. Most subjects can lie quietly in the scanner for 5 or more minutes [159] (for more advanced analyses, up to 100 minutes of data may be required for best results [160]). Of course, careful investigations have pointed out 'details' that influence the data collected during the resting-state, such as whether the subject's eyes are open or closed [161], whether they are completely awake [124], what they are actually thinking about during the run [162–164], or what task they performed immediately preceding the run [165]. Changes in the strength and directionality of functional connections have been described between runs in the same session, but also at much faster timescales (seconds to minutes) during a run (reviewed in [166]). Functional connectivity can also be estimated from task fMRI data, usually following the removal of stimulus-evoked activity [72], and shares about half its variance with the functional connectivity estimated from RS-fMRI data [167]. The remaining half of unshared variance could complicate the interpretation of individual differences in functional connectivity [94].

It remains the case that RS-fMRI provides the easiest functional data to collect and aggregate, across subject populations and sites [145], as is now done in several large efforts, such as those of the Human Connectome Project [149,168]. Individual differences research requires large sample sizes (see also main text) and thus RS-fMRI, despite imperfection on several fronts, is likely to remain the workhorse of individual differences fMRI research for the years to come (for a glimpse of the future, see Box 4).

properly take into account. Both validity and reliability can be enhanced through the use of readily available tools; we review the latest advances with respect to these issues and provide some specific recommendations for how the field can best advance (Figure 1, Key Figure).

In addition, we discuss important statistical considerations on the path to a science of individual differences from fMRI. Though correlation analysis (between individual fMRI-derived statistics and other measures of the same individuals) is overwhelmingly used in the literature, it is subject to overfitting and findings do not always generalize to other samples; instead, the out-of-sample predictive value of an fMRI-derived statistic with respect to another individual difference measure must be established. The typical sample size for fMRI studies ($n = 10\text{--}50$) also needs to be scaled up (to $n > 100$) for individual differences research; larger sample sizes not only increase statistical power in general but also allow more complex models to be fit.

Validity: Are Individual Differences Attributable to Brain Function?

A Common Space for Mapping Function

How can we tell whether individual differences in metrics such as BOLD activations or functional connectivity are actually related to differences in the underlying neural activity or communication, respectively? One ubiquitous problem arises in matching different brains such that functionally meaningful comparisons across subjects are possible in the first place. In a typical fMRI analysis pipeline, both structural and functional data from individual brains are spatially warped to a common anatomical space. The most widely used common space [18] is the MNI152 atlas, to which subjects' brains are anatomically warped via a volumetric transform ([19] for review). Such registration is appropriate for subcortical structures which are inherently volumetric; by contrast, the cortex is a 2D structure and volumetric alignment does not properly align folding patterns across subjects. Although switching to cortical folding-based inter-subject alignment [20,21] has been shown to somewhat reduce functional mismatch [22,23] (but see [24]), this has not yet become common practice. There are several reasons for this, from the burden of generating accurate cortical surfaces for each brain to the additional complexities of working with surface data (which until recently were not handled well by leading fMRI analysis software). Recent improvements to Freesurfer's automated cortical surface reconstruction pipeline [25], together with the release of a new file format that combines surface and volumetric data (Connectivity Informatics Technology Initiative, CIFTI) and software for visualization and analysis

Glossary

Cluster: a contiguous set of voxels or vertices whose value in a statistical parametric map exceeds the cluster-forming threshold.

Echo planar imaging: any rapid gradient-echo or spin-echo sequence in which k-space is traversed in one (single-shot) or a small number of excitations (multi-shot). Gradient-echo EPI is the workhorse of fMRI.

General linear model: a generalization of the multiple linear regression model to the case of more than one dependent variable. The GLM attempts to explain the BOLD response at each brain location given known experimental manipulations.

Global signal: the average BOLD signal across the whole brain.

Independent component analysis: a computational method for separating a multivariate signal into additive subcomponents. This is done by assuming that the subcomponents are non-Gaussian signals and that they are statistically independent from each other.

Myelin: a mixture of proteins and phospholipids forming an insulating sheath around many nerve fibers, which increases the speed at which impulses are conducted. Myelin content of the cortex can be mapped using a combination of T1-weighted and T2-weighted MRI volumes [32].

Overfitting: occurs when a statistical model describes random error or noise instead of the underlying relationship.

Pre-registration: Authors submit plans for data collection and analysis to a journal for peer review prior to conducting the experiment. If the proposal passes peer-review, and the authors follow their proposed experimental plan, the results are then published irrespective of the statistical outcome (null or significant).

Preprocessing: operations performed on raw fMRI data before statistical analysis.

Realignment: one of the operations performed during preprocessing of fMRI data. It consists of spatially aligning all collected volumes to a reference volume chosen from the same experimental run.

Region of interest: a part of the brain that is singled out for further analysis, on the basis of anatomical or functional information.

Representational geometry: Each experimental condition or stimulus is

(the Connectome Workbench), could foster wider adoption of surface analysis. Nevertheless, there is still no guarantee that functionally similar **vertices** (see [Glossary](#)) correspond spatially across subjects after such alignment: individual differences may thus invalidly arise from incorrect alignment of function across subjects.

A Multimodal Common Cortical Topography

Watching a movie with a rich variety of visual, auditory, and social percepts elicits time-locked brain activity that is correlated across subjects in many brain regions (inter-subject correlation, ISC) [26]. After cortical folding-based inter-subject alignment, the cortical mesh can be further warped to maximize inter-subject correlations during movie viewing [27]. In the absence of brain activity that is time-locked to a shared external stimulus, it is also possible to instead improve the vertex-wise match of functional connectivity patterns across subjects [28] using resting-state data (see also [29,30]). A multimodal surface matching (MSM) framework [31] was recently introduced that not only performs similarly to, and faster than, state-of-the-art algorithms based solely on cortical folding [21], but that can also flexibly accept other types of data (e.g., fMRI) to further inform registration (Figure 2A). The framework can, for example, accept any combination of geometric (shape-based), **myelin** [32], task activation, **retinotopy** [33], and functional and structural connectivity features. In theory, the combination of all this information in the MSM framework should yield the best possible inter-subject alignment, with the constraint that neighboring vertices must remain neighbors. However, the optimal set of modalities to include, the associated cost functions, and the relative weights assigned to each modality remain under investigation, and further validation of an improved alignment of function will be required before widespread use can be prescribed [31].

From a Common Cortical Topography to a Common Representational Space

Another recently introduced technique, ‘hyperalignment’, goes one step further, ignoring topological constraints altogether and using only **representational geometry** across subjects – in other words matching multivariate spatial patterns of activity (Figure 2B) [34]. As before [27], the fMRI data used in this study [34] were obtained in response to a full-length movie. The selection of voxels/vertices that are fed into the hyperalignment algorithm is the sole spatial constraint; data from each subject are eventually projected into a common space that can violate topology. The original implementation of the technique relied on pre-selecting a **region of interest** (ROI) [34]; a more recent whole-brain version of hyperalignment uses a searchlight centered on each vertex, then combines the resulting transformation matrices into a large, whole-brain transformation matrix that can be used to project the cortical data from each individual subject onto a common representational space [35]. Ongoing work is extending the hyperalignment algorithm to make use of resting-state data, instead of movie data, in an effort to maximize applicability to extant datasets. In theory, whole-brain hyperalignment provides the most thorough correction for anatomical variability. Because hyperalignment is based on an invertible transformation of the data, individual differences are not lost at any stage of the process. Nevertheless, the validity of individual differences measured in the common space is not assured because individual differences of interest may also be mixed into the transformation matrix for each subject. We would encourage researchers to explore hyperalignment for their datasets, and to report results obtained with and without hyperalignment, such that we can accumulate further evidence for its most appropriate application. In general, we would recommend that investigators try more than one approach to alignment, and report all of them, so we can see which might work best for which kinds of questions.

ROI Analysis

Many of the problems with alignment can be largely circumvented if a functional ROI is isolated in each individual subject. Functional localizers have traditionally been used to improve validity [36]: in an independent MRI run, a task designed to activate a ROI is performed, and the ROI is

represented as a brain-activity pattern. The dissimilarity of two patterns corresponds to the distance between their points in the representational space. The geometrical arrangement of multiple points in representational space is a characteristic of the representation.

Retinotopy: mapping of visual input from the retina to neurons in the visual cortex. Techniques exist for non-invasive retinotopic mapping with fMRI.

Run: in the context of a fMRI experiment, a run designates an uninterrupted collection of fMRI data, resulting in a sequence of successive volumes.

Statistical parametric map: the result of a univariate (voxel-by-voxel, or vertex-by-vertex) statistical analysis of fMRI data, using any standard statistical test; it consists of the values of the statistic at each voxel or vertex.

Vertex: the cortical surface is represented as a tessellation of triangles. The vertices of the surface mesh are the vertices of its triangular tiles.

