



# TAPAS: An Open-Source Software Package for Translational Neuromodeling and Computational Psychiatry

**Stefan Frässle<sup>1</sup>, Eduardo A. Aponte<sup>1</sup>, Saskia Bollmann<sup>1,2,3,4,5</sup>, Kay H. Brodersen<sup>1,6</sup>, Cao T. Do<sup>1</sup>, Olivia K. Harrison<sup>1,7,8</sup>, Samuel J. Harrison<sup>1</sup>, Jakob Heinzle<sup>1</sup>, Sandra Iglesias<sup>1</sup>, Lars Kasper<sup>1,9</sup>, Ekaterina I. Lomakina<sup>1,6</sup>, Christoph Mathys<sup>1,10</sup>, Matthias Müller-Schrader<sup>1</sup>, Inês Pereira<sup>1</sup>, Frederike H. Petzschner<sup>1</sup>, Sudhir Raman<sup>1</sup>, Dario Schöbi<sup>1</sup>, Birte Toussaint<sup>1</sup>, Lilian A. Weber<sup>1</sup>, Yu Yao<sup>1</sup> and Klaas E. Stephan<sup>1\*</sup>**

## OPEN ACCESS

### Edited by:

Henry W. Chase,  
University of Pittsburgh, United States

### Reviewed by:

Antonio Di Ieva,  
Macquarie University, Australia  
Peggy Series,  
University of Edinburgh,  
United Kingdom

### \*Correspondence:

Klaas E. Stephan  
stephan@biomed.ee.ethz.ch

### Specialty section:

This article was submitted to  
Computational Psychiatry,  
a section of the journal  
Frontiers in Psychiatry

**Received:** 15 March 2021

**Accepted:** 10 May 2021

**Published:** 02 June 2021

### Citation:

Frässle S, Aponte EA, Bollmann S, Brodersen KH, Do CT, Harrison OK, Harrison SJ, Heinzle J, Iglesias S, Kasper L, Lomakina EI, Mathys C, Müller-Schrader M, Pereira I, Petzschner FH, Raman S, Schöbi D, Toussaint B, Weber LA, Yao Y and Stephan KE (2021) TAPAS: An Open-Source Software Package for Translational Neuromodeling and Computational Psychiatry. *Front. Psychiatry* 12:680811. doi: 10.3389/fpsy.2021.680811

<sup>1</sup> Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland, <sup>2</sup> Institute for Biomedical Engineering, ETH Zurich and University of Zurich, Zurich, Switzerland, <sup>3</sup> Centre for Advanced Imaging, The University of Queensland, Brisbane, QLD, Australia, <sup>4</sup> Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, United States, <sup>5</sup> Department of Radiology, Harvard Medical School, Charlestown, MA, United States, <sup>6</sup> Department of Computer Science, ETH Zurich, Zurich, Switzerland, <sup>7</sup> Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom, <sup>8</sup> School of Pharmacy, University of Otago, Dunedin, New Zealand, <sup>9</sup> Techna Institute, University Health Network, Toronto, ON, Canada, <sup>10</sup> Interacting Minds Center, Aarhus University, Aarhus, Denmark

Psychiatry faces fundamental challenges with regard to mechanistically guided differential diagnosis, as well as prediction of clinical trajectories and treatment response of individual patients. This has motivated the genesis of two closely intertwined fields: (i) Translational Neuromodeling (TN), which develops “computational assays” for inferring patient-specific disease processes from neuroimaging, electrophysiological, and behavioral data; and (ii) Computational Psychiatry (CP), with the goal of incorporating computational assays into clinical decision making in everyday practice. In order to serve as objective and reliable tools for clinical routine, computational assays require end-to-end pipelines from raw data (input) to clinically useful information (output). While these are yet to be established in clinical practice, individual components of this general end-to-end pipeline are being developed and made openly available for community use. In this paper, we present the **T**ranslational **A**lgorithms for **P**sychiatry-**A**dvancing **S**cience (TAPAS) software package, an open-source collection of building blocks for computational assays in psychiatry. Collectively, the tools in TAPAS presently cover several important aspects of the desired end-to-end pipeline, including: (i) tailored experimental designs and optimization of measurement strategy prior to data acquisition, (ii) quality control during data acquisition, and (iii) artifact correction, statistical inference, and clinical application after data acquisition. Here, we review the different tools within TAPAS and illustrate how these may help provide a deeper understanding of neural and cognitive mechanisms of disease, with the ultimate goal of establishing automatized pipelines for predictions about

individual patients. We hope that the openly available tools in TAPAS will contribute to the further development of TN/CP and facilitate the translation of advances in computational neuroscience into clinically relevant computational assays.

**Keywords:** TAPAS, Translational Neuromodeling, Computational psychiatry, Computational psychosomatics, computational assays, open-source, software

## INTRODUCTION

Contemporary psychiatry uses disease classifications that are almost entirely based on syndromes (i.e., patterns of symptoms and signs) as defined by the Diagnostic and Statistical Manual of Mental Disorders [DSM; (1)] or the International Classification of Diseases [ICD; (2)]. While these schemes are valuable in that they provide a stratification of mental illness that relates to the subjective phenomenology of patients, they are inherently limited as they do not rest on pathophysiological or aetiological concepts of diseases. As a consequence, clinical labels proposed by DSM or ICD (e.g., schizophrenia or depression) typically have limited predictive validity with regard to clinical trajectories and do not inform the optimal treatment selection in individual patients (3, 4). Furthermore, clinical and scientific evidence suggests that these labels do not describe distinct categorical entities, but rather spectrum disorders that are characterized by substantial heterogeneity and overlap (5–7).

This has motivated novel approaches to advance our understanding of the pathophysiological and psychopathological processes underlying diseases, and to ultimately inform differential diagnosis and treatment prediction in individual patients (8). In addition to the rise of (epi)genetic approaches, advances in computational neuroscience have fueled hopes that it may become possible to establish quantitative diagnostic and prognostic computational tools that significantly improve clinical practice in psychiatry. In particular, mathematical models of neuroimaging data, as obtained using functional magnetic resonance imaging (fMRI) and electro/magnetoencephalography (EEG/MEG), hold great promise as they might offer readouts of the symptom-producing physiological processes underlying brain disorders (9–14). Similarly, advances in computational models of human behavior may enable inference on psychopathological processes at the computational (information-processing) level (15–20).

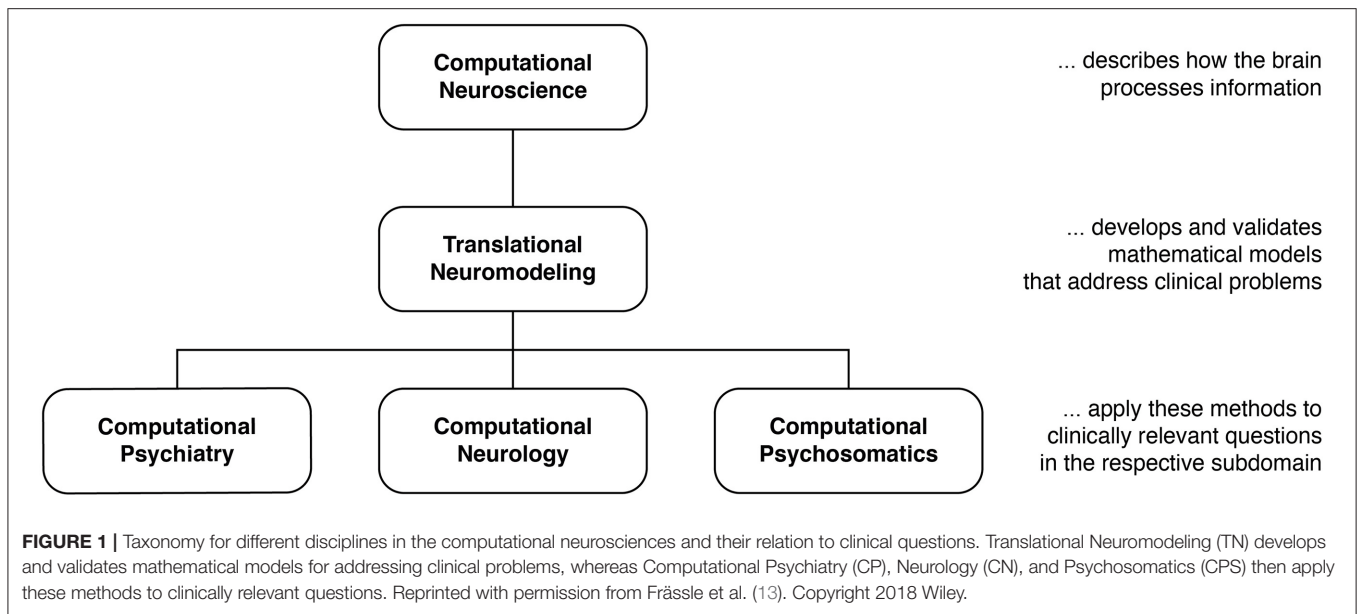
Efforts to exploit these scientific advances can be grouped into two separate yet overlapping approaches. Primarily methodological efforts toward the development of “computational assays” for inferring brain disease processes from neuroimaging, electrophysiological, and behavioral data are referred to as Translational Neuromodeling (TN); by contrast, Computational Psychiatry (CP), Computational Neurology (CN), and Computational Psychosomatics (CPS) are concerned with concrete applications in the respective clinical domains, with the ultimate goal of incorporating computational assays into routine clinical decision-making (Figure 1). Although many of the tools in TAPAS are equally useful for CN and CPS, here we focus on CP as this is arguably the most developed of the computational clinical neurosciences (16, 21–29, 114, 119).

Developments of computational assays are often based on generative models of neuroimaging or behavioral data. Generative models describe how measured data may have been caused by a particular (neuronal or cognitive) mechanism; their inversion allows computational assays to operate on inferred states of neural or cognitive systems (16, 29). This mechanistic interpretability is crucial in many clinical contexts. Additionally, traditional machine learning (ML) plays a central role in TN/CP, for example by translating the inferences from computational assays into patient-specific predictions, an approach referred to as “generative embedding” (30).<sup>1</sup>

While early TN/CP proposals date back over a decade [e.g., see Stephan et al. (31)], computational assays are yet to enter into routine clinical practice in psychiatry. In order to achieve translational success, computational assays will have to build on automatized and validated end-to-end pipelines and tools for optimal data acquisition. These pipelines need to support a complete analysis stream that takes raw data as input, and outputs clinically actionable results that are derived from inferred latent (hidden) computational quantities with pathophysiological or psychopathological relevance. Such an end-to-end pipeline will incorporate a series of fundamental steps (Figure 2): (i) Design, (ii) Conduct, (iii) Check and Correct, (iv) Preprocessing, (v) Inference, (vi) Clinical application. Each of these components poses significant challenges given the complex nature of the acquired data and of the computational tools. Here, we review the current state of development toward such an end-to-end pipeline, with a particular focus on our own work and software.

Considerable efforts have recently been made to develop standardized and user-friendly software packages that could serve as individual components for computational assays. For instance, in the context of neuroimaging data, packages like Statistical Parametric Mapping [SPM; (32)], FMRIB Software Library [FSL; (33)], or Analysis of Functional NeuroImages [AFNI; (34)] are widespread tools and cover several aspects of the aforementioned end-to-end pipeline, including preprocessing of neuroimaging data and statistical inference (typically in the framework of General Linear Models). Similarly, for behavioral data, software packages like the VBA Toolbox (35), hBayesDM (36), KFAS (37), COMPASS (38), or HDDM (39) allow inference on computational (information processing) quantities. This list is non-exhaustive and more software packages could be mentioned. While all of these packages have proven highly valuable to study behavior and brain function in humans, none of them has been

<sup>1</sup>ML is also used “on its own” in CP and applied directly to measured data, e.g., for producing patient-specific predictions or discovering structure in heterogeneous populations (222–224, 226–228). In this paper, however, we focus on approaches where ML operates on estimates provided by generative models.



designed with the specific goal of constructing a pipeline for clinically useful computational assays.

In this paper, we focus on the Translational Algorithms for Psychiatry-Advancing Science (TAPAS) software package which represents a collection of toolboxes that, collectively, aim to advance computational modeling of neuroimaging and behavioral data. While applicable to study human behavior and brain function in health, TAPAS differs from the aforementioned software packages in that its designated purpose is to provide clinically useful tools at every stage of the aforementioned end-to-end pipeline in order to advance translational success of computational approaches to psychiatry. TAPAS is primarily written in MATLAB (with some components in C and Python) and distributed as open-source code under the GNU General Public License 3.0 (<https://www.translationalneuromodeling.org/tapas>). It does not represent a single unified piece of software but rather a collection of toolboxes, each of which addresses a specific problem that arises in TN/CP approaches to neuroimaging and/or behavioral data analysis (**Figure 2**). More specifically, the development of each toolbox has been motivated by the general goal of TN/CP: to develop end-to-end pipelines that derive clinically actionable outputs from measured data, as illustrated in **Figure 2**.

In brief, TAPAS contains: (i) tailored experimental paradigms (tasks) that probe psychopathologically and/or pathophysiologically relevant processes, (ii) tools for optimization and monitoring of data quality in the specific context of fMRI, (iii) model-based physiological noise correction techniques for fMRI data, and (iv) generative models and associated statistical techniques that enable inference on latent (hidden) neurophysiological or cognitive quantities from neuroimaging or behavioral data. The latter range from network/circuit models that infer effective (directed) connectivity from fMRI and EEG/MEG data to behavioral

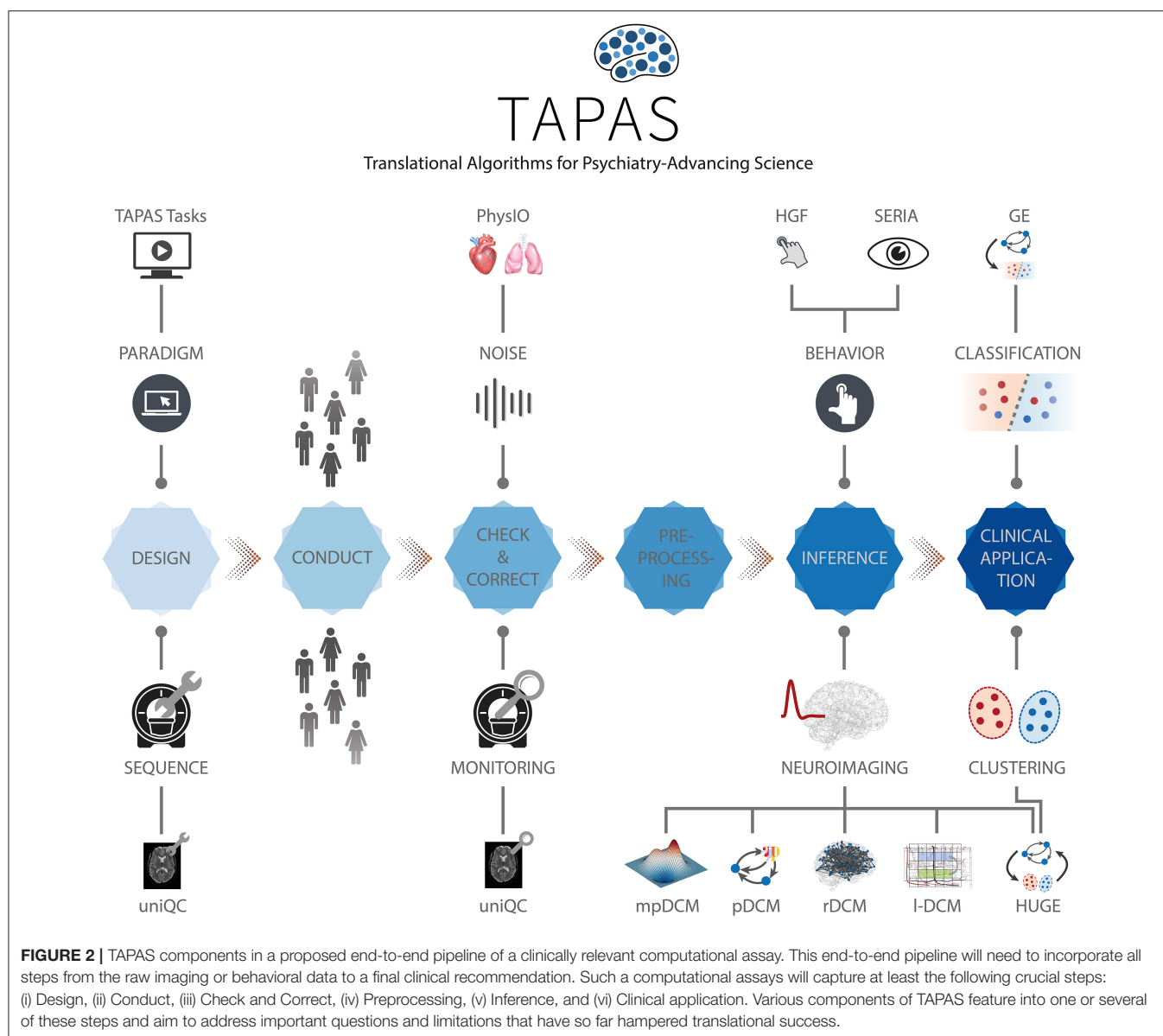
models that extract computational quantities from observed actions (e.g., decisions or eye movements). Importantly, TAPAS is not meant to provide a comprehensive collection of all tools that may potentially contribute to the development of end-to-end pipelines for computational assays. Instead, TAPAS represents a collection of toolboxes that are of particular strategic and practical relevance for advancing TN/CP and for supporting translational applications of computational approaches to problems in psychiatry.

To facilitate usability of our software, TAPAS is complemented with comprehensive documentation for each toolbox, as well as an active forum where users can seek help (<https://github.com/translationalneuromodeling/tapas/issues>).

Here, we provide a general overview of the different software toolboxes included in TAPAS and highlight how these may support the development of clinically useful computational assays for psychiatry. The paper is not meant to provide a comprehensive description of each toolbox, but instead offers a high-level perspective on how the different tools relate to each other in order to jointly advance TN/CP. For readers interested in a more in-depth treatment of a particular toolbox, references will be provided in the respective sections.

## DESIGN

The development of carefully designed experimental manipulations and the acquisition of high-quality data is paramount for (clinical) modeling. This is because any conclusion—whether a scientific or clinical one—fundamentally rests on the underlying data. The goal of tailored experimental paradigms and optimized data acquisition is to increase both the sensitivity and specificity of clinical tests; this necessitates optimizations at different stages of data collection.



(I) *Prior to data collection*: Tailored experimental paradigms have to be designed that capture relevant processes of interest. This may relate to physiological and cognitive aspects in health, or to pathophysiological and psychopathological mechanisms in disease. Furthermore, optimization of the data acquisition process is vital to ensure high quality measurements. This is particularly important in the context of fMRI data, where it is common that project-specific MR sequences have to be developed. These aspects will be covered in the current section.

