algorithmic operations of mutually aligning the fMRI images to one another, as well as matching them to other imaging modalities and external image-based templates, researchers create a dataset that is "compatible with already-established centres of calculation."[360] Importantly, the output of these transformations are 4D functional datasets that are still illegible—when preprocessed fMRI datasets are submitted to visual inspection, even experts cannot 'read' them. In short, by looking at these images, it is still impossible to determine which voxels exhibit task-induced activity and which do not. Nevertheless, thus standardised, the images can now finally undergo statistical analysis that will translate them into legible brain maps. Hence, as shown by my analysis, the purpose of preprocessing is to construct the analysability of the fMRI datasets while at the same time preserving their indexicality via a chain of traceable mathematical operations.

3.4 Statistical Analysis: Articulating the Task-Induced Neural Activity of Interest

Preprocessed functional 4D datasets remain illegible because the pertinent information concerning the brain activity of interests they entail is still spread across multiple brain volumes and buried under random noise. To construct the legibility of their fMRI data, researchers must determine which areas of the subjects' brains can be declared active. They do this by using statistical analysis, which enables them to make judgments about the "underlying patterns in the data" ridden with random noise.[361] Instead of more commonly known descriptive statistics that merely summarise the data, fMRI studies apply inferential statistics. This type of statistics permits researchers to use the datasets from their subject sample to make claims about a larger population.[362]

Inferential data analysis is based on the process called hypothesis testing. Generally speaking, this type of statistical analysis starts with the formulation of two opposing claims—the null hypothesis and the alternative hypothesis.[363] In the subsequent step, statistical tests are used to evaluate which of the two hypotheses describes the data with a higher probability. In fMRI, the null hypothesis amounts to the claim that the task had no effect on the data, or in other words, that there is no temporal correlation between the variation in the BOLD time series and the different experimental conditions. The alternative hypothesis states that the measured differences in the BOLD signal's average intensities between the task and the control condition are temporally correlated with the experimental intervention.[364]

During hypothesis testing, the analysis software executes automated statistical tests for each voxel independently. This voxel-by-voxel approach is known as mass

360 Latour, 71–72.
361 Worsley, "Statistical Analysis," 251.
362 Worsley, 251.
363 Huettel, Song, and McCarthy, *Imaging*, 331.
364 Huettel, Song, and McCarthy, 331.

univariate analysis.[365] It aims to identify the voxels in which the data provide sufficient empirical evidence to reject the null hypothesis.[366] If the numerical value of the resulting statistical test at a given voxel is below a predetermined threshold value, the null hypothesis has to be rejected, and that voxel is declared active.[367] The joint outcome of all tests performed across the brain is a statistical activation map—a 3D image whose voxels contain numerical values of test statistics. Only those voxels within this map that have been declared active are visualised in bright colours and superimposed on an anatomical brain image (see figs. 3.12 and 3.13). Conversely, all inactive voxels within this map remain invisible to the observer. It transpires from my description that such a map does not provide information about the neural activity of interest in absolute terms. Instead, the map shows in which voxels the probability that the task-induced response was due to chance is sufficiently low to declare these voxels active.

To apply hypothesis testing to fMRI data, researchers must first create a model that provides the basis for the alternative hypothesis. In most studies, this model is built within the theoretical framework called the general linear model (GML) and it entails researchers' detailed estimation of how the task intervention affected the subjects' brains during the experiment.[368] Put simply, by drawing on the GLM,[369] researchers create a study-specific model—called design matrix—that is tailored to their experiment. As we are about to see, by using a study-specific model, researchers can reconstruct from the fMRI data the information about the task-induced brain activity. Thus, in what follows, I will argue that study-specific models play crucial roles in producing the legibility of fMRI data.

My analysis in the upcoming sections is informed by Margaret Morrison's and Mary S. Morgan's notion of models as instruments of enquiry. Morrison and Margaret have argued that due to their "ability to effect a relation between scientific theories and the world," models can be used both as "a means to and a source of knowledge."[370] According to Morrison and Morgan, models can function as instruments because of their following features. First, their partial independence from both theory and data; second, their ability to fulfil diverse tasks ("functional autonomy"); and third, the flexible ways in which they can relate to both theory and data ("representational power").[371] Importantly, Morrison and Morgan have insisted that to understand the productive roles of models, we must look at how they are created and used in actual scientific practice.

---

365 Poldrack, Mumford, and Nichols, *Handbook*, 70.

366 Huettel, Song, and McCarthy, *Imaging*, 331.

367 Huettel, Song, and McCarthy, 331–32.

368 Poldrack, Mumford, and Nichols, *Handbook*, 70.

369 Friston, "Statistical Parametric Mapping," 16.

370 Morrison and Morgan, "Models as Mediating Instruments," 35.

371 Morrison and Morgan, 32. In fact, Morrison and Morgan have argued that these three characteristics allow models to function as autonomous agents in scientific research. See ibid., 10. Since I find that this term overstates the degree of partial independence both in the models' construction and use, I will refrain from calling models autonomous agents in the concrete cases I analyse here. I will talk instead about the productive roles of models in fMRI research.

Following this dictum, I will return to the case study from the previous sections to analyse how de Lange, Roelofs, and Toni transformed the model suggested by theory into a study-specific model that they then deployed to create multiple statistical activation maps. After that, I will examine a later study by the same group of authors to demonstrate how researchers can make the same dataset yield an entirely different type of analytical outcome called a connectivity map by using an alternative theoretical model of brain function. In the following four sections, I will discuss the chain of modelling decisions that determine what becomes visible and thus legible in brain maps as the output of statistical analysis. My aim is to show that despite their reliance on automated algorithms to transform the fMRI data into brain maps, researchers actively shape statistical analysis by deciding how many and what kinds of maps to create from a single dataset.

3.4.1 Building the Design Matrix as a Tool of Enquiry

Having collected and preprocessed fMRI data from eight patients with one-sided hysterical arm paralysis, de Lange, Roelofs, and Toni then moved on to the main stage of processing to identify the task-induced neural activities in the data. Using the SPM software, they performed a two-stage statistical analysis based on the general linear model (GLM). They first conducted separate first-level analyses for each subject. Next, during the second-level analysis, they combined the outputs from all single-subject analyses to compute group-level functional activation maps.[372] Since most studies use this approach, both in hysteria research and in neuroimaging in general, the de Lange, Roelofs, and Toni study is representative of fMRI data analysis and is treated as such throughout my discussion.[373] This and the following sections will focus mainly on examining the epistemic implications of the first-level analysis because, as I will show, this stage entails crucial modelling decisions that inform all subsequent processing steps.

But before we can examine the modelling decisions that de Lange, Roelofs, and Toni made, we must first take a brief look at the conceptual framework underlying their analysis. At its most basic, the GLM is an equation that defines a mathematical relationship between the signal intensity registered at a single voxel throughout the measurement and the experimental conditions that temporally coincided with this measurement. The underlying assumption of the GLM is that all factors contributing to the neural activity in a particular voxel linearly add up to form an overall BOLD response.[374] Based on this assumption of linearity, the GML describes the BOLD

372 Ashburner et al., "SPM12 Manual," 63; and Poldrack, Mumford, and Nichols, *Handbook*, 70.

373 Ashburner et al., "SPM12 Manual," 63; and Huettel, Song, and McCarthy, *Imaging*, 345.

374 The presumed linearity of the fMRI BOLD response is based on experimental findings. See Boynton et al., "Linear Systems Analysis"; and Dale and Buckner, "Selective Averaging." However, it should be noted that the linearity of the haemodynamic response is first and foremost a theoretical approximation, which is neither universally applicable to all study designs nor is it unchallenged as a concept. For experimental findings that have challenged the assumption of linearity, see Friston et al., "Non-Linear Responses"; and Vazquez and Noll, "Non-Linear Aspects." Nevertheless, most fMRI studies use the assumption of linearity as an acceptable approximation that considerably

response measured in a single voxel across various time points as a scaled sum of known contributing factors—referred to as explanatory variables—with the addition of unknown random noise.[375] Consequently, during statistical analysis, fMRI data are not processed in their spatial form—as a collection of brain slices. Instead, they are processed in their temporal form—as a set of time courses, one for each voxel.

The segment of the GLM equation that contains all explanatory variables together with the specifications of how each variable changes over time is known as the design matrix. This particular segment of the equation represents the study-specific model I referred to above. The random noise in the equation accounts for the difference between the values predicted by this model and the actual fMRI data.[376] Significantly, each explanatory variable in the design matrix is scaled by a parameter called the effect size. The effect size defines the relative contribution of the respective variable to the overall BOLD response measured at a given voxel.[377] In essence, effect sizes quantify the relative magnitude of the neural responses induced by particular experimental conditions at a single location. The crucial point is that the value of effect sizes is unknown before analysis. Hence, the very purpose of statistical analysis is to compute from the fMRI data the effect size estimates—and their standard errors—for each experimental condition specified in the design matrix.[378] But to be able to do this, researchers first have to use the GLM to build a study-specific design matrix. To examine how this is done in practice, let us now turn to our case study.

To create a design matrix, researchers must first define those mutually independent components of their experimental task that, according to their assumptions, added up to produce the neural activity behind the measured BOLD response in each voxel.[379] This means that the GLM provides researchers with an abstract template with which they can flexibly decompose the measured BOLD responses into a set of components. To perform such decomposition, researchers have to make judgments about the expected neural effects that different components of their experimental task elicited simultaneously. This step would be straightforward in an imaginary experiment that used a single stimulus. Yet, we have seen earlier in the chapter that de Lange, Roelofs, and Toni used a mixture of factorial and parametric experimental designs by employing multifaceted stimuli whose several aspects varied at once. In what follows, my analysis will demonstrate that translating such a complex experimental task into a design matrix entails multiple interpretational decisions.

