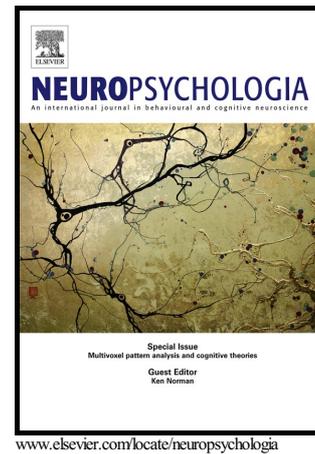


Author's Accepted Manuscript

A hitchhiker's guide to lesion-behaviour mapping

Bianca de Haan, Hans-Otto Karnath



PII: S0028-3932(17)30396-2
DOI: <https://doi.org/10.1016/j.neuropsychologia.2017.10.021>
Reference: NSY6540

To appear in: *Neuropsychologia*

Received date: 6 June 2017
Revised date: 16 October 2017
Accepted date: 17 October 2017

Cite this article as: Bianca de Haan and Hans-Otto Karnath, A hitchhiker's guide to lesion-behaviour mapping, *Neuropsychologia*, <https://doi.org/10.1016/j.neuropsychologia.2017.10.021>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A hitchhiker's guide to lesion-behaviour mapping

Bianca de Haan^{1,1}; Hans-Otto Karnath^{1,2}

¹ Center of Neurology, Division of Neuropsychology, Hertie-Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany

² Department of Psychology, University of South Carolina, Columbia, USA

¹ Present address: Division of Psychology, Department of Life Sciences, Centre for Cognitive Neuroscience, Brunel University London, Uxbridge, UK

Corresponding author:

Bianca de Haan

Division of Psychology, Department of Life Sciences

Brunel University London

Kingston Lane, Uxbridge, UB8 3PH, UK.

E-mail: bianca.dehaan@brunel.ac.uk

Tel: 0044 (0)1895 265797

Abstract

Lesion-behaviour mapping is an influential and popular approach to anatomically localise cognitive brain functions in the human brain. Multiple considerations, ranging from patient selection, assessment of lesion location and patient behaviour, spatial normalisation, statistical testing, to the anatomical interpretation of obtained results, are necessary to optimize a lesion-behaviour mapping study and arrive at meaningful conclusions. Here, we provide a hitchhiker's guide, giving practical guidelines and references for each step of the typical lesion-behaviour mapping study pipeline.

Key words:

Lesion analysis; voxel based lesion symptom mapping; VLSM; VLBM; brain behaviour inference; human

Accepted manuscript

1 Introduction

In a classical lesion analysis, the aim is to infer the cognitive function of an area of the human brain by observing the behavioural consequences of damage to that brain area. In the early days, the lesion method was the only approach available to study the functional architecture of the brain and these early studies have contributed tremendously to our understanding of a wide variety of cognitive functions (e.g. Damasio & Damasio, 1989). Nowadays, functional brain imaging and transient neuroinhibition/neurostimulation techniques have complemented this traditional approach. Nevertheless, the lesion method continues to be an essential and influential approach for neuroscientists aiming to study the functional architecture of the brain (see Rorden and Karnath, 2004 for a review). This continued importance of the lesion method as an approach to aid understanding of cognitive function is also illustrated by the steadily increasing number of scientific publications featuring this method over the years (see Figure 1).

--- Figure 1 about here ---

Surprisingly, despite this continued and increasing popularity of the lesion method, there currently are virtually no papers or books to assist scientists interested in using this method (see Wilson, 2016 for a notable exception). This is in stark contrast to the many excellent papers and books available to guide scientists interested in using other neuroscientific methods. Thus, inspired by similar recent guides for diffusion tensor imaging (DTI; Soares et al., 2013) and functional magnetic resonance imaging (fMRI; Soares et al., 2016), we here compile a hitchhiker's guide detailing the necessary considerations at each step of the typical lesion study pipeline (see Figure 2), with a primary focus on 'classical' univariate lesion analysis approaches (Bates et al., 2003; Rorden et al., 2007). New multivariate lesion analysis approaches have also been proposed (Smith et al., 2013; Mah et al., 2014a; Zhang et al., 2014; Pustina et al., in press, this issue) and, in light of known drawbacks associated with the univariate approach, such as limited statistical power and potential spatial bias (Mah et al., 2014a; Inoue et al., 2014; for a review, see also Sperber & Karnath, in press, this issue), it has been suggested that these multivariate approaches should be preferred over univariate lesion analysis approaches (Mah et al., 2014a). However, multivariate approaches have their own issues requiring future improvements, such as the feature selection method (Yourganov et al., 2015; Rondina et al., 2016). Moreover, the spatial bias in univariate approaches can be

reduced considerably by correcting for lesion volume and ensuring sufficient minimum lesion overlap (Sperber & Karnath, 2017; see also section 5.2.1 below), and the possibility that multivariate approaches are not likewise affected by a spatial bias has not been ruled out yet. As such, a more nuanced view would be to consider univariate and multivariate lesion analysis approaches as complementary (Karnath et al., in press).

--- Figure 2 about here ---

2 Patient selection

The first decision a researcher planning a lesion-behaviour mapping study has to make, is which patients to select for the study. Patient assessment is time-consuming and, as such, it is generally most efficient to restrict patient assessment to those patients that will allow both meaningful conclusions on the functional architecture of the brain and a meaningful investigation of the research hypothesis.

2.1 Lesion aetiology

One frequently used patient selection criterion is lesion aetiology. Of the 181 studies depicted in Figure 1 (i.e. the publications between 1995 and 2015 that used the lesion method found searching Pubmed), the vast majority (66.3%) were conducted with stroke patients.