Key Figure

Proposed Generic Analytical Pipeline for Individual Differences Research in fMRI

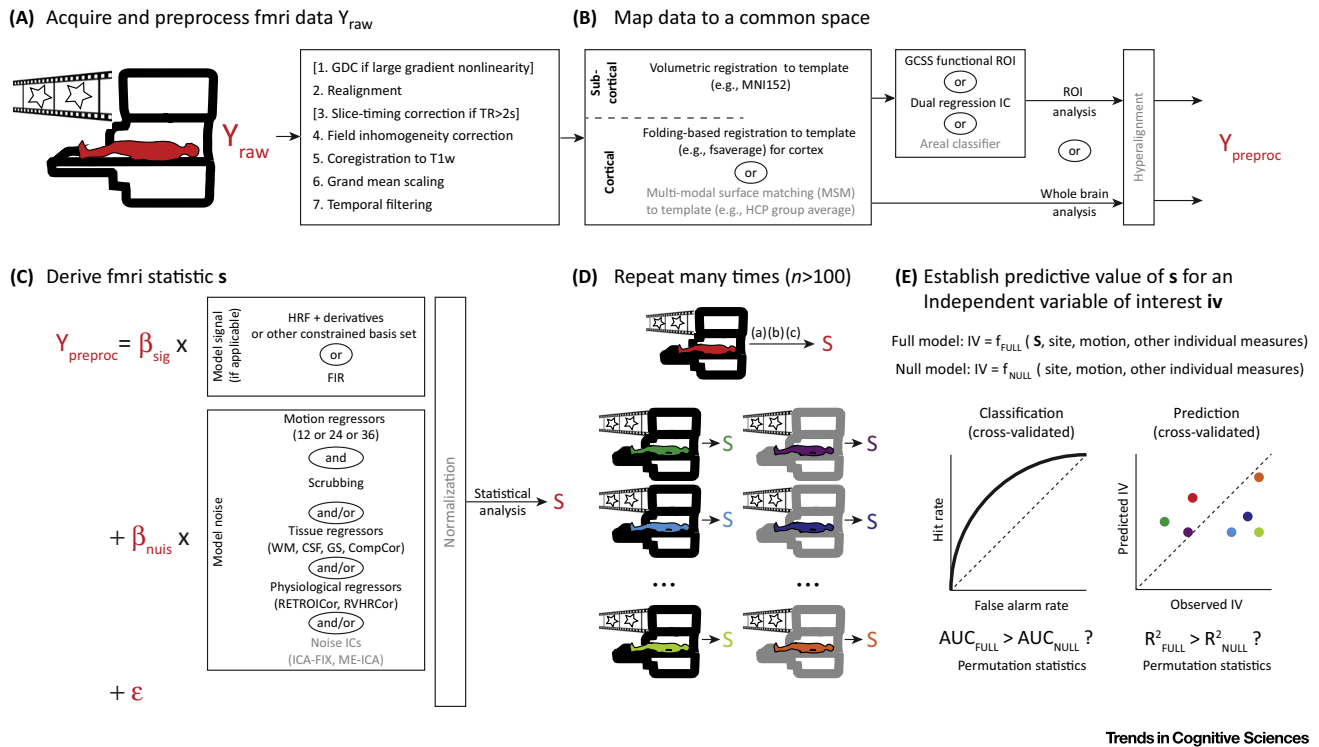
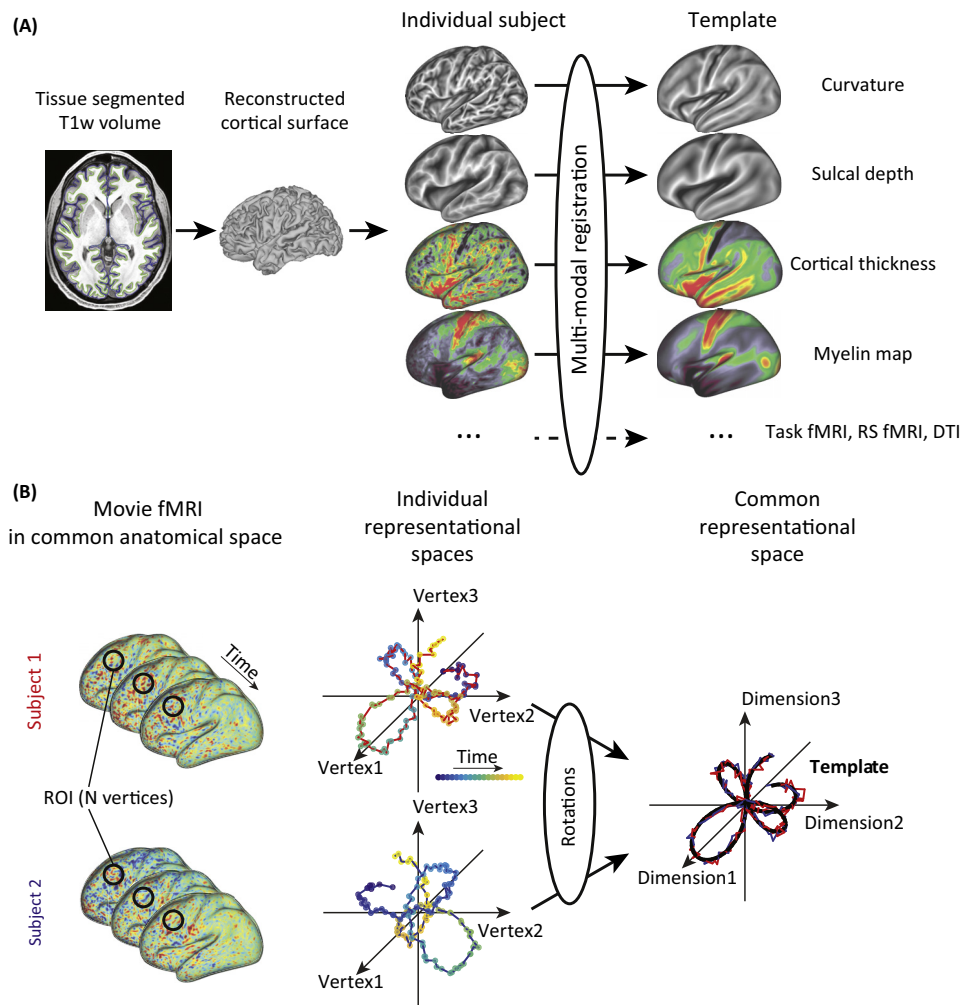


Figure 1. This pipeline may not fit all situations, given the multidimensionality and multi-component nature of fMRI data. (A) Data Y_{raw} are acquired while a subject undertakes a particular task (movie with stars). Minimal preprocessing is applied (e.g., HCP pipelines [25] shown in the box to the right). (B) Subcortical areas are volumetrically warped to match the MNI template, while the cortex is registered to a surface template using cortical folding patterns or a multi-modal surface matching approach. A choice is then made between conducting a region-of-interest or whole-brain analysis. Several options for group-informed definitions of functional parcels in individual subjects are available. Finally, hyperalignment can be performed to project data into a common representational space. (C) The preprocessed data $Y_{preproc}$ are modeled as a weighted linear combination of known signal (except for RS-fMRI) and confounds of no interest, plus unmodeled noise. All parameter estimates are normalized for vascular differences before further analysis. A statistic S (possibly multidimensional) is derived for this subject. (D) More than 100 subjects participate in the experiment, possibly with slight variations (e.g., at two different sites, black and gray scanners). (E) The fMRI-derived individual statistic is used together with confound variables to predict an individual measure of interest IV , for example the subject's IQ (prediction) or patient/control (classification). The same analysis is conducted without the fMRI-derived statistic. Performance for the full model is compared to performance for the null model, using a permutation test, to establish the unique predictive power of the fMRI-derived statistic. Text in gray: methods awaiting further validation.

defined in each individual from a **statistical parametric map** (SPM) as one of the **clusters** that exceeds a given threshold. However, the choice of a statistical threshold in single subjects is often a thorny issue; recent approaches can avoid arbitrary thresholds by using mixture models that fit the noise in each subject's data [37] or by using multiple spatial scales to define clusters of activation (threshold-free cluster enhancement, TFCE) [38]. Another issue is of ensuring that the individually-defined ROIs are indeed well-matched across subjects, a problem for which approaches such as the group-constrained subject-specific (GCSS) algorithm [39] provide elegant solutions. Functional localizers have a long and successful track-record and can provide functional ROIs in individual subjects with only a few minutes of scanning, for processes ranging from face perception [40] to theory of mind [41,42]. However, they quickly become inefficient if one wants to map many different processes in a study [43]. At present we would encourage investigators to continue using well-vetted localizers when studying specific psychological



Trends in Cognitive Sciences

Figure 2. Multimodal Surface Matching and Hyperalignment. (A) The multimodal surface matching framework [31] accepts multiple sources of information to align subjects: not only sulcal patterns but also cortical thickness, myelin maps, and RS-fMRI-derived functional connectivity, etc. (B) After selecting a region of interest in a common anatomical space, hyperalignment [34] aligns representational spaces across subjects; it is usually based on movie data.

processes; in other contexts, emerging methods for single-subject functional parcellation from resting-state data and other modalities (Box 2) should be considered. Pending further validation against known function, these methods may replace functional localizers altogether in the not-too-distant future.

The Ever-Lurking Plumbing Issue

Differences in vasculature cause differences in the hemodynamic response. A review of all the potential effects of vasculature differences is outside the scope of this article (see [44–46]); we focus here instead on existing solutions at the level of acquisition and data analysis that are rarely implemented but may prove crucial for valid individual differences research.

Capturing Response Shape Variability

The hemodynamic response function (HRF) is a model of the BOLD activity triggered by a neural event of infinitesimal duration. Most fMRI analyses assume a fixed-shape HRF (the canonical HRF).

Box 2. Functional Parcellation of the Brains of Individual Subjects

It is now well established that resting-state functional connectivity (RSFC) can extract networks in the brain that subserve shared psychological functions [145]. Accordingly, there has been interest in using RSFC to define functional parcels or ROIs in the brains of single subjects. There are two schools of thought regarding functional parcellation: those which allow the parcels to overlap spatially, and those that prefer to tile the brain with non-overlapping parcels. The first school typically uses independent component analysis (ICA) to decompose the RS-fMRI signal into statistically independent non-Gaussian sources (spatial maps) via a linear and instantaneous mixing process corrupted by additive Gaussian noise components [169]. The second school typically uses some variation of a clustering or region growing algorithm that results in parcels that have homogeneous patterns of whole-brain connectivity [148,170–173], or an algorithm that detects transitions in whole-brain RSFC [174,175].

There are two approaches to establishing a one-to-one correspondence between the parcels in different subjects that enhance validity beyond whole-brain alignment techniques. One is to perform RSFC-based parcellation in individual subjects independently with no prior information [148,171], then use some clustering or matching algorithm to establish one-to-one correspondence between parcels; this can prove a rather difficult problem, with for example splitting of parcels. The other, preferred approach is to start from a set of group-level parcels and discover their instantiation in individual subjects. The initial group parcellation is of course critical, and may come from analyzing averaged or concatenated individual data, or from a previously published result. Projecting it into individual brains is typically done using dual regression in the case of IC components [176], and different approaches have been proposed for other parcellation schemes – for example a template-matching procedure which assigns each voxel or vertex from an individual subject to a parcel (e.g., [177,201]), or an iterative assignment procedure as in [178].

Including information from multiple modalities has proved beneficial to refine inter-subject whole-brain alignment, as demonstrated in the MSM framework [31]. Using multiple modalities can also be instrumental in parcellating an individual brain, as demonstrated recently: after defining a multimodal group parcellation from the HCP data (including architecture, such as thickness and myelin content, task fMRI, RS-fMRI connectivity, and RS-fMRI derived topographic maps), a machine-learning approach can be used to find the instantiation of each group-defined parcel in single subjects, correctly accounting for variations in anatomy [200].