As discussed previously, the stimuli in our case study comprised thirty-two drawings of the left and right hands, presented in eight different degrees of rotation, either with the palm up or down. The patients were instructed to judge the laterality

---

simplifies the data analysis. For a more detailed discussion of the linearity of the BOLD response and the limits to this assumption, see Huettel, Song, and McCarthy, *Imaging*, 229–37.

375 Friston, "Statistical Parametric Mapping," 16.

376 In mathematical terms, the GLM is a matrix equation that takes the following form: $Y = X\beta + \varepsilon$. Y denotes the fMRI data, X the design matrix, ε the residual error, and β the effect sizes. For details, see Friston et al., "General Linear Approach," 191.

377 Friston et al., 191–92.

378 See Ashburner et al., "SPM12 Manual," 73; and Huettel, Song, and McCarthy, *Imaging*, 343–5.

379 Ashburner et al., "SPM12 Manual," 63–68; and Huettel, Song, and McCarthy, *Imaging*, 345–51.

of the presented hand. De Lange, Roelofs, and Toni chose to isolate only two factors of their experimental task in the first-level analysis—whether the motor imagery engaged the affected hand; and which level of biomechanical complexity the imagined movement entailed.[380] In effect, the researchers thus hypothesised that the overall activity in each voxel depended on two factors: first, whether the drawing corresponded to the patient's paralysed hand; and second, the degree of rotation of the presented image relative to the body.[381] Since half of the patients had left- and the other half right-hand paralysis, the researchers disregarded the laterality of the stimuli at the level of single-subject analyses.[382] Moreover, in building their matrix, the researchers also decided to ignore whether a particular hand stimulus was shown with the palm up or down.[383]

So far, we have seen how de Lange, Roelofs, and Toni defined the factors of the design matrix by choosing the components of their experimental manipulations whose effects on the data they wanted to explore. Next, the researchers turned to modelling the respective levels of these factors. This meant that they had to determine how the values of each component of interest changed during the experiment. The first factor could only have two different levels by referring to the affected or the healthy hand. However, regarding the increasing biomechanical complexity of the task (i.e., its parametric component), the researchers had several modelling options. They could assume a linear link between the increasing degree of rotation of the stimuli and the increasing intensity of the neural response. Alternatively, they could also allow for different types of non-linear relations.[384] Based on the analysis of the behavioural data,[385] de Lange, Roelofs, and Toni concluded that the relation was non-linear. Therefore, they chose to model the effect of each particular degree of rotation separately.[386] Finally, by conflating the clockwise and anti-clockwise orientations of the stimuli, they divided the eight degrees

380 De Lange, Roelofs, and Toni, "Self-Monitoring," 2053.

381 De Lange, Roelofs, and Toni, 2053.

382 In other words, at this stage, it did not matter which side of the patient's body was affected. But we will see later that the laterality of the hand drawings played a significant role in the subsequent group-level analysis.

383 The researchers provided no justification for this decision. Hence, it remains an open question why they included this stimulus variation in their task if they had no intention of analysing its effects. One possible explanation is that the inclusion of this particular aspect merely served to increase the variability of the presented images and thus prevent the patients from feeling bored or habituating to the stimuli. As discussed in section 3.1.2, it is vital to avoid or at least reduce the experimental subjects' habituation to stimuli, as it results in unwanted confounds that, in turn, lead to the production of potentially invalid fMRI maps.

384 For a theoretical explanation of different ways in which a parametric experimental design can be translated into a design matrix, see Worsley, "Statistical Analysis," 259–60.

385 As mentioned earlier, in many fMRI studies, researchers not only collect the imaging data but also measure various aspects of the participants' task performance, such as response times and error rates.

386 De Lange, Roelofs, and Toni, "Self-Monitoring," 2053. This decision was significant because if the researchers had chosen to assume either a linear or a less complex non-linear link, their factor would have contained fewer levels, thus resulting in a simpler but potentially less precise matrix. For details on the alternative options, see Worsley, "Statistical Analysis," 259–60.

of rotation into five different levels.[387] In the end, de Lange, Roelofs, and Toni thus created a complex 2-by-5 factorial design matrix. Hence, the columns of this matrix contained ten explanatory variables of interest altogether.

Each modelling decision discussed above is significant as it selectively imposed a specific interpretational framework on the data while foreclosing possible alternatives. Crucially, choosing into how many and which particular components to partition the experimental task determines what can be made legible in the fMRI data. This is because only the components that have been laid out in the matrix as separate explanatory variables can be taken into account when calculating activation maps. By deciding to omit an aspect of the experimental task from their design matrix, researchers essentially declare it epistemically insignificant and relegate its effects to random noise. Conversely, by explicitly including specific aspects of the task in the design matrix, researchers ascribe to them an active role in providing potential insights into the presumed neural mechanisms of hysteria. Hence, it is not the experimental design that determines what counts as a variable of interest and what as noise. Instead—and this is a crucial point—what is a variable of interest and what is noise in a particular study depends on how researchers decide to build their study-specific model.

The next step in building the design matrix entails modelling random noise. To this end, de Lange, Roelofs, and Toni included in their matrix the six motion parameters—three translations and three rotations—to filter out the residual effects of the subjects' head motion.[388] As discussed previously, during preprocessing, fMRI data had already been submitted to motion correction to erase the spatial misalignment caused by the subjects' minimal head movements during the acquisition. However, this preprocessing step was unable to remove the unwanted signal changes that also arose from the subjects' head movements. Such signal changes represent a significant problem for statistical analysis. Specifically, "even a very small [head] motion (< 0.3 mm) in a functional series can induce signal changes in the order of 10 percent," whereas "the typical changes in the neuronal signals of interest" amount to "only about 1 percent."[389]

Since subjects' head movements tend to temporally correlate with their performance of experimental tasks, such unwanted changes in the signal can be mistaken during analysis for the actual BOLD effects of interest and thus lead to the production of invalid fMRI maps.[390] To circumvent this problem, de Lange, Roelofs, and Toni included the six motion parameters in their design matrix so that, during the computer-based analysis, the motion-induced changes in the signal could be identified as noise and discarded. Moreover, de Lange, Roelofs, and Toni also added to their matrix the patients' incorrect responses to the experimental task, which had been registered as behavioural data during the measurement. In doing so, the researchers defined as noise the patients'

387 Specifically, the researchers assumed that the stimulus-induced imagined movement away from the body at an angle of 45 degrees had the same neural effects as the movement towards the body at the same angle. De Lange, Roelofs, and Toni, "Self-Monitoring," 2053.

388 These motion parameters were analysed in detail in section 3.3.2.

389 Jenkinson and Chappell, *Neuroimaging Analysis*, 201.

390 Jenkinson and Chappell, 115–16.

BOLD responses that temporally coincided with their false responses. Consequently, these effects were also excluded from further analysis.

It follows from my analysis that the additional columns in the design matrix jointly referred to as confounds serve to designate those changes in the BOLD signal that were not intentionally induced by the experimental manipulation. Although not actively used in the analysis, such confounds have an important auxiliary function. By clearly defining various sources of noise, the confounds help improve the fit between the measurement and the values predicted by the design matrix. In doing so, they increase the validity with which the effect sizes of the explanatory variables can be estimated from the data.[391] Hence, it can be said that modelling random noise is just as important a step in constructing the design matrix as is defining the variables of interest.

In principle, the inclusion of additional explanatory variables, both those of interest and confounds, allows researchers to construct a model that matches the predicted signal to the signal measured with increasing accuracy. Nevertheless, there is one caveat. Each additional explanatory variable lowers the potential validity with which subsequent statistical tests can detect task-induced brain activations.[392] This caveat is due to the very nature of statistical testing—the higher the amount of information one estimates from the noisy data, the less probable such estimates are.[393] Thus, when building their study-specific model, researchers have to establish a trade-off. On the one hand, they need to use a sufficient number of variables to describe their experimental effects with sufficient precision. On the other hand, however, they must also avoid having too many variables, which would lead to overfitting the data and thus inadvertently declaring noise for the information of interest.

In addition to deciding which explanatory variables to include in their design matrix, researchers must also make judgments about the temporal pattern of the neural activity that each variable elicited during the experiment. This is necessary because the design matrix has two dimensions. Whereas its columns contain individual explanatory variables, its rows describe the expected intensity of the neural activity arising from each of these variables at a specific point in time.[394] Thus, to fill in the rows of their design matrix, researchers must predict the onset, intensity, and duration of the neural responses induced by each explanatory variable. In most studies, the onset of the task-induced neural activity is assumed to coincide with the onset of the stimulus.[395] It is

391 Huettel, Song, and McCarthy, *Imaging*, 349.

392 Huettel, Song, and McCarthy, 349.

393 Specifically, each "additional column in the design matrix reduces the number of degrees of freedom available. In the limiting case, one could reproduce perfectly any set of n time points with a combination of n – 1 different model factors. Since the significance of any individual factor is evaluated as a function of the number of available degrees of freedom, it is in the researcher's interest for the number of factors to be as small as possible." Huettel, Song, and McCarthy, 349. The term degrees of freedom refers to the "number of independent observations within a data set. For many statistical tests, there is n – 1 degrees of freedom associated with n data points." Ibid., 335.

