Statistical power: implications for planning MEG studies

Maximilien Chaumon[1], Aina Puce[2], Nathalie George[1]

1. *Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS UMR 7225, Sorbonne Université, Centre MEG-EEG, Centre de NeuroImagerie Recherche (CENIR), Paris, France.*
2. *Department of Psychological & Brain Sciences, Programs in Neuroscience, Cognitive Science, Indiana University, 1101 East 10th St, Bloomington, IN 47405, USA.*

Corresponding author:
Maximilien Chaumon
Institut du Cerveau et de la Moelle Epinière
Hôpital de la Pitié-Salpêtrière
47 Boulevard de l'hôpital
CS 21414
75646 PARIS CEDEX 13 France

Email: maximilien.chaumon@icm-institute.org

Running title: Statistical power in MEG studies

Abstract

Statistical power is key for robust, replicable science. Here, we systematically explored how numbers of trials and subjects affect statistical power in MEG sensor-level data. More specifically, we simulated "experiments" using the MEG resting-state dataset of the Human Connectome Project (HCP); we divided the data in two conditions, injected a dipolar source at a known anatomical location in one condition, but not the other, and detected significant differences at sensor level with classical paired t-tests across subjects.

The group-level detectability of these simulated effects varied drastically with their anatomical origin. We thus examined in detail which spatial properties of the sources affected detectability, looking specifically at the distance and orientation of the source with respect to the closest sensor, and at the variability of these parameters across subjects. In line with previous single-subject studies, we found that the most detectable effects originate from source positions that are closest to the sensors and oriented tangentially with respect to the sensor array. In addition, cross-subject variability in orientation also affected group-level effect detectability, boosting detection in regions where this variability was small and hindering detection in regions where it was large. Incidentally, we observed a considerable covariation of source position, orientation, and their cross-subject variability in the individual brain anatomical space, making it difficult to assess the impact of each of these variables independently of one another. Therefore, we performed an additional set of simulations with the same simulation strategy as before, but this time, controlling spatial properties independently of individual anatomy. This allowed us to disentangle the effects of position, orientation, and cross-subject variability of the sources on detectability, and confirmed the strong impact of distance and orientation on detectability. Moreover, it emphasized that

orientation variability across subjects also affects detectability, whereas position variability does not.

Importantly, our study indicates that strict unequivocal recommendations as to the ideal number of trials and subjects for any experiment cannot be realistically provided for neurophysiological studies. Rather, it highlights the importance of considering the spatial constraints underlying expected sources of activity while planning experiments. This is particularly true when performing network level analyses in cognitive neuroscience studies, where nodes in a network shall not be considered equally detectable. Finally, we note that this variation in detectability may also account, in part, for the over-representation of some fields of study over others in the MEG domain (e.g. sensorimotor and auditory studies).

Keywords

statistical power, MEG, source modelling, simulation, distance, orientation

Introduction

Statistical power is a key component of robust, replicable science. An important variable affecting statistical power is sample size and it has been shown that previous studies have been undermined by too small sample sizes (Button et al., 2013; Szucs & Ioannidis, 2017). Additionally, suboptimal scientific practices such as experimental designs and analysis approaches inappropriate to answering the posed scientific question have been recently highlighted (Gelman & Loken, 2013; Kerr, 1998; Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Luck & Gaspelin, 2017). Overall, data gathering and analytical procedures that once were widely used are now being identified as being flawed, while reporting procedures are codified (Keil et al., 2014), and their endorsement critically assessed (Clayson, Carbine, Baldwin, & Larson, 2019; Larson & Carbine, 2017). Emphasis is being put on improving the robustness of statistical inference (Groppe, 2017; Kappenman & Keil, 2017; Kilner, 2013; Luck & Gaspelin, 2017; Simonsohn, Nelson, & Simmons, 2013), as well as on the design, preparation and documentation of carefully planned experiments (Chambers, Dienes, McIntosh, Rotshtein, & Willmes, 2015; Foster & Deardorff, 2017; Luck, 2005). The current study targets the confluence of these two trends, aiming to aid researchers in cognitive, affective, and social neuroscience make principled decisions on how many trials and subjects to include in their experiments to achieve adequate levels of statistical power. More precisely, we highlight some important variables that one should pay attention to when considering how many repetitions of experimental conditions and how many subjects one should test to achieve robust statistical inference in an MEG experiment.

The question of knowing how many trials and subjects to include in an MEG or EEG experiment has to date been mostly a matter of “know-how” or “rules of thumb” (Gross et al., 2013; Luck, 2012). Indeed, this topic has been discussed for decades without reaching any

definitive conclusion (Pivik et al., 1993; Picton et al., 2000; Duncan et al., 2009; Keil et al., 2014; Kane et al., 2017; Hari et al., 2018). However, the above-mentioned concerns about power and reproducibility call for a systematic evaluation of variation in statistical power. Which variables critically affect statistical power? Understanding these, how can the necessary (and sufficient) number of trials be planned in advance? Recently, Boudewyn and colleagues took a first step at answering this question in a principled manner. They used EEG recordings from 40 participants to examine how the number of observations included in their analysis affected the probability of finding a significant effect in Event-Related Potential (ERP) measures. As expected, large effects (e.g., an error-related negativity, which produces a 5-15 μV difference wave in the EEG) were detected at sensor level with fewer trials and participants than smaller effects were (e.g., a finer ~ 1 μV amplitude modulation in the lateralized readiness potential). We believe Boudewyn et al. (2018) to be the first EEG study that directly related the number of observations (trials and subjects) to statistical power. In a very recent study, Baker and colleagues (2019) took this approach one step further, by acknowledging the difference between within-sample (i.e. inter-trial) and between-sample (i.e. between subjects) variability in a number of assessment modalities, including MEG and EEG. They introduced so-called "power contours"—plots that depict different levels of statistical power as a relationship between number of trials against number of subjects. These power contours reveal the level of statistical power reached for given trial and subject numbers. We illustrate the findings of the present study using this plotting approach.

Here, we used MEG data from a large sample of subjects to reliably visualize the behavior of power measures. We simulated “experiments” with the large open resting-state MEG dataset of the Human Connectome Project (HCP, Larson-Prior et al., 2013), varying the numbers of trials and subjects to assess the effects of within- and between-subject sample sizes, respectively. This allowed us to simulate power in the context of real physiological

background activity. We simulated effects by injecting dipolar sources of fixed amplitude at known anatomical locations in one half the data and not in the other half (Figure 1). We then detected the effects observed at sensor level with classical paired t-tests across subjects. In an initial exploration, we used detailed individual source models and observed that detectability varies drastically according to the anatomical location of the sources. Accordingly, we explored different spatial properties at each location, namely their distance and orientation with respect to the closest sensor, and the cross-subject variability of these parameters. For this, we undertook two types of simulations. First, we examined changes of detectability across different anatomical areas. In doing so, we observed that the measured properties were impossible to properly disentangle due to anatomical constraints. Therefore, we ran a second set of simulations, where we imposed spatial properties of the sources independently of individual brain anatomy while using the same simulation strategy as in the first step. In this way, we could assess separately the effect of each spatial property of the sources (position, orientation, and their cross-subject variability) on detectability.

Material & Methods

Software

All data were processed using MATLAB (The Mathworks, Natick, MA, USA; Version 2018a) and the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011, github release of October 25 2019). All analysis and visualization scripts and toolboxes, including the FieldTrip toolbox, and megconnectome toolbox used, are available online at https://gitlab.com/icm-institute/meg/solid_meg.

Data

Input data

We used MEG data from 89 subjects from the Human Connectome Project (HCP) open data repository, described in detail in Larson-Prior et al. (2013). Of the 95 original subjects in the latest release (V3.0) of the MEG data, 6 subjects with missing data in one or another of the components described below were discarded, leaving the final number of 89 subjects for our analyses.

A whole series of preprocessing steps are already performed on the HCP data. For complete details, see the descriptions for the pre-processing pipelines for MEG (Larson-Prior et al., 2013), structural MRI (Glasser et al., 2013), as well as the online resources available at https://www.humanconnectome.org/software. We henceforth refer to these two preprocessing pipelines as the HCP-MEG pipeline and the HCP-structural pipeline. Note that these pipelines are part of different releases of the HCP data that were downloaded separately from https://db.humanconnectome.org. Subject identifier codes being unique across these releases, we used them to merge data from both pipelines.

Briefly, these MEG data were acquired at 2043.5101 Hz (bandpass filtering: DC-400 Hz) with a whole-head 248 magnetometer device (MAGNES 3600, 4D Neuroimaging, San Diego, CA). The HCP includes data from several activation tasks and three 6-minute resting state periods (eyes open, fixating). We used the latter in this study. MEG fiducials were obtained with a Polhemus system for coregistration to structural MRI data. Each subject was scanned with a high-resolution structural MRI on a 3T Siemens Skyra MRI scanner (Van Essen et al., 2012). To define anatomical labels in our study, we used the output of the HCP-structural pipeline, where a high-definition segmentation and anatomical labelling (Destrieux Atlas) of the

cortical mantle was performed using Freesurfer (Version 5.2, Fischl, 2012). Within the HCP-MEG pipeline, the structural MRI was used to create a single-shell volume conduction model using FieldTrip.

All of the material described above is available from the HCP database (individual MEG sensor space time courses, magnetometer definition, individual source space and individual head model) and formed the input data for our study.

Head model, source space, and forward model

We used the head models as provided for each subject by the HCP-MEG pipeline. We used the 4D Neuroimaging/BTi coordinate system to identify source positions within these models throughout the paper[1] (with mm units for clarity). We created a forward leadfield with FieldTrip using the magnetometer description and the provided head models, and generated the sensor signals by projecting sources (see below) with a 10 nA.m amplitude through the leadfield to the sensors (Figure 1a).

We used two different types of source models, where the source dipoles were either constrained to be orthogonal to the individual cortical mantle, or free from any anatomical constraint, as explained below.

Source models constrained by anatomy

We performed a first set of simulations constrained by individual subjects' anatomy, where the signal to be added to the resting-state data was generated from one source (i.e. one vertex), in the mesh provided by the HCP. This source was oriented orthogonally to the cortical mesh. Across participants, we identified sources of similar areal origin by means of their index in the cortical mesh provided by the HCP. It is worth mentioning that we do not

[1] Described here: http://www.fieldtriptoolbox.org/faq/how_are_the_different_head_and_mri_coordinate_systems_defined/#details-of-the-4dbti-coordinate-system

imply that these indices map strictly homologous portions of cortex across subjects. The mesh topology is however identical in all 89 subjects (all 64,984 vertices are connected in the same way in every subject), and it is used in practice to make cross-subject source-level comparisons and averaging (Fischl, 2012). No within-subject variability other than the one already present in the resting-state data was added. Between-subject variability thus occurred because of variations in position and orientation of the source vertices across subjects.

For high-resolution rendering of the cortical surfaces in Figures, we used the high-definition segmentation with 163,842 vertices per hemisphere found in the HCP-structural pipeline (Glasser et al., 2013). These segmentations are linearly coregistered to MNI coordinates (same origin, scale, and orientation), but not "warped" (i.e. non-linearly transformed to closely map to the MNI template brain) as is often performed for mapping individual source reconstructions across participants in MEG. This non-linear mapping would have been inappropriate here since we were interested in inter-subject anatomical variability.

Note that we ignore any explicit effect of signal amplitude in this study. All our brain anatomy-constrained simulations use an equivalent current dipole of 10 nA.m as the source of difference between the two conditions. In a recent study, Murakami and Okada (2015) showed that current density due to local neuronal currents, expressed as equivalent current dipole per unit of surface (in nA.m/mm^2), is remarkably constant throughout a range of vertebrate species. They argued that a value of current dipole moment density of 0.16-0.77 nA.m/mm^2 may serve as an effective physiological constraint for human neocortical MEG/EEG sources. As an indication, according to these estimates the current dipole value of 10 nA.m we chose in this study could correspond to an activation surface ranging between 13 and 62 mm^2. However, complexities arise due to the spatial extent of the sources (e.g. changing orientations around convexities of the cortical mantle). Accounting for these

complexities is beyond the scope of this paper, but it has been tackled in previous studies (Ahlfors, Han, Lin, et al., 2010; Fuchs, Kastner, Tech, Wagner, & Gasca, 2017).

Source models unconstrained by anatomy

As will become clear, measures of source properties that affect detectability tend to covary in ways that are not readily predictable. We thus also used a simulation approach where dipoles were placed independently of cortical anatomy. For these simulations, we took an initial source located around the precentral region (x=-12, y=33, z=70; source position illustrated below in Figure 8), oriented it normally to the mean cortical surface at that location, and systematically varied the spatial source properties starting from that location (see "Source properties description", below).

All other parameters of the simulations were kept identical to those of the anatomically constrained simulations described above, except that the amplitude of the injected sources was reduced to a 5 nA.m to avoid strong saturation of the power contour plots due to the very weak variability in simulations where individual parameters were held constant (see details below).

Simulations

In all of our simulations, we used the same procedure based on a Monte Carlo resampling strategy. The Monte Carlo procedure uses repeated random selection from a sample of data to approximate a property of the general population, that is, here, the probability of finding a statistically significant effect in the general population (aka. statistical power).