2.1.1 Use acute stroke patients to investigate the functional architecture of the brain

In the acute phase, strokes are associated with clear behavioural consequences that can, for the large part, be directly linked to the original function of the functionally impaired part of the brain (as strokes are sudden and as such the brain has not yet had time to functionally reorganise, see also Shahid et al. [2017]). As long as patients with mass shifts due to extensive haemorrhage or extensive oedema are excluded, all brain structures are typically still at their original locations and the parts of the brain affected by stroke can be reliably visualised in computed tomography (CT) (Mohr et al., 1995; Mayer et al., 2000) and/or magnetic resonance (MR) (Genovese et al., 2002) images (see section 3.1.1 below). As such, acute stroke patients are highly suitable for studies where the aim is to investigate the functional architecture of the brain.

2.1.2 Use chronic stroke patients to investigate the neural correlate of chronic dysfunction

Unfortunately, acute stroke patient data is not always easy to acquire. Access to stroke units and acute stroke patients may be restricted and the behavioural assessment of acute stroke patients is frequently difficult due to their often poor general state of health. As a consequence, many lesion-behaviour mapping studies have relied on more readily available chronic stroke patient data. However, in the chronic stroke phase, functional reorganization of the brain in the course of normal recovery can complicate the investigation of the functional architecture of the brain (Karnath and Rennig, 2016). This is mostly due to the fact that during lesion-behaviour mapping analyses, chronic stroke patients who have recovered from their initial cognitive deficit are grouped together with chronic stroke patients that never had a deficit. As a consequence, the lesion-behaviour mapping analysis (incorrectly) assumes that the parts of the brain damaged in chronic stroke patients who have recovered are not, or less critically, associated with the cognitive function of interest. Karnath and Rennig (2016) recently tested the consequences of this effect, comparing the three most common combinations of structural imaging data and behavioural scores used in previous lesion-behaviour mapping studies. Only the combination of acute behavioural scores and acute structural imaging precisely identified the targeted brain areas. In contrast, lesion-behaviour mapping analyses based on chronic behaviour, in combination with either chronic or acute imaging, hardly detected any of the targeted substrates.

Moreover, in the chronic stroke phase, the precise determination of the parts of the brain that are functionally impaired in CT and/or MR images is complicated by secondary morphological changes to the brain due to tissue resorption following brain damage, such as structural distortions, sulcal widening, and ventricle enlargement (Karnath and Rorden, 2012). As such, chronic stroke patients are less suitable than acute stroke patients for studies where the aim is to investigate the functional architecture of the brain. However, if combined with CT and/or MRI data obtained in the acute stroke phase, the behavioural data from chronic stroke patients has been shown to be highly suitable for studies where the aim is to investigate the neural correlates of chronic cognitive deficits, i.e. studies where the aim is to determine where in the brain acute damage results in a cognitive deficit that is still present in the chronic stroke stage 6 months or more following stroke onset (Karnath et al., 2011; Abela et al., 2012; Wu et al., 2015). This is a question with a high clinical relevance, as it ultimately has the potential to enable long-term clinical predictions based on the location of the acute brain damage.

2.1.3 Avoid combining acute and chronic patients in the same lesion-behaviour mapping analysis

Importantly, combining both acute and chronic stroke patients in the same lesion-behaviour mapping study entails the risk of ending up with the worst of two worlds: Even if CT and/or MRI data is obtained in the acute stroke phase for all patients, the different amounts of cortical reorganization in different patients are likely to confound the interpretation of the lesion-behaviour mapping results, regardless of whether the aim was to study the functional architecture of the brain, or to study the neural correlates of chronic cognitive deficits. This problem is most straightforwardly demonstrated with a thought experiment (see Figure 3): Imagine that both lesions depicted on the top of Figure 3 equally affect the brain area crucially related to a certain cognitive function of interest. Thus, directly following stroke onset, the behavioural deficit in these two patients is maximal and equal (i.e. the behavioural deficit score is '54' for both cases). In both patients the severity of the behavioural deficit decreases equally over time due to spontaneous recovery. Now imagine that both patients are recruited for a lesion-behaviour mapping study. However, whereas one of the patients is recruited and behaviourally assessed in the acute stroke phase (and thus shows a behavioural deficit score of '54' in Figure 3), the other patient is first seen and assessed in the intermediate/chronic stroke phase following considerable spontaneous recovery of his/her behavioural deficit (let us assume by 50%, leading to a measured behavioural deficit score of '27' in Figure 3). Although both brain lesions equally affect the brain area crucially related to the cognitive function of interest, the lesion analysis now erroneously weighs the lesion location of the patient with the behavioural deficit score of '27' as being less relevant for the cognitive function of interest than the lesion location of the subject with the behavioural deficit score of '54'. As such, the overall contribution of this brain area to the cognitive function of interest is ultimately underestimated. The main underlying issue is that lesion-behaviour mapping analyses assume that each patient is assessed at the same point in time following stroke onset and that thus the contribution of a certain brain area to a certain cognitive function is directly reflected in the behavioural scores in each patient. Combining both acute and chronic stroke patients in the same lesion-behaviour mapping study violates this assumption.

---- Figure 3 about here ----

2.1.4 Lesion aetiologies other than stroke

Beyond strokes, lesion-behaviour mapping analyses have also been conducted with other lesion aetiologies. In the 181 lesion-behaviour mapping studies depicted in Figure 1, the second and third most popular lesion aetiologies were traumatic brain injury and brain tumour (10.5% and 5.5% of the studies respectively). Moreover, a further 9.9% of these studies combined patients with different lesion aetiologies in the same study (e.g. included patients with stroke, traumatic brain injury and brain tumour). There is, however, considerable debate concerning the suitability of these patients with lesion aetiologies other than stroke for lesion-behaviour mapping analyses. Specifically, a considerable body of work suggests that traumatic brain injury and tumour patients might be less suitable than stroke patients for lesion-behaviour mapping analyses that aim to study the functional architecture of the healthy brain.