However, the shape of the HRF is well known to vary across subjects and brain regions [44,47]. Factors such as vasodilatory signaling, blood vessel stiffness, neurovascular coupling delay, venous transit time, and the time constant of autoregulatory feedback contribute to this variability [48,49]; these factors are particularly affected by aging and disease [50]. Inferences about individual differences in BOLD response magnitude are not valid when the true shape of the HRF differs across subjects. To account for these variations, the use of multiple, orthogonal basis functions to more accurately model the HRF has been proposed (reviewed in [51]). Popular constrained basis sets include the canonical HRF plus its temporal and dispersion derivatives [52], or a basis function set based on singular value decomposition [53]. Releasing all constraints, the finite impulse response (FIR) basis set has one free parameter for every time-point following stimulation for every event type modeled [54]. A caveat of increasing the number of free parameters is the concurrent increase in variance of the HRF produced, as well as the cost to compute it.

When the shape of the HRF is fixed, the amplitude of the BOLD response is easily compared across subjects on the basis of a single parameter estimate. Modeling HRF shape faithfully complicates the comparison of response amplitude across subjects: the BOLD response is now represented across several parameters, which must be jointly compared. Some solutions have been proposed for this joint comparison, but none is widely accepted yet [51,55]. Recent work combining estimation of the HRF shape with detection of activation in the same optimization seems a step in the right direction [56,57], although these approaches currently assume that HRF shape is similar across subjects in a given brain region, which may not be a valid assumption when investigating individual differences.

HRF shape variability is problematic even for resting-state fMRI because differences in response shape across subjects can result in differences in functional connectivity estimates. With current analytical practices of low-pass filtering the data (considering only fluctuations slower than 0.1 Hz), the effects of small differences in response shape should be small [45]. However, as investigators start looking at higher-frequency fluctuations [58], HRF shape variability may limit

validity in interpreting individual differences. Methods are emerging to deconvolve the spontaneous signal [59,60] and thus possibly account for these variations. In our view, further development of how best to quantify and incorporate HRF variability should be of the highest priority for individual differences research because all subsequent measures depend on getting this right in the first place.

Accounting for Vascular Differences across Subjects: Calibration and Normalization

The BOLD signal is a function of changes in cerebral blood flow (CBF), cerebral blood volume, and the cerebral metabolic rate of oxygen (CMRO₂) [61]; it also depends on the baseline physiological state (hematocrit, oxygen extraction fraction, blood pressure). Both hemodynamic coupling and baseline physiology differ across subjects [45,62]. Two main approaches have been put forward to try to control for differences in these parameters across subjects: calibration and normalization.

The calibrated fMRI technique has been around for many years, and can in theory control for differences in both baseline physiology and hemodynamic coupling across subjects. The technique has not seen widespread adoption, chiefly because it is difficult to implement (e.g., requiring concurrent measurement of BOLD and CBF and inhalation of CO₂ [63,64]) and is still imperfect [65].

Several approaches have been proposed to normalize for differences in baseline physiology across subjects. Many rely on a companion scan. For example, the BOLD response to hypercapnia, induced through administration of CO₂ [66] or by using a breath-hold challenge [67], can be used as a normalization factor (Figure 3A). Alternatively, whole-brain venous oxygenation levels can be measured with a special pulse sequence and used to normalize the BOLD response [68]. A more easily applicable option is to use the amplitude of low-frequency fluctuations in resting-state fMRI data (RS-ALFF) [69,70] as a normalization factor; indeed RS-ALFF reflects naturally-occurring variations in cardiac rhythm and in respiratory rate and depth [71], and approximates the BOLD response to a hypercapnic challenge (Figure 3A). In fact, one does not even need to acquire a separate resting-state scan. In the same way that functional connectivity can be derived from the residuals of a **general linear model** (GLM) for task-based fMRI data [72], the amplitude of low-frequency fluctuations in the residuals of task-based fMRI data (GLMres-ALFF) can also be used to rescale the BOLD signal change; this ‘vascular autorescaling’ (VasA) technique was even shown to outperform RS-ALFF-based normalization [73] (Figure 3B). Using data from the same fMRI run, as VasA does, is also desirable because it avoids contamination of the data with noise from a separate run.

While all these methods were developed with the aim of improving group statistics by reducing inter-subject vascular variance, they should be considered for valid assessment of neural individual differences (Figure 3C) [74]. As with any manipulation of the data, there are potential caveats: calibration and normalization methods may remove individual differences of neural origin, for instance if baseline blood oxygenation were linked to neural activity [62]. Before such techniques are routinely applied, more examples of their successful application will be necessary, ideally vetted by an independent measure of neural function [74].

Reliability: Individual Differences or Unmodeled Noise?

Once validity is maximized (inasmuch as current technology allows), it is also crucial to ensure that individual differences measured with fMRI are not merely attributable to unaccounted-for noise in the measurements. The reliability of fMRI has been inspected closely in recent years [75–77]. Of course there is no such thing as ‘the’ reliability of fMRI because different derived statistics are differentially affected by noise in the raw data and thus have different reliabilities [78] (for an overview of studies addressing the reliability of specific fMRI-derived measures see Table S1 in

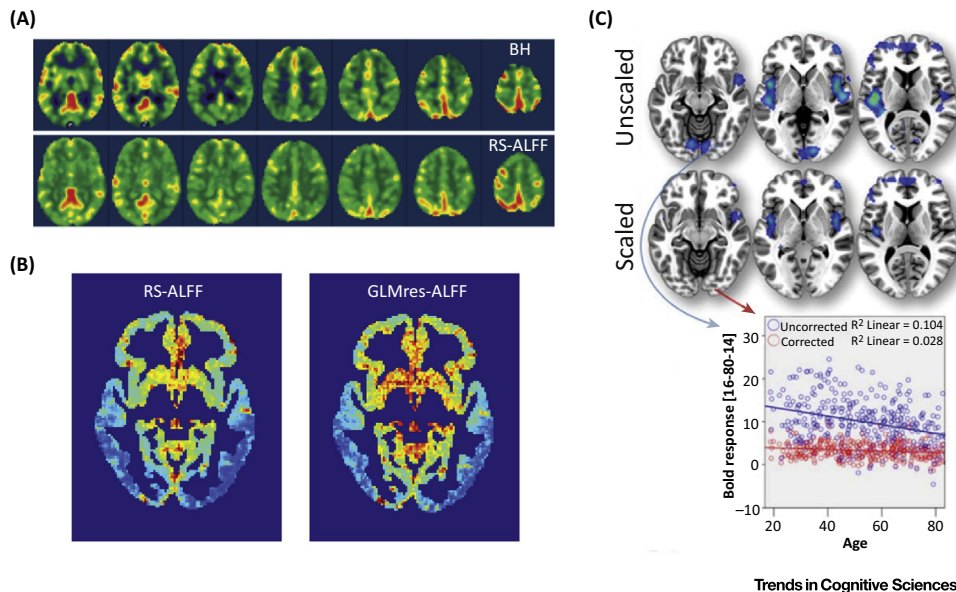


Figure 3. Post-Hoc Vascular Normalization Techniques. (A) Traditionally, hypercapnia has been used to elicit a BOLD signal change with no change in neural activity or consumption of oxygen; this response can be used to normalize the BOLD response measured in task runs. Breathholding (BH, top) can be used to induce hypercapnia. It was recently found that the amplitude of low-frequency fluctuations from RS-fMRI (RS-ALFF, bottom) mimics the hypercapnic BOLD response [69], and thus may be used instead of a hypercapnic run. Reproduced from [155] with permission from Oxford University Press. (B) The ALFF of the residuals of task fMRI data (GLMres-ALFF) also looks very similar to RS-ALFF (left) and can be used to rescale the BOLD signal change (vascular autoregulation, or VasA). Figure modified with permission from [73]. (C) Decrease in BOLD response to a sensorimotor task as a function of increasing age in the occipital lobe (blue) is abolished after scaling the sensorimotor task with RS-ALFF (red), where each point represents an individual. Figure adapted with permission from [74].

the supplemental information online). Moreover, although a general sense of the reliability of the most widely used fMRI measures can be derived from extant literature, details of the experimental procedures [79] and of the **preprocessing** applied to the raw data [80,81] matter. For these reasons, we believe that it is crucial to provide an assessment of reliability in each and every study that looks at individual differences, under the specific pipeline used for the analysis. At present, individual differences in fMRI are always studied with respect to an independent measure of the same individuals. The reliability of interest is that of the relationship between the fMRI-derived statistic and the independent variable (see also [78,82]). There are several flavors of reliability which are relevant to individual differences research, with direct analogs in behavioral testing. Test-retest reliability quantifies how variable the established relationship is in the same sample of subjects, under the same conditions (stimulus, scanner, time of day, analysis) at an appropriate time-interval (Figure 4A). It is also important to ensure that the relationship is robust to the exact preprocessing performed on the raw data, which can be conceived of as 'inter-rater' reliability (Figure 4B). In addition, because we are interested in brain function rather than the low-level properties of a given stimulus, the relationship must also be robust to the exact experimental conditions, known in psychology as 'parallel forms' reliability (Figure 4C). The relationship should further hold for a different sample of subjects (from the same population) (Figure 4D), and across scanners.