(II) *During data collection*: Measures have to be taken that allow maintaining a consistently high level of data quality across acquisitions; for instance, across different patients, scanners or sites. This is vital in order to ensure that comparable (clinical) conclusions can be drawn from the data. Tools that address this

aspect of data quality control will be covered in section Conduct, Check and Correct.

(III) *After data collection*: Post-hoc assessment of data quality is important to identify datasets that need to be excluded or extra analysis steps to deal with artifacts. Poor data quality might be due to severe artifacts and/or high noise levels in the data, induced by both the MR system and the participant itself (motion, physiological noise). To identify such cases, tools are required that allow for quantitative assessment of data quality and that facilitate the decision process as to whether satisfactory data quality can be restored or not. Finally, user-friendly tools are needed that enable automatized corrections to clean-up data as best as possible. We will address these aspects in section Preprocessing.



All these efforts ideally interact seamlessly with each other in order to maximize the sensitivity and specificity of diagnostic/prognostic tests that build on the acquired data. Suboptimal data acquisition and quality control can result in a high proportion of datasets that have to be excluded from further analyses or lead to false conclusions, which is particularly problematic in the context of clinical applications.

## Harmonization of Experimental Design

For probing disease-relevant cognitive processes, a plethora of experimental tasks have been proposed that frequently only differ in small details. While some variations are valuable as they address somewhat different aspects of a cognitive process, this diversity complicates exact comparison of findings across studies and often gives rise to “approximate replications” where an initial finding is not replicated exactly but some (vaguely) related finding is linked to the original observation (4, 40). A prominent example of this in the context of clinical neuroimaging is the frontal dysfunction hypothesis in schizophrenia. Since the original report (41), several studies have re-examined this question using somewhat different experimental approaches and have reported a variety of different outcomes, ranging from hyper- to hypoactivation, to no obvious alterations at all [cf. (4, 42)]. While it is possible that these differences could be of pathophysiological relevance, potentially referring to different subgroups in the schizophrenic spectrum, inconsistencies in the utilized experimental manipulations render any differences difficult to interpret. Hence, until such variations are properly explained or controlled for, approximate replications do not provide a solid basis for clinical tests.

One way to address this challenge is by openly sharing established experimental tasks. Notably, while the call for open sharing of data has been very prominent in recent years (43–45), this is less the case for sharing the experimental tasks themselves (but see, for instance, the task protocols utilized by the Human Connectome Project (46) which have been shared openly).

To this end, TAPAS comprises the module “TAPAS Tasks” which represents a collection of experimental paradigms that have been designed and thoroughly tested (**Figure 3, top left**). TAPAS Tasks comprises several tasks that cover both the exteroceptive and interoceptive<sup>2</sup> domains (for a complete list, see **Table 1**). For all paradigms, the stimulus code is provided as well as detailed documentation. This includes a comprehensive description of (i) the experimental task, (ii) software requirements, (iii) experimental set-up (including a list of necessary peripheral devices), and (iv) information on how to run the task.

Here, as an example, we describe the *Heartbeat Attention* (*HbAttention*) task in more detail (**Figure 3, bottom left**). The task implements a novel paradigm to probe purely attentional differences of the heartbeat evoked potential between exteroceptive and interoceptive conditions (47). The paradigm consists of alternating conditions where participants are

asked to focus attention either on their heart or on a simultaneously presented auditory stimulus (white noise). Importantly, in both conditions the sensory stimulation is identical. Using this paradigm, Petzschner et al. (47) found an increased heartbeat evoked potential during interoceptive compared with exteroceptive attention. A non-invasive readout of the attentional modulation of interoceptive processes could potentially be of high clinical relevance, since alterations in interoceptive processing have been recognized as a major component of various psychiatric conditions, including mood and anxiety disorders, eating disorders, drug addiction, as well as depression (48).

TAPAS Tasks represents work in progress and not all experimental paradigms listed above are available yet (for more details, see **Table 1**). Newly devised experimental paradigms will be added in the future. We hope that by making these tasks openly available to the community, TAPAS Tasks may contribute to growing a collection of standardized experimental paradigms in TN/CP.

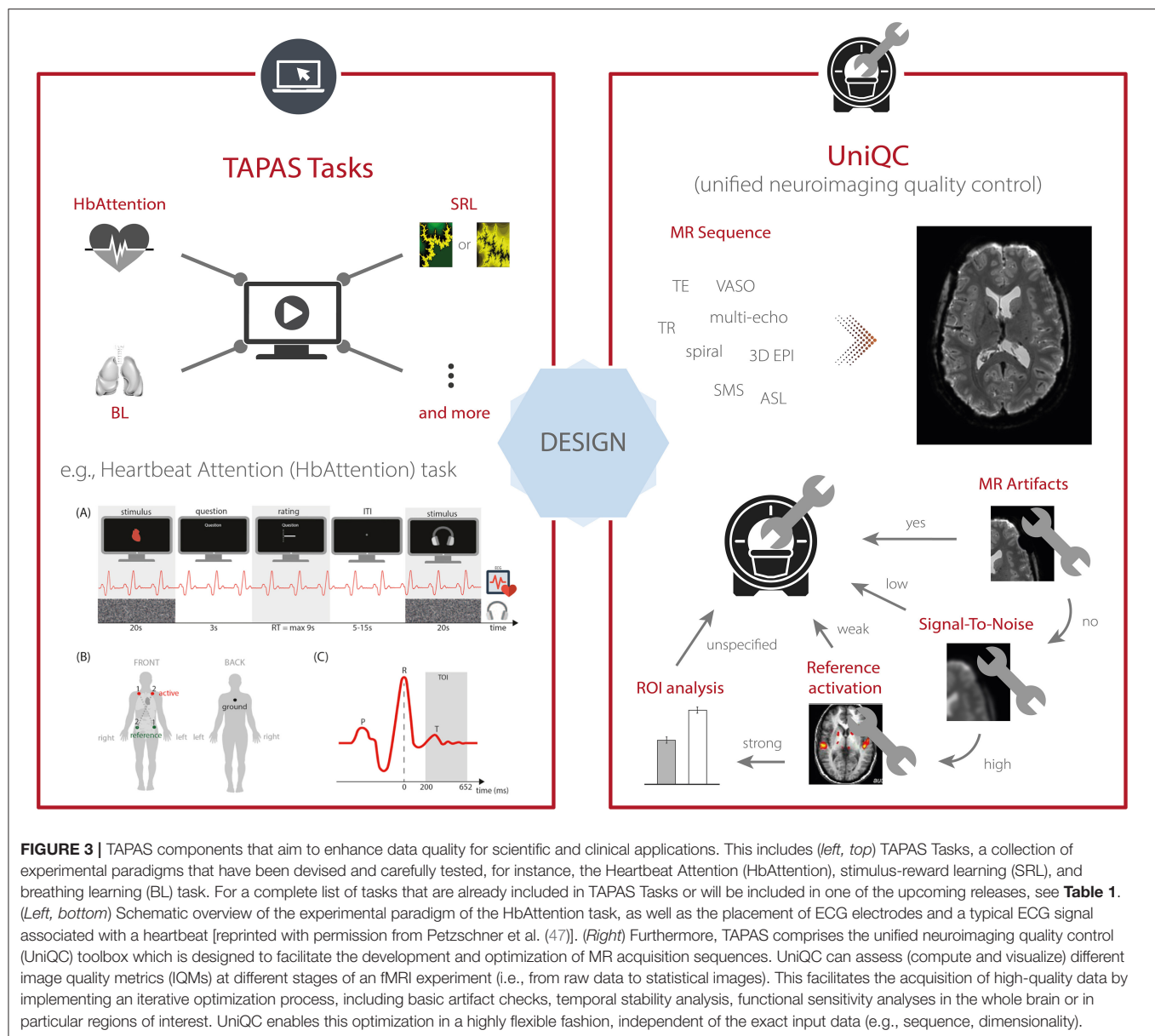
## Optimization of MRI Protocols

In the context of neuroimaging, data quality also depends heavily on the MR scanner settings and acquisition sequence. Carefully crafted sequences with optimized parameter choices can offer considerable gains in functional sensitivity and specificity for the targeted research question (49–51). However, due to the large variety of available parameters and their interdependency, optimization of MR protocols is challenging and suboptimal acquisition choices might reduce data quality, e.g., low signal-to-noise ratio or pronounced artifacts like ghosting, ringing, signal dropouts and distortions due to magnetic field inhomogeneities [for an overview, see (52–55)]. Hence, the development, optimization and validation of robust and powerful MR protocols prior to data acquisition is critical and tools are needed that ease this process.

To this end, TAPAS includes the *unified neuroimaging quality control* (UniQC) toolbox (56), which provides a framework for flexible, interactive and user-friendly computation and visualization of various quality measures in neuroimaging data (**Figure 3, right**). UniQC facilitates fast prototyping and optimization of acquisition sequences by providing tools for artifact detection and sensitivity analyses across the entire image or tailored toward specific regions of interest.

As sequence development constitutes an iterative process, feedback on image quality has to be immediate and specific to the protocol changes, so that the performed quality control (QC) query informs the operator on how to adjust parameters for the next scan (**Figure 3, right**). For example, if an unexpected bias field occurs in the mean image, both excitation and receiver channels could be compromised. In this case, fast display of individual coil images is critical, which would be omitted if the mean image were inconspicuous. Thus, QC during sequence development resembles a decision tree, where the outcome of one image quality metric (IQM) determines the selection of the next one, with varying display options. This necessitates an interactive, fast and flexible way to compute and visualize IQMs. Typically, such a decision tree starts from basic artifact

<sup>2</sup>Exteroception refers to perception of sensations originating from the external world, whereas interoception refers to perception of sensations originating from the own body or “internal world”.



**FIGURE 3 |** TAPAS components that aim to enhance data quality for scientific and clinical applications. This includes (*left, top*) TAPAS Tasks, a collection of experimental paradigms that have been devised and carefully tested, for instance, the Heartbeat Attention (HbAttention), stimulus-reward learning (SRL), and breathing learning (BL) task. For a complete list of tasks that are already included in TAPAS Tasks or will be included in one of the upcoming releases, see **Table 1**. (*Left, bottom*) Schematic overview of the experimental paradigm of the HbAttention task, as well as the placement of ECG electrodes and a typical ECG signal associated with a heartbeat [reprinted with permission from Petzschner et al. (47)]. (*Right*) Furthermore, TAPAS comprises the unified neuroimaging quality control (UniQC) toolbox which is designed to facilitate the development and optimization of MR acquisition sequences. UniQC can assess (compute and visualize) different image quality metrics (IQMs) at different stages of an fMRI experiment (i.e., from raw data to statistical images). This facilitates the acquisition of high-quality data by implementing an iterative optimization process, including basic artifact checks, temporal stability analysis, functional sensitivity analyses in the whole brain or in particular regions of interest. UniQC enables this optimization in a highly flexible fashion, independent of the exact input data (e.g., sequence, dimensionality).

checks over temporal stability all the way to functional sensitivity analyses in particular regions of interest, with the occasional return to the scanner, if a QC step fails (**Figure 3, right**).

To achieve this functionality, UniQC exploits the object-oriented programming capabilities in MATLAB. Importantly, UniQC is not restricted to 4-dimensional neuroimaging data (i.e., space and time) like most other software packages, but generalizes operations to  $n$ -dimensional data. This generalization to arbitrary numbers of dimensions comes in useful when handling data from multiple receiver coils (57), as well as multi-echo (58, 59) or combined magnitude/phase fMRI data (60–62) in a single unified framework. Similarly, prominent non-BOLD fMRI contrasts rely on an additional tag/control dimension, for instance, Vascular Space Occupancy [VASO;

(63, 64)] or Arterial Spin Labeling [ASL; (65, 66)], with the former being particularly relevant for depth-dependent fMRI. Furthermore, UniQC allows seamless integration with SPM and other MATLAB toolboxes in order to benefit from image and fMRI processing algorithms that are already available in those packages.

In summary, UniQC provides a flexible, interactive and user-friendly toolbox for evaluating MR pulse sequence development and quality control of  $n$ -dimensional neuroimaging data. These efforts carry over from the design stage to the data collection, in that UniQC allows utilizing the processing and visualization pipeline established here directly as a quality control protocol—enabling unique QC toward the aims of each study.

**TABLE 1** | List of tasks (to be) included in TAPAS Tasks.

Task	Description
Heartbeat attention ( <i>HbAttention</i> )	The HbAttention task probes differences in neural responses to heartbeats due to changes in attentional focus. The paradigm consists of alternating conditions where participants focus attention either on their heart (interoceptive condition) or on an external sound stimulus (exteroceptive condition), while keeping the sensory stimulation identical (47). The HbAttention task requires the acquisition of an ECG and can be run in the context of an EEG experiment, measuring a Heartbeat Evoked Potential (HEP), or an fMRI experiment. The task is programmed such that the stimulus timing can be easily adjusted to the experimental modality.
Heartbeat feedback ( <i>HbFeedback</i> )	The HbFeedback task presents auditory-visual stimuli that are either locked to an individual's online detected heartbeat (veridical feedback about the heartbeat) or presented at a rate that is faster or slower than the individual's heartrate. The task assesses the effects of veridical vs. false feedback on physiological and neural signals related to heartbeats. It requires the simultaneous recording of EEG and ECG signals.
Heartbeat mismatch ( <i>HbMMN</i> )	The HbMMN task consists of an auditory omission paradigm where a "standard" tone is presented shortly after each heartbeat, but occasionally omitted ("deviant"). In different conditions, the delay between heartbeat and tone is varied. This allows to measure changes in stimulus-evoked and heartbeat-evoked potentials between standards and omissions. In a control condition, tone presentation times are unrelated to heartbeats. EEG and ECG signals are simultaneously recorded during the task.
Filter detection ( <i>FD</i> )	The FD task is a perceptual threshold breathing task where participants have to indicate on each trial whether a very small resistance (i.e., filter) or sham (i.e., empty filter) was applied to the breathing system (yes/no version) (213), or in which interval resistance was applied (two-interval forced choice version). The task is tailored to assessing respiratory interoceptive accuracy and metacognition in individual participants. Behavioral responses are recorded.
Breathing learning ( <i>BL</i> )	The BL task represents an associative learning task where participants learn the association between visual cues and the subsequent presence/absence of an inspiratory resistive load. Respiratory load is applied using a novel MRI-compatible breathing system that allows for remote administration and monitoring of resistive loads and whose construction plan has been published (214). fMRI signals are recorded during the task.
Stimulus-reward learning ( <i>SRL</i> )	The SRL task requires participants to predict which of two simultaneously presented visual stimuli (i.e., fractals) would yield a monetary reward. The association strengths between the visual cues and monetary outcomes change over the course of the experiment, introducing volatility. fMRI or EEG data can be acquired during the task.
Auditory mismatch negativity ( <i>aMMN</i> )	The aMMN task is a variant of the auditory oddball paradigm in which the degree of volatility in the auditory stream varies over time. While engaging in a visual distraction task, participants passively listen to repeated presentations of a high and a low tone. During stable phases of the experiment, one stimulus reliably serves as the "standard" (more frequent) tone and the other one as the "deviant" tone. During volatile phases, the roles of standard and deviant switch more rapidly. Deviance processing can be compared between different levels of stability/volatility. Task versions are available for both EEG or fMRI recordings.
Visual mismatch negativity ( <i>vMMN</i> )	The vMMN task implements the identical probabilistic stimulus sequence as the aMMN task. However, instead of auditory stimuli, Gabor patches of different orientations are used to probe mismatch responses in the visual domain. Task versions are available for both EEG and fMRI recordings.
Antisaccades ( <i>AS</i> )	The AS task asks participants to perform antisaccades which are a type of voluntarily controlled eye movements. In TAPAS, code is available to run two different versions of the AS task (215, 216) using an EyeLink (SR Research, Ottawa, ON, Canada) eye tracking system. The versions of the task differ in the timing (with or without a delay before the eye movement) and position (at center or at peripheral target position) of the presentation of the task cue.

TAPAS Tasks will include a variety of different paradigms that probe exteroceptive and/or interoceptive processes. So far, the Filter detection (FD) and Breathing learning (BL) tasks are included; the other tasks will follow as soon as the respective papers are published.

## CONDUCT, CHECK AND CORRECT

Besides optimization steps prior to data acquisition, further steps are necessary during and after the measurement to ensure adequate data quality. Specifically, ongoing monitoring of data quality during image acquisition is critical, because both the MR system and the study participant constitute significant potential noise sources. On the system side, data quality across different time points and scanning sites may vary due to potential malfunctions or alterations in the scanner hardware, that must be detected in a timely manner. On the participant's side, even under ideal circumstances, with tailored experimental designs, optimized acquisition sequences and thorough quality control measures, fMRI data is still subject to artifacts outside

the experimenter's control [e.g., motion, physiology; (67, 68)]. Adequately correcting for these artifacts is essential to avoid bias in subsequent data analyses and to ensure that conclusions are not confounded [e.g., (69–71)]. In what follows, we elaborate on these points and describe tools in TAPAS that address these challenges.

## Quality Monitoring

fMRI analyses rely on the content of brain (or spinal cord) images as information source. Thus, the visual inspection of raw images by one or multiple experts is still often considered a gold standard for quality control. However, visual assessment depends on individual rater experience and visualization choices (e.g., slice orientation, windowing) which may reduce the apparent

information content of an image to detect artifacts or improper acquisition parameters (72) and generally aggravates inter-rater reliability. Furthermore, limited time resources and fatigue make the naïve visual inspection of every raw image quickly unfeasible, as even a single fMRI dataset contains hundreds of volumes with dozens of slices each. This challenge is exacerbated for large-scale datasets, like the Human Connectome Project [HCP; (46)] or UK Biobank (73), where thousands of participants are measured.

Automation of quality control is therefore required, and can in principle target both aspects of the manual image classification by raters. A first approach is to replace raters by a machine learning algorithm working on derived IQMs to reduce the high-dimensional feature space of image time series. A second option is that the expertise of the rater can be harnessed more efficiently by providing flexible tools for image manipulation and visualization with intuitive interfaces to derive and inspect relevant IQMs, thereby reducing operator fatigue and inconsistency.