In traumatic brain injury patients, a major neuropathological component, beyond focal brain damage and regardless of traumatic brain injury severity (mild, moderate or severe) or mechanism (closed or penetrating), is diffuse axonal injury (Gennarelli et al., 1982; Povlishock and Katz, 2005; Büki and Povlishock, 2006; Su and Bell, 2016). Moreover, this diffuse axonal injury has been suggested to contribute significantly to the cognitive impairments (and their recovery) observed following traumatic brain injury (Povlishock and Katz, 2005; Levine et al., 2013). However, while areas of diffuse axonal injury of sufficient size can be detected in MRI (Su and Bell, 2016), the full extent of diffuse axonal injury can only be detected histopathologically (Adams et al., 1991; Povlishock, 1993; Johnson et al., 2013). This presents a significant problem for lesion-behaviour mapping analyses, as these analyses require an accurate *in vivo* determination of which areas of the brain are functionally impaired and which areas of the brain are functionally intact. Additionally, traumatic brain injury patients are typically investigated in a chronic disease phase, in which case the same caveats as noted above for chronic stroke patients hold.

In brain tumour patients, there is likewise evidence that an accurate determination of functionally impaired and functionally intact areas of the brain can be problematic. The most common type of malignant primary brain tumour is glioma, accounting for 74.6% of all malignant brain and other central nervous system tumours (Ostrom et al., 2016). In gliomas, however, the precise spatial extent of the tumour is impossible to determine. The vast majority of gliomas are characterised by diffuse infiltration of surrounding tissue (Scherer,

1940) that can extend considerably beyond the tumour border visible in conventional T1 or T2 MRI images (Burger et al., 1988; McKnight et al., 2002; Swanson et al., 2004). While more recent imaging modalities such as diffusion tensor imaging and proton MR spectroscopy may improve the visualisation of the tumour (Claes et al., 2007), many areas of tumour infiltration occur at a spatial scale that cannot be detected even with these newer imaging modalities. Critically, and presenting a significant problem for lesion-behaviour mapping analyses, it is unclear whether or to what extent brain function is impaired in these (in vivo not detectable) areas of diffuse tumour infiltration (Karnath and Steinbach, 2011). In the case of preoperative tumour patients, this problem in accurately determining the functionally impaired and functionally intact areas of the brain is further exacerbated by observations suggesting that brain function can sometimes be preserved within the tumour, particularly (but not exclusively) in patients with a slow-growing low-grade glioma (Ojemann et al., 1996; Skirboll et al., 1996; Schiffbauer et al., 2001).

An additional problem with using glioma patients in lesion-behaviour analyses is that these tumours tend to develop on a relatively long time-scale. That is, unlike strokes or traumatic brain injury, gliomas do not have a sudden onset. Instead, gliomas slowly grow and are only diagnosed when they are both large enough to be detected in CT or MR images and clinically symptomatic (Swanson et al., 2003; Pallud et al., 2013). This relatively long time-scale of development means that, by the time the tumour is diagnosed, most glioma patients' brains have undergone a considerable amount of (compensatory) functional reorganization (Wunderlich et al., 1998; Fandino et al., 1999; Thiel et al., 2001; Holodny et al., 2002; Meyer et al., 2003; Taniguchi et al., 2004; Shaw et al., 2016). This means that in tumour patients, the behavioural consequence of brain damage may no longer reflect the original function of the damaged part of the brain, which presents a serious problem for lesion-behaviour mapping analyses that aim to study the functional architecture of the brain.

2.2 Lesion location

Beyond lesion aetiology, another patient selection criterion is lesion location. For example, based on a convincing body of previous findings, we might a priori expect certain language functions (e.g. language production) to be associated with a certain part of the brain (e.g. the left hemisphere). As such, when interested in, e.g., language production, our research question might be „which areas of the left hemisphere causally contribute to language production?“. In this case, it would make little sense to also assess patients with right

hemispheric brain damage. Likewise, on a smaller spatial scale, we might know from various previous investigations that a certain function of interest is located in a particular region of the brain, e.g., that a specific executive function is governed by prefrontal cortex. As such, when interested in describing where exactly *within* the prefrontal cortex this specific function is located, it would make little sense to also assess patients with posterior brain damage. However, while restricting patient selection to patients with damage to only a certain part of the brain is valid in these cases, it is also important to realise that this then by definition means that no inferences can be made about the potential (and perhaps even larger) contributions of not-investigated areas of the brain. Moreover, caution should be used when extending this logic to multiple parts of the brain (see also section 5.3 below). In patients with larger strokes, brain lesions are likely to encompass more than just one of these areas of interest. This creates a significant problem during classification of these cases. Exclusion is no solution, as this would create a bias towards smaller lesions and potentially milder cognitive symptoms, ultimately leading to different anatomical conclusions.

2.3 General criteria

Finally, beyond these main patient selection criteria, there are a few typical general exclusion criteria. Firstly, patients with evidence of clinically relevant cognitive impairments such as dementia or mental retardation and/or evidence of psychiatric disorders are usually excluded. In these patients, a valid assessment of the behaviour of interest would be difficult. Secondly patients with evidence of additional (pre-existing) neurological disorders beyond the neurological disorder of interest, such as Parkinson's disease, infections of the central nervous system, or older and/or additional diffuse brain lesions due to, e.g., previous strokes or chronic hypertension, are also usually excluded (although up to 2-5 pre-existing silent lacunes are typically allowed). In these patients it would be difficult to determine which aspect(s) of the behavioural deficit (if observed) can be attributed to the neurological disorder of interest and which aspect(s) of the behavioural deficit might instead be due to these additional neurological disorders. Finally, patients with mass shifts due to extensive haemorrhage or oedema should be excluded. In these patients, brain areas are no longer at their original positions, which potentially confounds the interpretation of lesion-behaviour mapping analysis results. Importantly, however, *absence of the behavioural deficit of interest should not be used as an exclusion criterion*. The inclusion of patients that do not have the behavioural deficit of interest (control patients) is essential, as this allows us to differentiate between areas of the brain where damage is associated with the deficit of interest and areas of

the brain where damage merely reflects increased vulnerability to injury (Rorden and Karnath, 2004). Moreover, restricting patient selection solely to patients that show the behavioural deficit of interest reduces variance in the behavioural data and so ultimately reduces statistical power to detect an effect in lesion-behaviour mapping analyses (see also section 3.2 below). Instead, researchers should ideally a priori decide on a reasonable patient recruitment time period and unselectively include all suitable (i.e. matching all inclusion criteria and none of the exclusion criteria) patients that present during that time period in the study.