Though many measures of reliability have been proposed over the years [75], some directly addressing the ratio of inter- to intra-subject variance [83], a predictive framework inspired by machine-learning is best suited to establish the reliability needed for individual differences research, as we discuss further below (see 'Choosing Prediction over Correlation'). Reliability is ensured when a relationship discovered in a sample of subjects at site 1 using stimulus 1 and

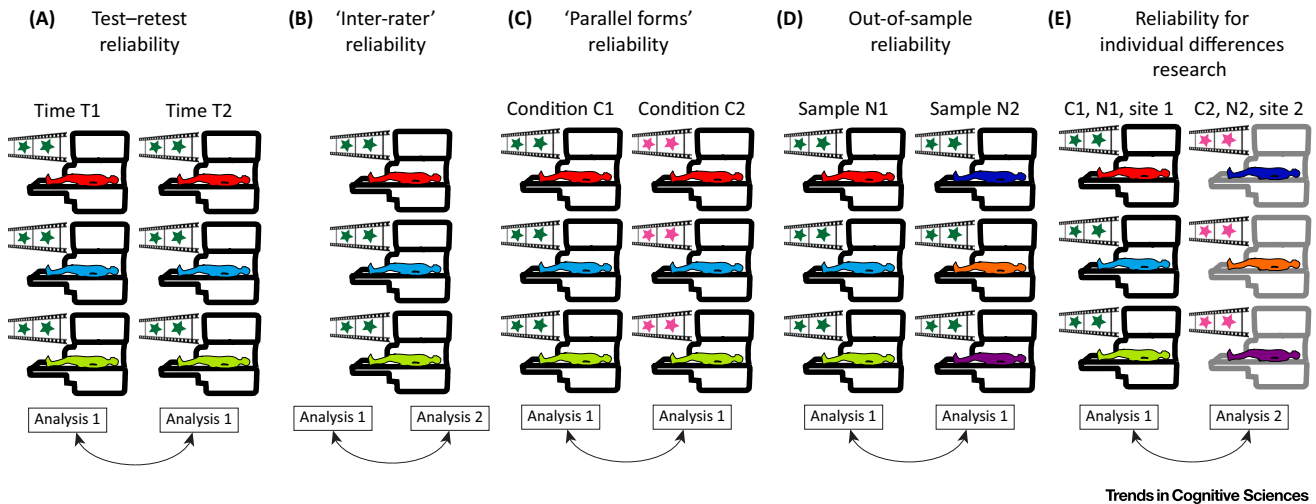


Figure 4. Reliability for Individual Differences Research. (A) Test–retest reliability. The same subjects are tested with the same stimuli in the same scanner, at an appropriate time-interval given the function of interest. (B) Inter-rater reliability. The same data are preprocessed in slightly different ways which are thought to be interchangeable (e.g., using physiological regressors or ICA to remove physiological noise). (C) Parallel forms reliability. The same subjects view slightly different stimuli which are thought to involve the same brain processing. (D) Out-of-sample reliability. A different set of subjects undergo the same experiment in the same scanner, and the data are analyzed in the same way. (E) Putting it all together for individual differences research. A result (e.g., relationship between fMRI-derived statistic and neuropsychological score) should be reliably obtained for different subjects at different sites, possibly using slightly different stimuli and preprocessing steps.

analysis 1 generalizes to a different sample of subjects at site 2 using stimulus 2 and analysis 2 (Figure 4E).

Sources of Within-Subject Variance We Acknowledge and Strive To Correct For

There are several sources of within-subject, inter-session variance that are well known, and which fMRI researchers strive to eliminate at acquisition time and/or correct for during preprocessing of the data [84,85]. Scanner-related noise, artifacts, and drift are unavoidable but are fairly simple to address through artifact rejection and temporal filtering. We briefly review here two main sources of within-subject variance which are more problematic and the state-of-the-art in addressing them: subject motion and subject body physiology.

Motion

Subject motion in the scanner leads to poor data quality, for obvious reasons: in fast **echo planar imaging** (EPI) sequences, motion not only disrupts spatial encoding but also disrupts the physical phenomena that the MRI signal relies on (e.g., resonance frequency, relaxation time between samples). Artifacts occur with frame-to-frame movements of a few tenths of a millimeter or less [86], meaning that they affect all datasets to some extent. The classical approach to correcting for motion artifacts has been to first estimate subject motion through rigid-body **realignment** of brain volumes, then to regress the calculated head motion parameters out of the fMRI timecourse to correct for any residual effects of motion [87]. Other researchers also include the **global signal** timecourse, the white matter timecourse, and the cerebrospinal fluid (CSF) timecourse as confound regressors [88]. This regression approach was recently shown to incompletely correct for motion artifacts in the context of resting-state fMRI analyses [86,89,90]. Additional corrections have been proposed, for example the removal of **independent components** related to motion [91], or the complete removal of motion-affected frames, known as scrubbing [86] ([92] for comprehensive review).

Thoroughly correcting for motion artifacts is important to ensure test–retest reliability: a given subject may move more in one session than in another. Nonetheless, a discussion of motion would have been equally appropriate in the previous section on validity: some subjects are

generally more fidgety than others in the scanner (some have argued that this constitutes a trait in and of itself, with a neurobiological basis [93]). Motion arguably contributes more to inter-subject variance than to intra-subject variance (e.g., men tend to exhibit more head movements than women [90], older people move more than younger people [94], and people with autism move more than controls [95]). Motion artifacts have complex effects on fMRI statistics, and incompletely correcting for them can lead to erroneous conclusions in individual differences research [86,95–97].

Physiology

In a previous section we pointed out how differences in vasculature may affect the validity of inter-individual comparisons. To complicate matters further, breathing and heart rate affect fMRI measurements. For instance, motion of the chest wall with respiration results in magnetic field changes [98], normal alterations of the depth or rate of breathing lead to variations in arterial CO₂ [99] and subsequent bloodflow changes [100], and pulsatile bloodflow due to heartbeat leads to fluctuations in signal intensity in arteries, arterioles, and other large vessels [101].

A common approach to correcting for these artifacts is to collect independent measurements of the heart rate (with a pulse oximeter) and respiration (with a respiratory belt), and regress them out of the fMRI signal using one of several available algorithms (reviewed in [85]); however, these corrections are not routinely applied. Interestingly, it has been found that many of these model-based corrections actually reduce test–retest reliability for functional connectivity analyses [102,103]. This has been interpreted to mean that a large fraction of functional connectivity reflects basic physiological signals [102] (another validity issue, e.g., [104]). A more optimistic take is that these model-based approaches still need some tweaking – either the models are not physiologically accurate or the measurements of heart rate and respiration are suboptimal (too noisy, or not measuring the variable of interest). Further development of the models (e.g., [105]) and of new and better MR-compatible measurement devices such as a continuous blood pressure monitoring device is needed [85].

Other approaches only rely on the fMRI data without the need for separate measurements (reviewed in [85]; see also [106–108]). Most of these are based on decomposing the data into components, for example, using independent component analysis (ICA) (but see [109]). Some of these methods have been shown to improve reliability [110]. However, a fair comparison with model-based physiological regression has not been conducted yet – typically, decomposition-based denoising techniques target a wider range of artifacts, including motion.

In decomposition-based methods, identifying noise components is not always trivial [111]; some components may be a mixture of signal and noise, and there is a risk of ‘throwing the baby out with the bathwater’. Automated algorithms, such as ICA-FIX [108], rely on prior manual classification of components, and are therefore not immune to this criticism. A promising new approach to teasing apart signals of neural origin from artifacts was recently introduced: using multi-echo EPI [112] it appears possible to distinguish a genuine BOLD effect from artifacts by looking at the timecourse of MR decay. The technique could lead to improvements in both the validity and reliability of measurements [113]. Finally, its combination with recent developments in simultaneous multi-slice acquisition [114] could make it a viable imaging protocol in terms of temporal and spatial resolution [115]; we are thus eagerly awaiting further validation and improvements of this technique (see also [116] for a related approach).

Other Sources of Intra-Subject Variance

Another Physiological Rhythm: Vasomotion

Vasomotion refers to the spontaneous changes in tone of blood vessels, independently of heart rate and respiration. Vasomotion leads to low-frequency oscillations in the BOLD signal [117].

There is some evidence that vasomotion may be localized to specific regions of cortex [118], making it a potentially serious confound. As with previous artifacts, vasomotion may differ systematically across subjects (validity). It is as yet unclear how vasomotion should be dealt with.

Baseline Physiological State

Baseline physiological state was already discussed in the validity section, but baseline physiology is also variable for a given subject. Going for a run an hour or less before scanning may lead to alterations of baseline physiology [119,120]. Anxiety or stress [121], extended exposure to high altitudes [122], recent sleep quality [123,124], or phase of the menstrual cycle (for women) [125] are all examples of poorly understood vascular and neural factors that may hinder the reliability of fMRI measurements if they are not properly measured and incorporated into a full model [126].

Neuromodulators and Vasoactive Substances

The effects of caffeine on fMRI measurements have been studied extensively. While it was once held that a dose of caffeine was a way to enhance the BOLD response [127], subsequent studies have muddied this picture [128,129]. Many complex effects of caffeine have been reported over the years, such as enhanced linearity of visually evoked BOLD responses [130]. The complexity likely stems from the double action of caffeine on neural activity and on hemodynamics ([131] for review).

Several other routinely consumed substances such as alcohol, nicotine, illicit drugs, and prescription medication, and even foods and food supplements, also have complex neural and vascular effects on fMRI measurements, and these are beyond the scope of this review. Using arterial spin labeling (ASL) and fMRI to obtain a better understanding of these effects, a field of research known as pharmacological MRI [132], may be a further important ingredient towards a reliable science of individual subjects from fMRI.

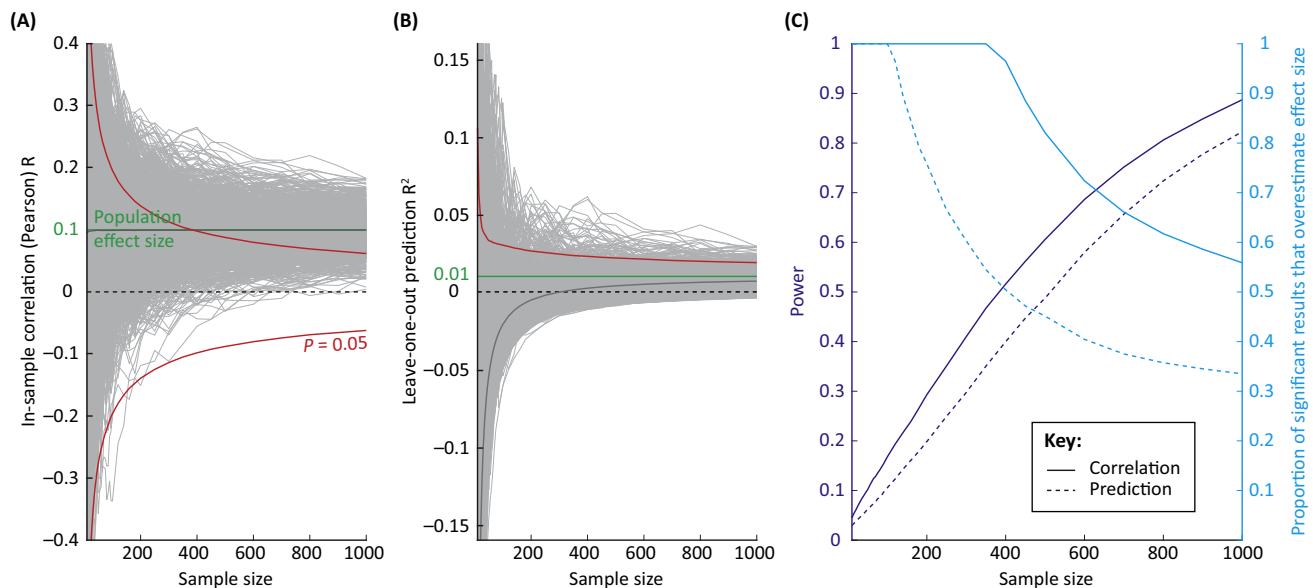
Further Considerations for an fMRI Science of Individual Differences

Choosing Prediction over Correlation

Currently the only way to interpret a fMRI-derived statistic is to relate it to another individual measure in the same set of subjects, such as their age, gender, test scores indexing aspects of intelligence and personality, or other measures of behavior (Table S2). By far the majority of fMRI studies of individual differences use correlation analysis to establish such a relationship. Correlation analysis relies on in-sample population inference and does not directly ensure the generalizability of the established relationship to out-of-sample individual subjects [133,134]. Shifting to a predictive framework is necessary to ensure generalizability and to interpret fMRI-derived statistics at the individual subject level [17,135] (for a more in-depth discussion of adopting a predictive machine-learning inspired framework, and the proper use of training, validation, and test datasets, see [136]). To fully control for remaining confounds, fMRI-derived statistics should be included in a full model alongside other potential predictors, and the unique predictive power of fMRI features should then be assessed through their selective removal from the model (as demonstrated in [16]) (Figure 1E).