The first approach was explored early on for anatomical T1-weighted images (74) and demonstrated good discriminability of undistorted, noisy, and distorted images based on a subset of 239 IQMs. Since then, various additional tools for automatic quality control have been proposed, introducing additional IQMs to assess image quality (75–77). These methods have been refined and extended into scalable QC frameworks for large-scale fMRI studies, most notably in the form of the MRI Quality Control tool [MRIQC; (78)] and within the UK biobank study (79). Their key advancement lies in the ability to classify images in a binary fashion (“good” vs. “problematic”) or even categorizing multiple artifact classes (79). Standardized quality reports then provide guidance, as to whether data from a given participant should be included in subsequent analyses or not. In order to ensure accurate classification, these algorithms are typically trained on large curated datasets that include both patients and healthy controls [e.g., ABIDE (80) and DS030 (81) for MRIQC].

While such fully automatized approaches with minimal visual output and manual assessment might be the only viable solution in studies with thousands of participants, the required high degree of standardization of the acquisition protocol as well as the need for large, representative training datasets poses limitations on its utility. In TN/CP, less well-studied clinical populations and novel technologies can pose challenges when trying to exploit established mappings between IQMs and image quality. In particular, this can occur when employing advanced imaging hardware or acquisition sequences to maximize sensitivity for individual subject measurements, IQMs may fall outside the standard range or become inapplicable. For example, higher magnetic field strengths and customized high-density or surface coils (82, 83) induce atypical image intensity variations (bias fields). Similarly, Nyquist ghosts manifest differently for spiral readouts than in conventional Cartesian echo-planar imaging. Thus, fMRI data may require different IQMs or thresholds when deviating from standard acquisition choices.

In these domains, the second option for QC automation is preferable. This approach empowers the rater to determine which IQMs to inspect, how to visualize them, and at which stage of the analysis stream to assess them. While parts of this approach have

been implemented by providing standardized visual reports (e.g., MRIQC, fMRWhy) or interactive QC visualization tools (e.g., visualQC), a comprehensive framework that integrates all these functionalities has been missing. To address this, we designed the UniQC toolbox to meet these demands by offering flexible, interactive and user-friendly assessment of fMRI data (**Figure 4, left**). Importantly, the quality control pipeline derived during the sequence design stage (section Design) can be readily deployed to cover QC automation, and once quality issues are detected, UniQC also provides a framework to “interrogate” the data efficiently and identify potential causes of the problem.

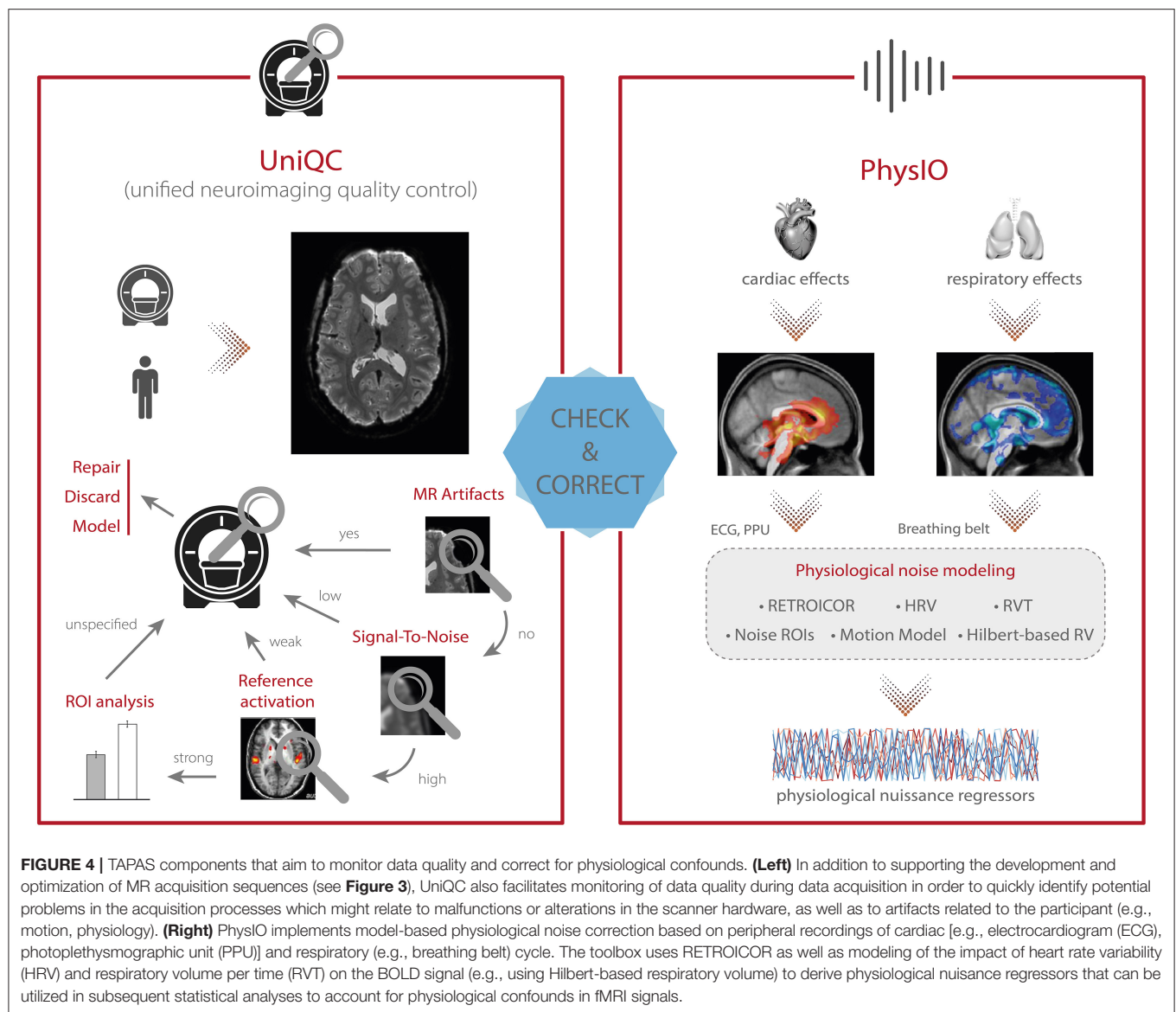
In principle, this fast and precise identification can lead to quality improvements in three ways (**Figure 4, left**). First, the information can be used to isolate and repair hardware malfunction (e.g., of certain coil elements) to swiftly restore quality levels for the next scan or participant. Second, the quality of the affected dataset can be increased by modeling the impact of distinct noise sources, as identified by the QC decision tree [e.g., electrostatic spike artifacts in the images or interactions between subject motion and magnetic field; (84–86)]. Third, selectively discarding the low-quality data only, as isolated by the customized QC interrogation, salvages quality levels for the remainder of the data (**Figure 4, left**).

Furthermore, unlike generic QC pipelines, the customization afforded by UniQC facilitates testing whether any given dataset shows a functional response relevant for the research question. For example, due to the seamless integration with other MATLAB toolboxes such as SPM, UniQC can analyze statistical maps from study-specific GLMs, as well as provide region-of-interest (ROI) statistics. With this functionality, task-fMRI performance can be validated using robust expected activation patterns as a sanity check, for example, sensory-motor activation during a learning task, before proceeding to more complex analyses. Thus, UniQC offers flexible, interactive and study-specific quality control of the image acquisition system and the imaged participant. Thanks to its modularity, UniQC can be further integrated to monitor quality throughout the preprocessing stage (section Preprocessing), independent of the concrete pipeline.

## Physiological Noise Modeling

One of the main confounds in fMRI is physiological noise as it perturbs blood oxygen level dependent (BOLD) signals (68, 87)—which can substantially hamper both classical fMRI analysis as well as computational modeling of the data. The two primary sources of physiological noise are the cardiac and respiratory cycles (88). The respiratory cycle introduces confounds by distortions of the magnetic field due to the movement of the participant’s chest (89) as well as bulk susceptibility variation in the lungs (90). Additionally, the respiratory cycle alters the pressure of blood CO<sub>2</sub> (which is a vasodilator) over longer time periods, thereby inducing slow signal fluctuations (91). The cardiac cycle, on the other hand, modulates blood volume and vessel diameter during systole and diastole, leading to small deformations of brain tissue and brainstem displacement, causing periodic motion of the cerebrospinal fluid (92). Furthermore, variability in heart rate





induces alterations in the oxygen level in the blood (93), and consequent low frequency signal fluctuations (94). Finally, interactions between the cardiac and respiratory cycles, such as in respiratory sinus arrhythmia (i.e., accelerated heartbeat during inhalation), induce additional non-trivial physiological fluctuations (95).

Various physiological noise correction methods for fMRI exist, either based solely on the fMRI time series and prior assumptions of spatiotemporal noise properties, or modeling the noise from independent physiological recordings (e.g., using electrocardiogram (ECG), photoplethysmographic unit (PPU), and breathing belts) (88, 96). Arguably, for TN/CP applications with clinical populations and pharmacological interventions, methods based on independent recordings might be preferable: In clinical populations, physiological processes that impact on the

BOLD signal may differ from priors that were defined based on the general population.

Several freely available implementations for model-based physiological noise correction are available, including AFNI 3DRETROICOR (67), FSL Physiological Noise Modeling (97), and PhLEM (98); however, relatively few studies have capitalized on these tools, in particular for task-based fMRI. One important practical challenge—that is exacerbated in clinical populations with less compliant subjects—is the variable data quality of the peripheral recordings which these models are based on. Reduced data quality may be due to subject motion or (partial) detachment/saturation of the peripheral devices. In most implementations, preprocessing of these recordings is minimal, and identification of the physiological cycles typically relies on peak detection provided by the MR scanner vendor



(96). Alternatively, manual intervention to correct erroneous detection is offered, hampering the development of automatized pipelines and the translation of physiological noise modeling into routine application.

The *PhysIO* toolbox (96) in TAPAS offers methods for model-based physiological noise correction based on peripheral recordings of the cardiac (e.g., ECG, PPU) and respiratory cycles (e.g., breathing belt). *PhysIO* utilizes these peripheral measures to model the periodic effects of pulsatile motion and field fluctuations using RETROICOR (67). Furthermore, the toolbox accounts for end-tidal CO<sub>2</sub> changes and heart rate-dependent blood oxygenation by convolving respiratory volume per time (RVT) and heart rate variability (HRV) with a respiratory and cardiac response function, respectively (91, 99, 218). For these methods, emphasis is placed on robust preprocessing of the input time series via reliable peak detection in low-SNR regimes, as well as a novel method for RVT estimation using the Hilbert transform (100). As a more data-driven alternative for noise correction, *PhysIO* also allows the extraction of signals from pre-defined regions of interest (“noise ROIs”) as additional confound regressors—for instance, signal related to white matter or cerebrospinal fluid (CSF). Finally, the toolbox also incorporates various strategies for correcting motion-related artifacts by implementing, for instance, the Volterra expansion confound set (101) or censoring strategies based on the framewise displacement (69). A schematic illustration of the modeling process in *PhysIO* is provided in **Figure 4, right**. *PhysIO* provides both command-line operation for de-noising multiple subjects conveniently, as well as a user-friendly graphical interface within the SPM Batch Editor. Thereby, physiological noise correction can be integrated with complete fMRI preprocessing pipelines, minimizing the need for manual interventions or custom programming (96). Additionally, *PhysIO* ensures robust preprocessing even for low-quality data and provides simple diagnostic tools to assess the correction efficacy in individual subjects. This renders *PhysIO* an accessible noise correction tool for preprocessing pipelines both in basic neuroscience studies as well as for clinical purposes.

## PREPROCESSING

Data preprocessing is closely intertwined with quality control and artifact correction. Thorough preprocessing is particularly important for complex data acquired using neuroimaging techniques such as fMRI or EEG/MEG (102). To this purpose, researchers typically create *ad hoc* preprocessing workflows for each study individually (103), building upon a large inventory of available tools. Broadly speaking, preprocessing steps can be separated into two main categories: (i) preprocessed time series are derived from the original data after application of retrospective signal corrections, spatiotemporal filtering, and resampling in a target space (e.g., MNI standard space), and (ii) confound-related information in the data can be modeled or taken into account through nuisance regressors (i.e., regressors of no interest) in subsequent statistical analyses using the general linear model (GLM). These confounds may include motion

parameters, framewise displacement, physiological (cardiac or respiratory) signals, or global signals (88, 104).

These and additional procedures for preprocessing neuroimaging data are available in various software packages including, SPM (32), FSL (105), FreeSurfer (106), AFNI (34), or Nilearn (107). The plethora of different preprocessing tools and workflows manifests in the absence of a current gold standard for preprocessing neuroimaging data, despite several attempts to establish best-practice guidelines (102, 108–110).

In an attempt toward common preprocessing standards, large-scale consortia like the HCP or the UK Biobank provide access not only to the raw data, but also to already preprocessed versions of the data. For instance, in the HCP database, researchers have access to the version of the data which have been subjected to a common minimal preprocessing pipeline (111). However, these workflows are usually tailored toward the particular dataset's idiosyncrasies and do not readily translate to other datasets. A first attempt toward such a universally applicable preprocessing workflow is fMRIPrep (112), which represents a pipeline that combines tools from several of the above-mentioned software packages. fMRIPrep autonomously adapts the workflow to the present data, rendering the approach robust to data idiosyncrasies and potentially applicable to any dataset without manual intervention.

While no toolbox dedicated to data preprocessing is currently available in TAPAS, our tools from the previous step (i.e., “Conduct, Check and Correct”) integrate well with most of the third-party software packages highlighted above. For instance, *PhysIO* integrates seamlessly with the batch editor system of SPM to facilitate the derivation of nuisance regressors related to physiological confounds that can be utilized in GLM analyses. Similarly, UniQC is designed to integrate with SPM and other MATLAB toolboxes. Importantly, *PhysIO* and UniQC are also designed to integrate well with other neuroimaging software packages. For instance, *PhysIO* stores all physiological noise regressors in a dedicated text file which can be inputted into the first-level analyses in any software package (e.g., FSL, AFNI).

## INFERENCE

Once neuroimaging and/or behavioral data have been preprocessed and artifacts have been corrected, the question arises how best to interrogate the data in order to gain insights into the functioning of the human brain and alterations thereof in disease. Concerning clinically oriented studies, it has been pointed out (4, 25) that focusing on differences in descriptive measures—such as BOLD activation, functional connectivity patterns or task performance—between patients and healthy controls is unlikely to result in improvements of clinical practice. This is because these analyses do not provide an understanding of the symptom-producing mechanisms and they do not easily inform the development of biologically grounded clinical tests. Consequently, these measures have not yet led to routine applications in clinical practice (4, 113).

To address this shortcoming, mathematical models of neuroimaging and behavioral data that capture putative

physiological and cognitive disease mechanisms may represent a promising avenue. This line of thinking is at the core of clinically-oriented modeling disciplines like Computational Psychiatry (15, 16, 22–24, 27), Computational Neurology (15, 114), and Computational Psychosomatics (115). For example, in Computational Psychiatry, a major goal is to move from the current syndromatic nosology to disease classifications based on computational assays that may improve differential diagnosis and treatment prediction for individual patients (7, 116).

TAPAS contributes to this endeavor by providing a collection of computational tools that can be applied to neuroimaging (fMRI) or behavioral data (decisions, eye movements). All of these approaches are so-called generative models (117). Generative models specify the joint probability  $p(y, \theta|m)$  over measured data  $y$  and model parameters  $\theta$ , which—according to probability theory—can be written as the product of the likelihood function  $p(y|\theta, m)$ , representing the probability of the data given a set of model parameters, and the prior distribution  $p(\theta|m)$ , encoding the *a priori* plausible range of parameter values. Together, likelihood and prior yield a probabilistic forward mapping from latent (hidden) states of a system (e.g., neuronal dynamics) to observable measurements (e.g., BOLD signal). Model inversion enables inference on the parameters and latent states of the system from measured data, and can be accomplished using a variety of approximate Bayesian techniques (e.g., variational Bayes, Markov chain Monte Carlo or Gaussian process optimization). Their ability to reveal latent mechanisms underneath the visible data and their natural connection to hypothesis testing through model selection procedures (118) have established generative modeling as a cornerstone of TN/CP (13, 119).

In brief, the generative models included in TAPAS comprise: (i) models of effective (directed) connectivity among neuronal populations, (ii) models of perception in the light of an uncertain and volatile environment, as well as (iii) models of inhibitory control. In what follows, we briefly describe the different generative models in TAPAS and highlight how each of them might be useful for clinical (neuro)modeling.

## Generative Models of Neuroimaging Data

Models of effective connectivity describe the mechanisms by which neuronal populations interact and how these mechanisms give rise to measured data (e.g., fMRI or EEG/MEG). By inverting these generative models, it is possible, in principle, to infer on the directed (synaptic) influences neuronal population exert on one another (120, 121). This differs from measures of functional connectivity (e.g., Pearson's correlation) which are essentially descriptive and undirected statistical indices. Models of effective connectivity hold particular promise for TN/CP, since global dysconnectivity has been proposed as a hallmark of various mental disorders (10, 122), including schizophrenia (31, 123–125, 164), autism (126–128), and depression (129–131).