We first used this procedure to get a general estimate of statistical power variability throughout the cortex for an example sample size. We then used this strategy again to document the impact of several spatial properties of the sources, simulating each time

experiments with a range of trial and subject numbers. In these latter simulations, the source dipole was either realistically constrained to be orthogonal to the individual cortical mantle, or free from any anatomical constraint in order to assess each spatial property of the source independently of one another.

Procedure: Monte Carlo simulations

In all our simulations, for a given number of subjects and trials, we first randomly chose the required number of subjects from the HCP database, sampling without replacement. For each subject, we then randomly selected our 'trials'. Each trial consisted in 25-ms (50 samples) time segments of the continuous resting state data, at least 2 s apart from each other. We then split these trials randomly in two sets of equal size, added signal (according to the procedure explained above) in one set and averaged the data across time points and trials. We then ran a paired t-test at each sensor across subjects and noted significance ($p<0.05$, uncorrected) for each sensor. In order to summarize the whole MEG helmet with one value, we used the state-of-the-art cluster-mass based correction for multiple comparisons (Maris & Oostenveld, 2007, 1000 permutations, cluster and significance thresholds both at $p<0.05$) and considered a given comparison significant only when the peak sensor (i.e. the sensor at which the projection of the source signal peaked) was included in a significant cluster. We repeated this procedure 500 times for each trial-by-subject number pair. This so-called Monte Carlo statistical power estimate approximates the probability of finding a significant effect if we were to run an experiment with the given parameters (number of subjects, trials and signal properties) in the general population. In result figures, we represent Monte Carlo power estimates as power contours (Baker et al., 2019).

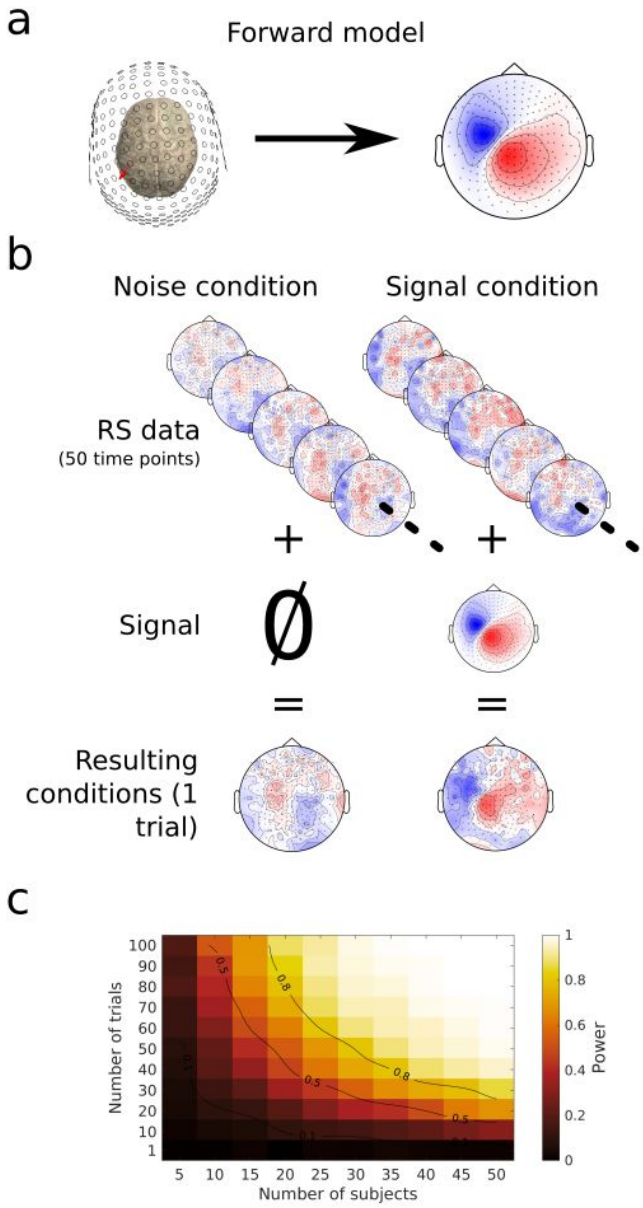

Figure 1. Schematic of simulation method. a. Forward model. *Head and source model provided by the HCP allowed a hypothetical signal to be projected from a given source to MEG sensors by using a forward model.* ***b. Example of simulated data for one subject and one trial.*** *Twenty-five ms (50 time points) of resting state data were extracted, averaged, and added to the forward modeled signal in the signal condition (the amplitude of the signal is exaggerated here for clarity). This procedure was run 1 to 100 times per subject on 5 to 50 subjects picked at random from the available pool of 89 in the HCP MEG database. We computed the average across trials within subject, and performed two-tailed paired t-tests across subjects between the two conditions. We repeated this procedure 500 times per Ntrials x Nsubject pairs to obtain a so-called Monte Carlo estimate of power for each sensor.* ***c. Power estimates and contours.*** *These estimates were plotted for each simulated experiment (number-of-subjects by number-of-trials pair), with spline-interpolated power values shown also as contour lines for visualization purposes. These plots are used throughout this paper to illustrate the detectability of a given effect.*

In an initial simulation, we assessed overall variations in statistical power for simulated sources throughout the brain with a given number of subjects and trials (25 subjects, 50 trials each). This is reported in the initial results section, 'Overall variations of statistical power throughout the brain'. In subsequent simulations we ran Monte Carlo simulations for all combinations of trial and subject numbers, from a single trial to 100 trials (in steps of 10) in each condition, and with 5 to 50 subjects (in steps of 5) in order to assess the impact of different source properties on statistical power, as described below.

Source properties description

We used the simulation approach described above to evaluate how changes in the spatial properties of the source dipole affect signal detection at sensor level. We examined what we will call first- and second-order properties of the sources. First-order properties refer to the position and orientation parameters of the sources. For better interpretability, we define these first-order properties in relation to the sensor array. We defined position as the distance between the source and the closest sensor, and orientation as the orientation of the source with respect to the closest sensor. Second-order properties refer to the variability in first-order properties across subjects. We examined separately the cross-subject variability in position, and the cross-subject variability in orientation, in two sets of simulations. Each property was first examined in simulations using individual brain anatomy, then manipulated in simulations unconstrained by anatomy.

First-order properties: position and orientation of the source with respect to the sensor array

The following two measures allowed us to explore whether and how source detectability depends on their distance and orientation with respect to the sensor array.

Distance to the closest sensor

The distance to the closest sensor was measured for a given source in each subject on their individual anatomy. The average of this distance measured in individual subjects is shown in Figure 3, in the Results section below. In the simulations not constrained by anatomy, we imposed this distance by shifting the position of the initial source (described in the "Models unconstrained by anatomy" section above) by a target distance inwards, or outwards, along the radius passing through the sensor closest to the initial source (labeled A44) and the origin of the coordinate system, between the ears, so that the distance to the closest sensor ranged from 20 to 100 mm.

Orientation with respect to the closest sensor

The orientation with respect to the closest sensor was computed as the arccosine of the product of the orientation vector of the sources (normal to the cortex in simulations using individual brain anatomy) with the orientation vector of the sensor. The average of this orientation measured on individual brains is shown in Figure 7 in the Results section below. This orientation was imposed in simulations not taking anatomy into account by rotating dipoles around the y-axis (passing through both ears). Note that this measure is close to a measure of the extent to which sources may be considered radial versus tangential to the sphere encompassing the sensors' helmet for the initial source considered. This measure ranged between 0 degrees for sources pointing directly at the sensor, up to 90 degrees for sources pointing orthogonal to the sensor (towards the front).

Second-order properties: variability of the sources across subjects

Variability in source position and orientation are thought to decrease detectability at the group level, and were thus examined here.

Position variability

Position variability was defined as the standard deviation of the position of each source across subjects (Figure 11). In simulations not taking anatomy into account, this standard deviation varied in the 0-10 mm range and was sampled at random from a trivariate normal distribution.

Orientation variability

Orientation variability was measured for each source location as the length of the averaged orientation vector across subjects (Figure 15). In simulations not taking anatomy into account, orientation variability was imposed by adding a random bivariate (azimuth and elevation) normally distributed angle to individual sources with a standard deviation that ranged from 0 to 180 degrees.

Results

Overall variations of statistical power throughout the brain

Before getting into the description of the explored spatial properties, we describe here the overall variations of statistical power for detecting sources throughout the brain. As an initial illustration, we calculated the average statistical power for detecting a source at each cortical location for a highly detailed source model following individual brain anatomy for a “classical” dataset of 25 human subjects and 50 trials per condition.

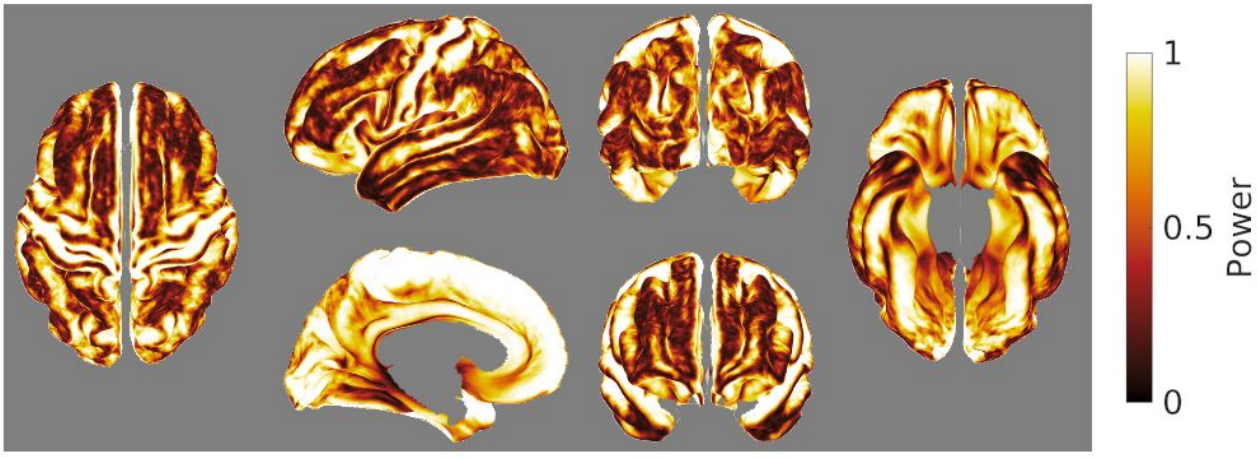

Figure 2. Estimated statistical power across the cortical surface. *Estimated statistical power for detecting (at MEG sensor level) the activity generated by a single dipole placed in turn at every possible neocortical position in a simulated experiment with 25 subjects with 50 trials each. Images of superior, left lateral, left medial, anterior, posterior, and inferior brain views are depicted. Note the higher power values within major sulci, as well as on the medial wall.*

Figure 2 shows the results of this initial analysis. Note that the absolute statistical power values *per se* are somewhat arbitrary, because we simulated the effect of dipoles with a given, constant amplitude. Yet, what is important, is the striking heterogeneity of statistical power across the brain that this figure reveals. Some of this heterogeneity may appear trivial with regard to what is known on MEG signal sensitivity to source orientation and depth. For instance, sources at gyral crests tend to be detected with very low statistical power, probably because they are radially oriented. In contrast, sources on the walls of the central sulcus, and on the medial surface of the frontal lobe - which are tangentially oriented - are detected with high power. However, some other observations may appear to be far less intuitive. For instance, the ventral surface of the temporal lobes shows relatively high power in comparison to the lateral surfaces, even though it is further away from the sensors. This observed heterogeneity of statistical power across the brain calls for better understanding of the influence of source properties on detectable signals at the sensor level. For this, we examined

the first- and second-order spatial properties of the sources, i.e. the spatial parameters defining source position and orientation, and their variability across subjects.

In the following sections, for each examined property, the results unfold as follows. We first describe how the property varies across the entire brain. We then use two example locations in the brain to illustrate realistic variations in this property. These realistic variations being always concomitant with changes in other properties, we then finish with simulations unconstrained by anatomy exploring an imposed range of values to evaluate the selective impact of this property on statistical power.

Distance to sensors

Figure 3a shows the average distance to the closest sensor for every location on the cortical surface, plotted on the average brain of the 89 subjects in the HCP data. The histogram of this distance across all vertices is plotted in Figure 3b. This distance varies from about 4 mm in cortical regions close to the sensor array (central sulcus, occipital cortex) to about 10 mm for the deepest structures.

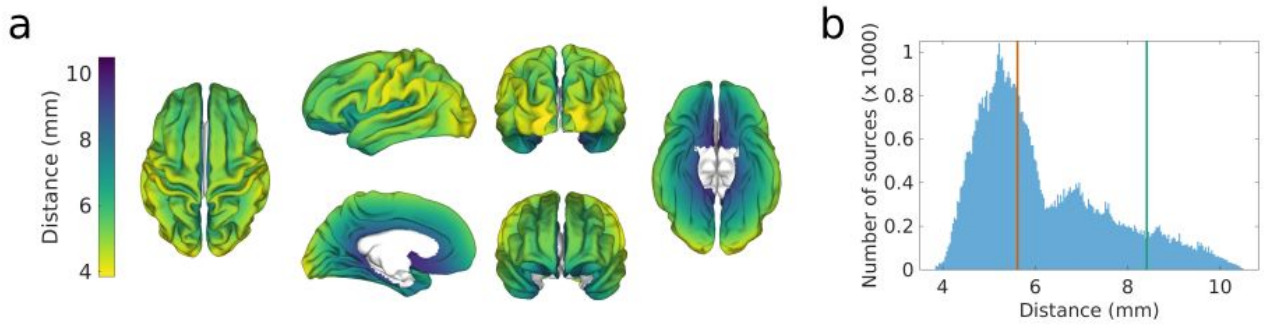

Figure 3. Mean distance to closest sensor across subjects. *Data calculated from 89 subjects in the HCP MEG dataset and presented on the averaged brain.* ***a. Mean distance to the closest sensor.*** *Mapped on average brain views arranged from left to right and top to bottom: superior, left lateral, left medial, posterior, anterior, and inferior.* ***b. Histogram of the distribution of the mean distance to sensors across all cortical vertices.*** *The red and green vertical lines indicate the*

distance to sensors of the two dipoles that we selected to illustrate the effect of this parameter in Figures 4 and 5, respectively.