Increasing Sample Size

It is now widely recognized that the small sample sizes routinely used in fMRI studies ($n = 10\text{--}50$) have low statistical power; given current reporting practices, this leads to inflated estimates of effect size [137,138] (Figure 5C) and thus poor replicability (see Box 3 for an overview of replicability-enhancing practices). fMRI research is not the first to face this challenge [143]. The current heuristic for studies using correlation analysis is a sample size of at least 100 (see [139] and Figure 5A for a more quantitative recommendation). In addition, replication in an independent sample is a worthy precaution, and this should be a requirement for studies that are not based on a strong prior hypothesis.



Trends in Cognitive Sciences

Figure 5. A Predictive Framework for a More Reproducible Science. These plots are provided for illustrative purposes only. They are based on simple simulations described in the supplemental material. (A) The ‘cone of confidence’ for correlation analysis, as in [139]. The true effect size in the population is $r = 0.1$ (green), with $\text{SNR} = 0.1$. The light-gray traces (1000 shown, of 100 000 generated) are bootstrap samples from this population. As sample size increases, the range of correlation values observed in the samples decreases. In red, the 95% confidence interval for the null hypothesis $r = 0$. (B) Same representation for the R^2 of a leave-one-out prediction analysis, using the same simulated samples. The average of all bootstrap samples is shown in dark gray; it asymptotes to the true effect size (green line) for larger sample sizes. (C) Power of the correlation and prediction analyses (dark blue), for $\alpha = 0.05$. Proportion of significant results ($P < 0.05$) which overestimates the effect size (light blue). Prediction analysis is less likely to overestimate effect size than correlation analysis; it even tends to underestimate it, making it a conservative analysis.

Larger samples are also beneficial in a predictive framework: a larger number of examples guards somewhat against **overfitting**. As to replication in an independent sample, it is a defining feature of a predictive framework. Model selection is conducted on the basis of training data, and independent testing data is reserved until the model is finalized. The fully trained model is tested once, and once only (as in machine learning competitions [140]), on the test data to establish generalizability (but see the recently introduced ‘reusable holdout’ algorithm for a viable alternative [202]).

Sample sizes in excess of 100 are still difficult to achieve for a small research group funded by a typical research grant; thus, despite this recommendation, underpowered correlation studies without internal replication will continue to be published. While it might seem that the cumulative output from many underpowered studies should eventually converge to a reliable conclusion through meta-analysis, this is not the case because of the strong bias to publish only significant findings [136,141] (Figure 5C). The reporting of all results regardless of their significance in the null hypothesis significance testing (NHST) framework, which could be implemented through initial **pre-registration** to ensure the quality of the methods [142], would be a way for studies with small sample sizes to contribute unbiased information for meta-analysis.

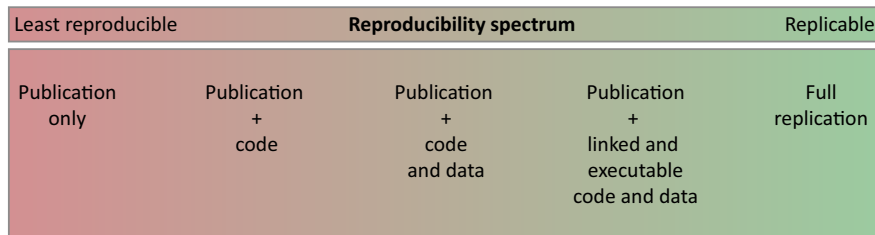
To achieve larger datasets in fMRI, data collected at different institutions can be aggregated in a shared online resource [144], as pioneered by the 1000 Functional Connectomes Project [145]. Several data-sharing platforms are currently available (see [146]), for example openfMRI [147]. For the accumulation of data across sites, standardized procedures must however be implemented. Resting-state fMRI is an obvious candidate for data aggregation given minimal instructions and requirements but, as we noted earlier, ‘small’ details must still be carefully

Box 3. Increasing the Reproducibility of fMRI Research

The replicability of fMRI research has recently been called into question. Culprits for the lack of replicability are small sample sizes [138], analytical flexibility [179] which fosters p-hacking [180] without appropriate control for the overall risk of false positives [181], and the bias to publish only statistically significant results (also known as the file drawer problem [182]). This means that we are often building new research on fragile grounds (previous false-positive results), and thus wasting a large amount of resources [183]. A shift in research practices is necessary.

Reproducibility consists in obtaining the same results using the same code and data [184] (Figure 1) (see also [183] for a slightly different terminology); it is a lower scientific standard than fully independent replication using a different dataset and similar analysis, but ensuring reproducibility in the first place is a surefire way to enhance replicability (although see [185] for a differing view). We note that failure to replicate a finding need not always imply that the finding was a false positive: a difference in outcome may also stem from any number of effects in execution or analysis that, unknown to the investigators attempting the replication, have actually changed the psychological process under investigation. Replication attempts help to establish the conditions under which a given finding is robust.

Following earlier recommendations for reporting fMRI methods [186], a recent white paper draft (from the OHBM Committee on Best Practices in Data Analysis and Sharing, also known as COBIDAS) is likely to be adopted by the imaging community. It identifies transparency, detailed reporting of methods, and importantly sharing of data and analysis pipelines (Figure 1) as key ingredients for reproducible fMRI research. Data-sharing platforms were discussed in the main text [146]; there are also platforms to share code, and guidelines on how to use scripting and pipelining to foster reproducibility [183]. An example of how to implement these recommendations is to release all the data together with a virtual machine environment to allow others to execute identical analyses [126]. At this stage such diligence still requires a large amount of technical overhead and cannot be made mandatory. The enforcement of these recommendations will initially largely depend on journals following suit and requiring these ingredients for publication, or on reviewers taking them into account when evaluating their peers' research. However, a growing set of resources may soon make this effort much more feasible and thus more widely adopted; in particular, we are closely following the efforts of the Stanford Center for Reproducible Neuroscience (<http://reproducibility.stanford.edu>) and of the Scientific Transparency Project (<http://post.stanford.edu>), both aiming at providing web-based platforms to generate reproducible workflows and leverage high-performance computing for the analysis of neuroimaging data.



Trends in Cognitive Sciences

Figure 1. From Reproducibility to Replicability: Sharing of Data and Executable Analysis Pipelines. Adapted from [184].

controlled (Box 1). Although simple task fMRI paradigms (e.g., finger tapping [148]) would also be good candidates, they may be too inefficient in that they are restricted to a narrow set of processes. We view movie fMRI as particularly well suited for data aggregation, given its ability to capture richer representational information while better matching subjects' mental states compared to the resting-state (Box 4), and at the same time preserving similarly minimal instructions and requirements.

Standardized procedures are of course easiest to implement in the context of large-scale projects such as the IMAGEN project (2000 subjects) [149], the WashU–UMinn Human Connectome Project (1200 subjects) [150], or the Cambridge Center for Ageing and Neuroscience (approximately 700 subjects) [151]. The latter two projects are aimed at acquiring state-of-the-art data and distributing it as a data-mining resource for fMRI researchers around the world. Projects such as these are akin to the accelerators used by particle physicists or the large telescopes used by astronomers: a few sites in the world

Box 4. Individual Assessment with Naturalistic, Condition-Rich Designs.

One promising class of stimuli, which affords tighter control on internal states than the resting-state while still being scalable (e.g., for aggregation of data across centers), comprises movies [26] and narrated stories [187]. These are naturalistic, highly-engaging stimuli which are very effective at driving brain activations and encompass a great diversity of mental states. Public datasets relying on story/movie data are in the making: for example, the studyforrest project recently released a high-quality 7T fMRI dataset of subjects listening to the narrated movie 'Forrest Gump' (1994) [188]. Efforts to extract and distribute annotations that can be used in analyses have also begun [189]. We present here three main methods to study the representations triggered by such stimuli across subjects.

Inter-Subject Correlation

Because all subjects watch the same movie, the timecourse of stimulus-related activity in their brains should look similar [190]. As one would expect, activity in early sensory cortices is similar across subjects, whereas activity in association cortices is less similar [26]. Dynamic inter-subject correlation analysis in light of a rich featural description of the stimuli is a promising avenue of research to capture rich information about individual differences in brain representations.

Representational Geometry

Different stimuli are represented in the brain along several dimensions, and representational geometry denotes the arrangement of stimuli in that space and the distances between them [191]. While representational geometry and representational similarity analysis have so far been exploited mostly to assess different models of brain function [192–194], they can also be leveraged to compare representational geometry across individual subjects [195,196]. Representational geometry can be derived from movie or story data, provided that the events of interest are properly labeled [189].

Voxel-Wise Modeling

Voxel-wise modeling consists in using high-dimensional featural models of natural stimuli as encoding models to make sense of fMRI data [197]. When a model is found to satisfactorily predict the data out-of-sample, it can then be interpreted. A recent study using movies with annotated semantic features generated detailed semantic maps of each subject's brain [198]. These maps represented those movie features that were most predictive of the activations observed at each voxel in the brain. These semantic maps capture a tremendous amount of information about the brain activity of each subject, and thus interrogating them with respect to individual differences is a very exciting endeavor. Such work is already well underway [199].

acquire the best possible data, and these data are subsequently probed (for many years) by the best analysts around the world.

Towards Normative fMRI Research?