A frequently used generative modeling framework for inferring effective connectivity from neuroimaging data is dynamic causal modeling [DCM; (132)]. In brief, DCM describes changes in neuronal activity as a function of the directed interactions among neuronal populations and

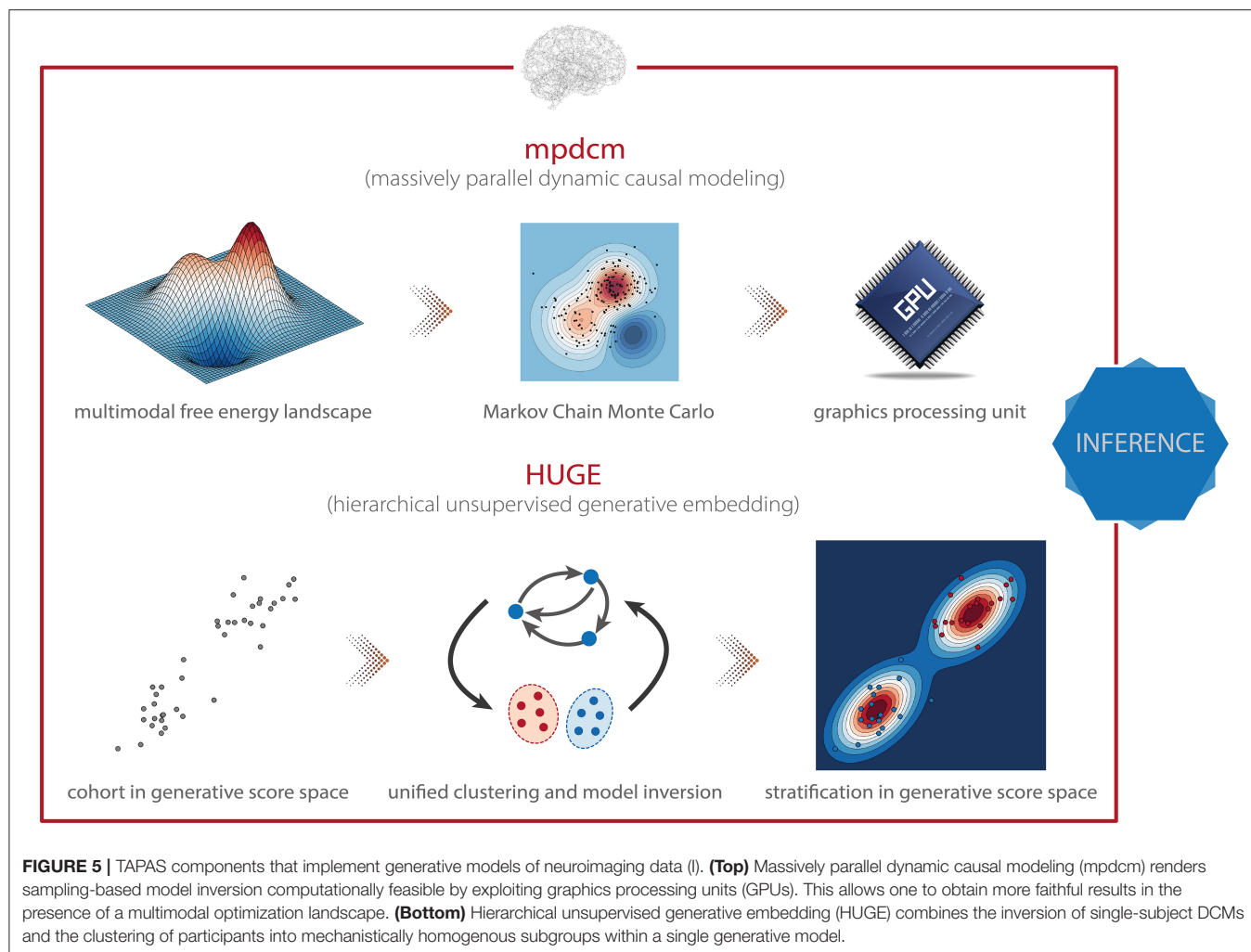
experimental manipulations that can perturb the system. DCM was initially introduced for fMRI (132) and later extended to electrophysiological data (133). Comprehensive reviews on DCM can be found elsewhere [e.g., (121, 134–136)]. DCM is freely available as part of SPM and has found widespread application. For example, with regard to clinical applications, DCM has been used to study schizophrenia (137–141), autism (142, 143), and depression (144, 145). Despite these promises, the classical DCM approach is also subject to several limitations—which may become particularly relevant in the context of TN/CP, where the goal is to develop computational assays that inform prediction of clinical trajectories and treatment responses in individual patients. In what follows, we highlight some of these limitations and outline how the methodological advances in DCM included in TAPAS aim to address these challenges.

## Global Optimization

When translating computational advances like DCM into computational assays, the robustness of the inference procedure and the reliability of the parameter estimates become paramount (146). Standard model inversion in DCM rests on variational Bayes under the Laplace approximation [VBL; (147)] which is computationally efficient, yet subject to several limitations (134): First, VBL rests on maximizing the negative free energy (which serves as a lower bound approximation to the log model evidence) using gradient ascent and is thus inherently susceptible to local maxima if the objective function is multimodal. Second, even when the global maximum is found, the distributional assumptions (i.e., Laplace and mean-field approximations) might not be justified, potentially rendering the approximate posterior distribution a poor representation of the true posterior. Third, when the distributional assumptions of the Laplace approximation are violated, the negative free energy is no longer guaranteed to represent a lower bound on the log model evidence (148).

Sampling-based model inversion schemes, typically based on Markov chain Monte Carlo (MCMC) methods, do not require any distributional assumptions about the posterior and are guaranteed to be asymptotically exact (i.e., converge to the global extremum in the limit of infinite samples). This renders sampling-based methods an appealing alternative to VBL. However, they come at the cost of other challenges: First, sampling-based routines are computationally expensive. Second, convergence is only guaranteed in the limit of infinite samples; detecting convergence in practice thus rests on heuristics. Third, unlike VBL, sampling-based methods do not readily provide an estimate of the (log) model evidence, but require additional strategies, which further aggravate the computational burden. For instance, the current gold standard for sampling-based estimates of the model evidence, thermodynamic integration [TI; (149–151)], requires running multiple MCMC chains at different “temperatures” (i.e., at different positions along a path from prior to posterior). Until recently, these reasons have been prohibitive for the use of sampling-based model inversion for DCMs.

The *massively parallel dynamic causal modeling* (mpdcm) toolbox (152) implemented in TAPAS renders sampling-based model inversion in the context of DCM for fMRI



computationally feasible (Figure 5, top). This is achieved by exploiting the power of graphics processing units (GPUs) for the evaluation of the likelihood function, which represents the computationally most expensive operation, as it requires integration of differential equations in the neuronal and hemodynamic models. Importantly, mpdcm even makes the evaluation of the model evidence via thermodynamic integration computationally feasible. In a recent preprint, Aponte et al. (217) demonstrated that TI provides more accurate and robust estimates of the model evidence than VBL, while computational demands are kept at a moderate level.

Beyond the mpdcm toolbox, which is designed to support DCM for fMRI, other gradient-free and gradient-based MCMC sampling schemes have also been introduced to DCM for electrophysiological data (153, 229). However, these tools have not yet been made publicly available.

### Empirical Bayes for DCM

Another challenge concerns the specification of prior distributions in DCM, which have been found to profoundly

impact the posterior estimates and their reliability (154). Notably, in the context of hierarchical Bayesian models, there is a principled way of estimating priors by exploiting measurements from multiple subjects: empirical Bayes [EB; (155–157)]. In brief, in EB, the posterior density at any given level is constrained by the level above. For instance, in a two-level hierarchical model, observed data  $y = \{y_1, y_2, \dots, y_n\}$  are assumed to be generated from a set of latent (hidden) parameters  $\theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  according to the likelihood  $p(y|\theta, m)$ . In turn, the parameters  $\theta$  are considered to represent samples from a population density  $p(\theta|\eta, m)$ , where  $\eta$  refers to the hyperparameters (117, 158). Consequently, under the hierarchical structure of a multi-subject or mixed-effects model, inference on the single-subject level is constrained by the group-level information. These constraints are then referred to as empirical priors since they are informed by the empirical data (of the entire group). A special case of EB is referred to as parametric empirical Bayes (PEB), where the hyperparameters  $\eta$  are approximated using the maximum likelihood estimate or a moment expansion, which allows one to express the hyperparameters in terms of the empirical mean and

variance (159). One particular variant of PEB is the Gaussian-Gaussian model, where single-subject data are assumed to be generated by adding Gaussian noise to the group mean (160).

The *hierarchical unsupervised generative embedding* (HUGE) toolbox contained in TAPAS implements EB in the context of DCM for fMRI (161, 162). HUGE combines the inversion of single-subject DCMs and the clustering of subjects into mechanistically homogenous subgroups into a single generative model (**Figure 5, bottom**). This is achieved by combining the non-linear DCMs at the individual level with a mixture-of-Gaussians clustering model at the hierarchically higher level (117, 158). In other words, HUGE assumes that each individual from a population of  $N$  subjects belongs to one of at most  $K$  subgroups or clusters. The DCM parameters  $\theta$  for all subjects from one cluster  $k$  are then assumed to be normally distributed with distinct mean  $\mu_k$  and covariance matrix  $\Sigma_k$ . This cluster-specific normal distribution effectively means that different prior distributions apply over subjects, depending on which subgroup they belong to, and that these priors are learned from the data (i.e., subgroup-specific EB). Hence, in principle, the framework is capable of stratifying heterogeneous spectrum disorders, as defined by DSM/ICD, into subgroups that share common pathophysiological mechanisms (for more details, see below). Importantly, HUGE also implements “pure” EB by fixing the number of clusters to one and merely exploiting the hierarchical dependencies in the data. This effectively switches off the clustering model. The utility of this mode of operation has been demonstrated in simulations by Yao et al. (162), highlighting the expected shrinkage effect (reduced variability) of the posterior parameter estimates toward the population mean observed in EB (117). Parameter estimation in HUGE can be performed by employing either a sampling-based MCMC inversion scheme (161, 163), which is asymptotically exact yet (relatively) slow, or a VB implementation (162), which is computationally more efficient yet might be vulnerable to local extrema. Notably, at the moment, only the VB implementation of HUGE is available in TAPAS. The sampling-based variant will be published as part of an upcoming release of the toolbox.

### Whole-Brain Effective Connectivity Analysis

Apart from the computational and statistical challenges mentioned above, a conceptual concern is that DCMs are typically restricted to relatively small networks in order to keep model inversion computationally feasible. While this may be advantageous in some cases by enforcing a theory-driven analysis of high-dimensional and noisy fMRI data, it can also represent a limiting factor. Specifically, many cognitive processes, as well as the “resting state” (i.e., unconstrained cognition in the absence of experimental manipulations), engage a widespread network that cannot be captured faithfully by a handful of nodes. Furthermore, in the context of Computational Psychiatry, putative pathophysiological processes underlying various mental disorders have been linked to global (large-scale) alterations of functional integration in brain networks [e.g., (31, 123, 125–127, 129, 130, 164)]. This calls for the development

of computational models that are capable of inferring effective (directed) connectivity in whole-brain networks (122).

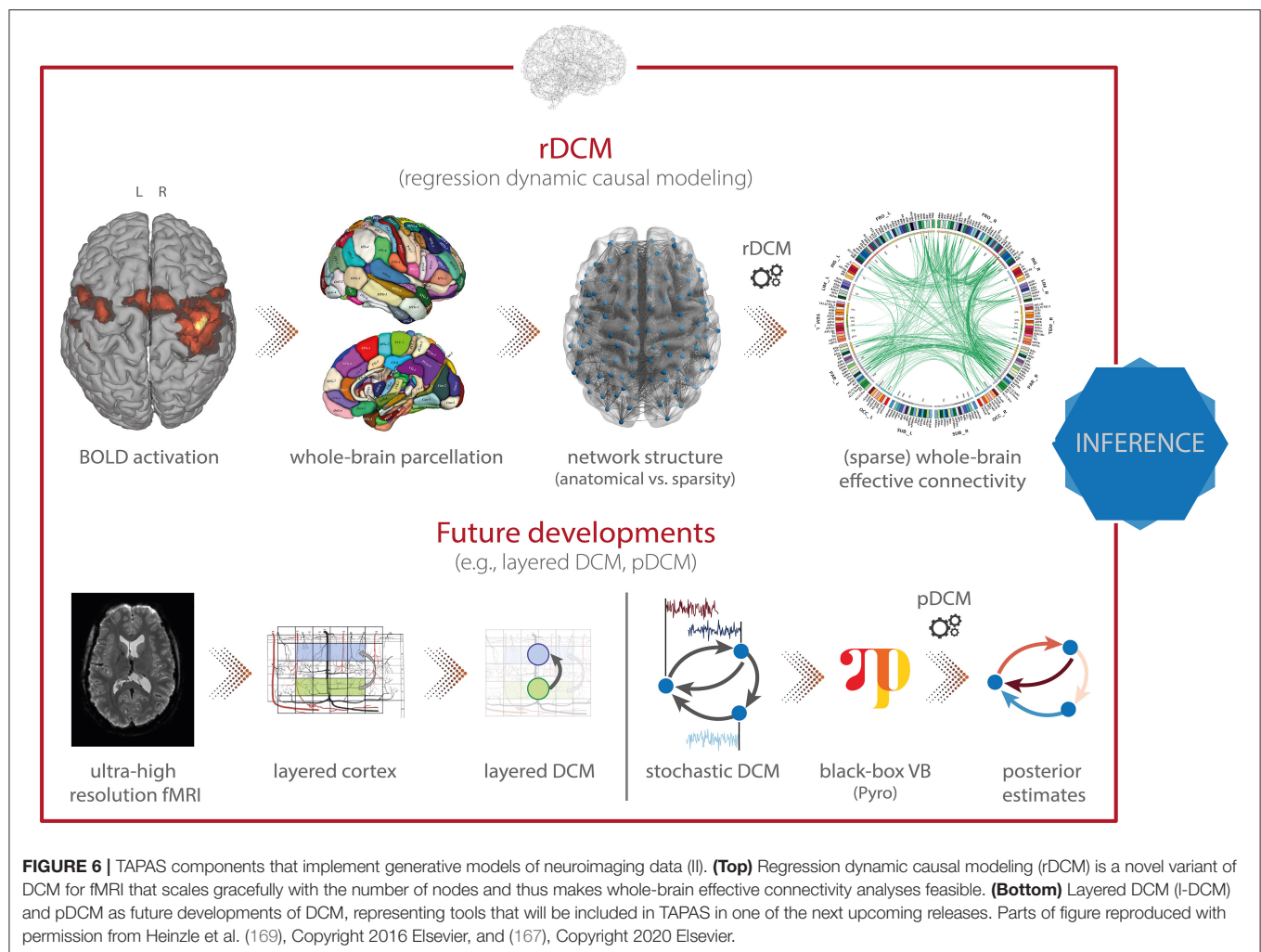
*Regression dynamic causal modeling* [rDCM; (165, 166)] represents a recent variant of DCM that renders model inversion extremely efficient. This is achieved by converting the numerically costly estimation of coupling parameters in differential equations of a linear DCM in the time domain into a Bayesian linear regression model in the frequency domain (**Figure 6, top**). Under a suitably chosen mean-field approximation, analytically solvable VB update equations can be derived for this model. The ensuing computational efficiency allows rDCM to scale gracefully to large-scale networks that comprise hundreds of regions. Furthermore, rDCM has recently been augmented with sparsity constraints to automatically prune fully connected networks to an optimal (in terms of maximal model evidence) degree of sparsity (165). This is achieved by introducing binary indicator variables into the likelihood function, which essentially serve as feature selectors. For this generative model, comprehensive simulation studies demonstrated the face validity of rDCM with regard to model parameter and model architecture recovery. Furthermore, we have provided initial demonstrations of the construct validity of the approach in applications to empirical data. For instance, using ultra-high field (7T) fMRI data from a simple hand movement paradigm with the known relevant connections, we demonstrated that rDCM inferred plausible effective connectivity patterns in whole-brain networks with more than 200 regions (167). Furthermore, we have recently demonstrated that rDCM can not only be applied to task-based, but also to resting-state fMRI data (168). Notably, inversion of whole-brain models with rDCM is computationally highly efficient on standard hardware: even for whole-brain networks with more than 200 regions, it takes only a couple of minutes for fixed network architectures, and a few hours when pruning fully connected networks.

### Future Developments

Besides the toolboxes mentioned above, which are already part of TAPAS, additional variants of DCM for fMRI will be released soon. Specifically, this includes: (i) layered dynamic causal modeling [layered DCM; (169)], and (ii) pDCM, a Python-based DCM implementation focused on amortized inference of stochastic DCMs using novel probabilistic-programming techniques. Here, we briefly outline these two advances.

First, layered DCM addresses challenges in effective connectivity analyses that become relevant when moving toward high-resolution fMRI measurements at the sub-millimeter scale (**Figure 6, bottom left**). This allows differentiating BOLD signals from different cortical layers, an important aspect for testing desiderata of modern theories of brain structure and function. Specifically, prominent “Bayesian brain” theories like predictive coding (170, 171) postulate that supragranular and infragranular cortical layers convey different signals via their efferent cortico-cortical connections. Testing these theories might not only further our understanding of the functioning of the human brain in health, but also has important implications for delineating pathophysiological processes in disease.





However, for layered fMRI data, the spatial layout of cortical blood supply—in particular, the venous blood draining back from lower layers to the cortical surface—confounds responses in different layers and thus renders interpretations non-trivial. Accounting for such draining effects is thus important, yet not readily possible within contemporary hemodynamic models, such as the Balloon model currently implemented in DCM (172–174) or more recent hemodynamic models that strive for increased biological plausibility (175). Layered DCM addresses this limitation by extending the classical Balloon model with a phenomenological description of blood draining effects. This rests on including a delayed coupling of the relative blood volume and deoxyhemoglobin concentration across layers. For this framework, Heinze et al. (169) demonstrated the face validity using simulation studies, as well as the practical utility in an application to empirical fMRI data from a simple visual paradigm. While a detailed dynamic model of layered blood flow effects was published recently (176), this model requires more parameters and full model inversion was not explored in this paper.

Second, a novel inversion scheme for stochastic DCMs will be included in TAPAS, leveraging recent breakthroughs in black-box variational inference. In brief, the approach makes use of optimization algorithms from deep learning software packages to allow inference of general probabilistic models (Figure 6, bottom right). For inference, one directly infers the neuronal and hemodynamic parameters, but temporal convolutional neuronal networks are used to amortize the inference of the neuronal states themselves. This allows inferring the hidden (neuronal) states for any length of input data while keeping the number of parameters fixed (instead of growing linearly with the number of time points). The advantage of using black-box variational inference algorithms is that the framework is highly flexible, allowing for easy modifications and extensions of the underlying generative model without any changes to the inference machinery. This renders the model very promising for clinical applications where generative models might need to be tailored toward specific diseases. A publication on pDCM is currently in preparation and the toolbox will be released in TAPAS as the paper is published.



## Generative Models of Behavioral Data

Neuroimaging provides functional readouts from disease-relevant neural circuits and thus delivers data for models of pathophysiology. However, these data and models are usually not suitable for drawing direct conclusions about cognition and its disturbances. By contrast, behavioral data can be used for inference on an agent's internal processes at the algorithmic (information processing) level [for a review of guidelines for the computational modeling of behavioral data, please refer to (177)]. Importantly, acquisition of behavioral data is often easier, cheaper and more patient-friendly than neuroimaging data, and computational models of behavior thus hold great promise for establishing clinically useful computational assays—on their own or in combination with neuroimaging data (16, 29). At present, TAPAS contains two different generative models of behavioral data which will be discussed next: (i) the Hierarchical Gaussian Filter [HGF; (178)], and (ii) the Stochastic Early Reaction, Inhibition and late Action model [SERIA; (179)].

### Hierarchical Gaussian Filter (HGF)

The *Hierarchical Gaussian Filter* [HGF; (178)] is a hierarchical Bayesian framework for individual learning under the various kinds of uncertainty which arise in realistic non-linear dynamic systems (e.g., perceptual uncertainty, environmental volatility). Importantly, the hierarchy implemented in the HGF is not to be confused with the hierarchy that was discussed in the context of empirical Bayes and, more specifically, HUGE. While in empirical Bayes, the hierarchy (typically) refers to a multi-subject or mixed-effects structure where the levels represent single-subject and group-level information, the HGF implements a hierarchy in which the levels represent the temporal evolution of latent states. More specifically, it consists of two parts: a generative model, the HGF-GM, which describes the stochastic evolution of the non-linearly coupled hidden states of a dynamic system; and the HGF proper, a set of deterministic update equations resulting from the variational inversion of the HGF-GM. The HGF proper contains the Kalman filter as a special case, but is also suited for filtering inputs generated by non-linear environments. Combined with an observation model, the HGF proper represents a particular implementation of the “observing the observer” framework developed by Daunizeau et al. (220, 221). This framework is based on the separation of two model components: (i) a perceptual model which describes an agent's inference on the environment (in this case the HGF proper), and (ii) a response model which describes how inferred latent states of the agent translate into the agent's observed actions, such as, decisions or responses (180).