Based on physics, the distance to the sensors is expected to have a large effect on signal amplitude, and thus on detectability at sensor level (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993; Hillebrand & Barnes, 2002; J. Malmivuo, Suihko, & Eskola, 1997). We now illustrate how group-level statistical power changes with the numbers of subjects and trials for a 10 nA.m dipole placed in two locations - one in the middle frontal gyrus and the other in the orbitofrontal gyrus.

The first location in the middle frontal gyrus (x=71, y=21, z=41 mm) is at a distance of 5.6 cm from the closest sensor (on average across subjects). Figure 4 shows simulation results for this dipole. An identical figure structure will be repeated for illustration of other source locations. Figure 4a displays the chosen dipoles across the 89 subjects of the HCP MEG dataset as well as the average dipole across all subjects. The position variability across subjects can be well seen in these top panels, as well as in Figure 4b, which shows the histogram of dipole positions relative to the average position across subjects, indicating that dipoles were spread within +/- 10 mm from the average position in most subjects. Similarly, Figure 4c shows the distribution of dipole orientations with respect to the closest sensor across subjects. This plot indicates that the orientations for these middle frontal gyrus dipoles were on average 69° away from the radial orientation (i.e. away from the orientation that points exactly in the same direction as the closest sensor). Figure 4d shows the average projection of this dipole to the sensor array, revealing a high signal amplitude at frontal sensors. This amplitude, when added to the resting state data could be detected with a standard paired t-test across subjects (corrected for multiple comparisons across sensors with a nonparametric permutation test using 1000 permutations) against resting state data alone, with the estimated statistical

power level shown in Figure 4e for varying numbers of subjects and trials. In this particular case, achieving 80% power to detect the dipole signal required at least 20 subjects with ~50 trials per subject, or up to 50 subjects with only 10 trials per subject.

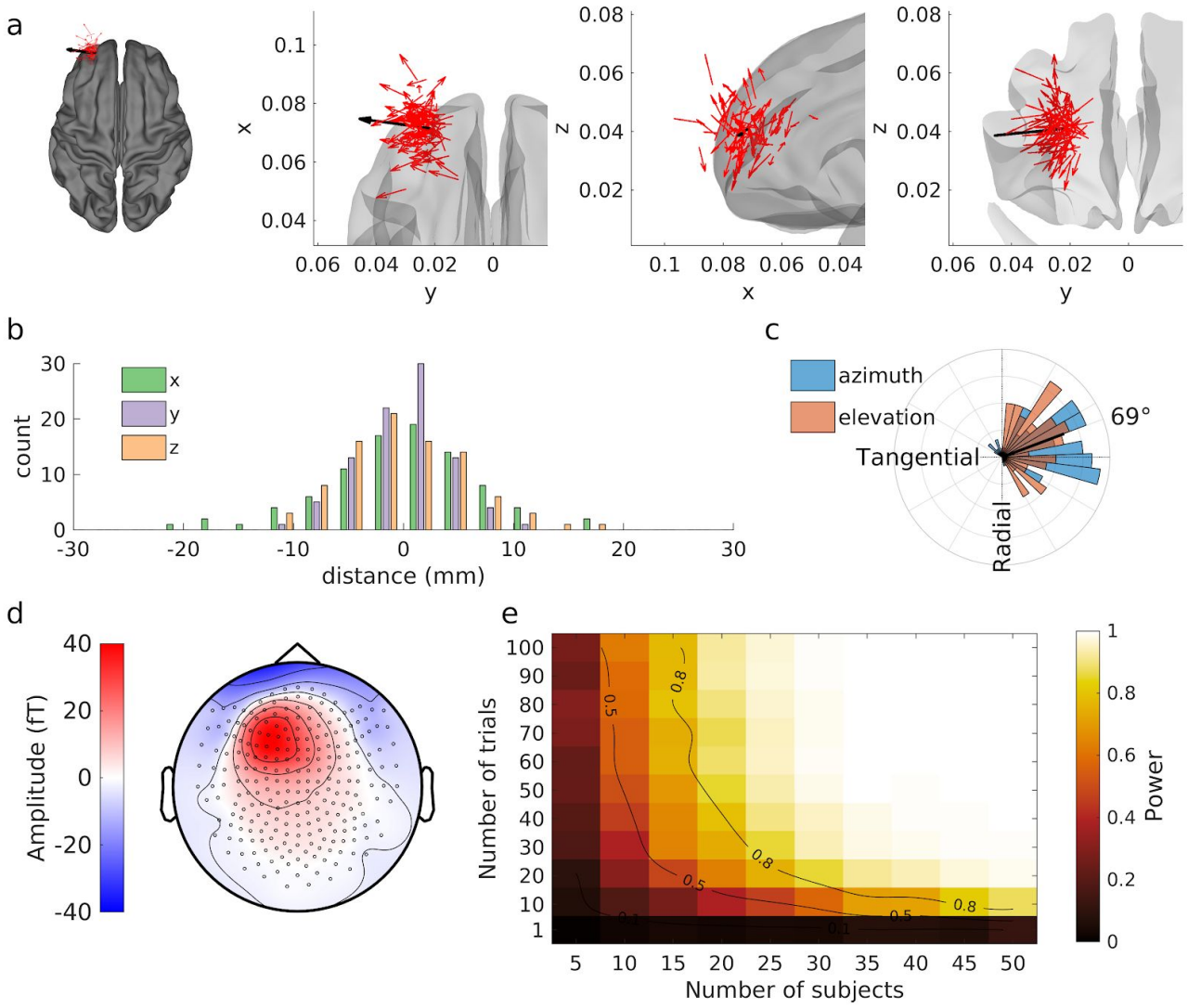

Figure 4. Detecting a source close to the sensors (56 mm to closest) in the middle frontal gyrus. a. *positions of individual dipoles in 89 subjects (red arrows) are represented on the average brain in a superior overall view, and zoomed superior, lateral, and posterior partial brain views. The bold black arrow in each panel shows the resultant vector or average dipole across 89 subjects.* ***b. Histogram of the distance of each dipole to the mean position of all dipoles in the three cartesian dimensions. c. Polar histogram of the orientations of each dipole relative to the orientation of the closest sensor.*** *Azimuth and elevation are referenced to the sensor orientation vector, so an orientation orthogonal to that of the closest sensor is shown as tangential (90°), and the parallel orientation is shown as radial (0°). The black line coming outwards from the center of the plot represents the orientation and length of the resultant vector scaled so that a resultant vector of length 1 (if all dipoles are strictly collinear and pointing in the same direction) spans the whole radius of the plot.* ***d. Average projection of the dipoles in sensor space.*** *Black circles identify sensor positions on the topographical view where the nose is at the top of the plot, and left*

hemisphere appears on the left. The color bar indicates the strength of magnetic field exiting (red) and entering (blue) the head. ***e. Monte Carlo simulations at this location.*** *Color represents estimated statistical power, i.e. the number of significant tests divided by the number of simulations (500) for a given combination of trials and subjects. Black isocontour lines on the plots highlight interpolated power estimates of 0.5 and 0.8.*

After examining the effects on statistical power of this relatively close-to-sensors dipole, we now turn to the detection of a deeper dipole, situated in the orbitofrontal gyrus (Figure 5, x=44, y=13, z=7), at a distance of 8.4 cm from the closest sensor (on average). Compared to the previous location just described (Figure 4), in spite of its slightly more consistent location (Figure 5b) and orientation (Figure 5c) across subjects, and an almost tangential orientation (85° away from the radial orientation), the signal from this dipole is projected with a much weaker amplitude to the sensor array (Figure 5d), is less well detected than the previous dipole (Figure 5e). At least 25 subjects with 90 trials each, or 50 subjects with at least 40 trials each would be necessary to detect this dipole with 80% statistical power.

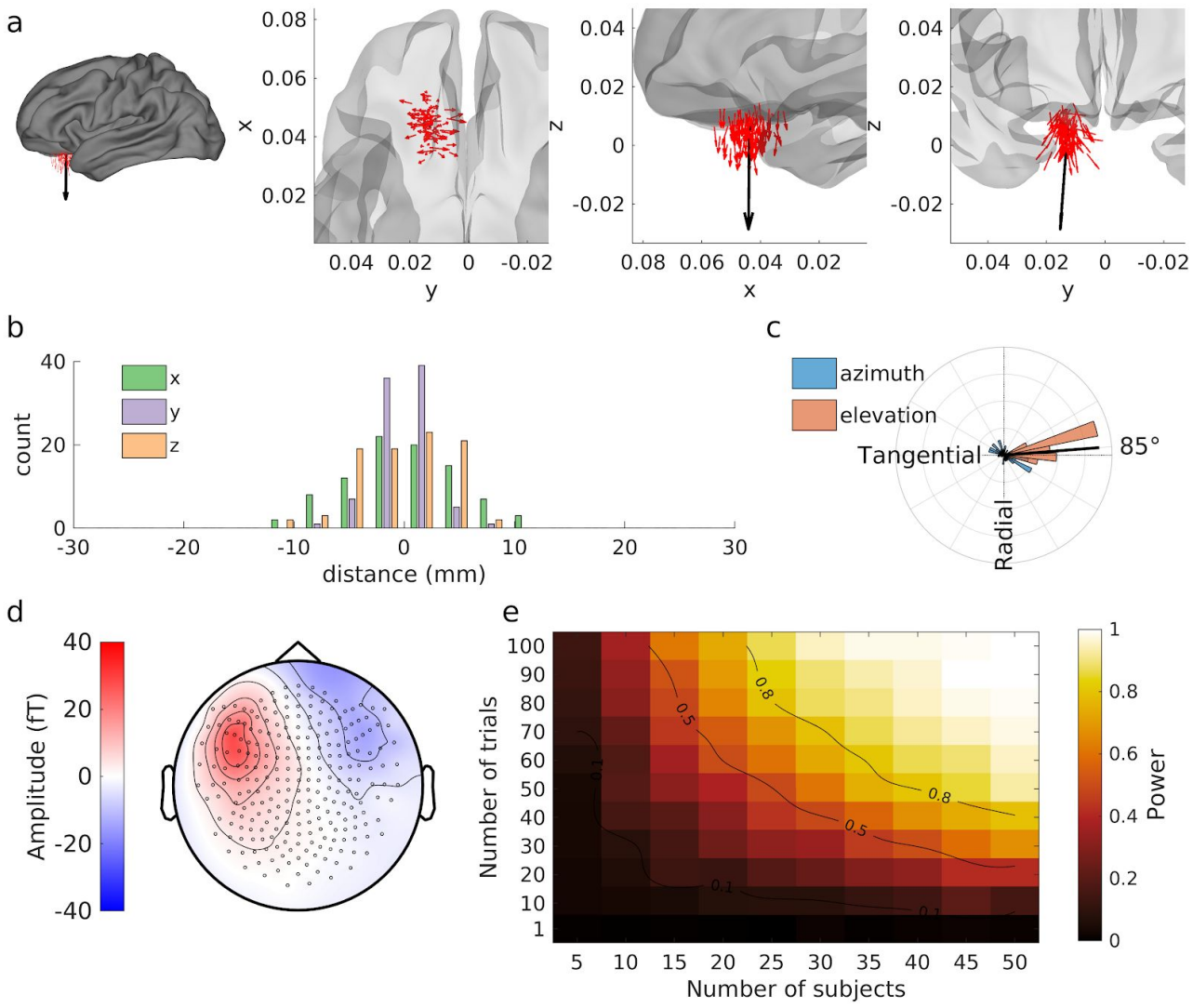

Figure 5: Detecting a deep source (84 mm to closest sensor) in orbitofrontal cortex. *The same conventions as for Figure 4 apply. See main text for further explanation.*

The two anterior frontal dipole locations presented above illustrate qualitatively the effect of distance to the sensors, but also highlight a clear limitation of this qualitative approach of describing the effect of source properties on statistical power. Given how dipole properties covary in realistic source models across brain regions, it is nearly impossible to disentangle the respective contributions of distance, orientation, and their variability across subjects with these realistic source models. Therefore, we turned to a more selective manipulation of the distance to sensors by eliminating individual variation in source positions in our simulations.