394 Huettel, Song, and McCarthy, 345–46.

395 Huettel, Song, and McCarthy, 351. This is why the synchronisation between the stimulus exposure and data acquisition is of critical importance for the analysability of the fMRI data.

this assumption that de Lange, Roelofs, and Toni made in their study.[396] Additionally, they judged that the duration of the induced neural responses corresponded with each patient's average response time, i.e., the period between the stimulus onset and the pressing of the button. Finally, they modelled the rotation-related increase in the intensity of the neural response as a non-linear process that had the same shape as the increase in the patients' reaction times. To determine the particular shape of this non-linear increase, de Lange, Roelofs, and Toni performed a separate statistical evaluation of the patients' behavioural data.[397]

Based on my analysis, it is apparent that the GLM, which the researchers used as the basic theoretical framework, did not determine their modelling decisions about the temporal structure of their study-specific matrix. Instead, we have seen that their modelling decisions were informed by the specific details of their experimental design, such as the timing of the stimuli. Just as importantly, the researchers also based their modelling decisions on the additional information about the participants' task performance (i.e., average response times) that they derived from the separately acquired behavioural data. Hence, I argue that the way in which de Lange, Roelofs, and Toni used non-imaging data to construct the legibility of their fMRI data represents a pertinent example of intermedial transcription.[398]

At this point, the design matrix that de Lange, Roelofs, and Toni had created contained the predicted neural responses for each explanatory variable over the course of the experiment. But, as discussed previously, the fMRI data that the matrix is meant to model contain the measurements of the correlated BOLD—i.e., haemodynamic—responses. Therefore, to create the matrix that the software can use to analyse the fMRI data, the prediction of the neural responses has to be mathematically combined with a model of the haemodynamic response.[399] The simplest option is to choose the software's default setting. This setting uses a canonical mathematical function to describe an average temporal course and a standard empirical shape of the BOLD response (see fig. 3.6).[400] This is the option that de Lange, Roelofs, and Toni chose to use. Yet, the canonical haemodynamic response function has its limitations. The generic function disregards physiological variations in the neurovascular coupling that result in different shapes and durations of the BOLD responses among different subjects and across different brain regions of the same individual.[401] Studies that

396 De Lange, Roelofs, and Toni, "Self-Monitoring," 2053.

397 De Lange, Roelofs, and Toni, 2053.

398 See Jäger, "Epistemology of Disruptions," 72.

399 Ashburner et al., "SPM 12 Manual," 68–69; and Huettel, Song and McCarthy, *Imaging*, 351–54.

400 Different analysis software packages offer their own generic model as a default setting. In the generic model used by the SPM, the BOLD response is described by a mathematical function whose visual representation is a curve. It has an onset delay of 1 to 2 seconds in relation to the short-duration neural activity that initiated it. This is then followed by a gradual rise to the peak at 6 seconds and a slow return to the baseline, including a prolonged undershoot. See Poldrack, Mumford, and Nichols, *Handbook*, 75–76.

401 Huettel, Song and McCarthy, *Imaging*, 352.

deploy the canonical function consider such variations as noise and are "biased to only find responses that are similar to that function."[402]

As an alternative, the SPM allows researchers to use more flexible models or even to calculate the characteristics of each subject's BOLD responses.[403] However, although the latter approaches are considered more precise than the use of the canonical haemodynamic response function, they are also more complex to compute and more challenging to interpret.[404] Consequently, the generic model is used in many studies as an acceptable approximation that significantly simplifies the analysis. On the whole, researchers' particular choice regarding which BOLD response model to implement in their study is a significant interpretational decision. As shown by my analysis, this choice has epistemic implications for the resulting functional maps.

Finally, before moving on to discuss how researchers deploy the design matrix, there is one more aspect to which I want to draw attention. While building the design matrix, researchers interact with the software's user interface and type commands that allow the software to implement their modelling decisions. Thus, the underlying structure of the resulting design matrix is a set of mathematical functions that informs the software-based statistical analysis. Significantly, such a design matrix, which consists of rows and columns, can also be displayed in the form of a table diagram. A single cell in this diagram refers to the intensity of the predicted neural response induced by a respective explanatory variable at a given time point of the experiment. This diagram is then visualised by encoding different intensities of the predicted neural responses in corresponding grey-scale values (fig. 3.9). The highest predicted neural response is indicated in white, its absence in black, and the intermediary values in various grey shades.[405]

It is important to note that the resulting diagrammatic visualisation is not requisite for the computer-based analysis. Instead, it specifically addresses the human eye and has a distinct utilitarian function. The diagram provides a highly effective overview of various modelling decisions that went into building the matrix by bringing them into explicit visual relations to one another. In other words, the results of the entire modelling process are thus summarised within a single image and can be viewed at a glance. In effect, it is in its diagrammatic form that the design matrix—as a mathematical representation of the predicted experimental effects—becomes graspable to its human creators. Also at this stage, the targeted use of a specifically designed visualisation plays an epistemically productive role in the working process. The key point here is that by scrutinising its diagrammatic visualisation, researchers can check the accuracy of their design matrix before putting it to work. Yet, as in all cases analysed so

402 Poldrack, Mumford, and Nichols, *Handbook*, 76.

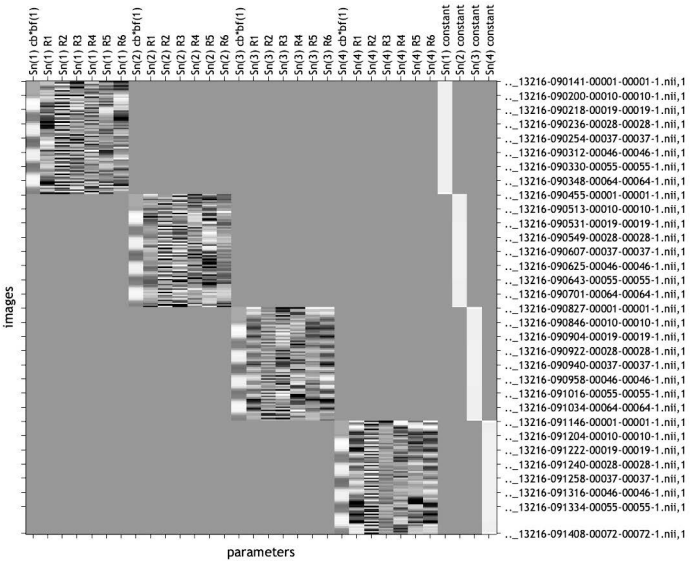
403 For details, see Ashburner et al., "SPM12 Manual," 68–69. Se also Poldrack, Mumford, and Nichols, *Handbook*, 76–81.

404 Huettel, Song, and McCarthy, *Imaging*, 352–54.

405 Huettel, Song, and McCarthy, 346.

far, being able to 'read' this diagram to assess its accuracy presupposes particular visual skills that researchers have to acquire through practice.[406]

Figure 3.9. Diagrammatic visualisation of a design matrix.



***

To sum up, we have seen how in the process of constructing a study-specific model, researchers actively and productively draw on the broader theoretical model provided by the GLM. My analysis has highlighted that one of the key features of this modelling process is its flexibility. On the one hand, this flexibility permits researchers to assemble a highly specific design matrix as a sufficiently accurate description of their particular experiment. On the other hand, it also allows them to inscribe a particular interpretational framework into their matrix. By this, however, I do not mean to imply that, in the process of constructing the design matrix, researchers already build the outcome of the analysis into their matrix.

Instead, the point I am making is that the researchers' modelling decisions limit the kinds of questions they can ask with the design matrix. I have analysed how in creating their study-specific model, researchers not only make judgments about the effects of their specific experimental task but also rely on a set of more general assumptions about the relations between the elicited neural and haemodynamic responses. All these choices add up to establish a particular epistemic framework that, while opening certain interpretational possibilities, also imposes constraints on what can be made

406 I am using the term reading here in the sense that Sybille Krämer has introduced. See Krämer, "Operative Bildlichkeit," 102.

legible in the fMRI data. My analysis thus allows us to draw the following conclusion. While the resulting study-specific model is a relatively accurate representation of the experimental intervention, it is also a representation explicitly built as a tool for selectively answering concrete research questions by filtering out brain activities of no interest from the data.

3.4.2 Deploying the Design Matrix to Compute Activation Maps from fMRI Data

Having built the design matrix, researchers then use it to translate the preprocessed fMRI data into statistical maps. As stated previously, statistical analysis is first performed for each subject separately. In the second stage, the results of single-subject analyses are used to draw statistical inferences at the group level. This two-stage process ends with the creation of group-level activation maps. Each of these stages entails multiple steps, during which algorithms execute massive amounts of black-boxed calculations. Two aspects of statistical analysis are of central concern for our discussion. First, in what follows, I will delineate the operations through which researchers close the gaps between the fMRI data and group-level activation maps. I will argue that the results of this process are indexical signs. Second, we will examine at which points of statistical analysis researchers actively shape the algorithmic operations.

In the previous section, we have discussed how researchers first build a design matrix by breaking up the experimental task into a set of conditions whose effects on the subjects' brains they want to explore in their fMRI data. As we have seen, each such condition of interest becomes an explanatory variable of interest in the study-specific design matrix. In the subsequent step, called model estimation, researchers put the design matrix to work.[407] During this step, researchers rely on automated algorithms to compare the study-specific model to the fMRI data. Based on the comparison, the algorithms calculate the extent to which each explanatory variable of interest contributed to the overall task-induced neural response at a given location. Model estimation is performed independently for each voxel.[408]

At the level of a single voxel, the result of this analytical step is a set of estimates of the unknown effect sizes—one for each explanatory variable of interest. To estimate the effect sizes that best explain the fMRI data at a given voxel, the algorithms match the time course of the BOLD response registered across different acquisition time points to the temporally correlated time course predicted by the design matrix.[409] Through a series of iterative steps, the algorithms then compute the best fit between these two time courses. For each effect size at each voxel, the algorithms calculate a single value. This value has been averaged across the subject's responses to multiple repetitions of

407 Huettel, Song, and McCarthy, *Imaging*, 343; and Poldrack, Mumford, and Nichols, *Handbook*, 191–94.

408 Huettel, Song, and McCarthy, *Imaging*, 343.

409 Expressed in mathematical terms, the algorithms have to solve the GLM equation by minimising the difference between the data and the value predicted by the design matrix. The difference is quantified by a cost function, which in this case is the so-called sum of least squares. For details, see Huettel, Song, and McCarthy, 336–37.

the same task over the course of the experiment.[410] The resulting combination of the estimated effect sizes necessarily varies from voxel to voxel. Such differences in the estimated effect sizes across voxels reflect the differences in the response magnitudes with which different parts of the subject's brain reacted to the same set of task conditions.