In the HGF, the perceptual model takes the form of a hierarchical Bayesian model where the temporal evolution of states at any level (except the first) are represented as Gaussian random walks or first-order autoregressive processes (Figure 7, top). Importantly, the step size of each walk (i.e., the variance of the Gaussian distribution) depends on the state at the next higher level. This coupling between levels is controlled by subject-specific parameters that shape the influence of uncertainty on learning. Under a VB approximation, one can derive efficient trial-by-trial update equations for this model that describe the

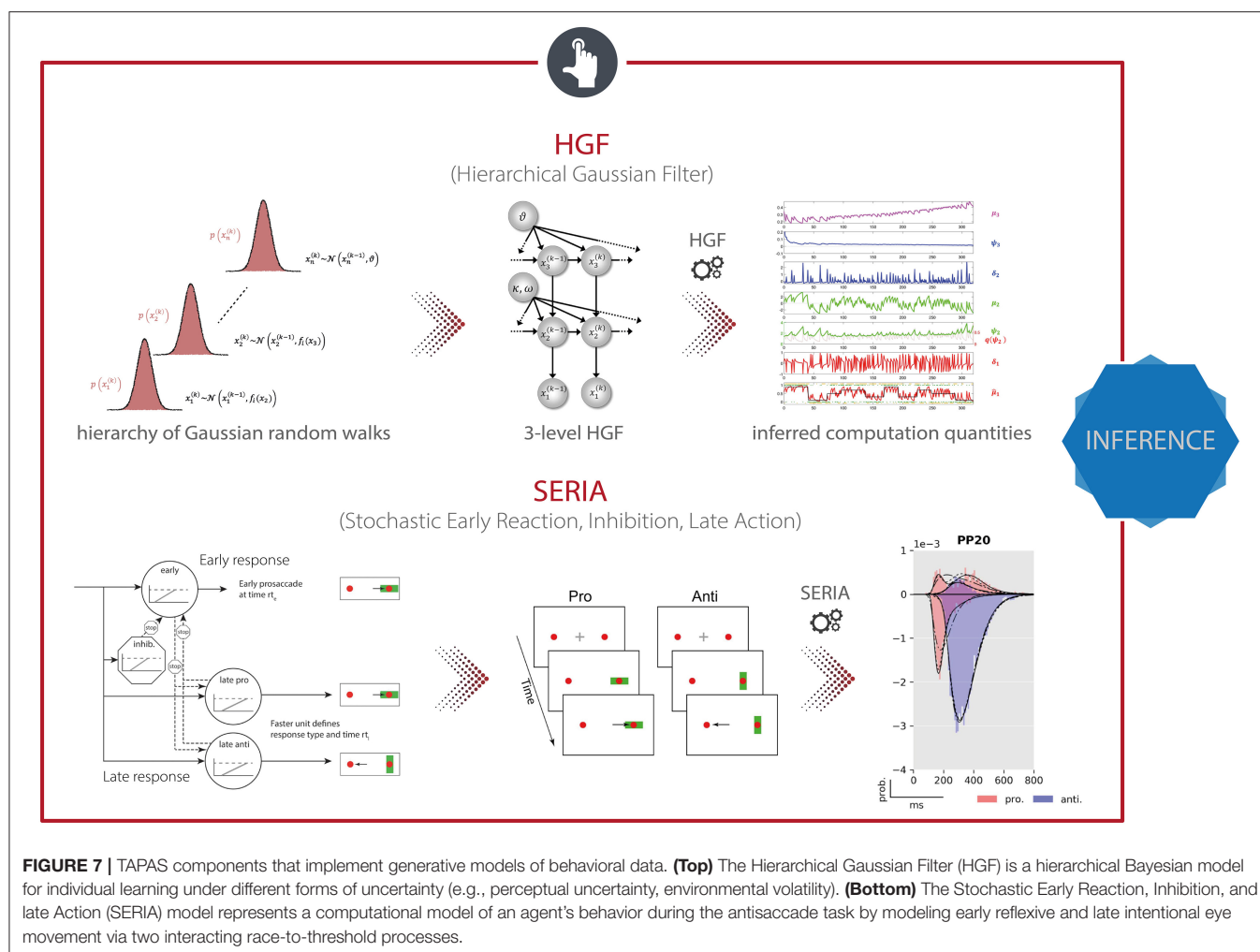
agent's belief updating. Importantly, these update equations rest on precision-weighted prediction errors (PE) at different levels of the hierarchy. In other words, the HGF tracks an agent's expression of (approximate) Bayesian learning in the presence of uncertainty under the assumption that the brain continuously updates a hierarchical generative model of sensory inputs, with PEs serving as the teaching signal. This perceptual model is then combined with a response model (e.g., unit-square sigmoid or softmax function; although a wide range of different response models is available) that links the agent's current estimates of the latent states to observed actions, such as motor or physiological responses (180). In combination with priors on the model parameters, this specifies a full generative model of observed responses that is inverted using maximum-a-posteriori (MAP) estimation. In summary, the HGF provides a generic approximation to subject-specific instantiations of hierarchical Bayesian learning. This generic form renders the HGF applicable to a wide range of scenarios, including discrete and continuous latent states, deterministic and probabilistic relations between environmental events and latent states, as well as learning under multiple forms of uncertainty such as perceptual uncertainty and environmental volatility (178).

Depending on the specific scientific question, the HGF can (or must) be compared to other models of learning. For example, the HGF toolbox already implements various other models of learning, including the Rescorla-Wagner model (181), Sutton model (182), and a Hidden Markov model (183). These models differ from the HGF in their proposed style of learning. For instance, while the updating in Rescorla-Wagner (RW) learning is structurally not dissimilar to that in the HGF (i.e., weighted prediction errors), the RW model differs fundamentally from the HGF in that prediction errors are weighted by a constant learning rate instead of time-dependent precision weights (178). These (and any other learning models of interest outside the toolbox) can be compared using Bayesian model selection [BMS; (117, 158)]. This is achieved by harvesting (an approximation to) the log model evidence from the different models and entering the estimates into existing tools for fixed-effects or random-effects BMS, e.g., SPM (184) or the VBA toolbox (35).

For applications in TN/CP, parameter estimates of the HGF (or any other learning model in the toolbox) can be used to characterize perceptual inference and decision-making in specific disorders [e.g., (18, 19)]; alternatively, and perhaps even more frequently, the estimated trajectories of precision-weighted PEs are used in trial-by-trial analyses of fMRI and EEG data [e.g., (20, 185, 186)]. The HGF can also be applied to other time series than behavioral ones. In an example from TN/CP, Brazil et al. (219) applied an HGF directly to BOLD signal time series from an fMRI experiment.

### Stochastic Early Reaction, Inhibition and late Action (SERIA) model

Eye movements represent a potentially very interesting functional readout for TN/CP. In addition to the experimental ease with which many data points can be measured, eye movements are disturbed in numerous psychiatric conditions (187–189). An experimental paradigm that has been used



frequently in this context is the antisaccade task (190) where participants are asked to suppress a reactive eye movement toward a visual cue and concurrently perform a saccade in the opposite direction (antisaccade). This task is of relevance for clinical applications since it has been widely used to study psychiatric and neurological diseases (188). Most prominently, the antisaccade task is hampered in schizophrenia (191, 192) where an elevated error rate has been proposed as an intermediate phenotype or endophenotype of the disease (193).

TAPAS comprises a computational model of an agent's behavior during the antisaccade task—the *Stochastic Early Reaction, Inhibition and late Action* (SERIA) model (179). Specifically, in order to model error rates and reaction times during the task, SERIA postulates two interacting processes (**Figure 7, bottom**): (i) a fast GO/NO-GO race between a prepotent response (prosaccade) toward the visual cue and a signal to cancel this erroneous action, and (ii) a slow GO/GO race between two units encoding the cue-action mapping, accounting for slow voluntary saccades. The parameters of this model, which are estimated

using a sampling-based hierarchical Bayesian scheme, are sensitive to dopaminergic and cholinergic manipulations and were found to allow for out-of-sample predictions about the drug administered to an individual with 70% accuracy (194).

### Future Developments

As for the generative models of neuroimaging data, extensions to the behavioral models will also be released as part of TAPAS in the future. In particular, an extension of the HGF is currently under development which aims to embed the classical HGF within an empirical Bayesian (EB) scheme. This hierarchical (in the sense of an EB scheme) version of the HGF combines HGFs at the individual level with a layer that represents group effects. Similar to HUGE, this formulation affords a principled way of estimating (empirical) priors from the data, effectively constraining single-subject estimates by group-level information. Inference in the H2GF rests on sampling-based MCMC methods that provide not only an estimate of individual model parameters, but also an approximation to the model evidence via thermodynamic integration (or other suitable techniques).

## CLINICAL APPLICATION

Computational assays are developed with the goal to solve concrete clinical problems. Here, we briefly consider 3 fundamental problems in psychiatry: (i) differential diagnosis (i.e., differentiation between several plausible conditions/mechanisms underlying a patient's symptoms), (ii) stratification of heterogeneous disorders into subgroups, and (iii) prediction of clinical trajectories or treatment responses (25). Different computational strategies are available to address each of these challenges.

First, differential diagnosis can be formalized as hypothesis testing, which—in a Bayesian framework—is equivalent to Bayesian model selection [BMS; (117, 158)] where different hypotheses (models) are compared in the light of observed neuroimaging and/or behavioral data. Specifically, by formalizing competing pathophysiological and/or psychopathological theories in terms of distinct models, we can assess the relative plausibility of these hypotheses by model comparison (119). This rests on comparing the model evidence, an index of model goodness that trades off accuracy and complexity. One early example of how model comparison can be used to support differential diagnosis was provided by a study on NMDA receptor antibody encephalitis (195). This disorder was only discovered relatively recently (196) and is poorly understood at the level of cortical circuit dysfunction. To disambiguate alternative circuit-level mechanisms how seizure activity may unfold during NMDA receptor antibody encephalitis, Cooray et al. (195) compared different implementations of a cortical microcircuit model (DCM). These alternatives differed in the type of synaptic connections that were allowed to change between seizure and non-seizure periods. Using artifact-free seizure data from two patients, the authors demonstrated that concomitant changes in excitatory and inhibitory connections, as well as the gain of inhibitory neurons, best explained the symptoms (seizure recordings).

Second, prediction of clinical trajectories and treatment outcome as well as stratification of spectrum disorders can be achieved by means of supervised and unsupervised generative embedding [GE; (197)], respectively. In brief, the key idea of GE is to perform (un)supervised learning in a feature space that is spanned by the posterior estimates obtained from a generative model fitted to the data. In the simplest way, this can be achieved by following a two-step procedure: First, a generative model of (neuroimaging or behavioral) data is used to infer the posterior densities over model parameters (e.g., neuronal connectivity, learning rate). Second, summary statistics of these posterior densities (e.g., maximum-a-posteriori estimates) enter a supervised (classification, regression) or unsupervised (clustering) machine learning technique. In doing so, the generative model serves as a theory-driven dimensionality reduction device which projects the high-dimensional and noisy data onto neurobiologically meaningful parameters that span a low-dimensional and interpretable space for (un)supervised learning. GE frequently yields more accurate results than conventional ML (30, 139, 198), likely because

the generative model separates signal (reflecting the process of interest) from (measurement) noise. For instance, in a recent study, Frässle et al. (198) utilized GE (combining DCM for fMRI and linear support vector machines) to predict the 2-year clinical trajectories of patients with major depressive disorder (MDD) from the NETHERLANDS Study of Depression and Anxiety (NESDA). Specifically, using GE, the authors could distinguish chronic patients from fast-remitting patients with 79% balanced accuracy. Similarly, gradually improving patients could be distinguished from fast-remitting patients with 61% balanced accuracy. This significantly outperformed classification based on conventional (descriptive) features, such as local activation or functional connectivity estimates, which were obtained from the same data. These results (in line with other studies) illustrate the potential of GE for clinical decision making. In what follows, we outline toolboxes included in TAPAS that can be used for supervised and unsupervised GE.

## Classification and Prediction

As outlined above, GE is typically implemented in terms of a two-step procedure: (i) generative models of measured data are inverted for each subject individually, and (ii) the summary statistics of the posterior estimates (e.g., the maximum-a-posteriori estimates) are used for supervised (classification, regression) or unsupervised (e.g., clustering) learning.

Here, we first focus on supervised GE as a formal way of performing differential diagnosis (via classification) or outcome prediction. To this end, TAPAS comprises the *Generative Embedding* (GE) toolbox, Python-based software that facilitates the generative embedding framework and allows for exploration and visualization of classification performance. The toolbox is a wrapper around scikit-learn (199), with the goal of providing a set of convenient functions and sensible defaults that form a suitable starting point for generative embedding analyses. The GE toolbox can use posterior parameter estimates from any of the generative models mentioned above as input features, and performs binary or multi-class classification. The toolbox utilizes logistic regression as its default classifier because it represents a simple linear model that protects against overfitting, is (relatively) simple to interpret, and the ensuing class probabilities are useful for interpreting classifier outputs. Furthermore, the toolbox implements a repeated k-fold cross-validation as the default procedure for both model selection (i.e., hyperparameter tuning) and model validation (i.e., estimating out-of-sample performance), including the possibility for within-fold confound correction (200). This choice is motivated by the fact that, in comparison to leave-one-out cross-validation, k-fold cross-validation has a lower variance and is therefore less prone to overfitting (201). Finally, significance testing of classification performance is done by default using permutation tests as they provide an unbiased estimate of error variance. This is in contrast to parametric tests (e.g., binomial confidence intervals, McNemar's test) which typically underestimate variance and are therefore overconfident (202).

## Stratification of Heterogeneous Psychiatric Disorders

The goal of stratifying heterogeneous disorders is to identify subgroups that share common pathophysiological or psychopathological mechanisms. The increased homogeneity in terms of underlying disease mechanisms increases the power of clinical trials and enhances predictions of clinically relevant outcomes (25). One way to achieve this goal is by using posterior parameters from a generative model for unsupervised learning (e.g., clustering). This approach has been utilized by Brodersen et al. (139) to identify distinct subgroups in a heterogeneous cohort of 41 patients with schizophrenia based on effective connectivity patterns during a working memory task (138). The authors showed that these purely physiologically informed and connectivity-based subgroups also differed clinically, as illustrated by significant differences in their negative symptom severity scores on the Positive and Negative Syndrome Scale (PANSS). An alternative to this two-step GE procedure is implemented in HUGE, a toolbox we already discussed in the context of generative models of neuroimaging data (161, 162). Specifically, HUGE casts unsupervised GE as a single hierarchical generative model that simultaneously describes individual data generation and assigns participants to clusters (**Figure 5, bottom**). Unifying these two steps has a couple of conceptual advantages: (i) the hierarchical nature of the model allows learning prior distributions from the data (i.e., empirical Bayes), (ii) model inversion at the single-subject level is regularized by (cluster-specific) group results, and (iii) clustering takes the uncertainty about individual connectivity parameter estimates into account.

## TAPAS IN ACTION

Finally, we briefly discuss a few selected examples from previous work that made use of different toolboxes from TAPAS. In particular, here we focus on studies that investigate pathophysiological/pathocomputational mechanisms and/or explore the role of neuromodulatory transmitter systems in these processes. The latter are a particularly prominent topic for clinical applications of generative models because the majority of available pharmacotherapeutic approaches in psychiatry targets synthesis, metabolism or receptors of neuromodulatory transmitters.

In order to non-invasively infer upon the status of neuromodulatory systems (e.g., dopamine, acetylcholine, serotonin, noradrenaline), various studies have combined experimental manipulations of different neuromodulatory systems with generative modeling of neuroimaging or behavioral data. In a first step, Iglesias et al. (185) have provided evidence for hierarchical belief updating during a sensory associative learning task under volatility and without rewards. Hierarchical belief updating via PEs plays a central role in “Bayesian brain” theories, such as predictive coding (170, 203). Iglesias et al. (185) utilized the HGF to infer upon subject-specific trajectories of precision-weighted PEs at different levels of the hierarchy which were then used in a GLM of fMRI data. They found that

low-level PEs, encoding the mismatch between prediction and actual visual stimulus outcome, were reflected by widespread BOLD activity in visual and supramodal areas, but also in the midbrain. Conversely, high-level PEs, encoding the mismatch between prediction and actual stimulus probabilities, were reflected by BOLD activity in the basal forebrain. Midbrain and basal forebrain contain dopaminergic and cholinergic neurons, respectively, suggesting that (i) dopaminergic midbrain neurons might signal PEs unrelated to reward and (ii) cholinergic neuron activity in the basal forebrain might reflect PEs about probabilities and may thus relate to “expected uncertainty” (204). Although a subsequent pharmacological study in human volunteers using the same paradigm did not support this notion (205), the finding that midbrain activity may reflect reward-unrelated prediction errors has since been replicated in several animal (206, 207) and human studies (208).

Importantly, neural correlates of computational quantities can not only be detected in fMRI data, but can also be found in EEG signals, where the superior temporal resolution allows for characterizing their precise temporal dynamics. For instance, Weber et al. (186) related HGF estimates to single-trial EEG data from participants who received ketamine in a placebo-controlled, double-blind, within-subject fashion. The authors demonstrated that PE-related activity was found in a temporal order consistent with hierarchical Bayesian theory. Additionally, they observed a significant impact of ketamine on the high-level PE about transition probabilities. Focusing on behavior only, further evidence has been provided for associations between computational quantities and the status of neuromodulatory systems. For instance, Vossel et al. (230) perturbed the cholinergic system using pharmacological interventions, and utilized the HGF to demonstrate that this led to an increase in the rate of belief updating about cue validity during a modified Posner’s task. Similarly, Marshall et al. (225) utilized pharmacological interventions in combination with the HGF to characterize the influence of noradrenergic, cholinergic and dopaminergic antagonists on individual estimates of uncertainty during a probabilistic serial reaction time task. The authors identified different roles for the different neuromodulatory systems, linking noradrenaline to unexpected uncertainty, acetylcholine to environmental uncertainty, and dopamine to uncertainty representations for fast, adaptive responses. Finally, Aponte et al. (194) demonstrated that computational quantities sensitive to neuromodulatory processes can also be derived from generative models of reflexive eye movements. Specifically, the authors conducted a double-blind placebo-controlled pharmacological study and found that computational quantities derived from an antisaccade task using the SERIA model can distinguish between dopaminergic and cholinergic effects on action selection and inhibitory control, allowing for out-of-sample predictions about the drug administered with 70% accuracy. In summary, both the HGF and SERIA comprise computational quantities that are sensitive to the functional status of different neuromodulatory systems.