3 Patient assessment

The next decision a researcher planning a lesion-behaviour mapping study has to make, is how to assess lesion location and behavioural status in each patient. Given the problems associated with lesion aetiologies other than stroke (see section 2.1.4 above), the following section and the rest of this manuscript will focus on stroke patients.

3.1 Assessing lesion location

As mentioned in section 2.1 above, imaging data should be obtained in the acute stroke phase, regardless of whether the aim is to study the functional architecture of the brain, or to study the neural correlates of chronic cognitive deficits. In acute stroke patients, the lesion can be visualised using either CT (Mohr et al., 1995; Mayer et al., 2000) or MRI (Neumann-Haefelin et al., 1999; Ricci et al., 1999; Schlaug et al., 1999). The development of CT templates for spatial normalisation of individual patient images (Rorden et al., 2012; see also section 4.1 below) has removed the main reason to disregard CT for lesion-behaviour mapping studies. Moreover, modern spiral CT scanners provide high resolution images and in many clinical institutions CT remains the dominant imaging modality of choice at admission. Importantly, the choice between administering CT or MRI to patients at admission is not random, but follows specific clinical criteria. As a consequence, the systematic exclusion of patients with CT images only implements a selection bias, typically influencing important factors such as lesion size, general clinical status, severity of cognitive deficits etc. (for a detailed discussion, see Sperber and Karnath, in press, this issue). As such, both CT and MRI data can and should be used for lesion analysis studies.

We suggest the following practical guidelines for the assessment of lesion location: In *patients with CT imaging only*, use noncontrast CT images to visualise the brain lesion. In noncontrast CT images, acute haemorrhagic strokes appear as hyperintense areas within minutes to hours following stroke onset (Bergström et al., 1977). Ischemic strokes, on the other hand, appear as hypointense areas between 24 to 36 hours following stroke onset (Mohr et al., 1995; von Kummer et al., 2001). In *patients with MR imaging only*, use diffusion-weighted images (DWI) to visualise the lesion if imaging is performed less than 48 hours following stroke onset, and use T2FLAIR images (ideally supplemented with DWI, particularly during the first 5 days following stroke onset [Ricci et al., 1999]) if imaging is performed more than 48 hours following stroke onset. In DWI, ischemic strokes appear as hyperintense areas within 2-6 hours following stroke onset (Warach et al., 1992; González et al., 1999), while the initial T2FLAIR infarct hyperintensity might be too subtle for accurate lesion visualisation within the first 48 hours (Lansberg et al., 2001). Finally, while not suitable for the visualisation of the lesion in the acute phase of a stroke, a T1 image might aid spatial normalisation (see section 4.2 below). In *patients with both CT and MR images*, the researcher is in the privileged situation to choose the best from both modalities, i.e. to use those images where the lesion is most conspicuous.

The information provided by structural imaging data could be meaningfully complemented with imaging data that allows visualisation of areas of the brain that are structurally intact, but may function abnormally (e.g. the ischemic penumbra and/or areas of diaschisis). That is, multimodal imaging of brain damage, where structural and functional information are combined, might provide a more accurate picture of the full extent of brain damage than structural imaging alone. In clinical settings, visualisation of areas of the brain that are structurally intact, but may function abnormally can be done using perfusion CT (Mayer et al., 2000; Koenig et al., 2001) or MR perfusion-weighted imaging (PWI; Schlaug et al., 1999; Schaefer et al., 2002; Zopf et al., 2012). Shahid et al. (2017) were recently able to show that, when patient assessment was performed within the first 48 hours following stroke onset, lesion analysis inferences were more accurate when based on both structural and perfusion imaging than when based on structural imaging alone. Unfortunately, however, effective usage of perfusion data is often difficult due to the fact that the precise relationship between the severity of hypoperfusion and the severity of the functional impairment is largely unknown. While guidelines have been posited concerning the degree of hypoperfusion likely to lead to a behaviourally relevant functional impairment for some areas (Neumann-Haefelin

et al., 1999; Hillis et al., 2001; Motta et al., 2014), it is currently unclear whether these guidelines are equally applicable to all areas of the brain. This is in contrast to areas that are lesioned, where we know that function is completely lost. Moreover, while there is evidence for the behavioural relevance of the ischemic penumbra (Shahid et al., 2017), evidence for the behavioural relevance of remote diaschisis is still mixed. While several studies suggest that subcortical damage may result in behaviourally relevant remote cortical hypoperfusion (Hillis et al., 2001, 2002, 2005; Karnath et al., 2005; Ticini et al., 2010), evidence that cortical damage results in behaviourally relevant remote cortical hypoperfusion is so far lacking (Zopf et al., 2009).