All the effect sizes estimated for a single experimental condition—one for each voxel—are stored as a 3D matrix.[411] This means that the output of model estimation is a new set of images. Each newly computed image encodes a subject-specific spatial distribution of the estimated effect sizes for a single task condition. It can thus be argued that model estimation categorically transforms fMRI data. Using a 4D fMRI dataset as its input,[412] model estimation produces a distinctly different kind of a 3D image. In the resulting images, the numerical voxel values no longer refer to signal intensities but to the estimated effect sizes.

For the sake of clarity, let me summarise a few points that I have made throughout this chapter. Researchers are interested in finding out the response magnitudes of the task-induced brain activity across voxels. However, as discussed previously, the scanner cannot measure this information directly. Instead, as a proxy for the information of interest, the scanner registers the correlated changes in the MR signal intensities.[413] The effect sizes researchers calculate from the MR signal intensities during model estimation are estimates of the not directly measurable response magnitudes of the task-induced brain activity. My analysis has shown that the design matrix—as the study-specific model of the estimated task-induced effects—plays a pivotal role in transforming a set of images that encode the measured signal intensities into a new set of images that encode the estimated effect sizes. As we have seen, the design matrix allows the black-boxed mathematical operations, which are hard-coded into the analysis software, to bridge the evidently sizeable gap between these two kinds of images.

We need to pay particular attention to two specific effects of this categorical transformation. First, model estimation results in massive compression of data since it displaces a large fMRI dataset with a small number of images. For example, in the de Lange, Roelofs, and Toni study, the fMRI dataset that comprised 547 brain volumes per subject was compressed into ten 3D images of the estimated effect sizes. Second, during model estimation, fMRI data undergo what I chose to designate as the elision of the temporal dimension. Specifically, whereas a 4D fMRI dataset encodes both the spatial distribution and the temporal development of the signal's intensity, images of the estimated effect sizes are devoid of any time-related information. In short, the input of model estimation is characterised by a temporal dimension, but the output is not. To understand why this elision happens, we need to remind ourselves that the automated

410 Earlier in this chapter, I have discussed how each task condition is repeated many times during an experiment. The very purpose of the repetition is to enable the averaging of the task-induced BOLD responses during the stage of model estimation.

411 Ashburner et al., "SPM12 Manual," 78.

412 As stated previously, a 4D fMRI dataset encodes the signal intensities registered not only at different spatial locations across the subject's brain but also throughout multiple task repetitions at various time points.

413 See section 3.2.1.

algorithms required the temporal correlation between the design matrix and fMRI data to compare the measured and predicted BOLD time courses. Based on this comparison, the algorithms computed the effect size estimates by averaging the BOLD responses across multiple repetitions of the same task. Hence, it can be said that the purpose of the temporal information was to enable the closing of the gap between the data and the model. Having fulfilled its purpose, the temporal information is no longer needed and, therefore, disappears from the rest of the analysis.

So far, we have discussed the process of model estimation. We now need to examine its output. In mathematical terms, the images of the estimated effect sizes are 3D matrices. Like fMRI data, these 3D matrices can also be visualised as a series of grey-scale brain slices.[414] At this point, a layperson might presume that large effect sizes contained in these images indicate voxels activated by a given task condition. Based on this assumption, the layperson might expect that researchers can identify active voxels by visually inspecting these images. This, however, is not the case. In fact, researchers do not even look at these images but merely use them as input for the next stage of algorithmic analysis. This is because these intermediary images are just as illegible as the fMRI data from which they were computed. Put simply, even in the images of the estimated effect sizes, the information of interest is still not encoded in ways that make it accessible to visual inspection. The problem is the following. Since they were computed from extremely noisy data, even numerically large effect sizes do not necessarily point to the presence of task-induced neural responses but could have instead occurred by mere chance.[415] To resolve this problem, in the next step of data analysis, researchers must evaluate whether an estimated effect size is significant compared to the residual noise in the data. To do this, researchers deploy inferential statistics.

As mentioned previously, inferential statistics entails testing the assumption called the null hypothesis. Generally speaking, the null hypothesis states that a task component of interest failed to elicit any brain activity in a given voxel. To submit the data to automated hypothesis testing, researchers must first specify the null hypothesis in relation to their particular experimental conditions and then decide which type of test to use to evaluate the thus defined null hypothesis. Depending on the analysis software they are using, researchers can choose among several types of test statistics. Each of the available tests implements a different mathematical model and makes different assumptions about the data.[416] Despite differences, the most commonly used statistics—such as t-tests and F-tests—share a key feature. They quantify the uncertainty of a task-induced response by evaluating its average estimated effect size relative to the extent to which this effect size randomly fluctuated during the experiment.[417] That is, both t- and F-tests measure if the task-induced effect is

414 Ashburner et al., "SPM12 Manual," 78.

415 Worsley, "Statistical Analysis," 251.

416 For details, see Worsley, 257–59.

417 For details, see Poldrack, Mumford, and Nichols, *Handbook*, 194–200. It is worth noting that to enable statistical testing, it is necessary first of all to calculate the level of noise fluctuation in the data. This is done by computing at each voxel the so-called error variance—the difference between the measured signal and the value predicted by the design matrix. See Worsley, "Statistical

sufficiently large compared to random noise so as not to have occurred by chance. As we will see later, based on the resulting numerical values of such test statistics, researchers differentiate between active and inactive voxels.

At this point, however, we still need to examine two significant aspects of hypothesis testing. First, in a group study, such as our case study, hypothesis testing is not done at the single-subject level, because the individual results are not of interest in themselves. Instead, the outputs of single-subject model estimations serve as the input for the second-level analysis. But before they can perform hypothesis testing at the group level, researchers must first use the software's algorithms to average the effect sizes across all subjects.[418] To this end, de Lange, Roelofs, and Toni used a so-called mixed-effect approach, whose underlying assumption is that the responses to the same task conditions vary randomly across subjects. In statistical terminology, this is expressed by saying that subjects are treated as a random effect.[419]

The mixed-effect approach is dominant in hysteria research and neuroimaging in general because it permits researchers to make inferences generalisable to a larger population.[420] The integral aspect of this approach is the estimation of a particular kind of noise, which is called inter-subject variance. Since this noise reflects the differences in the task-induced responses across subjects, it is not contained within a single dataset. Rather, this type of noise can be estimated only when fMRI data from various subjects are mathematically compared to one another. In this process, the individual subject's idiosyncratic task-induced neural responses are categorised as unwanted disturbances that could skew the results of the analysis. To eliminate such disturbances, during between-subject hypothesis testing, the algorithms quantify the magnitudes of the group-averaged task-induced responses relative to the variability in these responses across the subjects.[421] The resulting statistical group-level maps indicate only those task-induced neural responses that were shared across the subjects. Such responses are considered to be generalisable to all other hysteria patients with the same type of symptoms.

The second crucial aspect of hypothesis testing allows us to examine how human judgments shape algorithmic processes, as it entails researchers' decisions on how to specify concrete null hypotheses concerning their concrete experimental task. So far, we have seen how researchers first construct the design matrix and then

---

Analysis," 257–59; and Poldrack, Mumford, and Nichols, *Handbook*, 191–92. The computed values of error variance for all voxels are stored in a separate 3D image. See Ashburner et al., "SPM12 Manual," 78. Hence, the procedure of model estimation generates not only a set of images of estimated effect sizes but also an additional image that encodes the subject-specific spatial distribution of the estimated noise fluctuation across the brain.

418 The averaging entails building a second-level design matrix, which is then used during model estimation for calculating the means from the subject-specific effect size estimates. For details, see Poldrack, Mumford, and Nichols, *Handbook*, 102–4.

419 See Poldrack, Mumford, and Nichols, 100–2. An alternative approach, called fixed-effect analysis, assumes that all subjects reacted to the assigned task similarly. Yet, the fixed-effect analysis is viewed as less adequate since inferences based on it cannot be generalised beyond the sample. Ibid.

420 Ashburner et al., "SPM12 Manual," 63; and Poldrack, Mumford, and Nichols, *Handbook*, 100–5.

421 Poldrack, Mumford, and Nichols, *Handbook*, 102–4.

use hard-coded algorithms to estimate the contribution of each of its explanatory variables of interest to the BOLD responses measured across voxels. The next step of the analysis accommodates the fact that, as discussed previously, fMRI cannot measure the brain activity of interest in absolute terms. Instead, the acquired dataset only provides information about the relative MR signal changes across different experimental conditions.[422] For this reason, during hypothesis testing, researchers use test statistics to assess differential BOLD responses to various combinations of experimental conditions. In this context, a comparison of two or more experimental conditions—i.e., explanatory variables of interest—is called contrast. Working with such contrasts characterises the statistical analysis in most task-based fMRI studies.[423]

In fact, defining a set of null hypotheses in terms of testable contrasts represents the key step in implementing the design matrix as the study-specific model. This particular step enables researchers to combine multiple elements of their design matrix in various ways, both across different conditions within a single subject and among multiple subjects. Once they have used the design matrix to define contrasts, researchers can then look for the effects of these contrasts in the data. Crucially, through such use of contrasts, researchers explore their data in search of task-elicited brain responses. For example, researchers can search for voxels in which the activation either increased or decreased in response to a single task condition as opposed to baseline. Alternatively, research can choose to identify the locations of the voxels in which a particular explanatory variable of interest induced a greater average BOLD response than another variable.[424] For each of the contrasts thus defined, the algorithms calculate a separate activation map.

When defining contrasts for hypothesis testing, researchers can rely on the analysis software to automatically generate a range of mathematically possible contrasts based on the structure of the design matrix they created.[425] Yet, importantly, both the SPM and other analysis programmes permit researchers to flexibly define a variety of custom-made contrast. As we are about to see in the example of the de Lange, Roelofs, and Toni study, another significant point about hypothesis testing is that researchers do not compute activation maps for all calculable contrasts. Instead, researchers select only those contrasts they deem potentially meaningful. As my analysis will show, 'meaningful' contrasts are only those judged to be able to isolate a set of cognitive components of interest and map these onto the associated neural activity to deliver insights into the neural mechanism underlying the phenomenon under investigation.