A key concept in psychology is that of a 'norm', in other words the distribution of a score in the population of interest. Scores on psychological tests are usually standardized with respect to their distribution in a large normative sample (the size of which depends on the distribution of the score [152]), and this allows their direct interpretation without recourse to another measure. In the future, we may be able to interpret the fMRI-derived statistic of an individual subject in light of its distribution in a normative sample, provided that absolute reliability is achieved; this approach could lead, for clinical imaging, to a more biologically informed science of human neuropsychiatric disease [153,154] and a better basis for personalized medicine.

Concluding Remarks

Individual differences in brain function are key to understanding healthy differences based on personality, gender, age, or culture. They are also crucial for personalized medicine approaches to neuropsychiatric diseases. Recent technical advances have increased the sensitivity of functional MRI and set the stage for a characterization of brain activity at the level of momentary mental events in individual subjects. We now face key challenges of reliability and validity on the path to an fMRI-informed science of individual differences. We already have some tools to address these challenges: we propose a pipeline for the measurement and analysis of individual differences with fMRI as shown in Figure 1, on the basis of current knowledge (see Outstanding Questions). While interpretation of fMRI-derived measures currently relies on independent

Outstanding Questions

Will the newer methods that we describe here and recommend (highlighted in gray in Figure 1) stand the test of time?

Which of the concerns reviewed here (inter-subject alignment, hemodynamic variability across subjects, sources of noise) are the most important, and which can often be ignored in practice?

How do the many corrections we have reviewed here interact with one another? For example, does normalization become superfluous when the HRF is properly modeled?

How long will BOLD-fMRI still be the norm for non-invasive functional imaging of the whole brain in awake humans? Are other more valid non-invasive imaging techniques conceivable in the future?

What other modalities can be combined with fMRI to provide a richer set of measures and enhance validity, besides physiological recordings? Options include eye-tracking, EEG, and fNIRS.

Should pre-registration be required for fMRI studies of individual differences below a minimum sample size?

What role will behavioral and genotypic data play in the future? Will these always be required, or can we explore individual differences in a data-driven way based on fMRI data alone?

Will fMRI-derived individual differences lead us to revise clinical diagnostic categories?

measures of behavior, psychological scores, and neuropsychiatric diagnosis, we are hopeful that in the near future they will stand on their own in light of normative data. No matter the strategy, increasing sample size dramatically is necessary, and data-sharing efforts together with standardized procedures and reproducibility-enhancing practices will be key to the process. Beyond the current emphasis on resting-state fMRI data, which is easy to perform and aggregate, the use of naturalistic stimuli that capture rich inter-individual differences in cognition is a direction worth exploring.

Acknowledgments

This work was funded in part by a NARSAD grant from the Brain and Behavior Research Foundation (to J.D.) and a Conte Center from the National Institute of Mental Health (NIMH) (to R.A.). The authors thank Rebecca Schwarzlose, Michael Miller, Alex Huth, Swaroop Guntupalli and two anonymous reviewers for useful comments on the manuscript.

Appendix A Supplemental information

Supplemental information related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.tics.2016.03.014>.

References

- Miller, M.B. and Van Horn, J.D. (2007) Individual variability in brain activations associated with episodic retrieval: a role for large-scale databases. *Int. J. Psychophysiol.* 63, 205–213
- Van Horn, J.D. *et al.* (2008) Individual variability in brain activity: a nuisance or an opportunity? *Brain Imaging Behav.* 2, 327–334
- Kirilina, E. *et al.* (2016) The quest for the best: the impact of different EPI sequences on the sensitivity of random effect fMRI group analyses. *Neuroimage* 126, 49–59
- Dosenbach, N.U.F. *et al.* (2010) Prediction of individual brain maturity using fMRI. *Science* 329, 1358–1361
- Geerligns, L. *et al.* (2015) A brain-wide study of age-related changes in functional connectivity. *Cereb. Cortex* 25, 1987–1999
- Yarkoni, T. (2015) Neurobiological substrates of personality: a critical overview. In *Personality Processes and Individual Differences 4* (Shaver, M.M. *et al.*, eds), American Psychological Association, pp. 000–000
- van den Heuvel, M.P. *et al.* (2009) Efficiency of functional brain networks and intellectual performance. *J. Neurosci.* 29, 7619–7624
- Finn, E.S. *et al.* (2015) Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18, 1664–1671
- Smith, S.M. *et al.* (2015) A positive–negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* 18, 1565–1567
- Murphy, S.E. *et al.* (2013) The effect of the serotonin transporter polymorphism (5-HTTLPR) on amygdala function: a meta-analysis. *Mol. Psychiatry* 18, 512–520
- Arbabshirani, M.R. *et al.* (2013) Classification of schizophrenia patients based on resting-state functional network connectivity. *Front. Neurosci.* 7, 133
- Yahata, N. *et al.* A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nature Communications* 7, Article number: 11254 <http://dx.doi.org/10.1038/ncomms11254>
- Orrù, G. *et al.* (2012) Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci. Biobehav. Rev.* 36, 1140–1152
- Wolfers, T. *et al.* (2015) From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci. Biobehav. Rev.* 57, 328–349
- Williams, L.M. *et al.* (2015) Amygdala reactivity to emotional faces in the prediction of general and medication-specific responses to antidepressant treatment in the randomized iSPOT-D trial. *Neuropsychopharmacology* 40, 2398–2408
- Whelan, R. *et al.* (2014) Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature* 512, 185–189
- Gabrieli, J.D.E. *et al.* (2015) Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* 85, 11–26
- Laird, A.R. *et al.* (2010) Comparison of the disparity between Talairach and MNI coordinates in functional neuroimaging data: validation of the Lancaster transform. *Neuroimage* 51, 677–683
- Klein, A. *et al.* (2009) Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46, 786–802
- Fischl, B. *et al.* (1999) II. Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207
- Yeo, B.T.T. *et al.* (2010) Spherical demons: fast diffeomorphic landmark-free surface registration. *IEEE Trans. Med. Imaging* 29, 650–668
- Klein, A. *et al.* (2010) Evaluation of volume-based and surface-based brain image registration methods. *Neuroimage* 51, 214–220
- Frost, M.A. and Goebel, R. (2012) Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage* 59, 1369–1381
- Langers, D.R.M. (2014) Assessment of tonotopically organised subdivisions in human auditory cortex using volumetric and surface-based cortical alignments. *Hum. Brain Mapp* 35, 1544–1561
- Glasser, M.F. *et al.* (2013) The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80, 105–124
- Hasson, U. *et al.* (2010) Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* 14, 40–48
- Sabuncu, M.R. *et al.* (2010) Function-based intersubject alignment of human cortical anatomy. *Cereb. Cortex* 20, 130–140
- Conroy, B.R. *et al.* (2013) Inter-subject alignment of human cortical anatomy using functional connectivity. *Neuroimage* 81, 400–411
- Khullar, S. *et al.* (2011) ICA-fNORM: spatial normalization of fMRI data using intrinsic group-ICA networks. *Front. Syst. Neurosci.* 5, 93
- Çetin, M.S. *et al.* (2015) Enhanced disease characterization through multi network functional normalization in fMRI. *Front. Neurosci.* 9, 95
- Robinson, E.C. *et al.* (2014) MSM: a new flexible framework for multimodal surface matching. *Neuroimage* 100, 414–426
- Glasser, M.F. and Van Essen, D.C. (2011) Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* 31, 11597–11616
- Abdollahi, R.O. *et al.* (2014) Correspondences between retinotopic areas and myelin maps in human visual cortex. *Neuroimage* 99, 509–524