Beyond questions of pathophysiology and pharmacology, tools from TAPAS have also been used to characterize clinical populations. For instance, Powers et al. (19) studied conditioned



auditory hallucinations in four groups of people who differed both in their voice-hearing and treatment-seeking statuses. Utilizing the HGF to infer upon the participants' individual beliefs, the authors demonstrated that the weighting of prior beliefs was significantly larger in people with hallucinations than their non-hallucinating counterparts. This is consistent with the hypothesis that, in the context of a Bayesian brain, hallucinations may be explained by overly strong priors (209). Focusing on patients with autism spectrum disorder (ASD), Lawson et al. (18) utilized the HGF to provide evidence that ASD patients tend to overestimate volatility in the face of environmental changes. This leads to reduced learning about unexpected (surprising) events, which might serve as an explanation for the typical insistence on sameness and intolerance of change in ASD patients (210). Finally, Cole et al. (20) applied HGF estimates from an associative learning task to characterize brain responses to precision-weighted PEs in individuals at clinical high risk (CHR) for psychosis. Compared to a healthy control group, CHR individuals showed enhanced PE responses in several (particularly prefrontal) regions, consistent with the prediction from the dysconnection hypothesis of schizophrenia (125, 164) that (proneness to) psychosis is characterized by abnormal precision-weighted PE signaling in cortex. Furthermore, prefrontal PE activity was correlated with clinical status.

TAPAS tools that implement generative models of neuroimaging data have also been applied to clinical populations—although this is still rare. For instance, Yao et al. (162) applied HUGE to an fMRI dataset comprising aphasic patients (with a lesion in the left frontal and/or temporal cortex) and healthy controls (211) for an initial demonstration of the potential clinical utility of the model for patient stratification. In brief, the authors demonstrated that HUGE correctly identifies two clusters in the dataset, which mapped almost perfectly onto aphasic patients and healthy controls, yielding a balanced purity of 95.5%. While it is important to emphasize that diagnosing patients with aphasia does not yet represent a truly meaningful clinical problem, it demonstrates the practical utility of HUGE for stratification in a scenario where ground truth is known. Furthermore, regression DCM has been used to study alterations in whole-brain effective (directed) connectivity between psychotic patients, their first-degree relatives, as well as matched healthy controls. The authors demonstrate that patients showed distinctly different whole-brain connectivity patterns from healthy controls and first-degree relatives, and that the connectivity patterns allow for significant discrimination at the individual level. A publication on this work is currently in preparation.

Overall, the above studies illustrate the potential of generative models of behavioral and neuroimaging data for clinical applications. However, these studies do not yet implement the kind of end-to-end analysis pipeline that we outlined, at the beginning of the article, as a basis for future computational assays. Instead, the above studies simply used selected components from TAPAS at a time. Having said this, recent work by Harrison et al. comes close to the kind of end-to-end pipeline highlighted above, combining multiple

components from TAPAS (212). In brief, the authors aimed to investigate interoception and how anxiety relates to the perception of internal bodily states. To this end, Harrison et al. employed two paradigms available in TAPAS Tasks, namely the Filter Detection (FD) and Breathing Learning (BL) task. The FD task revealed differences in sensitivity to breathing perception and altered interoceptive metacognitive bias between low-anxiety and moderate-anxiety healthy controls. Furthermore, for the BL task, the authors acquired fMRI data using a high-field 7T MR scanner. PhysIO was employed for physiological noise correction based on measurements of cardiac and respiratory cycles. fMRI then underwent thorough preprocessing and artifact removal by combining tools from various software packages, including FSL and SPM. Brain activity coupled with dynamic changes in bodily states was then modeled using subject-specific trajectories of predictions and PE which have been inferred utilizing the HGF toolbox. This revealed the anterior insula to be associated with both interoceptive predictions and PEs, where the former was also differentially expressed in the low and moderate anxiety groups.

While this moves toward the kind of end-to-end analysis pipeline that we outline, it is important to note that the two groups tested by Harrison et al. do not represent clinical groups (but were recruited from the healthy population) and the study thus lacks the final module of the aforementioned pipeline (i.e., Clinical Application; see **Figure 2**). Furthermore, it is important to keep in mind that none of the studies mentioned above is yet of any direct clinical utility, in the sense that they do not address a practical clinical question, such as differential diagnosis or predicting outcomes/clinical trajectories. The latter in particular requires data from prospective studies that include information about future clinical outcomes—a critical condition for validating computational assays (16). Unfortunately, so far, these datasets are rare.

## CONCLUSION

In this article, we have described the Translational Algorithms for Psychiatry-Advancing Science (TAPAS) software package, an open-source collection of toolboxes (primarily written in MATLAB; with some components in C and Python) that aim to facilitate the acquisition and (computational) analysis of neuroimaging and behavioral data. Specifically, we reviewed the different toolboxes in TAPAS and highlighted how these might support the construction of end-to-end analysis pipelines—from raw data to clinical applications.

## SOFTWARE NOTE

The Translational Algorithms for Psychiatry-Advancing Science (TAPAS) software package, comprising all toolboxes described in this paper, is freely available as open-source code (<https://www.translationalneuromodeling.org/tapas>).



## AUTHOR CONTRIBUTIONS

SF and KS wrote manuscript. All authors helped revising the manuscript. Furthermore, several authors contributed software code to the TAPAS software package reviewed in this manuscript.

## FUNDING

This work was supported by the UZH Forschungskredit Postdoc, Grant No. FK-18-046, the ETH Zurich Postdoctoral Fellowship Program and the Marie Curie Actions for People

COFUND Program, Grant No. FEL-49 15-2 (SF), a Marie Skłodowska-Curie Postdoctoral Fellowship from the European Union's Horizon 2020 research and innovation program under the Grant Agreement No. 793580, as well as by the Royal Society of New Zealand (OKH), the Strategic Focal Area Personalized Health and Related Technologies (PHRT) of the ETH Domain, Grant No. 2017-403 (SJH), as well as the René and Susanne Braginsky Foundation, the Clinical Research Priority Program Pain of the University of Zurich, and the Swiss National Science Foundation, Grant No. 320030\_179377 (KES).

## REFERENCES

- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5 R)*. Arlington, VA: American Psychiatric Publishing. doi: 10.1176/appi.books.9780890425596
- World Health Organization (1990). *International Classification of Diseases*. Geneva: World Health Organization Press.
- Cuthbert B, Insel T. Toward new approaches to psychotic disorders: the NIMH research domain criteria project. *Schizophrenia Bull.* (2010) 36:1061–2. doi: 10.1093/schbul/sbq108
- Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry.* (2012) 17:1174–9. doi: 10.1038/mp.2012.105
- Krystal JH, State MW. Psychiatric disorders: diagnosis to therapy. *Cell.* (2014) 157:201–14. doi: 10.1016/j.cell.2014.02.042
- Owen MJ. New approaches to psychiatric diagnostic classification. *Neuron.* (2014) 84:564–71. doi: 10.1016/j.neuron.2014.10.028
- Stephan KE, Bach DR, Fletcher PC, Flint J, Frank MJ, Friston KJ, et al. Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. *Lancet Psychiatry.* (2016) 3:77–83. doi: 10.1016/S2215-0366(15)00361-2
- Stephan KE, Binder EB, Breakspear M, Dayan P, Johnstone EC, Meyer-Lindenberg A, et al. Charting the landscape of priority problems in psychiatry, part 2: pathogenesis and aetiology. *Lancet Psychiatry.* (2016) 3:84–90. doi: 10.1016/S2215-0366(15)00360-0
- Moran RJ, Jung F, Kumagai T, Endepols H, Graf R, Dolan RJ, et al. Dynamic causal models and physiological inference: a validation study using isoflurane anaesthesia in rodents. *PLoS ONE.* (2011) 6:e22790. doi: 10.1371/journal.pone.0022790
- Deco G, Kringelbach ML. Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. *Neuron.* (2014) 84:892–905. doi: 10.1016/j.neuron.2014.08.034
- Gilbert JR, Symmonds M, Hanna MG, Dolan RJ, Friston KJ, Moran RJ. Profiling neuronal ion channelopathies with non-invasive brain imaging and dynamic causal models: Case studies of single gene mutations. *Neuroimage.* (2016) 124:43–53. doi: 10.1016/j.neuroimage.2015.08.057
- Symmonds M, Moran CH, Leite MI, Buckley C, Irani SR, Stephan KE, et al. Ion channels in EEG: isolating channel dysfunction in NMDA receptor antibody encephalitis. *Brain.* (2018) 141:1691–702. doi: 10.1093/brain/aw y107
- Frässle S, Yao Y, Schöbi D, Aponte EA, Heinzle J, Stephan KE. Generative models for clinical applications in computational psychiatry. *Wiley Interdiscip Rev Cogn Sci.* (2018) 9:e1460. doi: 10.1002/wcs.1460
- Adams NE, Hughes LE, Phillips HN, Shaw AD, Murley AG, Nesbitt D, et al. GABA-ergic dynamics in human frontotemporal networks confirmed by pharmaco-magnetoencephalography. *J. Neurosci.* (2020) 40:1640–9. doi: 10.1523/JNEUROSCI.1689-19.2019
- Maia T, Frank M. From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci.* (2011) 14:154–62. doi: 10.1038/nn.2723
- Stephan K, Mathys C. Computational approaches to psychiatry. *Curr Opin Neurobiol.* (2014) 25:85–92. doi: 10.1016/j.conb.2013.12.007
- Iglesias S, Tomiello S, Schneebeli M, Stephan KE. Models of neuromodulation for computational psychiatry. *Wiley Interdiscip Rev Cogn Sci.* (2016) 8:e1420. doi: 10.1002/wcs.1420
- Lawson RP, Mathys C, Rees G. Adults with autism overestimate the volatility of the sensory environment. *Nat Neurosci.* (2017) 20:1293–9. doi: 10.1038/nn.4615
- Powers AR, Mathys C, Corlett PR. Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science.* (2017) 357:596–600. doi: 10.1126/science.aan3458
- Cole DM, Diaconescu AO, Pfeiffer UJ, Brodersen KH, Mathys CD, Julkowsky D, et al. Atypical processing of uncertainty in individuals at risk for psychosis. *Neuroimage Clin.* (2020) 26:102239. doi: 10.1016/j.nicl.2020.102239
- Jirsa VK, Sporns O, Breakspear M, Deco G, McIntosh AR. Towards the virtual brain: network modeling of the intact and the damaged brain. *Arch Ital Biol.* (2010) 148:189–205. doi: 10.4449/aib.v148i3.1223
- Montague P, Dolan R, Friston K, Dayan P. Computational psychiatry. *Trends Cogn Sci.* (2012) 16:72–80. doi: 10.1016/j.tics.2011.11.018
- Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry.* (2014) 1:148–58. doi: 10.1016/S2215-0366(14)70275-5
- Wang XJ, Krystal JH. Computational psychiatry. *Neuron.* (2014) 84:638–54. doi: 10.1016/j.neuron.2014.10.018
- Stephan KE, Iglesias S, Heinzle J, Diaconescu AO. Translational perspectives for computational neuroimaging. *Neuron.* (2015) 87:716–32. doi: 10.1016/j.neuron.2015.07.008
- Adams RA, Huys QJ, Roiser JP. Computational Psychiatry: towards a mathematically informed understanding of mental illness. *J Neurol Neurosurg Psychiatry.* (2016) 87:53–63. doi: 10.1136/jnnp-2015-310737
- Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci.* (2016) 19:404–13. doi: 10.1038/nn.4238
- Paulus MP, Huys QJ, Maia TV. A roadmap for the development of applied computational psychiatry. *Biol Psychiatry Cogn Neurosci Neuroimaging.* (2016) 1:386–92. doi: 10.1016/j.bpsc.2016.05.001
- Browning M, Carter CS, Chatham C, Den Ouden H, Gillan CM, Baker JT, et al. Realizing the clinical potential of computational psychiatry: report from the Banbury Center Meeting, February 2019. *Biol Psychiatry.* (2020) 88:e5–e10. doi: 10.1016/j.biopsych.2019.12.026
- Brodersen KH, Schofield TM, Leff AP, Ong CS, Lomakina EI, Buhmann JM, et al. Generative embedding for model-based classification of fMRI data. *PLoS Comput Biol.* (2011) 7:e1002079. doi: 10.1371/journal.pcbi.1002079
- Stephan K, Baldeweg T, Friston K. Synaptic plasticity and dysconnection in schizophrenia. *Biol Psychiatry.* (2006) 59:929–39. doi: 10.1016/j.biopsych.2005.10.005
- Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD. (2006) *Statistical Parametric Mapping: the Analysis of Functional Brain Images*. Burlington: Elsevier.
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage.* (2004) 23:S208–19. doi: 10.1016/j.neuroimage.2004.07.051

34. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res.* (1996) 29:162–73. doi: 10.1006/cbmr.1996.0014
35. Daunizeau J, Adam V, Rigoux L. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Comput Biol.* (2014) 10:e1003441. doi: 10.1371/journal.pcbi.1003441
36. Ahn WY, Haines N, Zhang L. Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Comput Psychiatr.* (2017) 1:24–57. doi: 10.1162/CPSY\_a\_00002
37. Helske J. KFAS: exponential family state space models in R. *J Statistical Software.* (2017) 78:1–39. doi: 10.18637/jss.v078.i10
38. Yousefi A, Paulk AC, Basu I, Mirsky JL, Dougherty DD, Eskandar EN, et al. COMPASS: an open-source, general-purpose software toolkit for computational psychiatry. *Front Neurosci.* (2018) 12:957. doi: 10.1101/377556
39. Wiecki TV, Sofer I, Frank MJ. HDDM: hierarchical bayesian estimation of the drift-diffusion model in python. *Front Neuroinformatics.* (2013) 7:14. doi: 10.3389/fninf.2013.00014
40. Maxwell SE. The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychol Methods.* (2004) 9:147–63. doi: 10.1037/1082-989X.9.2.147
41. Carter CS, Perlstein W, Ganguli R, Brar J, Mintun M, Cohen JD. Functional hypofrontality and working memory dysfunction in schizophrenia. *Am J Psychiatry.* (1998) 155:1285–7. doi: 10.1176/ajp.155.9.1285
42. Van Snellenberg JX, Torres IJ, Thornton AE. Functional neuroimaging of working memory in schizophrenia: task performance as a moderating variable. *Neuropsychology.* (2006) 20:497–510. doi: 10.1037/0894-4105.20.5.497
43. Van Horn JD, Grethe JS, Kostelec P, Woodward JB, Aslam JA, Rus D, et al. The Functional Magnetic Resonance Imaging Data Center (fMRIDC): the challenges and rewards of large-scale databasing of neuroimaging studies. *Philos Trans R Soc Lond B Biol Sci.* (2001) 356:1323–39. doi: 10.1098/rstb.2001.0916
44. Poline JB, Breeze JL, Ghosh S, Gorgolewski K, Halchenko YO, Hanke M, et al. Data sharing in neuroimaging research. *Front Neuroinform.* (2012) 6:9. doi: 10.3389/fninf.2012.00009
45. Poldrack RA, Gorgolewski KJ. Making big data open: data sharing in neuroimaging. *Nat Neurosci.* (2014) 17:1510–7. doi: 10.1038/nn.3818
46. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K, et al. The WU-minn human connectome project: an overview. *Neuroimage.* (2013) 80:62–79. doi: 10.1016/j.neuroimage.2013.05.041
47. Petzschner FH, Weber LA, Wellstein KV, Paolini G, Do CT, Stephan KE. Focus of attention modulates the heartbeat evoked potential. *Neuroimage.* (2019) 186:595–606. doi: 10.1016/j.neuroimage.2018.11.037
48. Khalsa SS, Adolphs R, Cameron OG, Critchley HD, Davenport PW, Feinstein JS, et al. Interoception and mental health: a roadmap. *Biol Psychiatry Cogn Neurosci Neuroimaging.* (2018) 3:501–13. doi: 10.1016/j.bpsc.2017.12.004
49. Poser BA, Setsompop K. Pulse sequences and parallel imaging for high spatiotemporal resolution MRI at ultra-high field. *Neuroimage.* (2018) 168:101–18. doi: 10.1016/j.neuroimage.2017.04.006
50. Bollmann S, Barth M. New acquisition techniques and their prospects for the achievable resolution of fMRI. *Prog Neurobiol.* (2020) 101936. doi: 10.1016/j.pneurobio.2020.101936
51. Weldon KB, Olman CA. Forging a path to mesoscopic imaging success with ultra-high field functional magnetic resonance imaging. *Philos Trans R Soc Lond B Biol Sci.* (2021) 376:20200040. doi: 10.1098/rstb.2020.0040
52. Bellon EM, Haacke EM, Coleman PE, Sacco DC, Steiger DA, Gangarosa RE. MR artifacts: a review. *AJR Am J Roentgenol.* (1986) 147:1271–81. doi: 10.2214/ajr.147.6.1271
53. Kirilina E, Lutti A, Poser BA, Blankenburg F, Weiskopf N. The quest for the best: The impact of different EPI sequences on the sensitivity of random effect fMRI group analyses. *Neuroimage.* (2016) 126:49–59. doi: 10.1016/j.neuroimage.2015.10.071
54. Bright MG, Murphy K. Cleaning up the fMRI time series: mitigating noise with advanced acquisition and correction strategies. *Neuroimage.* (2017) 154:1–3. doi: 10.1016/j.neuroimage.2017.03.056
55. Volz S, Callaghan MF, Josephs O, Weiskopf N. Maximising BOLD sensitivity through automated EPI protocol optimisation. *Neuroimage.* (2019) 189:159–70. doi: 10.1016/j.neuroimage.2018.12.052
56. Bollmann S, Kasper L, Pruessmann KP, Barth M, Stephan KE. Interactive and flexible quality control in fMRI sequence evaluation: the uniQC toolbox. In: *26th Annual Meeting of the International Society for Magnetic Resonance Imaging in Medicine.* Paris (2018). p. 2842. Available online at: <https://archive.ismrm.org/2018/2842.html>
57. Hesselmann V, Girnus R, Wedekind C, Hunsche S, Bunke J, Schulte O, et al. Functional MRI using multiple receiver coils: BOLD signal changes and signal-to-noise ratio for three-dimensional-RESTO vs. single shot EPI in comparison to a standard quadrature head coil. *J Magn Reson Imaging.* (2004) 20:321–6. doi: 10.1002/jmri.20101
58. Poser S, Wiese S, Gembris D, Mathiak K, Kessler C, Grosse-Ruyken ML, et al. Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. *Magn Reson Med.* (1999) 42:87–97. doi: 10.1002/(SICI)1522-2594(199907)42:1<87::AID-MRM13>3.0.CO;2-O
59. Poser BA, Versluis MJ, Hoogduin JM, Norris DG. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. *Magn Reson Med.* (2006) 55:1227–35. doi: 10.1002/mrm.20900
60. Menon RS. Postacquisition suppression of large-vessel BOLD signals in high-resolution fMRI. *Magn Reson Med.* (2002) 47:1–9. doi: 10.1002/mrm.10041
61. Rowe DB. Parameter estimation in the magnitude-only and complex-valued fMRI data models. *Neuroimage.* (2005) 25:1124–32. doi: 10.1016/j.neuroimage.2004.12.048
62. Kundu P, Voon V, Balchandani P, Lombardo MV, Poser BA, Bandettini PA. Multi-echo fMRI: a review of applications in fMRI denoising and analysis of BOLD signals. *Neuroimage.* (2017) 154:59–80. doi: 10.1016/j.neuroimage.2017.03.033
63. Huber L, Handwerker DA, Jangraw DC, Chen G, Hall A, Stuber C, et al. High-Resolution CBV-fMRI allows mapping of laminar activity and connectivity of cortical input and output in human M1. *Neuron.* (2017) 96:1253–63 e1257. doi: 10.1016/j.neuron.2017.11.005
64. Lu H, Hua J, van Zijl P. C. Noninvasive functional imaging of cerebral blood volume with vascular-space-occupancy (VASO) MRI. *NMR Biomed.* (2013) 26:932–48. doi: 10.1002/nbm.2905
65. Alsop DC, Detre JA, Golay X, Gunther M, Hendrikse J, Hernandez-Garcia L, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med.* (2015) 73:102–16. doi: 10.1002/mrm.25197
66. Koretsky AP. Early development of arterial spin labeling to measure regional brain blood flow by MRI. *Neuroimage.* (2012) 62:602–7. doi: 10.1016/j.neuroimage.2012.01.005
67. Glover GH, Li TQ, Ress D. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med.* (2000) 44:162–7. doi: 10.1002/1522-2594(200007)44:1<162::AID-MRM23>3.0.CO;2-E
68. Kruger G, Glover GH. Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magn Reson Med.* (2001) 46:631–7. doi: 10.1002/mrm.1240
69. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage.* (2012) 59:2142–54. doi: 10.1016/j.neuroimage.2011.10.018
70. Yendiki A, Koldewyn K, Kakunoori S, Kanwisher N, Fischl B. Spurious group differences due to head motion in a diffusion MRI study. *Neuroimage.* (2014) 88:79–90. doi: 10.1016/j.neuroimage.2013.11.027
71. Reuter M, Tisdall MD, Qureshi A, Buckner RL, van der Kouwe AJW, Fischl B. Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *Neuroimage.* (2015) 107:107–15. doi: 10.1016/j.neuroimage.2014.12.006
72. Gardner EA, Ellis JH, Hyde RJ, Aisen AM, Quint DJ, Carson PL. Detection of degradation of magnetic resonance (MR) images: comparison of an