---

422 See Huettel, Song, and McCarthy, *Imaging*, 354.

423 Hypothesis testing of single contrasts that entail a comparison of two conditions is performed with t-tests. Conversely, F-tests are used for contrasts that simultaneously compare multiple conditions. For details, see Poldrack, Mumford, and Nichols, *Handbook*, 194–200. Importantly, the contrast we are discussing here (in the sense of comparing the effects of two or more experimental conditions) is not to be confused with the image contrast we discussed earlier in this chapter.

424 The baseline condition is typically not included as a separate explanatory variable in the matrix, even when it plays a role in an experiment. If used in a contrast, the baseline is defined as the mere absence of all the other explanatory variables—i.e., a null condition. See Ashburner et al., "SPM12 Manual," 63.

425 See Ashburner et al., 88, 267, 269.

Contrasts that fail to unambiguously fulfil this function are disregarded. In effect, the choice for which particular contrasts to compute functional maps is guided by researchers' assumptions about the elementary cognitive components—and associated neural responses—that different aspects of their tasks were designed to induce. Let us now turn to our case study to see how this is done in practice.

As analysed previously, de Lange, Roelofs, and Toni constructed the first-level design matrix that contained ten explanatory variables of interest. Each variable referred to the presentation of either the affected or the unaffected hand in one of the five rotation levels. Although these variables could have been compared in many different ways, de Lange, Roelofs, and Toni chose to compute only two contrasts, which they then forwarded to the second-level analysis.[426] The first contrast entailed the comparison of the overall activity induced by the drawings of the affected as opposed to the unaffected hand, irrespective of their rotation levels. The other contrast isolated the increasing hand-independent BOLD response elicited by the increasing rotation level of the presented hand drawing as opposed to baseline.[427] These two contrasts allowed the researchers to isolate two mutually independent aspects of their task. The first contrast permitted them to search the data for the neural effects associated with hysterical paralysis. The second contrast enabled them to identify the neural responses elicited by the increasing task complexity. De Lange, Roelofs, and Toni chose to disregard all other possible contrasts at the single-subject levels, thus effectively declaring them meaningless.[428]

During group analysis, the researchers recombined the two single-subject contrasts from the first-level analyses to create more complex comparisons. By recombining the single-subject contrasts, de Lange, Roelofs, and Toni defined four different across-subject contrasts in the second-level analysis.[429] First, they computed the same two contrasts as they had done at the first-level analyses, only this time averaging them across all subjects. Additionally, they created a third group-level contrast to test if their two experimental factors (i.e., hand affectedness and rotation levels) mutually influenced each other. Notably, this new group-level contrast enabled them to search for the responses induced by the rotation-related differences between the affected and unaffected hands across subjects.

The choices de Lange, Roelofs, and Toni made so far were selective since they did not test all mathematically possible contrast but only those they deemed potentially meaningful from the cognitive perspective. Nevertheless, until this point, the researchers remained in the framework of standard contrasts that were pre-specified by the software. Yet, at this point, de Lange, Roelofs, and Toni decided to exploit the fact that half of their patients had a left-hand and the other half a right-hand paralysis. This fact permitted them to differently rearrange the single-subject contrasts

426 De Lange and colleagues selected the so-called main effects of each factor. See de Lange, Roelofs, and Toni, "Self-Monitoring," 2053.

427 As mentioned earlier, patients were looking at a fixation cross during the baseline condition.

428 For instance, the researchers chose to disregard the contrast between the affected hand and baseline, as well as multiple possible contrasts between each single rotation level and baseline.

429 See de Lange, Roelofs, and Toni, "Self-Monitoring," 2053.

between the affected and the unaffected hand at the level of group analysis. Specifically, the researchers used these single-subject contrasts to construct the fourth group-level contrast that compared the activations elicited by the left and the right hand. Since the software could not automatically generate this group-level contrast,[430] de Lange, Roelofs, and Toni had to define it manually. That is, their fourth group-level contrast was a custom-made one. Importantly, through this intervention, de Lange, Roelofs, and Toni were able to separate the task-induced neural effect of hysterical paralysis from those related to the hand laterality and thus calculate separate activation maps for each of these effects. It is safe to assume that this course of action was motivated by the researchers' active judgment that a separate analysis of these two particular experimental effects was relevant for providing potential insights into neural correlates of hysterical paralysis.

***

By way of summarising my analysis of statistical modelling of fMRI data, several points need to be emphasised. We have seen that a significant part of statistical analysis entails automated algorithmic operations such as model estimation and the computing of test statistics. Yet, I have foregrounded that the selective use of contrasts during hypothesis testing allows researchers to substantially shape the automated processes. By combining the explanatory variables of the design matrix into different contrasts, researchers can choose how to flexibly decompose the measured task-induced BOLD responses into multiple, separately analysable constituent parts. Each thus defined contrast enables researchers to isolate the neural effects that a particular aspect of their experimental intervention induced in the data. Therefore, I argue that while defining contrasts of interest, researchers reason with their study-specific model and use it as a tool with which they can actively explore an fMRI dataset from a variety of perspectives.

In the subsequent phase of hypothesis testing, automated algorithms analyse the data to identify the brain areas activated by the contrasts of interest, computing a separate statistical activation map for each contrast. Potential effects of other contrasts that could have been specified through alternative combinations of the elements of the design matrix are fully disregarded during hypothesis testing. The entire process is informed by researchers' selective judgments about which particular set of calculable contrasts is relevant for detecting the putative neural mechanisms of hysteria. Hence, the choice of pertinent contrasts is an act of interpretation a computer algorithm cannot make. Through this act of interpretation, researchers define which aspects of their

430 This is because de Lange, Roelofs, and Toni did not specify in the design matrix whether the presented stimulus was the right or the left hand. Instead, they only specified whether the stimulus referred to a patient's affected or unaffected hand. However, because the researchers knew which patient had an affected left instead of the affected right hand, they could easily intervene and instruct the software how to recombine the individual images to create the desired contrast between the left and the right hand. Since the study's authors did not respond to my attempts to communicate with them, the reconstruction I offer here is my own interpretation. This interpretation is based on the analysis of secondary literature and the insights I have gained while attending two SPM courses at the Department of Psychiatry and Psychotherapy, Charité Campus Mitte Berlin in March 2014 and January 2015.

experimental intervention will be made visible in the maps and which are relegated to noise.

Additionally, in this and the previous sections, I delineated how the operations of building and applying the study-specific statistical model to the fMRI data play a crucial role in constituting the activation maps' referential quality. Earlier in this chapter, I have shown that the measurement already establishes a physical link to the active brain. However, without the operations performed during statistical analysis, the task-induced neurophysiological effects of interest would remain buried under noise, as well as fragmented across fMRI images and datasets and, in effect, illegible. The consecutive steps through which the fMRI data are transformed into statistical maps thus articulate the traces of the neural effects of interests by isolating them from noise and synthesising them across multiple experimental conditions, time points and subjects.

As analysed above, this fMRI-specific process of articulation rests on a series of semantic operations that build a framework of interrelated comparisons and references.[431] Crucially, what follows from my analysis is that the resulting trace of the neural activity of interest does not exist independently of the process of its semantic articulation. In other words, my account challenges those neuroscientific narratives, which typically frame statistical analysis as a simple extraction of the information that had been inscribed into the fMRI data during the mutually synchronised experimental manipulation and data acquisition.[432] Contrary to this narrative, I claim that statistical analysis is best understood not as a passive reconstruction but as a medium-specific process of active interpretation. I have shown that statistical analysis relies heavily on the use of automated algorithms yet also necessitates researchers' active judgments to produce a new hybrid object. The resulting functional brain map is at once a fact and artefact,[433] a synthesis of measurement and modelling. Significantly, the process of computing epistemically valid functional brain maps is by no means arbitrary, as it is constrained by the evolving standards of the neuroimaging community about what constitutes acceptable methodological practice.[434] Thus constrained, this chain of interpretational operations provides an unbroken link between the resulting statistical maps and the indexical MR signals that went into the maps' construction.

---

431 For the sake of clarity, let me sum up the operations we have discussed in detail in this section. First, the model of the expected task-induced responses is compared to the data. Second, responses of a single subject to multiple task repetitions are compared to one another and averaged. Third, the average single-subject responses are compared across different individuals and again averaged. Fourth, different task conditions are mutually contrasted at both within- and across-subject levels and then compared to the level of noise. Importantly, the averaging across subjects is not based on merely calculating the arithmetic mean, since each subject is treated as a random variable. Hence, as previously mentioned, the averaging is based on the mixed-effects approach.

432 See, e.g., Worsley, "Statistical Analysis," 261.

433 See Latour, *Pandora's Hope*, 125.

434 Admittedly, the enormous flexibility with which researchers can analyse their data means that the process of statistical analysis is vulnerable to mishandling of the data by randomly trying out different analytical approaches and then selectively reporting only those that gave the best results (so-called p-hacking). However, such practices are considered bad science, producing epistemically questionable findings. See, e.g., Head et al., "P-Hacking," 1, e1002106.

Consequently, I argue that a statistical activation map is constituted as a highly mediated indexical sign. Its creation entails the combined effects of, first, the initial physical inscription of neurophysiological processes into the fMRI data; and second, the subsequent chain of semantic operations and mathematical transformations that articulate this trace in the data. My argument draws on Ludwig Jäger. He claims that to be instituted as an indexical sign of an object, a trace of some causal, physical contact with that object must undergo a medium-specific process of interpretation, which embeds this trace into a network of references to other signs and inscriptions. According to Jäger, both the indexical sign that points to an object and the object as the addressee of the sign's referential function are constituted through such semantic operations.[435]

My detailed analysis has shown that the indexicality of a functional activation map in the context of hysteria research does not consist in the map's ability to point to a single neural event or even to an individual subject's idiosyncratic, random brain activity. Instead, the indexicality of a functional map consists in its ability to point to, mostly group-averaged, brain activities of interest that were isolated during protracted statistical data analysis through a particular comparison of experimental conditions. Just as importantly, I have demonstrated that the indexicality of functional maps is as much a result of complex discursive and mathematical operations as it is of physical interventions. Therefore, the potential truth function of fMRI maps and, by extension, their epistemic efficacy in the scientific context cannot be divorced from the chain of the medium-specific operations that underpin their production.