34. Haxby, J.V. *et al.* (2011) A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404–416
35. Guntupalli, J.S. *et al.* (2016) A model of representational spaces in human cortex. *Cereb. Cortex* Published online February 26, 2106. <http://dx.doi.org/10.1093/cercor/bhw068>.
36. Saxe, R. *et al.* (2006) Divide and conquer: a defense of functional localizers. *Neuroimage* 30, 1088–1096
37. Gorgolewski, K.J. *et al.* (2012) Adaptive thresholding for reliable topological inference in single subject fMRI analysis. *Front. Hum. Neurosci.* 6, 245
38. Smith, S.M. and Nichols, T.E. (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44, 83–98
39. Nieto-Castañón, A. and Fedorenko, E. (2012) Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage* 63, 1646–1669
40. Kanwisher, N. and Yovel, G. (2006) The fusiform face area: a cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. Lond. B Biol. Sci* 361, 2109–2128
41. Saxe, R. and Powell, L.J. (2006) It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol. Sci.* 17, 692–699
42. Spunt, R.P. and Adolphs, R. (2014) Validating the why/how contrast for functional MRI studies of theory of mind. *Neuroimage* 99, 301–311
43. Friston, K.J. *et al.* (2006) A critique of functional localisers. *Neuroimage* 30, 1077–1087
44. Handwerker, D.A. *et al.* (2012) The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *Neuroimage* 62, 1017–1023
45. Liu, T.T. (2013) Neurovascular factors in resting-state functional MRI. *Neuroimage* 80, 339–348
46. Bandettini, P.A. (2014) Neuronal or hemodynamic? Grappling with the functional MRI signal. *Brain Connect.* 4, 487–498
47. Shan, Z.Y. *et al.* (2016) Genes influence the amplitude and timing of brain hemodynamic responses. *Neuroimage* 124, 663–671
48. Buxton, R.B. *et al.* (2004) Modeling the hemodynamic response to brain activation. *Neuroimage* 23 (Suppl. 1), S220–S233
49. Stephan, K.E. *et al.* (2007) Comparing hemodynamic models with DCM. *Neuroimage* 38, 387–401
50. D'Esposito, M. *et al.* (2003) Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. *Nat. Rev. Neurosci.* 4, 863–872
51. Lindquist, M.A. *et al.* (2009) Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. *Neuroimage* 45, S187–S198
52. Henson, R.N.A. *et al.* (2002) Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *Neuroimage* 15, 83–97
53. Woolrich, M.W. *et al.* (2004) Constrained linear basis sets for HRF modelling using variational Bayes. *Neuroimage* 21, 1748–1761
54. Glover, G.H. (1999) Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* 9, 416–429
55. Calhoun, V.D. *et al.* (2004) fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *Neuroimage* 22, 252–257
56. Degras, D. and Lindquist, M.A. (2014) A hierarchical model for simultaneous detection and estimation in multi-subject fMRI studies. *Neuroimage* 98, 61–72
57. Pedregosa, F. *et al.* (2015) Data-driven HRF estimation for encoding and decoding models. *Neuroimage* 104, 209–220
58. Chen, J.E. and Glover, G.H. (2015) BOLD fractional contribution to resting-state functional connectivity above 0.1 Hz. *Neuroimage* 107, 207–218
59. Wu, G.-R. *et al.* (2013) A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data. *Med. Image Anal.* 17, 365–374
60. Sreenivasan, K.R. *et al.* (2015) Nonparametric hemodynamic deconvolution of fMRI using homomorphic filtering. *IEEE Trans. Med. Imaging* 34, 1155–1163
61. Griffeth, V.E.M. and Buxton, R.B. (2011) A theoretical framework for estimating cerebral oxygen metabolism changes using the calibrated-BOLD method: modeling the effects of blood volume distribution, hematocrit, oxygen extraction fraction, and tissue signal properties on the BOLD signal. *Neuroimage* 58, 198–212
62. Liu, T.T. *et al.* (2013) An introduction to normalization and calibration methods in functional MRI. *Psychometrika* 78, 308–321
63. Pike, G.B. (2012) Quantitative functional MRI: concepts, issues and future challenges. *Neuroimage* 62, 1234–1240
64. Blockley, N.P. *et al.* (2015) Calibrating the BOLD response without administering gases: comparison of hypercapnia calibration with calibration using an asymmetric spin echo. *Neuroimage* 104, 423–429
65. Blockley, N.P. *et al.* (2013) A review of calibrated blood oxygenation level-dependent (BOLD) methods for the measurement of task-induced changes in brain oxygen metabolism. *NMR Biomed* 26, 987–1003
66. Bandettini, P.A. and Wong, E.C. (1997) A hypercapnia-based normalization method for improved spatial localization of human brain activation with fMRI. *NMR Biomed.* 10, 197–203
67. Handwerker, D.A. *et al.* (2007) Reducing vascular variability of fMRI data across aging populations using a breathholding task. *Hum. Brain Mapp.* 28, 846–859
68. Lu, H. *et al.* (2010) Improving fMRI sensitivity by normalization of basal physiologic state. *Hum. Brain Mapp.* 31, 80–87
69. Kannurpatti, S.S. and Biswal, B.B. (2008) Detection and scaling of task-induced fMRI-BOLD response using resting state fluctuations. *Neuroimage* 40, 1567–1574
70. Kalcher, K. *et al.* (2013) RESCALE: voxel-specific task-fMRI scaling using resting state fluctuation amplitude. *Neuroimage* 70, 80–88
71. Golestani, A.M. *et al.* (2015) Mapping the end-tidal CO₂ response function in the resting-state BOLD fMRI signal: spatial specificity, test-retest reliability and effect of fMRI sampling rate. *Neuroimage* 104, 266–277
72. Fair, D.A. *et al.* (2007) A method for using blocked and event-related fMRI data to study 'resting state' functional connectivity. *Neuroimage* 35, 396–405
73. Kazan, S.M. *et al.* (2016) Vascular autorescaling of fMRI (VasA fMRI) improves sensitivity of population studies: A pilot study. *Neuroimage* 124, 794–805
74. Tsvetanov, K.A. *et al.* (2015) The effect of ageing on fMRI: correction for the confounding effects of vascular reactivity evaluated by joint fMRI and MEG in 335 adults. *Hum. Brain Mapp* 36, 2248–2269
75. Bennett, C.M. and Miller, M.B. (2010) How reliable are the results from functional magnetic resonance imaging? *Ann. N. Y. Acad. Sci.* 1191, 133–155
76. Zuo, X.-N. and 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
77. Andellini, M. *et al.* (2015) Test-retest reliability of graph metrics of resting state MRI functional brain networks: a review. *J. Neurosci. Methods* 253, 183–192
78. Barch, D.M. and Mathalon, D.H. (2011) Using brain imaging measures in studies of procognitive pharmacologic agents in schizophrenia: psychometric and quality assurance considerations. *Biol. Psychiatry* 70, 13–18
79. Bennett, C.M. and Miller, M.B. (2013) fMRI reliability: influences of task and experimental design. *Cogn. Affect. Behav. Neurosci.* 13, 690–702
80. Churchill, N.W. *et al.* (2012) Optimizing preprocessing and analysis pipelines for single-subject fMRI. 2. Interactions with ICA, PCA, task contrast and inter-subject heterogeneity. *PLoS One* 7, e31147
81. Aurich, N.K. *et al.* (2015) Evaluating the reliability of different preprocessing steps to estimate graph theoretical measures in resting state fMRI data. *Front. Neurosci.* 9, 48

82. Raemaekers, M. *et al.* (2007) Test-retest reliability of fMRI activation during prosaccades and antisaccades. *Neuroimage* 36, 532–542
83. Caceres, A. *et al.* (2009) Measuring fMRI reliability with the intraclass correlation coefficient. *Neuroimage* 45, 758–768
84. Greve, D.N. *et al.* (2013) A survey of the sources of noise in fMRI. *Psychometrika* 78, 396–416
85. Murphy, K. *et al.* (2013) Resting-state fMRI confounds and cleanup. *Neuroimage* 80, 349–359
86. Power, J.D. *et al.* (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154
87. Power, J.D. *et al.* (2014) Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84, 320–341
88. Fox, M.D. *et al.* (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U. S. A* 102, 9673–9678
89. 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
90. Van Dijk, K.R.A. *et al.* (2012) The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage* 59, 431–438
91. Pruim, R.H.R. *et al.* (2015) Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. *Neuroimage* 112, 278–287
92. Power, J.D. *et al.* (2015) Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage* 105, 536–551
93. Zeng, L.-L. *et al.* (2014) Neurobiological basis of head motion in brain imaging. *Proc. Natl. Acad. Sci. U. S. A* 111, 6058–6062
94. Geerligns, L. *et al.* (2015) State and trait components of functional connectivity: individual differences vary with mental state. *J. Neurosci* 35, 13949–13961
95. Tyszka, J.M. *et al.* (2014) Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. *Cereb. Cortex* 24, 1894–1905
96. Satterthwaite, T.D. *et al.* (2013) Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. *Neuroimage* 83, 45–57
97. Turner, B.O. *et al.* (2015) One dataset, many conclusions: BOLD variability's complicated relationships with age and motion artifacts. *Brain Imaging Behav.* 9, 115–127
98. Raj, D. *et al.* (2001) Respiratory effects in human functional magnetic resonance imaging due to bulk susceptibility changes. *Phys. Med. Biol* 46, 3331–3340
99. Wise, R.G. *et al.* (2004) Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. *Neuroimage* 21, 1652–1664
100. Chang, C. and Glover, G.H. (2009) Relationship between respiration, end-tidal CO₂, and BOLD signals in resting-state fMRI. *Neuroimage* 47, 1381–1393
101. Dagli, M.S. *et al.* (1999) Localization of cardiac-induced signal change in fMRI. *Neuroimage* 9, 407–415
102. Birn, R.M. *et al.* (2014) The influence of physiological noise correction on test-retest reliability of resting-state functional connectivity. *Brain Connect.* 4, 511–522
103. Lipp, I. *et al.* (2014) Understanding the contribution of neural and physiological signal variation to the low repeatability of emotion-induced BOLD responses. *Neuroimage* 86, 335–342
104. Kassner, A. *et al.* (2010) Blood-oxygen level dependent MRI measures of cerebrovascular reactivity using a controlled respiratory challenge: reproducibility and gender differences. *J. Magn. Reson. Imaging* 31, 298–304
105. Särrkkä, S. *et al.* (2012) Dynamic retrospective filtering of physiological noise in BOLD fMRI: DRIFTER. *Neuroimage* 60, 1517–1527
106. Kay, K. *et al.* (2013) GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Front. Neurosci* 7, 00247
107. Churchill, N.W. and Strother, S.C. (2013) PHYCAA+: an optimized, adaptive procedure for measuring and controlling physiological noise in BOLD fMRI. *Neuroimage* 82, 306–325
108. Salimi-Khorshidi, G. *et al.* (2014) Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* 90, 449–468
109. Anderson, J.S. *et al.* (2011) Network anticorrelations, global regression, and phase-shifted soft tissue correction. *Hum. Brain Mapp.* 32, 919–934
110. Pruim, R.H.R. *et al.* (2015) ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* 112, 267–277
111. Kelly, R.E., Jr. *et al.* (2010) Visual inspection of independent components: defining a procedure for artifact removal from fMRI data. *J. Neurosci. Methods* 189, 233–245
112. Kundu, P. *et al.* (2012) Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *Neuroimage* 60, 1759–1770
113. Kundu, P. *et al.* (2015) Robust resting state fMRI processing for studies on typical brain development based on multi-echo EPI acquisition. *Brain Imaging Behav.* 9, 56–73
114. Moeller, S. *et al.* (2010) Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn. Reson. Med* 63, 1144–1153
115. Olafsson, V. *et al.* (2015) Enhanced identification of BOLD-like components with multi-echo simultaneous multi-slice (MESMS) fMRI and multi-echo ICA. *Neuroimage* 112, 43–51
116. Bright, M.G. and Murphy, K. (2013) Removing motion and physiological artifacts from intrinsic BOLD fluctuations using short echo data. *Neuroimage* 64, 526–537
117. Tong, Y. and Frederick, B.D. (2014) Studying the spatial distribution of physiological effects on BOLD signals using ultrafast fMRI. *Front. Hum. Neurosci.* 8, 196
118. Rayshubskiy, A. *et al.* (2014) Direct, intraoperative observation of ~0.1 Hz hemodynamic oscillations in awake human cortex: implications for fMRI. *Neuroimage* 87, 323–331
119. MacIntosh, B.J. *et al.* (2014) Impact of a single bout of aerobic exercise on regional brain perfusion and activation responses in healthy young adults. *PLoS One* 9, e85163
120. Rajab, A.S. *et al.* (2014) A single session of exercise increases connectivity in sensorimotor-related brain networks: a resting-state fMRI study in young healthy adults. *Front. Hum. Neurosci.* 8, 625
121. Keulers, E.H.H. *et al.* (2015) The association between cortisol and the BOLD response in male adolescents undergoing fMRI. *Brain Res.* 1598, 1–11
122. Levin, J.M. *et al.* (2001) Influence of baseline hematocrit and hemodilution on BOLD fMRI activation. *Magn. Reson. Imaging* 19, 1055–1062
123. Yoo, S.S. *et al.* (2007) The human emotional brain without sleep – a prefrontal amygdala disconnect. *Curr. Biol.* 17, R877–R878
124. Tagliazucchi, E. and Laufs, H. (2014) Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* 82, 695–708
125. Dreher, J.C. *et al.* (2007) Menstrual cycle phase modulates reward-related neural function in women. *Proc. Natl. Acad. Sci. U. S. A* 104, 2465–2470
126. Poldrack, R.A. *et al.* (2015) Long-term neural and physiological phenotyping of a single human. *Nat. Commun.* 6
127. Mulderink, T.A. *et al.* (2002) On the use of caffeine as a contrast booster for BOLD fMRI studies. *Neuroimage* 15, 37–44
128. Laurienti, P.J. *et al.* (2003) Relationship between caffeine-induced changes in resting cerebral perfusion and blood oxygenation level-dependent signal. *Am J Neuroradiol* 24, 1607–1611
129. Addicott, M.A. *et al.* (2012) The effects of dietary caffeine use and abstinence on blood oxygen level-dependent activation and cerebral blood flow. *Journal of caffeine research* 2, 15–22
130. Liu, T.T. and Liu, J. (2010) Caffeine increases the linearity of the visual BOLD response. *Neuroimage* 49, 2311–2317
131. Koppelstaetter, F. *et al.* (2010) Caffeine and cognition in functional magnetic resonance imaging. *J. Alzheimers. Dis.* 20 (Suppl. 1), S71–S84