We selectively placed sources at set distances from sensors, along a radius running from a source in the precentral sulcus (a region detected with high power, as we later discuss see Figure 8, x=-12, y=33, z=70) towards the center of the head, running the same simulations and examining the results as before. Figure 6 shows the effect of this manipulation on statistical power for distances ranging from 20 to 100 mm. For all distances explored, dipoles were arranged in strictly the same orientation and relative positions (Figure 6a). The projected signal at sensors is illustrated in Figure 6b, showing the effect of distance-to-sensors on signal amplitude. Finally, Figure 6c shows how estimated statistical power varied across distances in these simulations. Note that our goal here is not to make a specific claim about statistical power at any specific distance, but rather to show how power varies with distance. In fact, we used dipoles with an amplitude of 5 nA.m, i.e. half the amplitude used in the previous simulations so as not to completely saturate the power plots here. Under these conditions, our simulations show that distance can make dipoles that could be detected with 100% power for any number of trials above 10 in 10 or more subjects when placed at 20 mm almost undetectable when placed at 100 mm from the closest sensor.

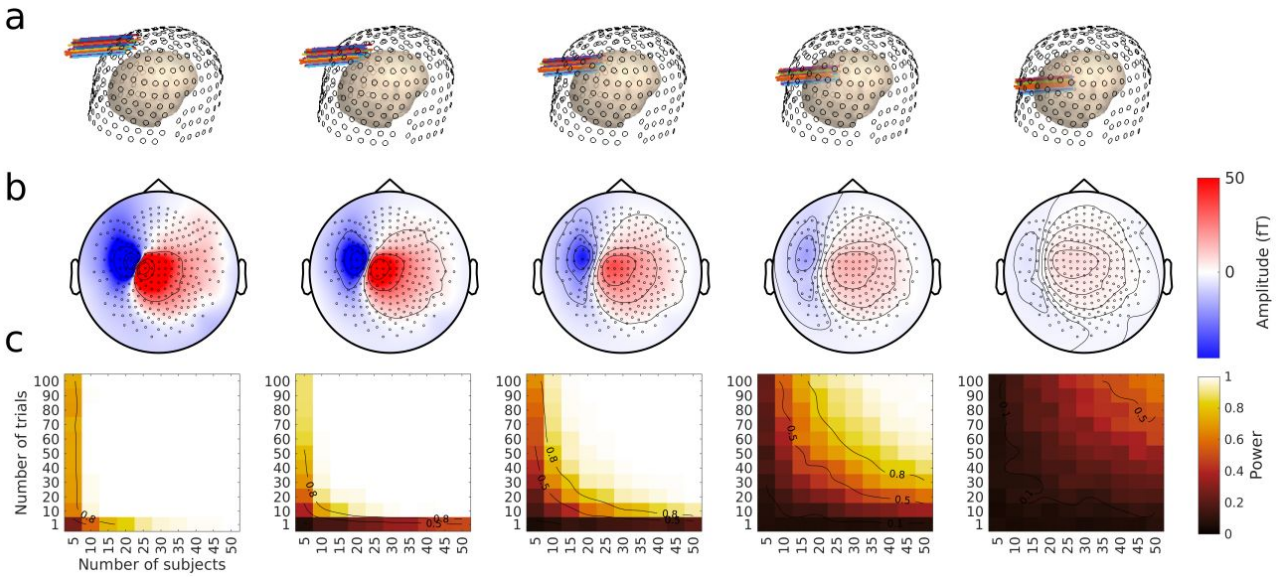

Figure 6. Detecting sources at varying distances from the sensors. a. Locations of individual dipoles. *Dipoles in 89 subjects (colored arrows) are represented on an individual head model (pinkish inner surface) relative to positions of sensors (black open circles on outer surface).*

Dipoles were placed at five equally spaced distances ranging from 20 to 100 mm from the sensor closest to an initial source in the precentral sulcus (see main text), towards the center of the head (origin of the coordinate system). ***b. Average projection of the dipoles in sensor space.*** *The same conventions as in Figure 4d apply. See main text for further explanation.* ***c. Results of the Monte Carlo simulations at each location.*** *The same conventions as in Figure 4e apply. See main text for further explanation.*

In summary, manipulating the distance to the sensors affected detectability drastically. A source with a given amplitude and spatial properties that is readily detected with very few trials and subjects when placed at a neocortical location close to the sensors is virtually impossible to detect if it occurs in the center of the head. However, other parameters, such as source strength, may compensate for a large distance, and all relevant parameters should thus be considered, when trying to predict statistical power for given sources in the real brain.

Orientation with respect to sensor

Figure 7a shows the average source orientation relative the closest sensor for every location on the cortical surface. Most sulcal walls and much of the inferior temporal cortex are close to tangential (0°), and the crest of gyri (e.g. superior frontal gyrus), and the depths of sulci (e.g. the insular cortex) are closer to radial (90°). The histogram of this orientation across all vertices is plotted in Figure 7b. It is evident that the distribution is skewed towards values below 90°, reflecting the overall convex topology of the cortical mantle.

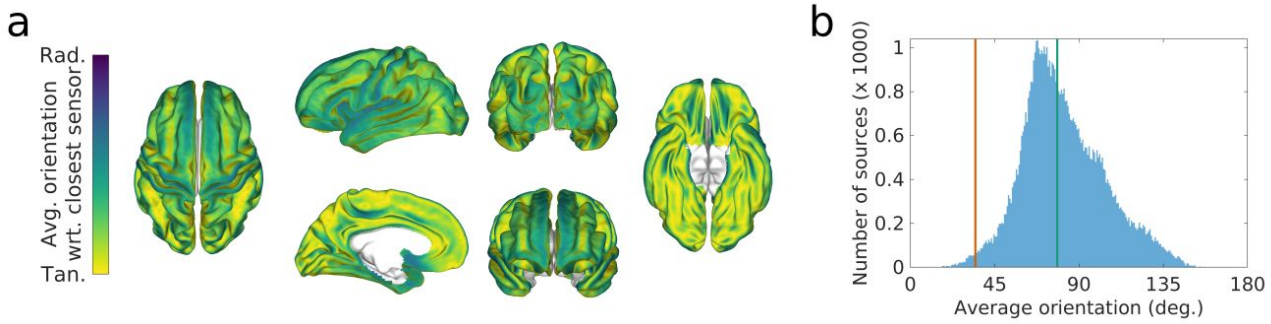

Figure 7. Average dipole orientation with respect to the closest sensor across subjects. *Data from 89 subjects in the HCP data plotted on the average brain.* ***a. Average orientation relative to the closest sensor.*** *This figure shows the mean of this angle mapped on superior, left*

lateral, left medial, posterior, anterior, and inferior views of the brain. For illustration here, we mirrored the computed angle (see methods) around the tangential axis (90°) for clarity. In other words we discarded information on whether the dipoles pointed more towards, or away from the sensors, and only consider orientation (computationally, the values displayed are the inverse cosine of the absolute value of the product of interest). Legend on color bar: Rad. = radial, Tan.= tangential. ***b. Histogram of the distribution of mean angles across all cortical vertices.*** *Contrary to panel a, values spread the full half circle. The red and green vertical lines indicate the orientation of the two dipoles that we selected to illustrate the effect of this parameter in Figures 8 and 9, respectively. deg. = degrees*

Based on physics, we expect the orientation of the sources relative to that of the sensor coils to have a large effect on signal amplitude (Hämäläinen et al., 1993; Hillebrand & Barnes, 2002; J. A. Malmivuo & Suihko, 2004), and thus on detectability. Here, using the same logic and display format as in the previous section on distance to sensors, we illustrate the effect of source orientation relative to sensors on group-level statistical power. We examine two closeby locations in the precentral region, one in the posterior bank of the precentral sulcus, and the second one on the gyral crest of the precentral gyrus. Each of these locations have very similar spatial properties, except for their orientation relative to the closest sensor. The sulcal source (x=-12, y=33, z=70) is close to a tangential orientation, while the gyral source (x=-4, y=37, z=78) is closer to a radial orientation. Our simulations show how the detectability of MEG signals at sensor level varies across numbers of subjects and trials for these two dipoles, as shown on the contour power plots of Figures 8 and 9. The effect of orientation is dramatic in this situation, with 80% power for as little as 15 subjects and 40 trials, or 10 trials in 40 or more subjects with the first (more tangential) dipole, and no sufficient number of subjects and trials in our simulations to reach 80% power with the second (more radial) dipole. This result illustrates the well-known effect of source orientation on MEG signal (e.g. Hillebrand & Barnes, 2002).

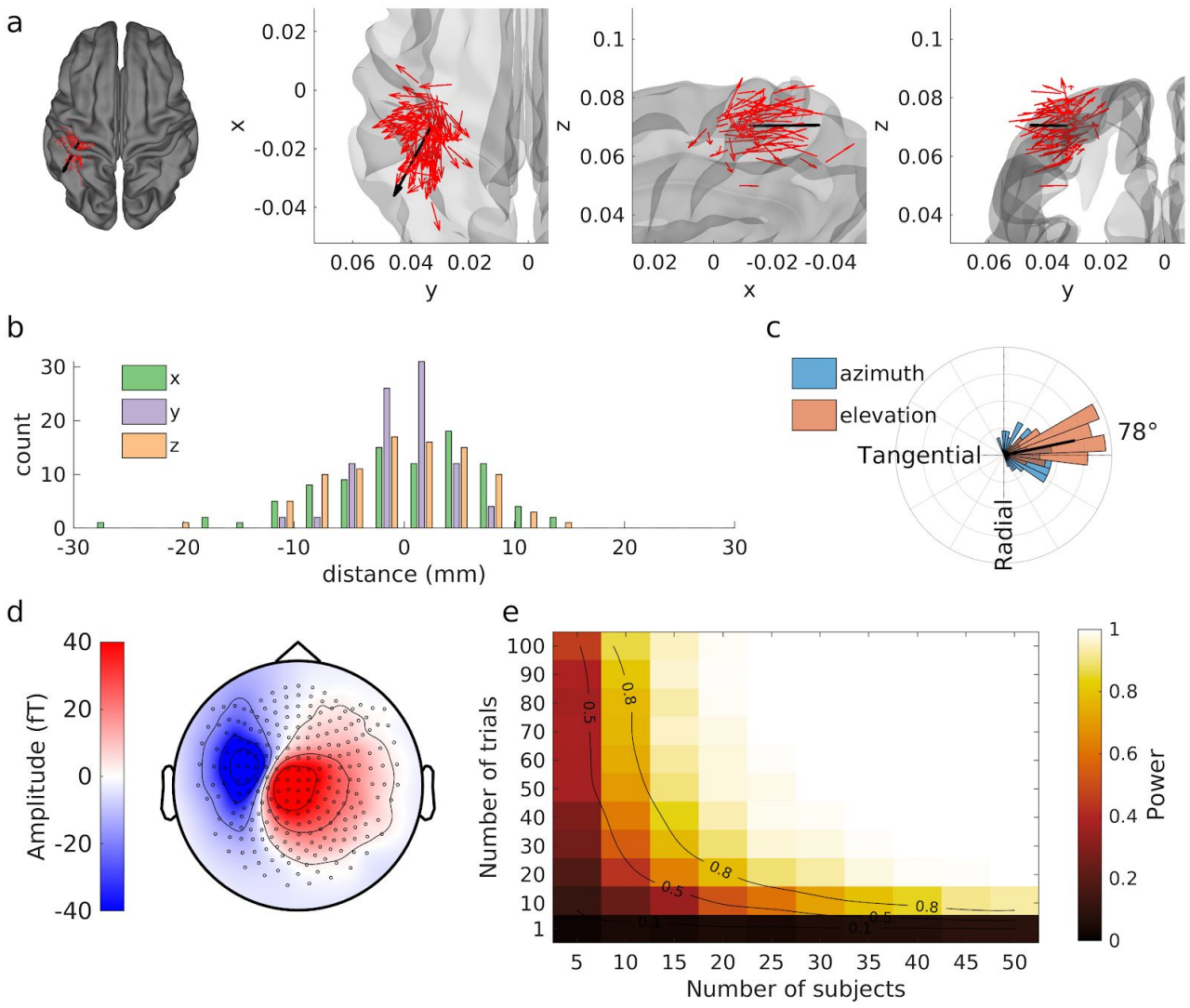

Figure 8. Detecting a tangential source in the precentral sulcus. *The same conventions as for Figure 4 apply. See main text for further explanation.*