However, before researchers can use the thus obtained fMRI activation maps to make judgments about possible neural mechanisms that underpin different hysterical symptoms, they must perform one additional step. As we will discuss in detail in the following section, this step addresses and aims to remedy the inherent limitation of statistical testing. Unless remedied, this limitation poses a serious threat to the carefully constructed indexicality of functional brain maps.

3.4.3 Disambiguating Active from Inactive Voxels

After the automated algorithms have calculated the chosen test statistics for a given contrast of experimental conditions across the entire brain, each voxel obtains a single numerical value. A large statistic value indicates a significant difference between the effects elicited by the experimental conditions contrasted at a given voxel. However, it is crucial to note that even a large statistic value in itself still does not provide sufficient reason to declare a voxel active. In fact, as we are about to see in what follows, even at this point, researchers have to make a few more crucial interpretational decisions before they can disambiguate active from inactive voxels.

435 See Jäger, "Indexikalität und Evidenz," 302–9. For similar positions that define indexicality not as a direct effect of the physical contact between an object and its sign but as a result of the subsequent process of interpretation, see Lefebvre, "Pointing," 220–44; and Olin, "Touching Photographs," 99–118.

To be able to reject the null hypothesis for a chosen contrast and thus declare a set of voxels active, researchers must first use the resulting test statistics to obtain the estimates of the so-called probability values (p-values). By definition, a p-value denotes the probability of observing under an identical replication of the experiment a test statistic as large as or larger than the one obtained, provided that the null hypothesis of no effect is true.[436] Expressed in simpler terms, the smaller the p-value is, the less likely it is that the reconstructed task-induced response is mere noise. By convention, the null hypothesis is rejected in a voxel whose p-value is below a predefined numerical level, called the significance threshold.[437] Voxels that fulfil this condition are considered to exhibit a statistically significant value. They are declared active and included in the statistical activation map. Conversely, all voxels with p-values above the threshold are labelled inactive and excluded from the map. Consequently, the resulting activation map does not display the presence of task-induced neural activations in absolute terms. Instead, and this is crucial, the map only shows the varying levels of probability that certain brain areas responded to a chosen contrast of experimental conditions.

A predefined threshold is used for distinguishing between active and inactive voxels so as to minimise the amount of what, in statistical terms, is referred to as the type I errors or false positives. Such errors arise when an inactive voxel is falsely declared active by rejecting the null hypothesis, although there was no actual experimentally induced effect in the data.[438] False positives are an inherent feature of statistical testing because there is always a chance of obtaining large statistic values by chance and thus mislabelling noise for an effect of interest. Such errors present a serious problem since they generate wrong information. To minimise the presence of false positives, in statistics in general and in fMRI in particular, the threshold is typically set at a nominal value of 0.05 for single test statistics. This means that a 5% rate of false positives is typically deemed to produce valid results.[439]

However, the problem concerning fMRI is that statistical tests are performed for each voxel separately across the whole brain volume. This approach entails an enormous number of tests, which inflate the number of false positives and result in what is known as the multiple comparisons problem.[440] For example, since a 3D fMRI image in our case study contained 64 x 64 x 32 voxels, approximately 50,000 to

436 Poldrack, Mumford, and Nichols, *Handbook*, 110.

437 Huettel, Song, and McCarthy, *Imaging*, 332–33.

438 Huettel, Song, and McCarthy, 332–33. See also Poldrack, Mumford, and Nichols, *Handbook*, 110.

439 Huettel, Song, and McCarthy, *Imaging*, 357. This arbitrary cut-off value "was originally developed by statistician Ronald Fisher in the 1920s in the context of his research on crop variance in Hertfordshire, England. Fisher offered the idea of *p*-values as a means of protecting researchers from declaring truth based on patterns in noise. In an ironic twist, *p*-values are now often used to lend credence to noisy claims based on small samples." Gelman and Loken, "Statistical Crisis in Science," 460. For discussions of the challenges and potential pitfalls of the current focus in the scientific research in general on a false-positive rate of 5% (i.e., $p \leq .05$) and how this can often lead to biased and unreproducible experimental results, see Gelman and Loken, 460–64; and Simmons, Nelson, and Simonsohn, "False-Positive Psychology."

440 Huettel, Song, and McCarthy, *Imaging*, 357.

75,000 independent statistical tests had to be performed for each contrast.[441] With the significance threshold set at 0.05 for every test in isolation, the resulting activation maps contained, on average, several thousand voxels that were falsely labelled active. The problem with false positives was humorously illustrated by a famous fMRI study by Bennett et al. in which the researchers 'demonstrated' the presence of brain activity in a dead salmon.[442] As a matter of fact, all activated voxels in the functional maps they computed for the dead salmon were false positives. Importantly, the very aim of the Bennett et al. study was to emphasise the necessity of adequately correcting such errors.

Multiple methods have been developed for addressing the multiple comparisons problem. Several of the most widely used methods are included in the SPM and comparable analysis software as available pre-programmed options.[443] The shared aim of all such options is to minimise the number of false-positive voxels in the resulting maps by calculating a corrected threshold value. What differs across the methods is how they calculate the corrected threshold value. Several particularly stringent correction procedures are jointly referred to as familywise error rate (FEW) methods. The FEW methods take into account the total number of statistical tests that have been performed across the brain volume during the analysis and then compute corrected maps, which, on average, have only a 5% chance of containing any false positives.[444] The newly calculated threshold value of 0.05 implies that only one in twenty corrected functional maps contains a false positive. The FEW methods are highly effective in controlling the false positives. Yet, their major drawback is that they considerably increase another type of intrinsic statistical error called false negatives.

False negatives are the direct opposite of false positives. Also known as the type II errors, false negatives arise when active voxels are falsely declared inactive by accepting the null hypothesis when there are actual effects in the data.[445] To avoid inflating the false-negative rate through the excessively stringent FEW methods, researchers may opt to use a more liberal correction approach, called the false discovery rate (FDR).[446]

---

441 Strictly speaking, a 3D fMRI image whose size is 64 x 64 x 32 entails 130,000 voxels. But the brain does not occupy the entire volume of this 3D image. Those portions of the image that do not contain brain tissue are referred to as "nonbrain voxels." Jenkinson and Chappell, *Neuroimaging Analysis*, 150. During the preprocessing step called the brain extraction, the intensity of nonbrain voxels is set to zero. Ibid. In a normalised 3D fMRI image, typically only 50,000–75,000 out of 130,000 voxels refer to the brain tissue. The rest are nonbrain voxels. Statistical testing is performed only on those voxels that contain brain tissue, whereas nonbrain voxels are entirely disregarded. See Ashburner et al., "SPM12 Manual," 69–70. Hence, the correction of the multiple comparisons problem only considers the number of tests performed on the within-brain voxels. I am grateful to Torsten Wüstenberg for drawing my attention to this fact.

442 Bennett et al., "Post-Mortem Atlantic Salmon," 39–41.

443 See Ashburner et al., "SPM12 Manual," 237–38; and Poldrack, Mumford, and Nichols, *Handbook*, 116–23.

444 Ashburner et al., "SPM12 Manual," 247–48. The three most widely used FEW procedures are the random field theory approach, the Bonferroni, and the Monte Carlo corrections. See also Poldrack, Mumford, and Nichols, *Handbook*, 117.

445 Poldrack, Mumford, and Nichols, *Handbook*, 111.

446 Poldrack, Mumford, and Nichols, 121–23.

However, while the FDR method increases the chance of detecting real effects in the data, its disadvantage is that it less effectively reduces the presence of false positives. This is due to the fact that the FEW and FDR methods not only deploy different mathematical models but also differently define what counts as an acceptable false-positive rate. By definition, in an FDR-corrected map with a significance value of 0.05, on average, 5% of all active voxels are false positives.[447] In effect, by choosing a specific correction method, researchers make crucial interpretational decisions about how to balance the reduction of false positives at the expense of increasing false negatives in their functional maps.

In essence, both false positives and false negatives present a major problem for fMRI analysis because a significant presence of either of these types of errors results in invalid statistical maps.[448] False positives lead researchers to make erroneous claims about non-existent effects in the data. False negatives are no less problematic as they cause researchers to miss potentially significant activations. The crucial problem, I suggest, is that both types of errors introduce a potential rupture into the thus far carefully constructed referential chain, which links statistical maps to the indexical MR signals. But these errors are the unavoidable price that researchers have to pay for using statistical analysis to translate the noisy, illegible fMRI data into legible functional maps.

It should be emphasised that fMRI maps can never be entirely purged of either false positives or false negatives. Nevertheless, we have seen that various correction methods allow researchers to reduce the rupture introduced by such errors. The principal goal of such correction methods is to achieve what members of the neuroimaging community consider an optimal balance between minimising the presence of both false positives and false negatives. If researchers manage to achieve this goal, the resulting maps are regarded to possess sufficient referential quality to point to the brain activities of interest and can thus serve as the basis for scientific judgments about these brain activities. It can, therefore, be argued that, if chosen adequately, the correction methods perform the operation of restoring the indexicality of fMRI maps. They do so by decreasing the presence of the elements that threaten to break the integrity of the referential chain which underpins the production of fMRI maps. My analysis has foregrounded that, on the one hand, this operation is material because it entails specific mathematical transformations to which fMRI maps are submitted. Yet, on the other hand, the restoration of the indexicality of fMRI maps is also a discursive operation, as it requires the authentication of the community of experts.