132. Wang, D.J.J. *et al.* (2011) Potentials and challenges for arterial spin labeling in pharmacological magnetic resonance imaging. *J. Pharmacol. Exp. Ther.* 337, 359–366
133. Whelan, R. and Garavan, H. (2014) When optimism hurts: inflated predictions in psychiatric neuroimaging. *Biol. Psychiatry* 75, 746–748
134. Lo, A. *et al.* (2015) Why significant variables aren't automatically good predictors. *Proc. Natl. Acad. Sci. U. S. A* 112, 13892–13897
135. Linden, D.E.J. (2012) The challenges and promise of neuroimaging in psychiatry. *Neuron* 73, 8–22
136. Yarkoni, T. and Westfall, J. (2016) *Choosing prediction over explanation in psychology: lessons from machine learning*, figshare Retrieved: 15 39, Apr 28, 2016 (GMT) <https://dx.doi.org/10.6084/m9.figshare.2441878.v1>
137. Yarkoni, T. and Braver, T.S. (2012) Cognitive neurosciences approaches to individual differences in working memory and executive control: conceptual and methodological issues. In *Handbook of Individual Differences in Cognition: Attention, Memory, and Executive Control* (Gruszka, A. *et al.*, eds), pp. 87–107, Springer
138. Button, K.S. *et al.* (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376
139. Schönbrodt, F.D. and Perugini, M. (2013) At what sample size do correlations stabilize? *J. Res. Pers.* 47, 609–612
140. Silva, R.F. *et al.* (2014) The tenth annual MLSP competition: schizophrenia classification challenge. In *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6 IEEE
141. Ioannidis, J.P.A. (2005) Why most published research findings are false. *PLoS Med.* 2, e124
142. Nosek, B.A. and Lakens, D. (2014) Registered reports. *Soc. Psychol.* 45, 137–141
143. Siontis, K.C.M. *et al.* (2010) Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. *Eur. J. Hum. Genet.* 18, 832–837
144. Van Horn, J.D. and Gazzaniga, M.S. (2013) Why share data? Lessons learned from the fMRIDC. *Neuroimage* 82, 677–682
145. Biswal, B.B. *et al.* (2010) Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U. S. A* 107, 4734–4739
146. Eickhoff, S. *et al.* (2016) Sharing the wealth: neuroimaging data repositories. *Neuroimage* 124, 1065–1068
147. Poldrack, R.A. and Gorgolewski, K.J. (2015) OpenfMRI: open sharing of task fMRI data. *Neuroimage* Published online June 3, 2015. <http://dx.doi.org/10.1016/j.neuroimage.2015.05.073>
148. Blumensath, T. *et al.* (2013) Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. *Neuroimage* 76, 313–324
149. Schumann, G. *et al.* (2010) The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol. Psychiatry* 15, 1128–1139
150. Van Essen, D.C. *et al.* (2013) The WU–Minn Human Connectome Project: an overview. *Neuroimage* 80, 62–79
151. Taylor, J.R. *et al.* (2015) The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *Neuroimage*. Published online September 12, 2015. <http://dx.doi.org/10.1016/j.neuroimage.2015.09.018>
152. Crawford, J.R. and Garthwaite, P.H. (2008) On the 'optimal' size for normative samples in neuropsychology: capturing the uncertainty when normative data are used to quantify the standing of a neuropsychological test score. *Child Neuropsychol.* 14, 99–117
153. Cuthbert, B.N. and Insel, T.R. (2013) Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* 11, 126
154. Insel, T.R. and Cuthbert, B.N. (2015) Brain disorders? Precisely. *Science* 348, 499–500
155. Di, X. *et al.* (2013) Calibrating BOLD fMRI activations with neurovascular and anatomical constraints. *Cereb. Cortex* 23, 255–263
156. Power, J.D. *et al.* (2014) Studying brain organization via spontaneous fMRI signal. *Neuron* 84, 681–696
157. Biswal, B. *et al.* (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541
158. Smith, S.M. *et al.* (2009) Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U. S. A* 106, 13040–13045
159. Birn, R.M. *et al.* (2013) The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage* 83, 550–558
160. Laumann, T.O. *et al.* (2015) Functional system and areal organization of a highly sampled individual human brain. *Neuron* 87, 657–670
161. 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
162. Christoff, K. *et al.* (2009) Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci. U. S. A* 106, 8719–8724
163. Delamilleure, P. *et al.* (2010) The resting state questionnaire: an introspective questionnaire for evaluation of inner experience during the conscious resting state. *Brain Res. Bull.* 81, 565–573
164. Hurlburt, R.T. *et al.* (2015) What goes on in the resting-state?. A qualitative glimpse into resting-state experience in the scanner. *Front. Psychol.* 6, 1535
165. Tung, K.-C. *et al.* (2013) Alterations in resting functional connectivity due to recent motor task. *Neuroimage* 78, 316–324
166. Hutchison, R.M. *et al.* (2013) Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* 80, 360–378
167. Cole, M.W. *et al.* (2014) Intrinsic and task-evoked network architectures of the human brain. *Neuron* 83, 238–251
168. Smith, S.M. *et al.* (2013) Resting-state fMRI in the Human Connectome Project. *Neuroimage* 80, 144–168
169. Beckmann, C.F. *et al.* (2005) Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Lond. B Biol. Sci* 360, 1001–1013
170. Yeo, B.T.T. *et al.* (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165
171. Craddock, R.C. *et al.* (2012) A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp* 33, 1914–1928
172. Wig, G.S. *et al.* (2014) Parcellating an individual subject's cortical and subcortical brain structures using snowball sampling of resting-state correlations. *Cereb. Cortex* 24, 2036–2054
173. Arslan, S. *et al.* (2015) Joint spectral decomposition for the parcellation of the human cerebral cortex using resting-state fMRI. *Inf. Process. Med. Imaging* 24, 85–97
174. Wig, G.S. *et al.* (2014) An approach for parcellating human cortical areas using resting-state correlations. *Neuroimage* 93, 276–291
175. Gordon, E.M. *et al.* (2016) Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* 26, 288–303
176. Beckmann, C.F. *et al.* (2009) Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. *Neuroimage* 47, S148
177. Hacker, C.D. *et al.* (2013) Resting state network estimation in individual subjects. *Neuroimage* 82, 616–633
178. Wang, D. *et al.* (2015) Parcellating cortical functional networks in individuals. *Nat. Neurosci* 18, 1853–1860
179. Carp, J. (2012) On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments. *Front. Neurosci.* 6, 149
180. Neuroskeptic (2012) The nine circles of scientific hell. *Perspect. Psychol. Sci.* 7, 643–644
181. Simmons, J.P. *et al.* (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci* 22, 1359–1366
182. Simonsohn, U. *et al.* (2014) P-curve: a key to the file-drawer. *J. Exp. Psychol. Gen.* 143, 534–547

183. Pernet, C. and Poline, J.-B. (2015) Improving functional magnetic resonance imaging reproducibility. *Gigascience* 4, 15
184. Peng, R.D. (2011) Reproducible research in computational science. *Science* 334, 1226–1227
185. Drummond, C. (2009) Replicability is not reproducibility: nor is it good science. In *Proceedings of the Evaluation Methods for Machine Learning Workshop at the 26th International Conference for Machine Learning*. Montreal, Quebec, Canada
186. Poldrack, R.A. *et al.* (2008) Guidelines for reporting an fMRI study. *Neuroimage* 40, 409–414
187. Regev, M. *et al.* (2013) Selective and invariant neural responses to spoken and written narratives. *J. Neurosci* 33, 15978–15988
188. Hanke, M. *et al.* (2014) A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Sci. Data* 1, 140003
189. Labs, A. *et al.* (2015) Portrayed emotions in the movie 'Forrest Gump'. *F1000Res.* 4, 92
190. Hasson, U. *et al.* (2004) Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640
191. Kriegeskorte, N. and Kievit, R.A. (2013) Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* 17, 401–412
192. Nili, H. *et al.* (2014) A toolbox for representational similarity analysis. *PLoS Comput. Biol* 10, e1003553
193. Skerry, A.E. and Saxe, R. (2015) Neural representations of emotion are organized around abstract event features. *Curr. Biol* 25, 1945–1954
194. Tamir, D.I. *et al.* (2016) Neural evidence that three dimensions organize mental state representation: rationality, social impact, and valence. *Proc. Natl. Acad. Sci. U. S. A.* 113, 194–199
195. Charest, I. *et al.* (2014) Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14565–14570
196. Charest, I. and Kriegeskorte, N. (2015) The brain of the beholder: honouring individual representational idiosyncrasies. *Language, Cognition and Neuroscience* 30, 367–379
197. Naselaris, T. *et al.* (2011) Encoding and decoding in fMRI. *Neuroimage* 56, 400–410
198. Huth, A.G. *et al.* (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76, 1210–1224
199. Huth, A.G. *et al.* (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* <http://dx.doi.org/10.1038/nature17637>
200. Glasser, M.F. *et al.* A Multi-modal parcellation of human cerebral cortex. *Nature* (in press).
201. Gordon, E.M. *et al.* (2015) Individual variability of the system-level organization of the human brain. *Cereb. Cortex* Published online October 13, 2015. <http://dx.doi.org/10.1093/cercor/bhv239>
202. Dwork, C. *et al.* (2015) STATISTICS. The reusable holdout: Preserving validity in adaptive data analysis. *Science* 349, 636–638