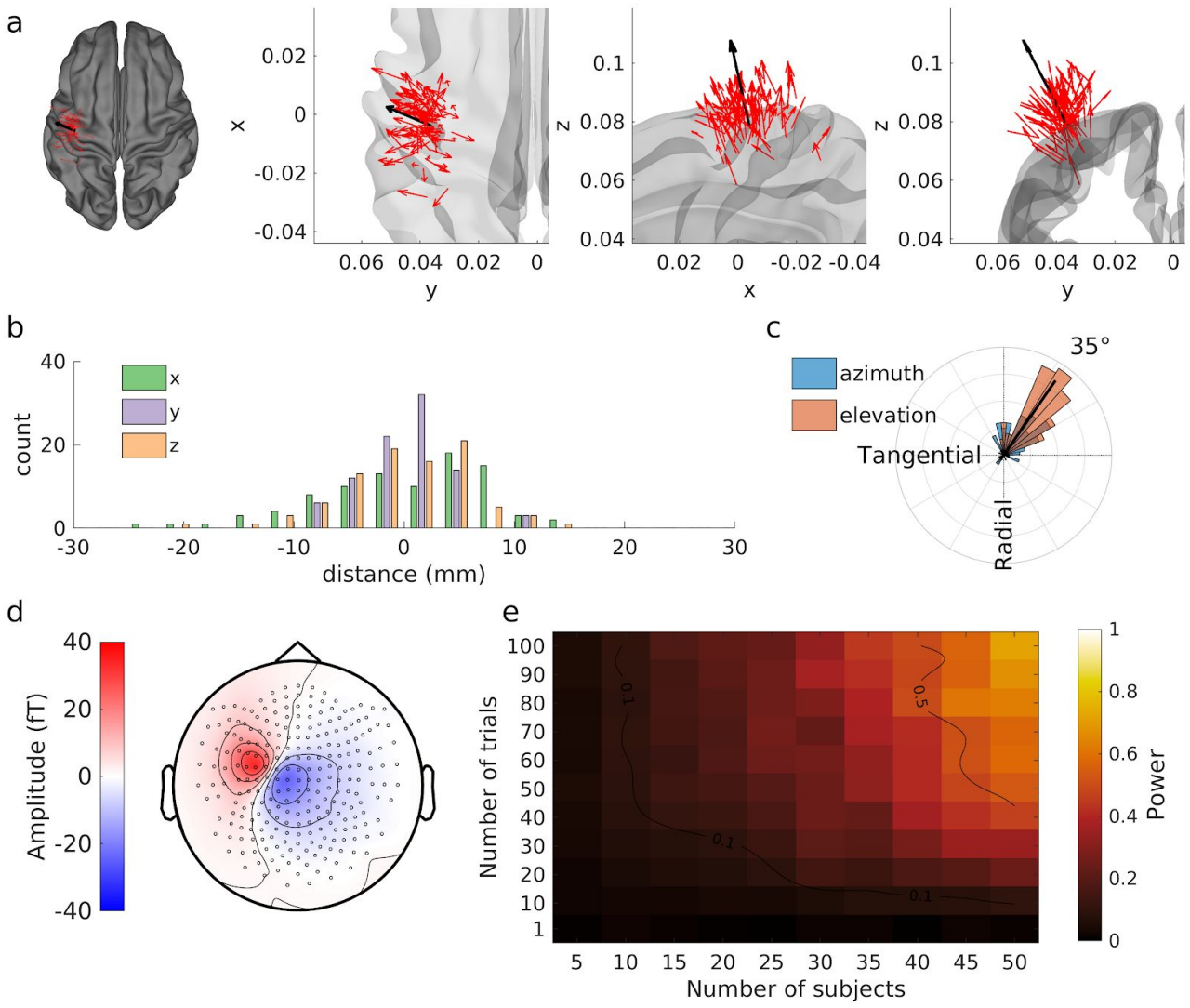

Figure 9. Detecting radial source source in the precentral gyrus. *The same conventions as for Figure 4 apply. See main text for further explanation.*

Similar to the previous section, we now turn to a more selective manipulation of orientation by eliminating individual anatomical variation, by selectively orienting sources at set angles with respect to the closest sensor, and running the same simulations as previously. Figure 10 shows the effect of this manipulation on statistical power for five equally spaced orientations ranging from 0° to 90°, starting from an original orientation strictly parallel to the closest sensor, and progressively tipping around the y axis (passing through both ears), up to an orthogonal orientation (approximately tangential to the head surface). For all orientations

explored, dipoles were placed at the exact same position (Figure 10a). The projected signal at sensors is illustrated in Figure 10b, showing the effect of orientation with respect to sensors on signal amplitude. The most radial sources (left) produce a simulated signal that is about 10 times smaller than the most tangential sources (right), although with similar topography. Finally, Figure 10c shows how estimated statistical power varied in these simulations from a situation where signal could never be detected above 40% with the number of subjects and trials we explored for radial orientations, to a situation where any number of trials above 10 to 20 in 15 or more subjects yielded 100% estimated power. Noteworthy, the decrease in detectability is narrowly focused at the radial orientation. A mere 22.5° shift away from this orientation (second orientation from the left in Figure 10) produces a major improvement in detectability, with 80% power reached with 50 trials in 20 subjects.

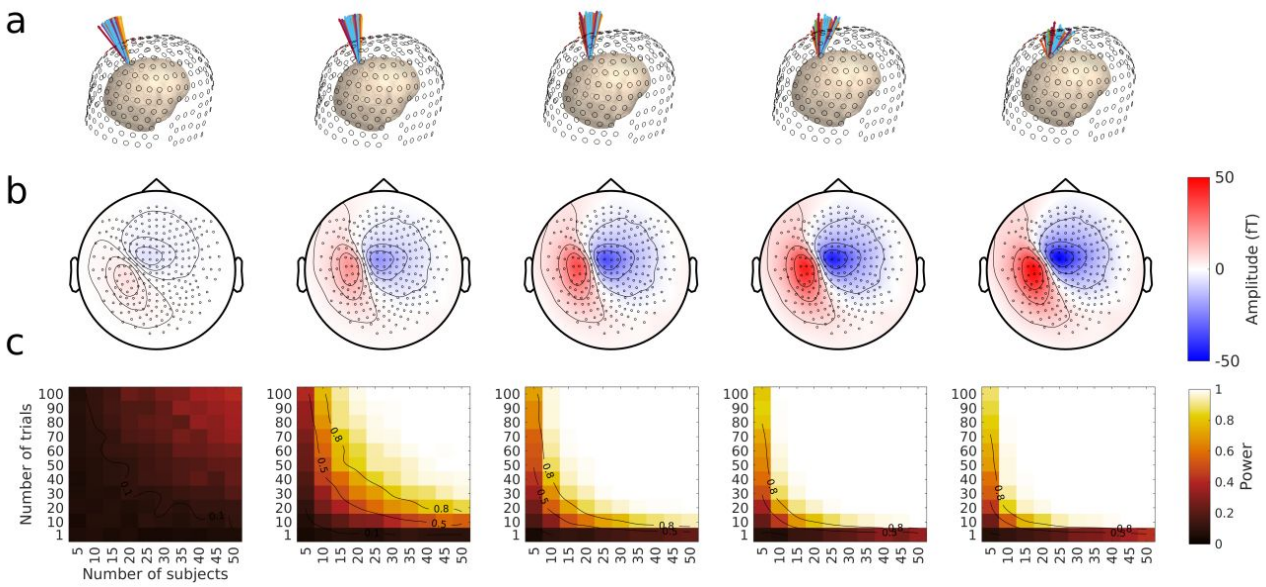

Figure 10. Detecting sources at varying orientations with respect to sensors. *The same conventions as in Figure 6 apply. Orientations vary from 0 degrees (pointing toward the closest sensor), i.e. radial orientation, on the left, to 90 degrees (pointing toward the front), i.e. tangential orientation in 5 equally spaced angles. See main text for further explanations.*

In summary, manipulating the orientation with respect to the sensors affected detectability drastically, especially when sources are pointing exactly in the same direction as the closest sensor (i.e. radially). However, this loss of power is recovered rapidly as soon as the average orientation of the sources shifts away from the radial orientation.

Taken together, the previous two sections allowed us to explore the effect of the position and orientation of dipolar sources - our so-called 'first-order properties'. Quite predictably, we have seen that these first-order properties have a strong impact on signal detectability. In the following two sections, we describe our explorations of second-order spatial properties derived from the cross-subject variability in source position and orientation, and their effect on group-level detectability.

Position variability across subjects

Figure 11a shows the average cross-subject standard deviation (across the three cartesian dimensions) in position for every location on the cortical surface. The variability histogram across all vertices is plotted in Figure 11b. Overall, values ranged from 3.4 to 8.1 mm, with maximal values occurring in areas where cortical folding was more variable across subjects, such as in occipital cortex, and minimal values occurring in anterior medio-ventral regions, as well as in insular cortex. Dipoles placed in the most variable regions could be up to 10 mm away from each other, whereas dipoles placed, for example in insular cortex, were on average within 5 mm from each other.

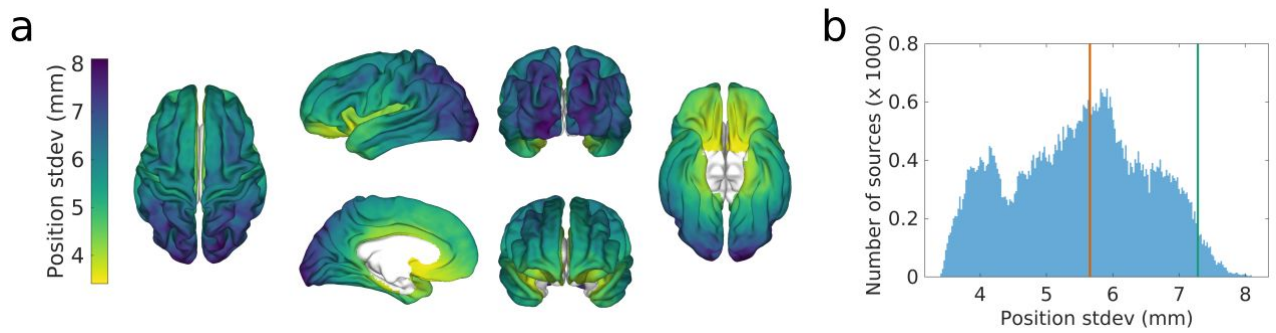

Figure 11. Average cross-subject variability in position *shown on the average brain of 89 subjects in the HCP data.* ***a.*** *The cross-subject variability in position is the standard deviation of position of the sources, shown on superior, left lateral, left medial, posterior, anterior, and inferior views of the brain.* ***b. Histogram of the distribution of average standard deviations across all cortical vertices.*** *The red and green vertical lines indicate the orientation of the two dipoles that we selected to illustrate the effect of this parameter in Figures 12 and 13, respectively.*

Position variability is expected to have an impact on detectability because dipoles at different locations, even if they share the same orientation, project a magnetic field that is topographically different. However, changes in position will rarely change the polarity of the magnetic field at a given sensor, and the similarity in the topographies created by two dipoles may override their differences. So, it is hard to predict the extent to which variability in position will affect detectability at sensor level. To address this question, we first selected two posterior locations in the individual brains of the HCP subjects, one with relatively low spatial variability in fusiform gyrus (x=-39, y=33, z=5), and the second one with large variability in superior occipital cortex (x=-69, y=12, z=48).

For the fusiform gyrus source, all dipoles were within a narrow +/- 10 mm region in the horizontal plane (x-y plane in Figure 12a, left zoomed view, and Figure 12b), and within +/- 15 mm in the vertical plane (x-z plane in Figure 12a, middle zoomed view, and Figure 12b). Their orientation was relatively consistent (Figure 12c). This source resulted in a signal with a strong amplitude at sensor level (Figure 12d), despite being at a distance from sensors, and

was thus detected with a relatively high-power across subjects, with 80% power with as little as 15 subjects with 60 trials each, or only 20 trials in 35 or more subjects (Figure 12e).

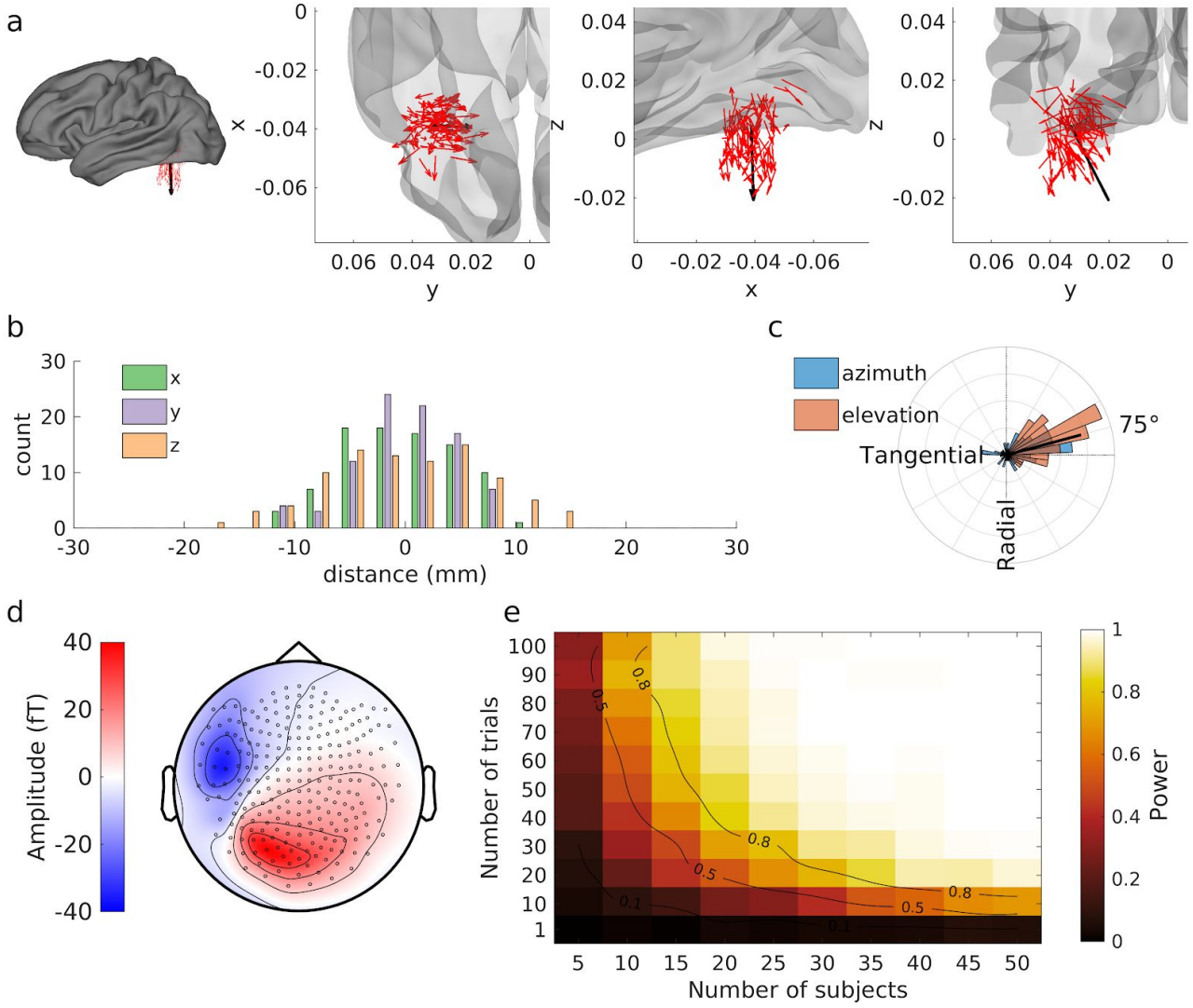

Figure 12. Detecting a source with relatively low cross-subject variability (5.6 mm). *The same conventions as for Figure 4 apply. See main text for further explanation.*

The source in the superior occipital gyrus, on the other hand, has a much larger variability in position across HCP subjects. This spatial variability is visible in Figure 13a (zoomed insets) and 13b. We note that, although relatively narrow compared to some other sources, orientation spread at this location was also larger than in the fusiform cortex (Figure 13c). Together, these differences yielded a relatively lower amplitude compared to the previous

dipole in the fusiform gyrus, and a poorer detection, with no tested number of subjects and trials reaching 50% power.