3.1.1 Lesion delineation

Following CT or MR data acquisition, the lesion needs to be delineated on each slice of the patient's brain image. The standard is manual lesion delineation, which can be done using programs like MRIcroN (<https://www.nitrc.org/projects/mricron>; Rorden and Brett, 2000), or ITK-SNAP (<http://www.itksnap.org>; Yushkevich et al., 2006). However, manual lesion delineation is time-consuming and potentially observer-dependent (Ashton et al., 2003). To address these disadvantages, both fully automated and semi-automated lesion delineation methods have been developed. Fully automated lesion delineation methods can roughly be divided in unsupervised (e.g. Gillebert et al., 2014; Mah et al., 2014b) and supervised (e.g. Griffis et al., 2016; Pustina et al., 2016) classification algorithms. As the name implies, fully automated methods do not require any user interaction. As such, these methods are substantially less time-consuming and observer-dependent than manual lesion delineation, which improves replicability and reproducibility across labs. A considerable downside of these fully automated methods, however, is that they may be more susceptible to imaging artefacts and thus less precise than the current gold standard, manual lesion delineation. Importantly, this reduced precision associated with fully automated lesion delineation methods may influence subsequent lesion-behaviour mapping results (Pustina et al., 2016). Given this downside, semi-automated lesion delineation methods that combine fully automated steps with mandatory user interaction might be able to provide an optimal compromise. While several semi-automated lesion delineation approaches exist (e.g. Wilke et al., 2011), the semi-automated lesion delineation approach *Clusterize* (<https://www.medizin.uni-tuebingen.de/kinder/en/research/neuroimaging/software/>; Clas et al., 2012) has recently been shown to be capable of significantly speeding up lesion delineation, without loss of either lesion delineation precision or lesion delineation

reproducibility in acute stroke patients scanned in both CT and a range of common MRI modalities (de Haan et al., 2015). The principle of *Clusterize* is simple: On the basis of local intensity maxima and iterative region growing, the whole CT or MR brain image is fully automatically clusterized, including the lesioned area. The user subsequently manually selects those clusters that correspond to the lesion. *Clusterize* may thus combine the best of two worlds: the presence of fully automated steps makes it less time-consuming than manual lesion delineation, while mandatory user interaction results in lesion delineation that is more precise and/or less error-prone (in the sense that they are closer to the results from the current gold standard, manual delineation) than that obtained by fully automated lesion demarcation methods (de Haan et al., 2015).

3.2 Assessing behavioural status

In a lesion-behaviour mapping analysis, the behavioural status needs to be assessed for each patient. Typically, this means that the cognitive function of interest needs to be operationalised. Several important considerations concerning this operationalisation of cognitive functions, such as taking care to distinguish between impaired performance on a test used to operationalise a cognitive function and the clinical syndrome of interest, and ensuring that the test used to operationalise a cognitive function is as specific to the cognitive function of interest as possible, are discussed in Sperber & Karnath (in press, this issue). Additionally, it is important to ensure that the test used to characterize the behavioural deficit has sufficient sensitivity. To help ensure this, the range of scores used in a test should be chosen to measure the full range of the underlying behaviour with both sufficient and meaningful resolution. That is, the optimal range of scores would allow both a measurement of the full possible range of behaviour and a measurement of the smallest meaningful difference in the behaviour of interest. Insufficient sensitivity both reduces the ability to detect a true effect in lesion-behaviour mapping analyses, and reduces the likelihood that an observed significant effect reflects a true effect (Button et al., 2013; Ingre, 2013). Moreover, given that cognitive performance is typically expressed on a continuous scale, dichotomization of this inherently continuous performance should be avoided. Dichotomization of continuous behaviour is known to result in a significant loss of information and thus of statistical power (Cohen, 1983). In the rare cases where dichotomization of continuous behaviour is unavoidable (for example when performing lesion subtraction analyses to study exceedingly rare deficits, see section 5.1 below), the dichotomization and subsequent classification of individual patients as ‘impaired’ or ‘non-

impaired' should be performed with proper statistical procedures (Crawford and Howell, 1998; Crawford and Garthwaite, 2005). Finally, when assessing the behavioural status of a patient, potential neuropsychological co-morbidity should be taken into account (see also Bonato et al., 2012; Sperber & Karnath, in press, this issue). Care should be taken to ensure that the test used to operationalise the cognitive function of interest does not introduce a systematic bias against patients with certain co-occurring deficits. Moreover, frequently co-occurring deficits (e.g. reduction of general cognitive status, language impairments) whose severity might correlate with the severity of the cognitive deficit of interest should ideally be assessed in addition to the cognitive function of interest, so that they can be controlled for during the lesion-behaviour mapping analyses (e.g. by including them as nuisance covariates, see section 5.2.1 below).

4 Spatial normalisation of patient brain and lesion map

Following patient imaging and lesion delineation, we have a 3D binary lesion map reflecting the voxels where brain function is impaired for each patient that we can use for lesion-behaviour mapping analysis. However, all brains differ in orientation, size, and shape. As such, before we can perform voxelwise (statistical) comparisons, we need to spatially normalise the patient brains and lesion maps to ultimately ensure that a given voxel (roughly) represents the same anatomical structure in each patient. Thus, the third decision a researcher has to make, is how to spatially normalise the patient brain and lesion map. Spatial normalisation can be performed with programs such as BrainVoyager (<http://www.brainvoyager.com/>; Goebel, 2012), SPM (<http://www.fil.ion.ucl.ac.uk/spm/>), FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>; Jenkinson et al., 2012), AFNI (<https://afni.nimh.nih.gov/>; Cox, 1996, 2012), or ANTs (<http://stnava.github.io/ANTs/>; Avants et al., 2011). The analysis package SPM is widely used due to its platform independence, free obtainability, and availability of many add-ons. As such, we here will focus on the spatial normalisation routines of SPM, as implemented in the Clinical Toolbox (<https://www.nitrc.org/projects/clinicaltbx/>; Rorden et al., 2012). This Clinical Toolbox provides specialised templates that allow spatial normalisation of both CT and MR brain images of elderly, stroke-aged populations (see section 4.4 below). As such, the Clinical Toolbox is ideally suited to be used in lesion-behaviour mapping studies where the patients included are typically older, and where different modalities, i.e. CT as well as MR images,

are present in different patients. Beyond routines for traditional spatial normalisation and unified segmentation and normalisation approaches (see sections 4.1 and 4.2 below), the Clinical Toolbox also provides processing steps that aid the spatial normalisation of scans from stroke patients, including corrections for the presence of a lesion (see section 4.3 below).