- automated MR image-quality analysis system with trained human observers. *Acad Radiol.* (1995) 2:277–81. doi: 10.1016/S1076-6332(05)80184-9
73. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* (2015) 12:e1001779. doi: 10.1371/journal.pmed.1001779
  74. Woodard JP, Carley-Spencer MP. No-reference image quality metrics for structural MRI. *Neuroinformatics.* (2006) 4:243–62. doi: 10.1385/NI:4:3:243
  75. Gedamu EL, Collins DL, Arnold DL. Automated quality control of brain MR images. *J Magn Reson Imaging.* (2008) 28:308–19. doi: 10.1002/jmri.21434
  76. Mortamet B, Bernstein MA, Jack CR Jr, Gunter JL, Ward C, et al. Automatic quality assessment in structural brain magnetic resonance imaging. *Magn Reson Med.* (2009) 62:365–72. doi: 10.1002/mrm.21992
  77. Pizarro RA, Cheng X, Barnett A, Lemaitre H, Verchinski BA, Goldman AL, et al. Automated quality assessment of structural magnetic resonance brain images based on a supervised machine learning algorithm. *Front Neuroinform.* (2016) 10:52. doi: 10.3389/fninf.2016.00052
  78. Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE.* (2017) 12:e0184661. doi: 10.1371/journal.pone.0184661
  79. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage.* (2018) 166:400–24. doi: 10.1016/j.neuroimage.2017.10.034
  80. Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry.* (2014) 19:659–67. doi: 10.1038/mp.2013.78
  81. Poldrack RA, Congdon E, Triplett W, Gorgolewski KJ, Karlsgodt KH, Mumford JA, et al. A phenotype-wide examination of neural and cognitive function. *Sci Data.* (2016) 3:160110. doi: 10.1038/sdata.2016.110
  82. Keil B, Wald LL. Massively parallel MRI detector arrays. *J Magnetic Resonance.* (2013) 229:75–89. doi: 10.1016/j.jmr.2013.02.001
  83. Hendriks AD, D'Agata F, Raimondo L, Schakel T, Geerts L, Luijten PR, et al. Pushing functional MRI spatial and temporal resolution further: high-density receive arrays combined with shot-selective 2D CAIPIRINHA for 3D echo-planar imaging at 7 T. *NMR Biomed.* (2020) 33:e4281. doi: 10.1002/nbm.4281
  84. Dymerska B, Poser BA, Barth M, Tractnig S, Robinson SD. A method for the dynamic correction of B0-related distortions in single-echo EPI at 7T. *Neuroimage.* (2018) 168:321–31. doi: 10.1016/j.neuroimage.2016.07.009
  85. Jezzard P, Balaban RS. Correction for geometric distortion in echo planar images from B0 field variations. *Magn Reson Med.* (1995) 34:65–73. doi: 10.1002/mrm.1910340111
  86. Jezzard P, Clare S. Sources of distortion in functional MRI data. *Hum Brain Mapp.* (1999) 8:80–5. doi: 10.1002/(SICI)1097-0193(1999)8:2<80::AID-HBM2>3.0.CO;2-C
  87. Hutton C, Josephs O, Stadler J, Featherstone E, Reid A, Speck O, et al. The impact of physiological noise correction on fMRI at 7 T. *Neuroimage.* (2011) 57:101–12. doi: 10.1016/j.neuroimage.2011.04.018
  88. Murphy K, Birn RM, Bandettini PA. Resting-state fMRI confounds and cleanup. *Neuroimage.* (2013) 80:349–59. doi: 10.1016/j.neuroimage.2013.04.001
  89. Brosch JR, Talavage TM, Ulmer JL, Nyenhuis JA. Simulation of human respiration in fMRI with a mechanical model. *IEEE Trans Biomed Eng.* (2002) 49:700–7. doi: 10.1109/TBME.2002.1010854
  90. Raj D, Anderson AW, Gore JC. Respiratory effects in human functional magnetic resonance imaging due to bulk susceptibility changes. *Phys Med Biol.* (2001) 46:3331–40. doi: 10.1088/0031-9155/46/12/318
  91. Birn RM, Diamond JB, Smith MA, Bandettini PA. Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *Neuroimage.* (2006) 31:1536–48. doi: 10.1016/j.neuroimage.2006.02.048
  92. Dagli MS, Ingeholm JE, Haxby JV. Localization of cardiac-induced signal change in fMRI. *Neuroimage.* (1999) 9:407–15. doi: 10.1006/nimg.1998.0424
  93. Huettel SA, Song AW, McCarthy G. *Functional Magnetic Resonance Imaging.* Sunderland, MA: Sinauer Associates (2009).
  94. Chang C, Cunningham JP, Glover GH. Influence of heart rate on the BOLD signal: the cardiac response function. *Neuroimage.* (2009) 44:857–69. doi: 10.1016/j.neuroimage.2008.09.029
  95. Hirsch JA, Bishop B. Respiratory sinus arrhythmia in humans - how breathing pattern modulates heart-rate. *Am J Physiol.* (1981) 241:H620–9. doi: 10.1152/ajpheart.1981.241.4.H620
  96. Kasper L, Bollmann S, Diaconescu AO, Hutton C, Heinze J, Iglesias S, et al. The PhysIO toolbox for modeling physiological noise in fMRI data. *J Neurosci Methods.* (2017) 276:56–72. doi: 10.1016/j.jneumeth.2016.10.019
  97. Brooks JC, Beckmann CF, Miller KL, Wise RG, Porro CA, Tracey I, et al. Physiological noise modelling for spinal functional magnetic resonance imaging studies. *Neuroimage.* (2008) 39:680–92. doi: 10.1016/j.neuroimage.2007.09.018
  98. Verstynen TD, Deshpande V. Using pulse oximetry to account for high and low frequency physiological artifacts in the BOLD signal. *Neuroimage.* (2011) 55:1633–44. doi: 10.1016/j.neuroimage.2010.11.090
  99. Chang C, Glover GH. Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. *Neuroimage.* (2009) 47:1448–59. doi: 10.1016/j.neuroimage.2009.05.012
  100. Harrison SJ, Bianchi S, Heinze J, Stephan KE, Iglesias S, Kasper L. A Hilbert-based method for processing respiratory timeseries. *bioRxiv [Preprint].* (2020). doi: 10.1101/2020.09.30.321562
  101. Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R. Movement-related effects in fMRI time-series. *Magn Reson Med.* (1996) 35:346–55. doi: 10.1002/mrm.1910350312
  102. Ashburner J. Preparing fMRI Data for Statistical Analysis. *Fmri Techniques and Protocols, 2nd Edition.* (2016) 119:155–81. doi: 10.1007/978-1-4939-5611-1\_6
  103. Carp J. The secret lives of experiments: methods reporting in the fMRI literature. *Neuroimage.* (2012) 63:289–300. doi: 10.1016/j.neuroimage.2012.07.004
  104. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage.* (2014) 84:320–41. doi: 10.1016/j.neuroimage.2013.08.048
  105. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. *Neuroimage.* (2012) 62:782–90. doi: 10.1016/j.neuroimage.2011.09.015
  106. Fischl B. FreeSurfer. *Neuroimage.* (2012) 62:774–81. doi: 10.1016/j.neuroimage.2012.01.021
  107. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossai J, et al. Machine learning for neuroimaging with scikit-learn. *Front Neuroinform.* (2014) 8:14. doi: 10.3389/fninf.2014.00014
  108. Strother SC. Evaluating fMRI preprocessing pipelines. *IEEE Eng Med Biol Mag.* (2006) 25:27–41. doi: 10.1109/EMEMB.2006.1607667
  109. Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data.* (2016) 3:160044. doi: 10.1038/sdata.2016.44
  110. Power JD, Plitt M, Laumann TO, Martin A. Sources and implications of whole-brain fMRI signals in humans. *Neuroimage.* (2017) 146:609–25. doi: 10.1016/j.neuroimage.2016.09.038
  111. Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage.* (2013) 80:105–24. doi: 10.1016/j.neuroimage.2013.04.127
  112. Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods.* (2019) 16:111–16. doi: 10.1038/s41592-018-0235-4
  113. Filiou MD, Turck CW. General overview: biomarkers in neuroscience research. *Int Rev Neurobiol.* (2011) 101:1–17. doi: 10.1016/B978-0-12-387718-5.00001-8
  114. Jirsa VK, Proix T, Perdikis D, Woodman MM, Wang H, Gonzalez-Martinez J, et al. The virtual epileptic patient: individualized whole-brain models of epilepsy spread. *Neuroimage.* (2016) 145:377–88. doi: 10.1016/j.neuroimage.2016.04.049
  115. Petzschner FH, Weber LAE, Gard T, Stephan KE. Computational psychosomatics and computational psychiatry: toward a joint framework for differential diagnosis. *Biol Psychiatry.* (2017) 82:421–30. doi: 10.1016/j.biopsych.2017.05.012



116. Friston KJ, Redish AD, Gordon JA. Computational nosology and precision psychiatry. *Comput Psychiatr.* (2017) 1:2–23. doi: 10.1162/cpsy\_a\_00001
117. Bishop CM. *Pattern Recognition and Machine Learning*. New York, NY: Springer. (2006).
118. Penny WD, Stephan KE, Mechelli A, Friston KJ. Comparing dynamic causal models. *Neuroimage.* (2004) 22:1157–72. doi: 10.1016/j.neuroimage.2004.03.026
119. Stephan KE, Schlagenhaut F, Huys QJ, Raman S, Aponte EA, Brodersen KH, et al. Computational neuroimaging strategies for single patient predictions. *Neuroimage.* (2017) 145:180–99. doi: 10.1016/j.neuroimage.2016.06.038
120. Friston KJ. Functional and effective connectivity: a review. *Brain Connect.* (2011) 1:13–6. doi: 10.1089/brain.2011.0008
121. Friston K, Moran R, Seth A. Analysing connectivity with Granger causality and dynamic causal modelling. *Current Opinion Neurobiol.* (2013) 23:172–8. doi: 10.1016/j.conb.2012.11.010
122. Menon V. Large-scale brain networks and psychopathology: a unifying triple network model. *Trends Cogn Sci.* (2011) 15:483–506. doi: 10.1016/j.tics.2011.08.003
123. Friston KJ, Frith CD. Schizophrenia: a disconnection syndrome? *Clin. Neurosci.* (1995) 3:89–97.
124. Bullmore ET, Frangou S, Murray RM. The dysplastic net hypothesis: an integration of developmental and dysconnectivity theories of schizophrenia. *Schizophr Res.* (1997) 28:143–56. doi: 10.1016/S0920-9964(97)00114-X
125. Friston K, Brown HR, Siemerkus J, Stephan KE. The dysconnection hypothesis (2016). *Schizophr Res.* (2016) 176:83–94. doi: 10.1016/j.schres.2016.07.014
126. Kennedy DP, Redcay E, Courchesne E. Failing to deactivate: resting functional abnormalities in autism. *Proc Natl Acad Sci USA.* (2006) 103:8275–80. doi: 10.1073/pnas.0600674103
127. Courchesne E, Pierce K, Schumann CM, Redcay E, Buckwalter JA, Kennedy DP, et al. Mapping early brain development in autism. *Neuron.* (2007) 56:399–413. doi: 10.1016/j.neuron.2007.10.016
128. Muller RA. The study of autism as a distributed disorder. *Ment Retard Dev Disabil Res Rev.* (2007) 13:85–95. doi: 10.1002/mrdd.20141
129. Mayberg HS. Limbic-cortical dysregulation: a proposed model of depression. *J Neuropsychiatry Clin Neurosci.* (1997) 9:471–81. doi: 10.1176/jnp.9.3.471
130. Greicius MD, Flores BH, Menon V, Glover GH, Solvason HB, Kenna H, et al. Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol Psychiatry.* (2007) 62:429–37. doi: 10.1016/j.biopsych.2006.09.020
131. Wang L, Hermens DF, Hickie IB, Lagopoulos J. A systematic review of resting-state functional-MRI studies in major depression. *J Affect Disord.* (2012) 142:6–12. doi: 10.1016/j.jad.2012.04.013
132. Friston K, Harrison L, Penny W. Dynamic causal modelling. *Neuroimage.* (2003) 19:1273–302. doi: 10.1016/S1053-8119(03)00202-7
133. David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ. Dynamic causal modeling of evoked responses in EEG and MEG. *Neuroimage.* (2006) 30:1255–72. doi: 10.1016/j.neuroimage.2005.10.045
134. Daunizeau J, David O, Stephan K. Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *Neuroimage.* (2011) 58:312–22. doi: 10.1016/j.neuroimage.2009.11.062
135. Moran R, Pinotsis DA, Friston K. Neural masses and fields in dynamic causal modeling. *Front Comput Neurosci.* (2013) 7:57. doi: 10.3389/fncom.2013.00057
136. Stephan K, Penny W, Moran R, den Ouden H, Daunizeau J, Friston K. Ten simple rules for dynamic causal modeling. *Neuroimage.* (2010) 49:3099–109. doi: 10.1016/j.neuroimage.2009.11.015
137. Dima D, Roiser JP, Dietrich DE, Bonnemann C, Lanfermann H, Emrich HM, et al. Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. *Neuroimage.* (2009) 46:1180–6. doi: 10.1016/j.neuroimage.2009.03.033
138. Deserno L, Sterzer P, Wüstenberg T, Heinz A, Schlagenhaut F. Reduced prefrontal-parietal effective connectivity and working memory deficits in schizophrenia. *J. Neurosci.* (2012) 32:12–20. doi: 10.1523/JNEUROSCI.3405-11.2012
139. Brodersen KH, Deserno L, Schlagenhaut F, Lin Z, Penny WD, Buhmann JM, et al. Dissecting psychiatric spectrum disorders by generative embedding. *Neuroimage Clin.* (2014) 4:98–111. doi: 10.1016/j.nicl.2013.11.002
140. Lefebvre S, Demeulemeester M, Leroy A, Delmaire C, Lopes R, Pins D, et al. Network dynamics during the different stages of hallucinations in schizophrenia. *Hum Brain Mapp.* (2016) 37:2571–86. doi: 10.1002/hbm.23197
141. Li B, Cui LB, Xi YB, Friston KJ, Guo F, Wang HN, et al. Abnormal effective connectivity in the brain is involved in auditory verbal hallucinations in schizophrenia. *Neurosci Bull.* (2017) 33:281–91. doi: 10.1007/s12264-017-0101-x
142. Grèzes J, Wicker B, Berthoz S, de Gelder. B. A failure to grasp the affective meaning of actions in autism spectrum disorder subjects. *Neuropsychologia.* (2009) 47:1816–25. doi: 10.1016/j.neuropsychologia.2009.02.021
143. Radulescu E, Minati L, Ganeshan B, Harrison NA, Gray MA, Beacher FD, et al. Abnormalities in fronto-striatal connectivity within language networks relate to differences in grey-matter heterogeneity in Asperger syndrome. *Neuroimage Clin.* (2013) 2:716–26. doi: 10.1016/j.nicl.2013.05.010
144. Schlösser RG, Wagner G, Koch K, Dahnke R, Reichenbach JR, Sauer H. Fronto-cingulate effective connectivity in major depression: a study with fMRI and dynamic causal modeling. *Neuroimage.* (2008) 43:645–55. doi: 10.1016/j.neuroimage.2008.08.002
145. Almeida JR, Versace A, Mechelli A, Hassel S, Quevedo K, Kupfer DJ, et al. Abnormal amygdala-prefrontal effective connectivity to happy faces differentiates bipolar from major depression. *Biol Psychiatry.* (2009) 66:451–9. doi: 10.1016/j.biopsych.2009.03.024
146. Woolrich MW, Stephan KE. Biophysical network models and the human connectome. *Neuroimage.* (2013) 80:330–8. doi: 10.1016/j.neuroimage.2013.03.059
147. Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W. Variational free energy and the Laplace approximation. *Neuroimage.* (2007) 34:220–34. doi: 10.1016/j.neuroimage.2006.08.035
148. Wipf D, Nagarajan S. A unified Bayesian framework for MEG/EEG source imaging. *Neuroimage.* (2009) 44:947–66. doi: 10.1016/j.neuroimage.2008.02.059
149. Calderhead B, Girolami M. Estimating Bayes factors via thermodynamic integration and population MCMC. *Comp Statistics Data Analysis.* (2009) 53:4028–45. doi: 10.1016/j.csda.2009.07.025
150. Kirkwood JG. Statistical mechanics of fluid mixtures. *J Chem Phys.* (1935) 3:300–13. doi: 10.1063/1.1749657
151. Lartillot N, Philippe H. Computing Bayes factors using thermodynamic integration. *Systematic Biol.* (2006) 55:195–207. doi: 10.1080/10635150500433722
152. Aponte EA, Raman S, Sengupta B, Penny WD, Stephan KE, Heinzle J. mpdcm: a toolbox for massively parallel dynamic causal modeling. *J Neurosci Methods.* (2016) 257:7–16. doi: 10.1016/j.jneumeth.2015.09.009
153. Sengupta B, Friston KJ, Penny WD. Gradient-free MCMC methods for dynamic causal modelling. *Neuroimage.* (2015) 112:375–81. doi: 10.1016/j.neuroimage.2015.03.008
154. Frässle S, Stephan KE, Friston KJ, Steup M, Krach S, Paulus FM, et al. Test-retest reliability of dynamic causal modeling for fMRI. *Neuroimage.* (2015) 117:56–66. doi: 10.1016/j.neuroimage.2015.05.040
155. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data, Second edition*. ed. Boca Raton: CRC Press, Taylor and Francis Group (2015).
156. Efron B, Morris C. Stein's estimation rule and its competitors - empirical bayes approach. *J Am Stat Assoc.* (1973) 68:117–30. doi: 10.1080/01621459.1973.10481350
157. Kass R, Steffey D. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J Am Stat Assoc.* (1989) 84:717–26. doi: 10.1080/01621459.1989.10478825
158. Gelman A, Charlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. Chapman and Hall (2004). doi: 10.1201/9780429258480
159. Maritz JS, Lwin T. *Empirical Bayes Methods, 2nd ed.* London; New York, NY: Chapman and Hall (1989).
160. Friston KJ, Litvak V, Oswal A, Razi A, Stephan KE, van Wijk BC, et al. Bayesian model reduction and empirical Bayes for group (DCM) studies. *Neuroimage.* (2016) 128:413–31. doi: 10.1016/j.neuroimage.2015.11.015
161. Raman S, Deserno L, Schlagenhaut F, Stephan KE. A hierarchical model for integrating unsupervised generative embedding and empirical Bayes. *J Neurosci Methods.* (2016) 269:6–20. doi: 10.1016/j.jneumeth.2016.04.022