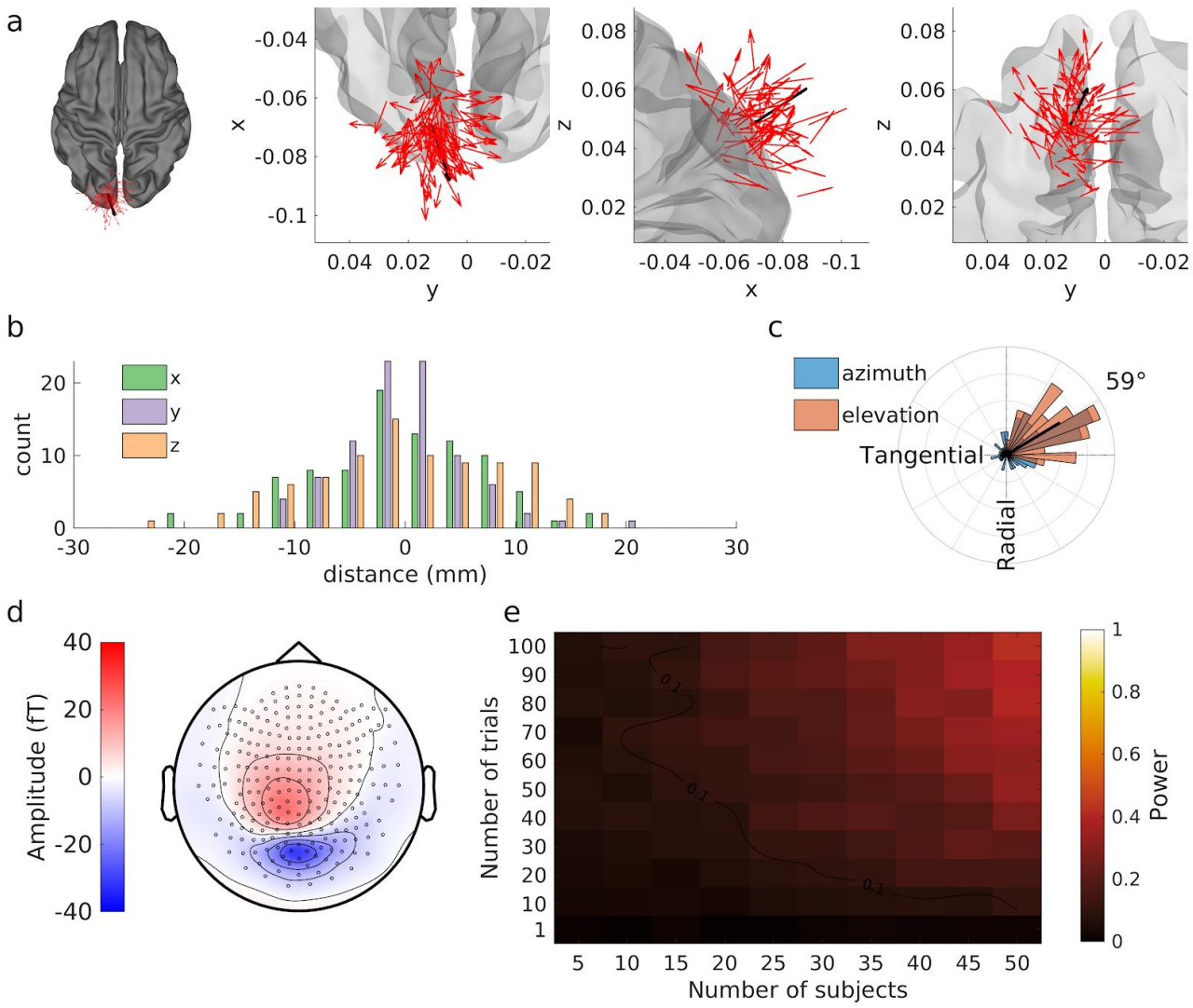

Figure 13. Detecting a 10 Na.m source with relatively high cross-subject variability (7.2 mm) in the superior occipital gyrus. *The same conventions as for Figure 4 apply. See main text for further explanation.*

The two dipole locations presented above illustrate qualitatively the effect of position variability on detectability. However, as already mentioned, the co-variability of the different spatial properties in anatomically constrained simulations is a confound that undermines robust conclusions specific to position variability. Therefore, we next moved to a more

specific manipulation of position variability to tightly controlling this variability, while maintaining other variables constant, in simulations unconstrained by anatomy. We selectively allowed dipoles to change position across subjects by sampling positions at random from a 3D normal distribution with a set standard deviations across subjects. Figure 14 shows the effect of this manipulation on statistical power for five equally spaced position variabilities ranging from 0 to 10 mm standard deviation.

For all explored position variabilities, the average position of dipoles across participants tended toward the same position in the precentral gyrus (Figure 14a). The projected signal at sensors is illustrated in Figure 14b, showing little effect on signal amplitude, and in turn little effect on estimated statistical power.

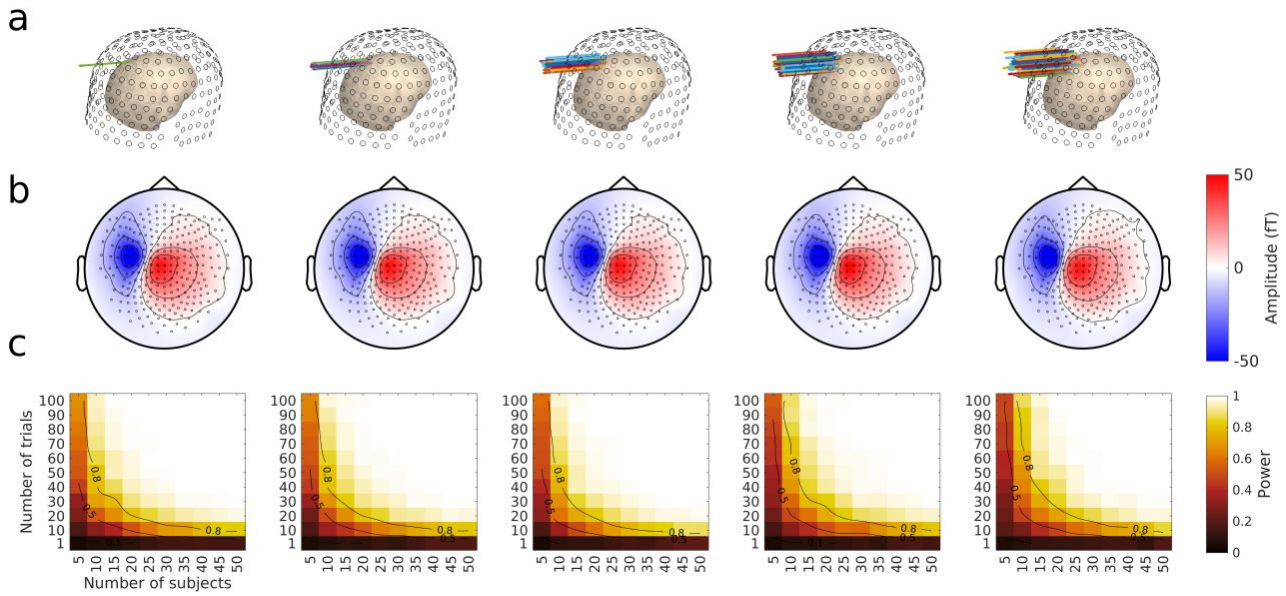

Figure 14. Detecting 5 Na.m sources with varying cross-subject position variability. *The same conventions as in Figure 6 apply. See main text for further explanations.*

In summary, manipulating position variability had little effect on detectability when the manipulation was selective. In realistic settings, regions with high position variability across

subjects in the occipital cortex tend to be also more variable in terms of orientation, as we will examine in the next section.

Orientation variability across subjects

Figure 15a shows the resultant vector length (i.e. the length of the average dipole across subject) for every location on the cortical surface, plotted on the average brain of the 89 subjects in the HCP dataset. The histogram of these vector lengths across all vertices is plotted in Figure 15b, which highlights the limits of the variability of this measure across the whole brain. The two sources that we selected to discuss below to illustrate the effect are highlighted with vertical lines. Overall, values range between 0.35 and 1, with minimal values in lateral occipital regions, and maximal values along the central sulcus, the ventral prefrontal cortex, dorsomedial prefrontal cortex and insular cortex. Dipoles in the most variable regions of the lateral occipital cortex can be oriented up to 180° apart (i.e. in opposite directions) across subjects, dramatically reducing the net contribution to the signal at individual sensor across a group of subjects.

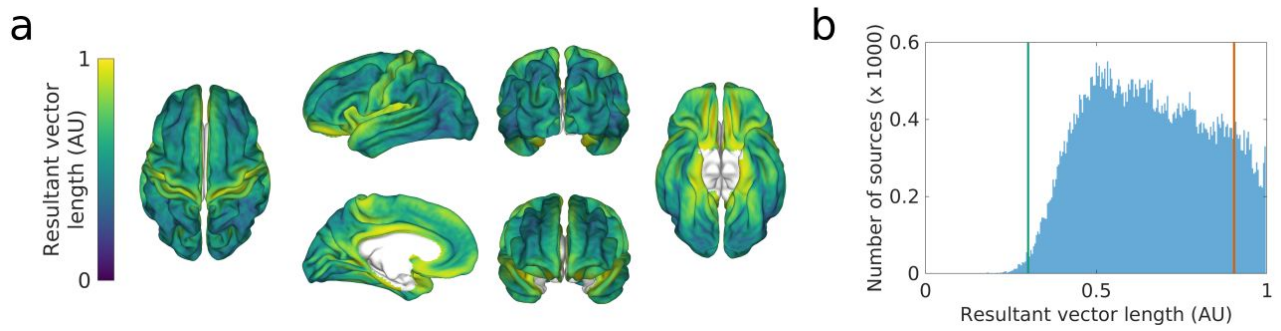

Figure 15. Average cross-subject variability in orientation *shown on the average brain of 89 subjects in the HCP data.* ***a.*** *The cross-subject variability in orientation is the average resultant vector length shown on superior, left lateral, left medial, posterior, anterior, and inferior views of the brain.* ***b. Histogram of the distribution of resultant vector length across all cortical vertices.*** *The red and green vertical lines indicate the orientation of the two dipoles that we selected to illustrate the effect of this parameter in Figures 16 and 17, respectively.*

Orientation variability is expected to have an impact on detectability because dipoles with different orientations, even if they share the same position, project a magnetic field that has a different topography. Contrary to position variability, changes in orientation can change the polarity of the magnetic field at a given sensor even with minimal orientation change. Here, we review two locations, one with relatively low orientation variability in the insula (x=6, y=36, z=30), and the second one with large variability in the posterior superior temporal sulcus (x=-47, y=42, z=30).

For the insula source, Figure 16a shows that most individual sources are parallel (see in particular the middle zoomed inset), resulting in a length of the resultant vector of 0.90, shown as a black line in Figure 16c, and an average projection to sensors with relatively high amplitude. The power to detect this signal is such that 20 subjects with 90 trials each, or 50

subjects with only 30 trials each allow detecting the 10 nA.m source 80% of the time.