4.1 Spatial normalisation of imaging data with low radiometric resolution

Imaging data with a low radiometric resolution is imaging data where the image intensity values offer a poor differentiation between different tissue types, particularly between grey and white matter brain tissue. In acute stroke patients, this typically applies to CT and T2FLAIR data. Here, the typical approach is to perform spatial normalisation by matching the orientation, size, and shape of each patient brain to the orientation, size, and shape of a template brain in standard stereotaxic space. This matching is done with an automated algorithm that aims to find the necessary image transformations that minimize the least mean square difference between the voxel intensities of the patient and template brain (Ashburner and Friston, 2003). In a first step, the patient brain is matched to the template brain using linear (affine) image transformations that can include translations, rotations, zooms, and/or shears. These affine transformations change the entire patient brain in the same way and so result in a global match between the patient brain and the template brain. Subsequently, in a second step, the patient brain is further matched to the template brain using nonlinear (nonaffine) image transformations consisting of cosine basis functions. These nonaffine transformations allow local changes to the patient brain and so improve the match between patient and template brain. To avoid overfitting during this second step, the difference in the amount of nonlinear transformations of adjacent areas is simultaneously minimized (known as 'regularisation'). As such, spatial normalisation matches the overall orientation, size and shape of the patient brain to that of the template brain, but not individual sulci. Once the necessary image transformations have been estimated, they can be applied to both the patient's brain image and the lesion map, bringing both in standard stereotaxic space.

4.2 Spatial normalisation of imaging data with high radiometric resolution

Imaging data with a high radiometric resolution is imaging data where the image intensity values offer a good differentiation between different tissue types, particularly between grey and white matter brain tissue. This typically does not apply to the imaging data used to visualise the lesion in acute stroke (see section 3.1 above), but does usually apply to an

additionally collected T1 image. Likewise, when DWI data is collected, this is usually collected with different b-values (typically b0, b500, and b1000). While the b1000 image best suited to visualise the lesion in acute stroke patients has a low radiometric resolution, the additionally collected b0 image often has a relatively high radiometric resolution (Mah et al., 2014b). In these cases, the typical approach is to first coregister the image with a low radiometric resolution (e.g. the T2FLAIR or the b1000 DWI), used to visualise the lesion, to the image with a high radiometric resolution (e.g. the T1 or the b0 DWI). Subsequently, the image with a high radiometric resolution is normalised using the unified segmentation and normalisation approach (Ashburner and Friston, 2005). This approach has been shown to be superior to the traditional normalisation approach discussed in section 4.1 above in both healthy (Crinion et al., 2007; Klein et al., 2009) and stroke (Crinion et al., 2007) populations, but requires an image with a high radiometric resolution. The unified segmentation and normalisation approach combines tissue classification (i.e. segmentation), bias correction, and image registration (i.e. spatial normalisation) in a single model. Estimation of the model parameters is done by repeatedly alternating between tissue classification (to assign each voxel to a tissue class on the basis of its intensity), bias correction (to correct for image intensity nonuniformity due to magnetic field inhomogeneity), and image registration (to bring patient image and template-based tissue probability maps into common space using regularised affine and non-affine transformations and so derive the image transformations necessary to bring the patient image into standard stereotaxic space) steps. As this model accounts for the conditional dependencies between the steps (i.e. tissue classification aids the bias correction and image registration steps, and bias correction and image registration aid the tissue classification step), the results are superior to those obtained following serial application of the same steps. Once the necessary image transformations have been estimated, they are applied to the patient's brain image(s) and the lesion map, bringing all images in standard stereotaxic space.

4.3 Correcting for the lesion during spatial normalisation

While both spatial normalisation approaches described above work well with imaging data from neurologically healthy subjects, the presence of a lesion in imaging data from stroke patients presents a challenge, as lesions are characterised by abnormal image intensity. This area of abnormal image intensity in the patient brain locally creates a large mismatch between the patient brain and the template brain / template-based tissue probability maps, ultimately leading to local overfitting in the lesion area during the minimization of this mismatch (Brett

et al., 2001; Andersen et al., 2010). The two dominant solutions to this overfitting problem are cost function masking (Brett et al., 2001) and enantiomorphic normalisation (Nachev et al., 2008). During cost function masking, lesioned voxels are excluded during spatial normalisation using (a typically slightly smoothed version of) the binary lesion map. As such, the image transformations necessary to bring the patient's brain image(s) and the lesion map in standard stereotaxic space are derived from intact areas of the brain only. During enantiomorphic normalisation, on the other hand, the lesion is 'corrected' by replacing it with brain tissue from the lesion homologue in the intact hemisphere of the brain. As such, the image transformations necessary to bring the patient's brain image(s) and the lesion map in standard stereotaxic space are effectively derived from a brain image without a lesion.

Logically, one would expect enantiomorphic normalisation to perform better than cost function masking when lesions are large and unilateral, as spatial normalisation with cost function masking becomes less accurate as lesion size increases (as the area from which the image transformations can be derived decreases with increasing lesion size), while enantiomorphic normalisation does not. Cost function masking would, however, be expected to perform better than enantiomorphic normalisation when lesions are bilateral and affect similar areas in both hemispheres, as enantiomorphic normalisation would in this case replace the lesion with the likewise lesioned homologue. Moreover, as enantiomorphic normalisation assumes that the brain is essentially symmetric, enantiomorphic normalisation might be suboptimal in areas known to be considerably asymmetric (e.g. the planum temporale).