There are two caveats, however. First, the general adequacy of even well-established and widely used correction methods is still debated in the neuroimaging community. In other words, what counts as the optimal approach to correcting the multiple comparisons problem continues to be re-negotiated among experts. While some researchers "feel that conventional approaches to multiple-comparison correction are too lax and allow too many false positives", others argue that most "thresholds are

447 Poldrack, Mumford, and Nichols, 121–23.

448 In specialist terms, maps with a high rate of false positives are said to lack specificity, whereas those with a large amount of false negatives lack sensitivity. Poldrack, Mumford, and Nichols, 122.

too conservative and risk missing most of the interesting effects."[449] Second, the level of balance between the rates of false positives and false negatives researchers can achieve in a particular study is also limited by the conditions of the data acquisition. Specifically, the rates of both types of errors do not only depend on the efficacy of statistical tests. Instead, they are also influenced by the relative size of the task-induced effects compared to the noise and, most problematically, the size of the subject sample.[450] Consequently, studies with a small number of participants—which, as discussed previously, are prevalent in fMRI hysteria research—suffer from what is known as low statistical power. This means that such studies are hampered by a significantly lower chance of discovering true effects of experimental intervention in the data and a higher likelihood that the nominally positive results are false.[451] In short, small-sized studies tend to have higher rates of both false positives and false negatives. Moreover, by extension, small-sized studies might struggle with the fact that hardly any of their active voxels survive either of the correction methods described above.[452]

To circumvent this problem and thus avoid producing empty maps, many fMRI studies employ an alternative correction method called clusterwise thresholding. This approach predominates in fMRI hysteria research and was also used in our case study.[453] Its underlying assumption is that the likelihood of a single voxel being active by chance is much higher than that of a group of neighbouring voxels called a cluster.[454] In essence, researchers ignore single voxels and instead ascribe statistical significance only to groups of voxels whose size is above a threshold that specifies a critical cluster size. This approach effectively minimises false positives while also allowing researchers to detect activations that would not survive more stringent correction methods.[455] Yet its drawback lies in the potential loss of spatial specificity. If the calculated clusters are particularly large—as was the case in the de Lange, Roloefs, and Toni study—such maps

449 Poldrack et al., "Scanning the Horizon," 121–22.

450 Poldrack, Mumford, and Nichols, *Handbook*, 111.

451 Button et al., "Power Failure," 366. See also Cremers, Wager, and Yarkoni, "Statistical Power." Strictly speaking, the sample size required for detecting an underlying neural activity with a particular experimental design can be calculated using the procedure called power analysis. See Poldrack, Mumford, and Nichols, *Handbook*, 126–29. The problem with this analysis is that it is, in effect, somewhat circular. To perform it, one has to be able to estimate the size of the expected neural activity by relying on previously conducted studies. But, as discussed earlier, most fMRI studies of hysteria have so far been performed on small samples. Hence, it is easy to conclude that there is currently not enough reliable data for adequate power analysis in fMRI-based hysteria research.

452 Poldrack, Mumford, and Nichols, *Handbook*, 121.

453 See, e.g., Baek, "Motor Intention," 1626; Espay et al., "Functional Tremor," 182; and Stone et al., "Simulated Weakness," 963.

454 In a two-step procedure, researchers first choose a liberal primary threshold arbitrarily. This allows them to identify groups of neighbouring voxels whose individual statistical values lie above this primary threshold. In the second step, only those clusters that are as large as or larger than the cluster-size threshold are declared to be statistically significant and thus active. This second, more stringent threshold is calculated "based on the estimated distribution of cluster sizes under the null hypothesis of no activation in any voxel in that cluster." Importantly, a more liberal primary threshold results in a larger critical size threshold. Woo, Krishnan, and Wager, "Cluster-Extent Thresholding," 412. Hence, choosing the primary threshold is an important epistemic decision.

455 Huettel, Song, and McCarthy, *Imaging*, 361.

merely provide somewhat vague information that some signal was present somewhere within a relatively extensive brain area. Put differently, "even when the cluster-level false positive rate is well controlled, large true positive clusters are likely to consist of mostly noise and render the positive findings useless because of its low informativeness."[456] Thus, although cluster-size thresholding allows researchers to translate fMRI data into visualisable maps without compromising their epistemic validity, the resulting maps are not always unambiguously interpretable in anatomical terms. The interpretational ambiguity is particularly pronounced if active clusters happen to spread across multiple brain areas.

***

In sum, only after the ascription of statistical significance entailed in thresholding and the correction of multiple comparisons problem are researchers finally able to distinguish between active and inactive voxels. Significantly, the ascription of significance is also an attribution of visibility since only those voxels that pass the corrected threshold are visualised in the resulting statistical maps. We have seen that researchers can choose among various commonly used thresholding methods, all of which have particular advantages but also carry potential pitfalls. To produce maps that are indexically linked to the brain activity of interest, researchers must find a trade-off between controlling both false positives and false negatives, while at the same time achieving sufficient spatial specificity. Moreover, researchers must not only comply with the standards of the neuroimaging community but also take into account the particular epistemic limitations of their study.

On the whole, I suggest that the ascription of significance represents a focal semantic operation in fMRI analysis. Depending on how optimally researchers are able to perform it, this operation either successfully perpetuates or ruptures the medium-specific construction of the functional maps' indexicality on which the potential epistemic validity of these images hinges. Notably, the indexicality of functional maps necessarily remains highly indirect. It amounts to pointing with sufficient statistical likelihood to the presence of task-induced activations, which researchers can finally visualise and interpret. But before we turn to discussing how researchers work with visualisations of functional activation maps, let us now take a step back and examine how an alternative statistical analysis can be used to produce an entirely different kind of brain map from the same fMRI dataset.

3.4.4 Modelling the Legibility of the Brain's Internal Interactions

In the previous sections, we have examined the operations through which scientists transform fMRI data into statistical activation maps to identify the spatial distribution of the brain areas activated by a chosen contrast of experimental conditions. Referred to as functional segregation or localisation, this approach parcellates the brain into separate, functionally specialised regions.[457] Despite its widespread

456 Woo, Krishnan, and Wager, "Cluster-Extent Thresholding," 418.

457 See Büchel and Friston, "Brain Connectivity," 295.

application, from the point of view of cognitive neuroscience, localisation has a major epistemic limitation. Based on activation maps alone, researchers cannot determine whether—and if then how—disparate brain regions interacted with one another to produce the task-induced responses.[458] To surpass this limitation, researchers can use an alternative approach that permits them to make inferences about "how spatially distant brain regions interact and work together to create mental function."[459] Known as functional integration, this approach comprises different analytical methods and different concepts of what counts as an interaction among brain regions.[460] Two key concepts that dominate this still relatively new approach are functional and effective connectivity.

Functional connectivity is defined as a correlation in temporal patterns of activity across remote brain regions. Its underlying assumption is that the temporal coherence of the spatially distributed brain activities indicates some level of mutual interaction among these activities.[461] Although mostly used in resting-state fMRI studies,[462] functional connectivity analyses can also be applied to task-based data. Yet the caveat is that such analyses provide neither information about the direction of the neural interactions nor about how such interactions arise.[463] Conversely, the alternative concept of effective connectivity comprises analyses aimed at determining the influence that one brain region exerts upon another, thus allowing researchers to "disambiguate correlations of a spurious sort from those mediated by direct or indirect neuronal interactions."[464] A variety of methods used for measuring effective connectivity deploy not only different models of neural influence but also ascribe different levels of causality to that influence. Furthermore, there is a disagreement in the neuroscientific literature about where to draw the demarcation line between functional and effective connectivity.[465]

Due to such competing approaches to both how connectivity is defined and analysed, functional integration is still considered "a less than a mature field."[466] Nevertheless, the use of connectivity analyses in cognitive neuroscience has surged in

458 An activation map neither provides information about the region-specific responses' temporal sequence nor their mutual causal relationships. Büchel and Friston, 295–56.

459 Poldrack, Mumford, and Nichols, *Handbook*, 130.

460 See Poldrack, Mumford, and Nichols, 130–59. For a detailed account, see Friston, "Functional Integration," 471–91.

461 Büchel and Friston, "Brain Connectivity," 296.

462 As mentioned previously, in the resting-state fMRI paradigm, the subject is not required to perform an explicit task, but instead instructed to lie still and not think about anything specific. Resting-state fMRI studies deploy various types of functional connectivity analyses to identify correlated patterns of intrinsic brain activities that are independent of any external stimuli. See, e.g., Raichle, "Restless Brain." I will discuss the application of the resting-state approach in contemporary hysteria research in section 4.4.1.

463 Büchel and Friston, "Brain Connectivity," 296.

464 Friston et al., "Psychophysiological Interactions," 219.

465 For an overview of methods, see Friston, "Functional and Effective Connectivity", 13–36. See also Poldrack, Mumford and Nichols, *Handbook*, 130–59. Interestingly, these two accounts differ in where they place the demarcation line between functional and effective connectivity.

466 Büchel and Friston, "Brain Connectivity," 307.

recent years.[467] This general trend has been mirrored by a gradual increase in both resting-state and task-based studies that aim to establish how spatially distributed brain areas interact to give rise to hysterical symptoms.[468] Interestingly, none other than de Lange, Toni, and Roelofs authored the first full-length fMRI group study that applied connectivity analysis to hysteria.[469] Three years after their initial fMRI paper on hysterical arm paralysis—which so far served as our case study—de Lange, Toni, and Roelofs returned to the same dataset. This time, they used a method called the psychophysiological interaction (PPI) to translate their initial fMRI data into a set of statistical connectivity maps. Since then, multiple task-based fMRI studies of hysterical symptoms have used the PPI to compute connectivity maps.[470] It can, therefore, be said that this type of functional map is playing an increasing role in recent attempts to elucidate potential neural correlates of hysterical symptoms. For this reason, in what follows, I will analyse the operations that determine the production of task-based connectivity maps by drawing on the example of the de Lange, Toni, and Roelofs study from 2010.