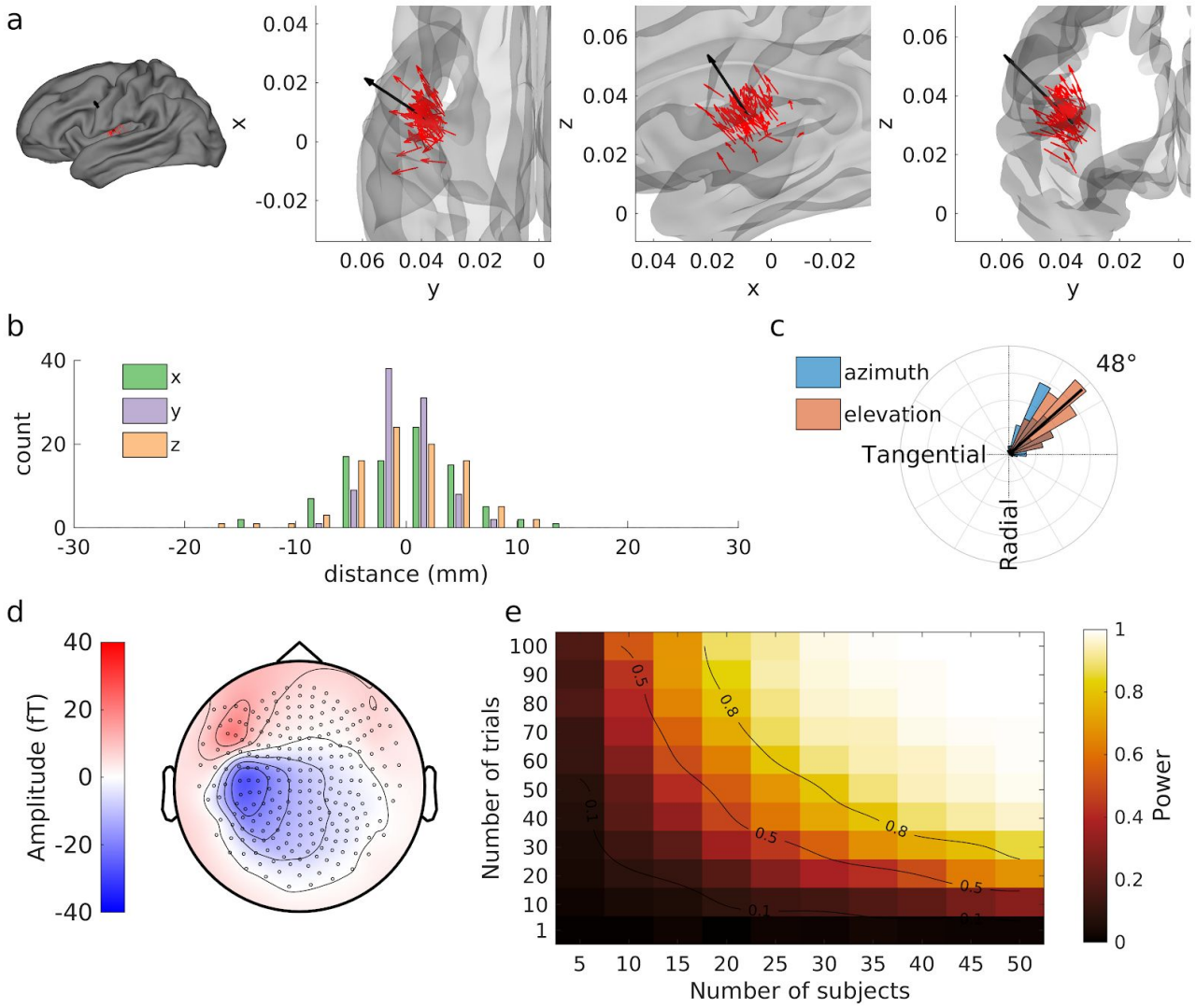

Figure 16. Detecting a 10 Na.m source with relatively low cross-subject orientation variability (resultant vector length of 0.9) in the insula. *The same conventions as for Figure 4 apply. See main text for further explanation.*

The source in the posterior superior temporal sulcus on the other hand shows a disorganized orientation across subjects (Figure 17a and c), a resultant vector length of 0.30, and a very weak average topography (Figure 17d). The signal is virtually undetectable within the range of explored subjects and trial numbers.

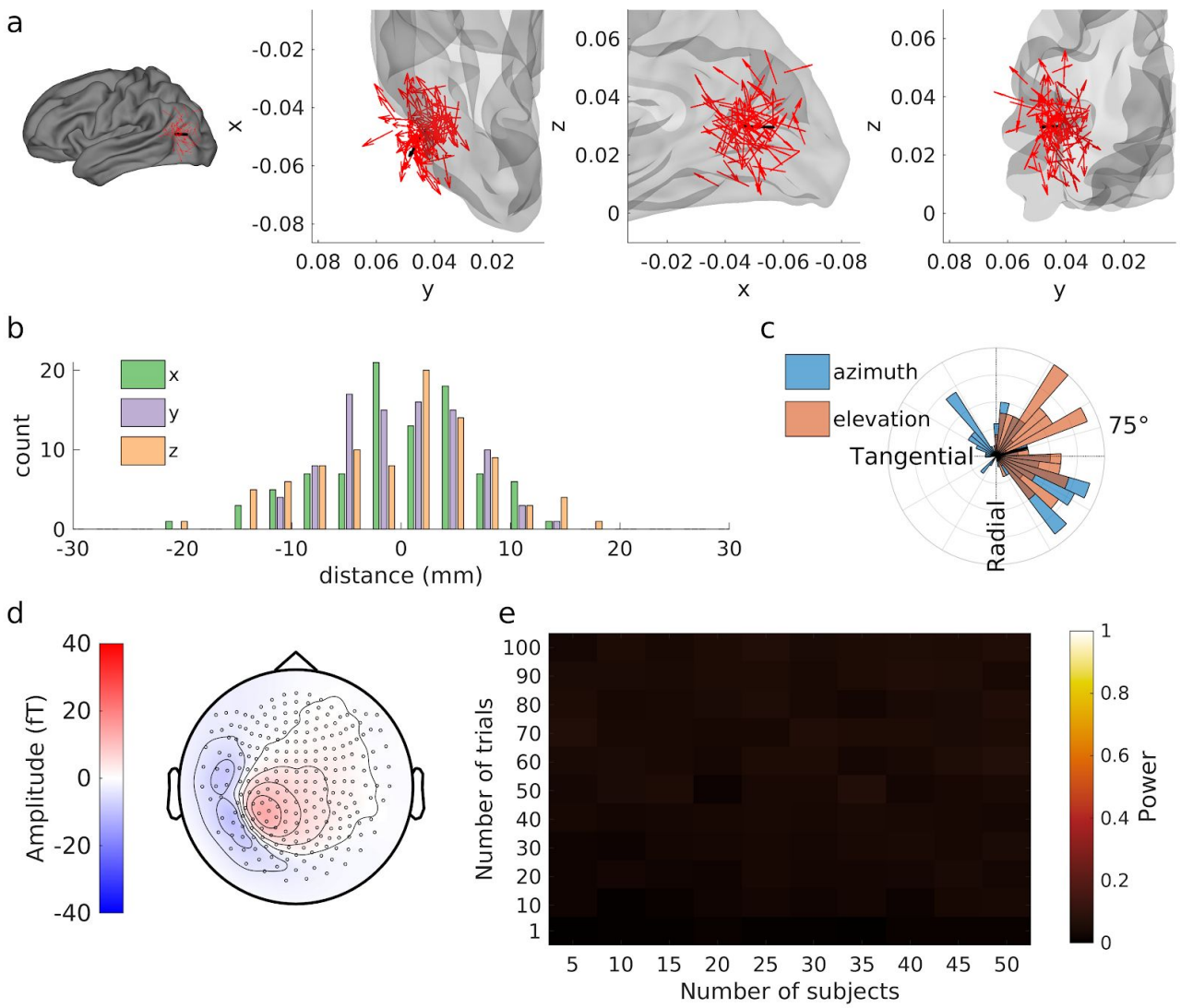

Figure 17. Detecting a 10 Na.m source with relatively high cross-subject orientation variability (resultant vector length of 0.3) in the posterior superior temporal sulcus. *The same conventions as for Figure 4 apply. See main text for further explanation.*

The two dipole locations presented above illustrate qualitatively the effect of orientation variability on detectability. To examine this effect more rigorously, we next moved to a specific manipulation of orientation variability. This time, position was held constant across subjects, and only orientation could vary across subjects. We sampled orientations at random, adding a normally distributed random azimuth and elevation to the original orientation (arbitrarily set to the tangential source in the precentral sulcus examined in Figure 8), with a

standard deviation set to five evenly spaced values between 0° (fixed orientation) and 180°. The effect of this manipulation on statistical power is illustrated in Figure 18. It is noteworthy that the effect of orientation variability is most visible with random orientations beyond a standard deviation of 90°, arguably due to the fact that dipoles cancelling each other out (i.e. oriented in opposite directions) can only occur when the distribution of added random orientations is sufficiently broad.

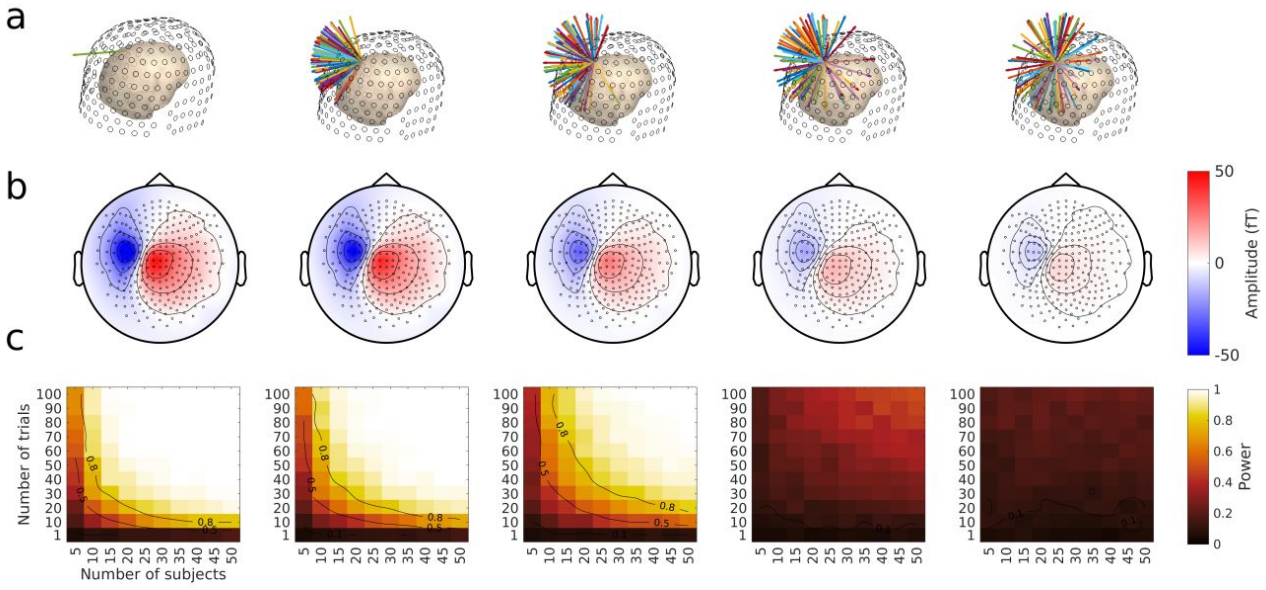

Figure 18. Detecting sources with varying cross-subject orientation variability. *The same conventions as in Figure 6 apply. See main text for further explanations.*

In sum, we have just shown that second-order properties of the sources can have very different effects on detectability. However, within the range of variabilities explored, orientation variability has a much larger effect than position variability. Together with the effect of the first-order properties examined above, these results highlight the importance of considering source properties when planning a well-powered MEG experiment. We discuss this and other issues in more detail below.

Discussion

Adequately powered MEG (and EEG) experiments require an appropriate sample size. In this study, we showed that the properties of the brain sources expected to contribute effects in a given experiment critically affect attainable statistical power in MEG. Specifically, we focused on the spatial properties of the sources and on their variability across subjects to examine how these factors affect group-level statistical power with classical tests at sensor level. We showed--quite expectedly--that the distance of the sources to the sensor array has a strong effect on detectability, with deeper sources being virtually undetectable with reasonable sample sizes for a single-session MEG experiment. We also observed another well-known effect, namely that sources oriented radially are almost undetectable in MEG data, as compared to other orientations. Looking next at the influence of cross-subject variability, we observed that position variability across subjects had, in fact, little effect on statistical power, whereas orientation variability (or the lack thereof) strongly affected statistical power. Signal detectability at sensor level thus depends not only on the source origin (position and orientation), but also on the less well predictable cross-subject variability of this origin. Therefore, one take home message for this study is that there is no simple solution for finding the optimal number of trials and subjects for all types of evoked MEG (or EEG) studies. Specifically, there will likely be variations in power across different brain structures within a single experiment that we believe make calculations of power challenging. Thus, we emphasize the importance of considering the expected brain sources of activity, their anatomical location, and their cross-subject variability while planning studies in cognitive, affective, and social neuroscience. Below we discuss some issues and considerations for planning MEEG studies, taking these variables into account.