4.4 Choosing the right template

Finally, an important consideration during spatial normalisation concerns the choice of template / template-based tissue probability maps. Firstly, when using the traditional spatial normalisation approach described in section 4.1 above, the template image should ideally have the same image modality as the patient image that is spatially normalised, as the accuracy of this approach depends on how similar the voxel intensities of a given brain area are between the patient and the template image. The unified segmentation and normalisation approach described in section 4.2 above, on the other hand, is modality-independent. Secondly, regardless of the spatial normalisation approach chosen, the population from which the template image or template-based tissue probability maps are derived should roughly match the population of the lesion mapping study. That is, if the lesion-behaviour mapping

study is performed in elderly stroke patients, the template or template-based tissue probability maps used should ideally have been derived from an elderly population.

For elderly adults (mean age 61.3 years), a template is available for CT imaging data, as well as template-based tissue probability maps for MR imaging data (Rorden et al., 2012). For young adults (mean age 25 years), templates are available for T1 and T2 imaging data, as well as template-based tissue probability maps (Mazziotta et al., 1995, 2001a, 2001b). For paediatric populations, various templates are available for T1 imaging data, as well as template-based tissue probability maps (e.g. Wilke et al., 2008; <http://jerlab.psych.sc.edu/neurodevelopmentalmriddatabase/>). Finally, a template derived from a wide range of adults (mean age 35.4, range 18-69 years) is available for T2FLAIR imaging data (<http://glahngroup.org> or <http://brainder.org>; Winkler et al.). Given that the average age of acute stroke patients included in a lesion-behaviour mapping study is typically over 60, the Clinical Toolbox (<https://www.nitrc.org/projects/clinicaltbx/>) includes the above-mentioned CT template and template-based tissue probability maps derived from elderly adults, as well as the T2FLAIR template. This toolbox can, however, easily be modified for use with other templates or template-based tissue probability maps.

5 Performing voxelwise (statistical) comparisons

Following lesion delineation and spatial normalisation, we have a spatially normalised binary lesion map for each patient. Moreover, we have a behavioural measurement for each patient. With these two sources of information, we are ready to perform a voxelwise lesion-behaviour mapping analysis to relate lesion location and patient behaviour. Over the years, the methods to perform voxelwise analyses have continuously improved, from early subtraction analyses, to voxelwise statistical analyses (with correction for multiple comparisons), and ultimately to voxelwise statistical analyses that account for nuisance covariates such as lesion volume. Each of these methods will be discussed in the sections below.

5.1 Lesion subtraction analyses

The simplest type of voxelwise analysis is a lesion subtraction analysis. Here, the lesion overlap map of patients without the cognitive deficit of interest is subtracted from the lesion overlap map of patients showing the cognitive deficit of interest. This can be performed with

programs such as MRIcroN (<https://www.nitrc.org/projects/mricron>). To account for potential sample size differences between the two patient groups, these subtraction analyses need to use proportional values. That is, for each voxel the percentage of patients without the cognitive deficit of interest that have a lesion at the voxel is subtracted from the percentage of patients with the cognitive deficit of interest that have a lesion at the voxel. The result of the subtraction analysis is then a map with the percentage relative frequency difference between these two groups for each voxel. For example, say we have 10 patients with the cognitive deficit of interest and 20 patients without the cognitive deficit of interest, where at a given voxel 9 of the 10 patients with the deficit have a lesion (i.e. 90%), while 10 of the 20 patients without the deficit have a lesion (i.e. 50%). In this case, the percentage relative frequency difference at that voxel would be $90\% - 50\% = 40\%$, indicating that this voxel is damaged 40% more frequently in patients with the cognitive deficit of interest than in patients without this deficit. These subtraction analyses are superior to simple overlap analyses that focus on only those patients that show the disorder of interest, because overlap analyses might simply highlight regions that reflect increased vulnerability of certain regions to injury (see Rorden & Karnath, 2004). In contrast, a lesion subtraction analysis highlights those areas of the brain where lesions are more frequent in patients with than in patients without the cognitive deficit of interest, and so distinguishes between regions that are merely often damaged in strokes and regions that are specifically associated with the deficit of interest (Rorden and Karnath, 2004). To control for neuropsychological co-morbidity, the two patient groups contrasted in a lesion subtraction analysis need to be comparable with respect to additional neurological impairments (of no interest) such as, e.g., paresis, visual field defects, etc. While subtraction analyses have merit in the study of exceedingly rare cognitive deficits (where it would be near to impossible to obtain a larger sample size), it is important to realise that these analyses are purely descriptive and allow no statistical inference. For statistical inference, voxelwise statistical analyses are necessary.

5.2 Voxelwise statistical lesion-behaviour mapping analyses

In a voxelwise statistical lesion-behaviour mapping analysis, we perform a statistical test at each voxel to relate voxel status (lesioned/nonlesioned) and patient behaviour. These voxelwise statistical analyses can be performed with programs such as VLSM (<https://langneurosci.mc.vanderbilt.edu/resources.html>; Bates et al., 2003), VoxBo (<https://www.nitrc.org/projects/voxbo/>, Kimberg et al., 2007), NPM (<https://www.nitrc.org/projects/mricron>; Rorden et al., 2007), NiiStat