162. Yao Y, Raman SS, Schiek M, Leff A, Frässle S, Stephan KE. Variational Bayesian inversion for hierarchical unsupervised generative embedding (HUGE). *Neuroimage*. (2018) 179:604–19. doi: 10.1016/j.neuroimage.2018.06.073
163. Yao Y, Stephan KE. Markov chain Monte Carlo methods for hierarchical clustering of dynamic causal models. *arXiv*. (2020) 2012.05744. doi: 10.1002/hbm.25431
164. Stephan K, Friston K, Frith C. Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia Bull.* (2009) 35:509–27. doi: 10.1093/schbul/sbn176
165. Frässle S, Lomakina EI, Kasper L, Manjaly ZM, Leff A, Pruessmann KP, et al. A generative model of whole-brain effective connectivity. *Neuroimage*. (2018) 179:505–29. doi: 10.1016/j.neuroimage.2018.05.058
166. Frässle S, Lomakina EI, Razi A, Friston KJ, Buhmann JM, Stephan KE. Regression DCM for fMRI. *Neuroimage*. (2017) 155:406–21. doi: 10.1016/j.neuroimage.2017.02.090
167. Frässle S, Manjaly ZM, Do CT, Kasper L, Pruessmann KP, Stephan KE. Whole-brain estimates of directed connectivity for human connectomics. *Neuroimage*. (2021) 225:117491. doi: 10.1016/j.neuroimage.2020.117491
168. Frässle S, Harrison SJ, Heinze J, Clementz B, Tamminga C, Sweeney J, et al. Regression dynamic causal modeling for resting-state fMRI. *bioRxiv*. (2020). doi: 10.1101/2020.08.12.247536
169. Heinze J, Koopmans PJ, den Ouden HEM, Raman S, Stephan KE. A hemodynamic model for layered BOLD signals. *Neuroimage*. (2016) 125:556–70. doi: 10.1016/j.neuroimage.2015.10.025
170. Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*. (1999) 2:79–87. doi: 10.1038/4580
171. Friston K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*. (2005) 360:815–36. doi: 10.1098/rstb.2005.1622
172. Buxton R, Wong E, Frank L. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med*. (1998) 39:855–64. doi: 10.1002/mrm.1910390602
173. Friston KJ, Mechelli A, Turner R, Price CJ. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, other hemodynamics. *Neuroimage*. (2000) 12:466–77. doi: 10.1006/nimg.2000.0630
174. Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ. Comparing hemodynamic models with DCM. *Neuroimage*. (2007) 38:387–401. doi: 10.1016/j.neuroimage.2007.07.040
175. Havlicek M, Roebroek A, Friston K, Gardumi A, Ivanov D, Uludag K. Physiologically informed dynamic causal modeling of fMRI data. *Neuroimage*. (2015) 122:355–72. doi: 10.1016/j.neuroimage.2015.07.078
176. Havlicek M, Uludag K. A dynamical model of the laminar BOLD response. *Neuroimage*. (2020) 204:116209. doi: 10.1016/j.neuroimage.2019.116209
177. Wilson RC, Collins AG. Ten simple rules for the computational modeling of behavioral data. *Elife*. (2019) 8:e49547. doi: 10.7554/eLife.49547
178. Mathys C, Daunizeau J, Friston KJ, Stephan KE. A bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci*. (2011) 5:39. doi: 10.3389/fnhum.2011.00039
179. Aponte EA, Schöbi D, Stephan KE, Heinze J. The Stochastic Early Reaction, Inhibition, and late Action (SERIA) model for antisaccades. *PLoS Comp Biol*. (2017) 13:e1005692. doi: 10.1371/journal.pcbi.1005692
180. Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, et al. Uncertainty in perception and the Hierarchical Gaussian Filter. *Front Hum Neurosci*. (2014) 8:825. doi: 10.3389/fnhum.2014.00825
181. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF, editors. *Classical Conditioning II: Current Research and Theory*. New York, NY: Appleton Century Crofts (1972). p. 64–99.
182. Sutton RS. Gain adaptation beats least squares. In: *Proceedings of the 7th Yale Workshop on Adaptive and Learning Systems* (1992).
183. Baum LE, Petrie T. Statistical inference for probabilistic functions of finite state markov chains. *Annals Mathematical Statistics*. (1966) 37:1554. doi: 10.1214/aoms/1177699147
184. Stephan K, Penny W, Daunizeau J, Moran R, Friston K. Bayesian model selection for group studies. *Neuroimage*. (2009) 46:1004–17. doi: 10.1016/j.neuroimage.2009.03.025
185. Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HE, et al. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron*. (2013) 80:519–30. doi: 10.1016/j.neuron.2013.09.009
186. Weber LA, Diaconescu AO, Mathys C, Schmidt A, Komater M, Vollenweider FX, et al. Ketamine affects prediction errors about statistical regularities: a computational single-trial analysis of the mismatch negativity. *J Neurosci*. (2020) 40:5658–68. doi: 10.1523/JNEUROSCI.3069-19.2020
187. Heinze J, Aponte EA, Stephan KE. Computational models of eye movements and their application to schizophrenia. *Current Opinion Behav Sci*. (2016) 11:21–9. doi: 10.1016/j.cobeha.2016.03.008
188. Hutton SB, Ettinger U. The antisaccade task as a research tool in psychopathology: A critical review. *Psychophysiology*. (2006) 43:302–13. doi: 10.1111/j.1469-8986.2006.00403.x
189. Rommelse NNJ, Van der Stigchel S, Sergeant JA. A review on eye movement studies in childhood and adolescent psychiatry. *Brain Cognition*. (2008) 68:391–414. doi: 10.1016/j.bandc.2008.08.025
190. Hallett PE. Primary and secondary saccades to goals defined by instructions. *Vision Res*. (1978) 18:1279–96. doi: 10.1016/0042-6989(78)90218-3
191. Fukushima J, Fukushima K, Chiba T, Tanaka S, Yamashita I, Kato M. Disturbances of voluntary control of saccadic eye-movements in schizophrenic-patients. *Biol Psychiatry*. (1988) 23:670–7. doi: 10.1016/0006-3223(88)90050-9
192. Curtis CE, Calkins ME, Grove WM, Feil KJ, Lacono WG. Saccadic disinhibition in patients with acute and remitted schizophrenia and their first-degree biological relatives. *Am J Psychiatry*. (2001) 158:100–6. doi: 10.1176/appi.ajp.158.1.100
193. Reilly JL, Frankovich K, Hill S, Gershon ES, Keefe RSE, Keshavan MS, et al. Elevated Antisaccade Error Rate as an Intermediate Phenotype for Psychosis Across Diagnostic Categories. *Schizophrenia Bull.* (2014) 40:1011–21. doi: 10.1093/schbul/sbt132
194. Aponte EA, Schöbi D, Stephan KE, Heinze J. Computational dissociation of dopaminergic and cholinergic effects on action selection and inhibitory control. *Biol Psychiatry Cogn Neurosci Neuroimaging*. (2020) 5:364–72. doi: 10.1016/j.bpsc.2019.10.011
195. Cooray GK, Sengupta B, Douglas P, Englund M, Wickstrom R, Friston K. Characterising seizures in anti-NMDA-receptor encephalitis with dynamic causal modelling. *Neuroimage*. (2015) 118:508–19. doi: 10.1016/j.neuroimage.2015.05.064
196. Dalmau J, Tuzun E, Wu HY, Masjuan J, Rossi JE, Voloschin A, et al. Paraneoplastic anti-N-methyl-D-aspartate receptor encephalitis associated with ovarian teratoma. *Ann Neurol*. (2007) 61:25–36. doi: 10.1002/ana.21050
197. Shawe-Taylor J, Cristianini N. *Kernel Methods for Pattern Analysis*. Cambridge: Cambridge University Press (2004).
198. Frässle S, Marquand AF, Schmaal L, Dinga R, Veltman DJ, van der Wee NJA, et al. Predicting individual clinical trajectories of depression with generative embedding. *Neuroimage Clin*. (2020) 26:102213. doi: 10.1016/j.nicl.2020.102213
199. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Machine Learn Res*. (2011) 12:2825–30. Available online at: <https://dl.acm.org/doi/10.5555/1953048.2078195>
200. Snoek L, Miletic S, Scholte HS. How to control for confounds in decoding analyses of neuroimaging data. *Neuroimage*. (2019) 184:741–60. doi: 10.1016/j.neuroimage.2018.09.074
201. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *Neuroimage*. (2017) 145:166–79. doi: 10.1016/j.neuroimage.2016.10.038
202. Varoquaux G. Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage*. (2018) 180:68–77. doi: 10.1016/j.neuroimage.2017.06.061
203. Friston K. The free-energy principle: a rough guide to the brain? *Trends Cogn Sci*. (2009) 13:293–301. doi: 10.1016/j.tics.2009.04.005
204. Yu AJ, Dayan P. Uncertainty, neuromodulation, and attention. *Neuron*. (2005) 46:681–92. doi: 10.1016/j.neuron.2005.04.026
205. Iglesias S, Kasper L, Harrison SJ, Manka R, Mathys C, Stephan KE. Cholinergic and dopaminergic effects on prediction error and uncertainty responses during sensory associative learning. *Neuroimage*. (2020) 226:117590. doi: 10.1016/j.neuroimage.2020.117590



206. Takahashi YK, Batchelor HM, Liu B, Khanna A, Morales M, Schoenbaum G. Dopamine neurons respond to errors in the prediction of sensory features of expected rewards. *Neuron*. (2017) 95:1395–405 e1393. doi: 10.1016/j.neuron.2017.08.025
207. Stalnaker TA, Howard JD, Takahashi YK, Gershman SJ, Kahnt T, Schoenbaum G. Dopamine neuron ensembles signal the content of sensory prediction errors. *Elife*. (2019) 8:e49315. doi: 10.7554/eLife.49315.010
208. Suarez JA, Howard JD, Schoenbaum G, Kahnt T. Sensory prediction errors in the human midbrain signal identity violations independent of perceptual distance. *Elife*. (2019) 8:e43962. doi: 10.7554/eLife.43962.032
209. Friston KJ. Hallucinations and perceptual inference. *Behav Brain Sci*. (2005) 28:764. doi: 10.1017/S0140525X05290131
210. Pellicano E, Burr D. When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends Cognitive Sci*. (2012) 16:504–10. doi: 10.1016/j.tics.2012.08.009
211. Schofield TM, Penny WD, Stephan KE, Crinion JT, Thompson AJ, Price CJ, et al. Changes in auditory feedback connections determine the severity of speech processing deficits after stroke. *J Neurosci*. (2012) 32:4260–70. doi: 10.1523/JNEUROSCI.4670-11.2012
212. Harrison OK, Nanz L, Marino S, Luchinger R, Hennel F, Hess AJ, et al. Interoception of breathing and its relationship with anxiety. *bioRxiv [Preprint]*. (2021). doi: 10.1101/2021.03.24.436881
213. Harrison OK, Garfinkel SN, Marlow L, Finnegan S, Marino S, Nanz L, et al. The Filter Detection Task for measurement of breathing-related interoception and metacognition. *bioRxiv [Preprint]*. (2020). doi: 10.1101/2020.06.29.176941
214. Rieger SW, Stephan KE, Harrison OK. Remote, automated, and MRI-compatible administration of interoceptive inspiratory resistive loading. *Front Hum Neurosci*. (2020) 14:161. doi: 10.3389/fnhum.2020.00161
215. Aponte EA, Stephan KE, Heinzle J. Switch costs in inhibitory control and voluntary behaviour: a computational study of the antisaccade task. *Eur J Neurosci*. (2019) 50:3205–20. doi: 10.1111/ejn.14435
216. Aponte EA, Tschan DG, Stephan KE, Heinzle J. Inhibition failures and late errors in the antisaccade task: influence of cue delay. *J Neurophysiol*. (2018) 120:3001–16. doi: 10.1152/jn.00240.2018
217. Aponte EA, Yao Y, Raman S, Frässle S, Heinzle J, Penny WD, et al. An introduction to thermodynamic integration and application to dynamic causal models. *bioRxiv*. (2020). doi: 10.1101/2020.12.21.423807
218. Birn RM, Smith MA, Jones TB, Bandettini PA. The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes in respiration. *Neuroimage*. (2008) 40:644–54. doi: 10.1016/j.neuroimage.2007.11.059
219. Brazil IA, Mathys CD, Popma A, Hoppenbrouwers SS, Cohn MD. Representational uncertainty in the brain during threat conditioning and the link with psychopathic traits. *Biol Psychiatry Cogn Neurosci Neuroimaging*. (2017) 2:689–95. doi: 10.1016/j.bpsc.2017.04.005
220. Daunizeau J, den Ouden HE, Pessiglione M, Kiebel SJ, Friston KJ, Stephan KE. Observing the observer (II): deciding when to decide. *PLoS ONE*. (2010) 5:e15555. doi: 10.1371/journal.pone.0015555
221. Daunizeau J, den Ouden HE, Pessiglione M, Kiebel SJ, Stephan KE, Friston KJ. Observing the observer (I): meta-bayesian models of learning and decision-making. *PLoS ONE*. (2010) 5:e15554. doi: 10.1371/journal.pone.0015554
222. Durstewitz D, Koppe G, Meyer-Lindenberg A. Deep neural networks in psychiatry. *Mol. Psychiatry*. (2019) 24:1583–98. doi: 10.1038/s41380-019-0365-9
223. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. (2018) 14:91–118. doi: 10.1146/annurev-clinpsy-032816-045037
224. Gillan CM, Whelan R. What big data can do for treatment in psychiatry. *Current Opinion Behav Sci*. (2017) 18:34–42. doi: 10.1016/j.cobeha.2017.07.003
225. Marshall L, Mathys C, Ruge D, de Berker AO, Dayan P, Stephan KE, et al. Pharmacological fingerprints of contextual uncertainty. *PLoS Biol*. (2016) 14:e1002575. doi: 10.1371/journal.pbio.1002575
226. Mourao-Miranda J, Reinders AATS, Rocha-Rego V, Lappin J, Rondina J, et al. Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. *Psychol Med*. (2012) 42:1037–47. doi: 10.1017/S0033291711002005
227. Portugal LCL, Schrouff J, Stiffler R, Bertocci M, Bebko G, Chase H, et al. Predicting anxiety from wholebrain activity patterns to emotional faces in young adults: a machine learning approach. *Neuroimage-Clinical*. (2019) 23:101813. doi: 10.1016/j.nicl.2019.101813
228. Schmaal L, Marquand AF, Rhebergen D, van Tol MJ, Ruhé HG, van der Wee NJ, et al. Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: a multivariate pattern recognition study. *Biol Psychiatry*. (2015) 78:278–86. doi: 10.1016/j.biopsych.2014.11.018
229. Sengupta B, Friston KJ, Penny WD. Gradient-based MCMC samplers for dynamic causal modelling. *Neuroimage*. (2016) 125:1107–18. doi: 10.1016/j.neuroimage.2015.07.043
230. Vossel S, Bauer M, Mathys C, Adams RA, Dolan RJ, Stephan KE, et al. Cholinergic stimulation enhances Bayesian belief updating in the deployment of spatial attention. *J Neurosci*. (2014) 34:15735–42. doi: 10.1523/JNEUROSCI.0091-14.2014

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Frässle, Aponte, Bollmann, Brodersen, Do, Harrison, Harrison, Heinzle, Iglesias, Kasper, Lomakina, Mathys, Müller-Schrader, Pereira, Petzschner, Raman, Schöbi, Toussaint, Weber, Yao and Stephan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.