Previous studies have explored the effects of spatial properties from a single-subject signal detection perspective, aiming to provide estimates of how much MEG signal can be detected given these properties and the amount of noise in an individual subject's data (Ahlfors, Han, Belliveau, & Hämäläinen, 2010; Goldenholz et al., 2009; Hillebrand & Barnes, 2002). It was therefore shown that source depth relative to sensors is the most important factor affecting source localizability, with additional effects due to source orientation (Hillebrand & Barnes, 2002). Others showed that MEG signal-to-noise varies considerably with source location (Goldenholz et al., 2009) and orientation (Ahlfors, Han, Belliveau, et al., 2010) in the brain. Pushing the analysis one step further, our group-level statistics approach takes a pragmatic stance on this issue, speaking directly to experimenters for whom changes in signal-to-noise ratio are particularly relevant for deciding on sample size, i.e. how many subjects or trials to include in an experiment.

We showed here that this question cannot be answered unequivocally. First, the large effects of spatial properties across brain structures on detectability that we observed (Figure 2) are a major challenge for predicting statistical power. Indeed, not only do the parameters that we examine greatly vary across the brain (as shown by Figures 3, 7, 11 and 15), but also these parameters are not independent of one another across the cortical mantle. For instance, sources with the shortest distances to sensors also tend to have a radial orientation, as gyral crests are generally found on the outer surface of the brain. Other regions with mostly tangential orientations, such as along the central sulcus, can be spatially highly consistent across subjects. In contrast, some brain regions have high degrees of inter-individual variability due to variations in cortical folding (e.g. the posterior STS, MT+/V5, inferior parietal cortex, Caspers et al., 2006) or high degrees of curvature (e.g. occipital pole) or hemispheric asymmetries (Croxson, Forkel, Cerliani, & Thiebaut de Schotten, 2018; Ochiai et

al., 2004). Therefore, variations of power across the brain for a given sample size are in practice difficult to predict.

Furthermore, statistical inference is usually made at the group-level. Predicting the adequate number of subjects for a given study also implies considering between-subject variability in different cortical regions. Although first-order spatial properties are straightforward to measure and their effect on signal to model with Maxwell's equations, second order properties' effects on group-level statistics are much more difficult to model. Our approach was to directly manipulate anatomical variations in our simulations free from anatomical constraints. Our first observation is that strictly parallel dipoles tend to combine across subjects, even when their position is relatively spread out, to create a more readily detectable net magnetic field. Randomly oriented dipoles on the other hand tend to cancel each other out. Beyond a certain orientation variability, statistical power becomes critically low, giving no opportunity to detect a signal even with a large number of subjects. This observation has an interesting implication: when comparing our whole brain power analysis for a sample of 25 subjects (Figure 2) with the whole brain signal-to-noise mapping for a single subject in MEG obtained by Goldenholz et al. (Goldenholz et al., 2009, their Figure 2, first row), at least two differences are apparent: First, some regions with high SNR e.g. the occipito-parietal cortex, remain hard to detect at the group-level. Second, other regions with low SNR, in particular on the medial and ventral surfaces of the brain, are detectable at the group level (e.g. in anterior cingulate cortex, or ventromedial prefrontal cortex). We hypothesize that cross-subject variability explains these discrepancies. On the one hand, it is thanks to a particularly consistent source orientation across subjects that some regions with low SNR in Goldenholz' study still show high power in our study, and on the other hand other regions with high SNR are less detectable at the group level due to high variability in source orientations across subjects. Together, although there is no unequivocal answer to the

question of an appropriate sample size in MEG (or EEG) experiments, our study highlights the importance of considering inter-subject variability in addition to the long-known spatial properties of sources when it comes to deciding how many subjects and trials to include in an experiment.

An important point to note regarding the decreased detectability with higher inter-subject variability is that it is standard procedure in MEG research to localize sources in individual subjects on their own anatomy before combining data for group analysis in source space (Hari & Puce, 2017; Jas et al., 2018). Going for source-level analysis can avoid a large portion of the problem of inter-subject variability, provided that data at the cortical level are properly aligned across subjects, which can itself be a challenge (REFS). Specifically, the cancellation effects we observe here across subjects when sources of different subjects point in opposing directions (Figure 17) should be largely avoided when sources are estimated in individual subjects first, then aligned across subjects on a template before being averaged (REFS). Further studies will have to examine the limits of this reasoning in detail. In particular, it will be important to examine group-level source detection in detail if we want to formulate more precise recommendations in the future.

A similar group-level approach has recently been used by Boudewyn et al. (2018) for EEG data in sensor space, exploring the dependency of statistical power on sample size (subjects and trials), while focusing on effect size at the electrode level for a set of ERP components. Our approach here is similar, but instead of starting from an expected ERP difference at sensor level, we started from the expected neural source. This complementary approach may allow principled investigations beyond known ERP effects, and could allow a better planning of studies targeted at activating specific brain regions. It also has potentially strong practical

implications as power analyses are usually mandatory in grant applications, or for preregistering studies.

Another important aspect to consider for future studies will be that of spatially extended sources. Indeed, although dipolar sources are easy to model and often used to appreciate detectability, more realistic extended patches of activity can have a difficult-to-predict net effect at sensors. For instance, such extended patches that occur in the homologous regions of the two hemispheres on the mesial wall, or are spread across opposite walls of a sulcus can show relatively lower detectability due to opposite sources cancelling each other out (Ahlfors, Han, Lin, et al., 2010; Fuchs, Kastner, Tech, Wagner, & Gasca, 2017; Hari & Puce, 2017). It should be noted, however, that functional area borders often tend to follow gyral and sulcal crests, making synchronous activation on both sides of a sulcus less likely (Destrieux, Fischl, Dale, & Halgren, 2010; Glasser et al., 2016). The detectability of extended sources at the group-level in such cortical regions needs to be explored in the future for additional practical recommendations.

High-temporal resolution is desirable in cognitive, clinical and social neuroscience studies, particularly where naturalistic and hyperscanning data are acquired. It will therefore be of importance to extend explorations to the time domain, and characterise changes in variability in the time domain, particularly before and immediately after stimulus onset and examine effects on group-level signal detectability at specific latencies. In the future, it will be important to consider investigations in sensor space, in addition to those in source space, as portable EEG and MEG studies are likely to use relatively low numbers of sensors, making it unrealistic to transform these data into source space.

Given the current emphasis on whole-brain data collection in cognitive neuroscience and the explosion of network science based data analyses (Bassett & Sporns, 2017) and the

modulation of activation as a function of perceptual and cognitive manipulations (Medaglia, Lynall, & Bassett, 2015), it is important to understand how statistical power can vary across the brain and the critical dimensions in which these variations occur. Critically, the statistical power for detecting a given network will *at most be only as good as its least detectable node*. It is thus important while planning an experiment entailing a network analysis to plan sample size accordingly.

Acknowledgements

Maximilien Chaumon and Nathalie George work on a platform that has received infrastructure funding from the programs "Investissements d'avenir" ANR-10-IAIHU-06 and ANR-11-INBS-0006 (specifically for the Centre de Neuroimagerie de Recherche [CENIR] of the ICM). Aina Puce received support from Indiana University, Bloomington (sabbatical leave).

References

Ahlfors, S. P., Han, J., Belliveau, J. W., & Hämäläinen, M. S. (2010). Sensitivity of MEG and EEG to Source Orientation. *Brain Topography*, *23*(3), 227–232. https://doi.org/10.1007/s10548-010-0154-x

Ahlfors, S. P., Han, J., Lin, F.-H., Witzel, T., Belliveau, J. W., Hämäläinen, M. S., & Halgren, E. (2010). Cancellation of EEG and MEG signals generated by extended and distributed sources. *Human Brain Mapping*, *31*(1), 140–149. https://doi.org/10.1002/hbm.20851

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2019). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *ArXiv:1902.06122 [q-Bio, Stat]*. Retrieved from

http://arxiv.org/abs/1902.06122

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. https://doi.org/10.1038/nn.4502

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6), e13049. https://doi.org/10.1111/psyp.13049

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn3475

Caspers, S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., & Zilles, K. (2006). The human inferior parietal cortex: Cytoarchitectonic parcellation and interindividual variability. *NeuroImage*, *33*(2), 430–448. https://doi.org/10.1016/j.neuroimage.2006.06.054

Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P., & Willmes, K. (2015). Registered Reports: Realigning incentives in scientific publishing. *Cortex*, *66*, A1–A2. https://doi.org/10.1016/j.cortex.2015.03.022

Clayson, P. E., Carbine, K. A., Baldwin, S. A., & Larson, M. J. (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. *Psychophysiology*. https://doi.org/10.1111/psyp.13437

Croxson, P. L., Forkel, S. J., Cerliani, L., & Thiebaut de Schotten, M. (2018). Structural Variability Across the Primate Brain: A Cross-Species Comparison. *Cerebral Cortex*, *28*(11), 3829–3841. https://doi.org/10.1093/cercor/bhx244

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Fischl, B. (2012). FreeSurfer. *Neuroimage*, *62*(2), 774–781.

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association : JMLA*, *105*(2), 203–206. https://doi.org/10/gfxvhq

Fuchs, M., Kastner, J., Tech, R., Wagner, M., & Gasca, F. (2017). MEG and EEG dipole clusters from extended cortical sources. *Biomedical Engineering Letters*, *7*(3), 185–191. https://doi.org/10.1007/s13534-017-0019-2

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., … Essen, D. C. V. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. https://doi.org/10.1038/nature18933

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., … Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Goldenholz, D. M., Ahlfors, S. P., Hämäläinen, M. S., Sharon, D., Ishitobi, M., Vaina, L. M., & Stufflebeam, S. M. (2009). Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography. *Human Brain Mapping*, *30*(4), 1077–1086. https://doi.org/10.1002/hbm.20571

Groppe, D. M. (2017). Combating the scientific decline effect with confidence (intervals). *Psychophysiology*, *54*(1), 139–145. https://doi.org/10.1111/psyp.12616

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., … Schoffelen, J.-M. (2013). Good practice for conducting and reporting MEG research. *NeuroImage*, *65*, 349–363. https://doi.org/10.1016/j.neuroimage.2012.10.001

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993).

Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, *65*(2), 413–497. https://doi.org/10.1103/RevModPhys.65.413

Hari, R., & Puce, A. (2017). *MEG-EEG Primer*. Oxford University Press.

Hillebrand, A., & Barnes, G. R. (2002). A Quantitative Assessment of the Sensitivity of Whole-Head MEG to Activity in the Adult Human Cortex. *NeuroImage*, *16*(3, Part A), 638–650. https://doi.org/10.1006/nimg.2002.1102

Jas, M., Larson, E., Engemann, D. A., Leppäkangas, J., Taulu, S., Hämäläinen, M., & Gramfort, A. (2018). A Reproducible MEG/EEG Group Study With the MNE Software: Recommendations, Quality Assessments, and Good Practices. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00530

Kappenman, E. S., & Keil, A. (2017). Introduction to the special issue on recentering science: Replication, robustness, and reproducibility in psychophysiology. *Psychophysiology*, *54*(1), 3–5.

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., … Yee, C. M. (2014). Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, *51*(1), 1–21. https://doi.org/10.1111/psyp.12147

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217.

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2013.03.024

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Larson, M. J., & Carbine, K. A. (2017). Sample size calculations in human electrophysiology (EEG and ERP) studies: A systematic review and recommendations for increased rigor. *International Journal of Psychophysiology*, *111*, 33–41. https://doi.org/10.1016/j.ijpsycho.2016.06.015

Larson-Prior, L. J., Oostenveld, R., Della Penna, S., Michalareas, G., Prior, F., Babajani-Feremi, A., … Snyder, A. Z. (2013). Adding dynamics to the Human Connectome Project with MEG. *NeuroImage*, *80*, 190–201. https://doi.org/10.1016/j.neuroimage.2013.05.056

Luck, S. J. (2005). Ten simple rules for designing ERP experiments. *Event-Related Potentials: A Methods Handbook*, *262083337*.

Luck, S. J. (2012). Event-related potentials. *APA Handbook of Research Methods in Psychology*, *1*, 523–546.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Malmivuo, J. A., & Suihko, V. E. (2004). Effect of Skull Resistivity on the Spatial Resolutions of EEG and MEG. *IEEE Transactions on Biomedical Engineering*, *51*(7), 1276–1280. https://doi.org/10.1109/TBME.2004.827255

Malmivuo, J., Suihko, V., & Eskola, H. (1997). Sensitivity distributions of EEG and MEG measurements. *IEEE Transactions on Biomedical Engineering*, *44*(3), 196–208. https://doi.org/10.1109/10.554766

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Medaglia, J. D., Lynall, M.-E., & Bassett, D. S. (2015). Cognitive Network Neuroscience. *Journal of Cognitive Neuroscience*, *27*(8), 1471–1491. https://doi.org/10.1162/jocn_a_00810

Ochiai, T., Grimault, S., Scavarda, D., Roch, G., Hori, T., Rivière, D., … Régis, J. (2004). Sulcal

pattern and morphology of the superior temporal sulcus. *NeuroImage*, *22*(2), 706–719. https://doi.org/10.1016/j.neuroimage.2004.01.023

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Intell. Neuroscience*, *2011*, 1:1–1:9. https://doi.org/10.1155/2011/156869

Simonsohn, U., Nelson, L., & Simmons, J. (2013). *P-Curve: A Key to the File Drawer* (SSRN Scholarly Paper No. ID 2256237). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=2256237

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, *15*(3), e2000797. https://doi.org/10.1371/journal.pbio.2000797

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., … Curtiss, S. W. (2012). The Human Connectome Project: A data acquisition perspective. *Neuroimage*, *62*(4), 2222–2231.