In general terms, the psychophysiological interaction analysis permits researchers to make inferences about how task-induced cognitive processes (i.e., the psychological factor) alter the influence that one brain region has on others (i.e., the physiological factor).[471] To perform the PPI analysis, researchers must first specify the task components whose modulatory effect is of interest to them. Next, they need to choose the area—called the seed region—whose influence on the rest of the brain they want to investigate. Since the seed region is necessarily an area activated by the task components of interest, researchers must first perform a standard GLM activation analysis to identify its location.[472] Put simply, the creation of a pertinent statistical activation map is a necessary precondition for the PPI analysis. For this reason, de Lange, Toni, and Roelofs used the PPI analysis to build directly upon their initial study in which they had pinpointed several areas of the prefrontal cortex that were differentially activated by the stimuli of the affected and the unaffected hand.[473] With the PPI analysis, the researchers could now use the same fMRI dataset to ask the following question: With which brain areas did the chosen seed regions interact differently depending on whether the patients were induced to imagine moving their affected or

467 In 2010, "the annual increase in publications on connectivity surpassed the yearly increase in publications on activations *per se*." Friston, "Functional and Effective Connectivity," 13 (emphasis in original).

468 Baek et al., "Motor Intention"; Otti et al., "Somatoform Pain"; and Voon et al., "Limbic Activity."

469 See de Lange, Toni, and Roelofs, "Altered Connectivity."

470 See, e.g., Aybek et al., "Life Events"; Hassa et al., "Motor Control"; and Voon et al., "Involuntary Nature."

471 Friston et al., "Psychophysiological Interactions," 223. Strictly speaking, the PPI is a method located at the intersection between functional and effective connectivity. Researchers use the PPI to establish a neural interaction that is stronger than a mere temporal correlation across brain regions. However, researchers cannot interpret the thus identified neural interaction in terms of any clear-cut causal relations. See Ashburner et al., "SPM12 Manual," 340.

472 See Ashburner et al., "SPM12 Manual," 341; and Poldrack, Mumford, and Nichols, *Handbook*, 134.

473 De Lange, Roelofs, and Toni, "Self-Monitoring," 2056. I will return in more detail to the researchers' interpretation of the activation maps later in the chapter.

unaffected hand? Significantly, this new question allowed the researchers to shift the focus away from identifying the direct effects that the external factors (i.e., the task conditions) had on the patients' brains and focus instead on examining the internal neural interactions.

Answering this question with the PPI analysis meant that de Lange, Toni, and Roelofs had to once more rely on the general linear model (GLM) to construct yet another study-specific design matrix.[474] They then used this matrix in the subsequent statistical testing to compute the connectivity maps. We now appear to be on familiar ground, as this process sounds similar to the standard GLM analysis discussed previously. However, there are several significant differences. We have seen that the standard GLM analysis allowed researchers considerable autonomy in defining the elements of the design matrix. This autonomy, as I have argued, was a necessary precondition that enabled researchers to pertinently model the expected effects of their experimental task on the data. By contrast, the PPI design matrix comprises three fixed types of explanatory variables that partition the BOLD response within each voxel into a combination of the experimental intervention and the brain's internal interactions. These variables include: first, the estimated local BOLD response to the task condition; second, the input from the seed region's BOLD response; and finally, the PPI term that models the additional task-modulated influence of the seed region.[475] Since the structure of the PPI design matrix is predefined, in this case, researchers have a considerably narrower modelling autonomy than in the activation analysis. In fact, the only modelling decisions they can make are choosing the location of the seed region and selecting the task condition of interest.

Despite its apparent structural simplicity, the construction of the PPI design matrix is far from straightforward, as it requires multiple intermediary modelling steps. First, the seed region's BOLD response must be computed using the classical GLM activation analysis.[476] This means that the PPI analysis is already implicitly informed by the theoretical assumptions, mathematical operations, and interpretational decisions inscribed into the preceding activation analysis. Moreover, the biggest challenge involves specifying the PPI term. It is worth noting that the PPI term is of central interest for the analysis as it models the predicted task-modulated neural interaction between the seed region and the rest of the brain. To define the PPI term, researchers must first estimate the neural activity in the seed region. Since fMRI cannot measure the neural activity directly, researchers rely on specifically developed deconvolution algorithms that use sophisticated mathematical modelling to compute the most likely neural signal underlying the BOLD response from the seed region.[477] Finally, to build the PPI term that predicts the BOLD responses across the brain, the estimated neural signal must be multiplied by the timing of the experimental task that induced it and a

474 De Lange, Toni, and Roelofs, "Altered Connectivity," 1783–84. See also Ashburner et al., "SPM12 Manual," 339–41.

475 Ashburner et al., "SPM12 Manual," 339–40; and Poldrack, Mumford, and Nichols, *Handbook*, 134–35.

476 Ashburner et al., "SPM12 Manual," 340; and Poldrack, Mumford, and Nichols, *Handbook*, 134.

477 Poldrack, Mumford, and Nichols, *Handbook*, 135–36.

model of the haemodynamic response.[478] Only after the SPM's black-boxed machinery has executed all these extensive mathematical transformations can the PPI matrix be put to work.

The deployment of the PPI matrix is performed in steps that are very similar to those we have analysed in the previous section. Hence, the outputs of single-subject model estimations are used for the voxelwise hypothesis testing of chosen contrasts at the group level.[479] However, the key difference is that, in the PPI analysis, the effect of interest is defined by comparing the PPI term and baseline.[480] This contrast allows researchers to identify the voxels in which the BOLD response temporally co-fluctuated with the experimentally induced response in the seed region.[481] Using this contrast, researchers can determine which spatially distant brain areas interacted differently with the seed region under the influence of the task. In other words, what is of interest in the PPI analysis is the indirect influence that the task-related neural activity in the seed region had on the task-related brain activities in the rest of the brain. Crucially, this means that what was considered noise in the standard activation analysis is now declared the signal of interest. Conversely, the direct effects of the task on the BOLD response in each voxel, which represented the information of interest in the standard activation analysis, are treated as noise by the PPI analysis and, therefore, disregarded during statistical testing. Thus, what counts as pertinent information and what is viewed as a disturbing factor is not fixed within a single fMRI dataset. Instead, such decisions depend entirely on the type of analysis researchers choose to perform on the data.

After the algorithms had executed hypothesis testing at each voxel, and the results underwent clusterwise thresholding as described previously, de Lange, Toni, and Roelofs were able to visualise their connectivity maps. The resulting maps displayed the brain areas whose neural interactions with the chosen seed regions in the prefrontal cortex either increased or decreased with sufficient statistical significance, depending on whether the patients were shown the imagery of the paralysed or the healthy hand. The PPI analysis thus enabled the researchers to use the same fMRI dataset as in the previous study, but this time to create maps that provided complementary insights into the hysteria patients' neural activities. In the initial study, the researchers used their fMRI dataset to identify those isolated brain areas whose localised dysfunction might have given rise to hysterical paralysis. Conversely, the connectivity analysis facilitated a substantial shift in the perspective. In the subsequent study, de Lange, Toni, and Roelofs used the same fMRI dataset to identify the aberrant interactions across spatially distant brain regions as the potential neural mechanism underlying hysterical paralysis.

478 Poldrack, Mumford, and Nichols, 136.

479 Ashburner et al., "SPM12 Manual," 350–54.

480 In sections 3.4.1 and 3.4.2, I have shown that researchers can flexibly define and test a variety of contrasts of interest during standard activation analysis.

481 Friston, "Functional and Effective Connectivity," 23.

***

In summary, the two consecutive studies by de Lange, Toni, and Roelofs generated categorically different imaging findings through the applications of two different analytical approaches to the same fMRI dataset. We have seen that each of the two approaches was informed by a substantially different model of the brain function. In one case, the focus was on strictly localised activations (functional segregation), whereas in the other, on the dynamic connections among spatially remote brain areas (functional integration). Just as significantly, each approach also rested on partly contrary definitions of what counted as the information of interest in the fMRI data instead of noise. Therefore, each approach required that the researchers deploy different kinds of mathematical transformations to obtain what they defined as pertinent information.

In effect, my analysis has shown that the kind of information that is articulated from a particular fMRI dataset and translated into a legible statistical map is, at the most basic level, predicated on the model of the brain's functional organisation which underpins the analytical approach chosen by researchers. Because these models are not mutually exclusive, they can be applied in separate analytical procedures to the same fMRI dataset to construct multiple, mutually complementary statistical brain maps. Through the use of such mutually complementary analyses, a single fMRI dataset is constructed as what I would like to designate as *semantically multipotent*. What I mean by this is that each fMRI dataset holds the potential to be made legible in multiple epistemically valid ways. As we have seen, it is up to researchers to decide which specific semantic potential of their fMRI dataset they want to articulate to answer their study-specific research questions. In each case, the result of such an articulation is a particular statistical brain map.

3.5 Visualising Functional Brain Maps: Ascribing the Symbolic Meaning

Only after they have completed all the steps entailed in the time-consuming data analysis and thus obtained the statistical maps of their choice can researchers finally turn to evaluating the empirical results of their experiment. To put it more plainly, it is not before this point that researchers can even see which brain areas were differentially activated—with sufficient statistical significance—by the comparisons of the experimental conditions they chose to test. Having invested weeks or even months into painstakingly constructing their functional maps, researchers can, at last, use them to answer two crucial questions. In which anatomical regions of the brain did the experimental intervention trigger neural responses? And, how do such patterns of brain activity relate to cognitive processes that play a role in the formation and manifestation of the hysterical symptom of interest, or more generally, any other phenomenon under investigation?

Answering these questions requires researchers to make sense of their statistical brain maps. Yet, there is one crucial point that I want to make. Although the statistical brain maps are legible, their exact informational content and medical meaning are far