(<https://www.nitrc.org/projects/niistat/>) and/or LESYMAP (Pustina et al., in press, this issue). When the *behavioural data is continuous*, the behavioural data of the group of patients in whom a given voxel is damaged is statistically compared to the behavioural data of the group of patients in whom that same voxel is intact. This is traditionally done with a two-sample t-test, which assumes that the behavioural data is normally distributed and measured on an interval scale. Unfortunately, however, behavioural data from patient populations is often not normally distributed. During behavioural assessment, patients without the deficit of interest will typically all demonstrate close to maximum performance, whereas performance in patients with the deficit of interest will typically be poorer and more variable over patients. As a consequence, the distribution of the behavioural data from patient populations is often negatively skewed. Moreover, behavioural data from patient populations is often not measured on an interval scale. Instead, many tests designed to assess cognitive function in patient populations measure patient behaviour on an ordinal scale, where the behavioural data is ordered (e.g. higher scores denote better performance), but the distances between the individual measurements are not known (e.g. the difference between a score of ,1‘ and ,2‘ is not necessarily the same as the difference between a score of ,2‘ and ,3‘). Unfortunately, the t-test tends to be overly conservative when its test assumptions are violated, resulting in a reduction of statistical power to detect an effect in lesion-behaviour mapping analyses. Instead, the assumption free rank order test proposed by Brunner and Munzel (2000) might be more appropriate in these situations. In a lesion-behaviour mapping analysis with simulated data, this so-called Brunner-Munzel test has been shown to have higher statistical power than the t-test, while offering similar protection against false positives, in situations where the distribution of the behavioural data is skewed (Rorden et al., 2007).

When the *behavioural data is binomial* (i.e. when the deficit is either present or absent, as in e.g. hemianopia), we statistically assess, for each voxel, whether the variables ‘voxel status’ (voxel lesioned vs. voxel intact) and ‘behavioural status’ (deficit present vs. deficit absent) are associated or independent. The statistical test typically used in these situations is the Pearson’s chi-squared test. In many lesion-behaviour mapping analyses, however, expected cell frequencies are lower than 5-10 in at least some voxels, resulting in inflated false positive rates when using Pearson’s chi-squared test. Traditional solutions to this problem are to use Yates’s correction for continuity or a Fisher’s exact test. These solutions, however, both assume fixed marginals, meaning that they assume that the column totals (in how many patients is this voxel lesioned and in how many patients is this voxel intact) and row totals

(how many patients have a deficit and how many patients do not) are fixed in advance before data collection starts. Obviously, this is not the case in a typical lesion-behaviour mapping analysis, and as a consequence, both Yates's correction for continuity and Fisher's exact test tend to be overly conservative. A statistical test that might be more appropriate in these situations is the quasi-exact test proposed by Liebermeister (1877). In a lesion-behaviour mapping analysis with simulated data, observed false positive rates for this so-called Liebermeister test closely approximated the set false positive threshold, whereas observed false positive rates for Fisher's exact test tended to be too low (Rorden et al., 2007).

5.2.1 Inclusion of nuisance covariates and ensuring sufficient minimum lesion overlap

One variable known to correlate strongly with the severity of behavioural deficit in stroke populations is lesion volume (the larger the lesion, the more likely it is that a patient will show a behavioural deficit). Thus, to avoid identifying brain areas where damage is related simply to lesion volume instead of patient behaviour, lesion volume should be included as a nuisance covariate in a voxelwise statistical lesion analysis. This can be done using regression approaches (e.g. logistic regression or the general linear model). Moreover, statistical power varies over voxels as a function of the amount of lesions that overlap at each voxel, with statistical power theoretically being maximal at voxels that are lesioned in half of the patient sample. Importantly, statistical power is absent at voxels that are damaged in none of the patients (as this would result in one empty group for the t-test and Brunner-Munzel test, or two empty cells for the Liebermeister test). Thus, to ensure sufficient statistical power, voxels damaged in a very low percentage of the patient sample should be excluded. Correcting for lesion volume as well as ensuring sufficient minimum lesion overlap has been shown to reduce the spatial bias and so improve the anatomical validity in univariate voxelwise statistical analyses (Sperber and Karnath, 2017). Finally, similarly as done for lesion volume, other nuisance covariates can additionally be included, such as the severity of frequently co-occurring deficits that may correlate with the cognitive function of interest (see also section 3.2 above), fiber tract disconnection likelihood (Rudrauf et al., 2008), etc.

5.2.2 Correcting for multiple comparisons

During a voxelwise statistical lesion-behaviour mapping analysis, the same statistical test is performed at many individual voxels. However, if each statistical test has the typical false positive probability of 5%, performing a statistical test at e.g. 100 voxels will be expected to result in 5 false positives. That is, as more and more statistical tests are performed, the

probability of observing at least one false positive increases. In fact, performing 100 independent statistical tests, each with the typical false positive probability of 5%, will increase the overall probability of at least one false positive to 99.4%. In these situations, we do not want to control the probability of observing a false positive in each individual voxel. Instead, we want to control the overall probability of observing a false positive (over all voxels tested), also known as the family-wise error rate. To do this, we need to correct for multiple comparisons.

The traditional method to correct for multiple comparisons is the *Bonferroni correction*. Here, we simply divide our desired false positive probability by the amount of tests that we perform. Thus, if we assess 100 voxels and want to ensure that the family-wise error rate does not exceed 5%, we would set the false positive probability threshold for each individual voxel at $5/100=0.05\%$. While this method offers excellent control of the family-wise error rate, it is also very conservative (particularly in voxelwise lesion-behaviour mapping analyses where the individual voxels are not truly independent), and thus severely reduces the statistical power to detect an effect. As such, considerable efforts have been made to develop alternative, less conservative, ways to correct for multiple comparisons.

A more exact way to correct for multiple comparisons and control the family-wise error rate, without sacrificing statistical power, is *permutation thresholding*. Permutation thresholding aims to determine whether an observed test statistic at a voxel (e.g. a t-test, Brunner-Munzel or Lieberman statistic) is truly due to the difference in voxel status (lesioned or non-lesioned) or not. The underlying logic is that if the observed test statistic is truly due to the difference in voxel status, similar or more extreme test statistics would be unlikely to arise in situations where the pairing of behavioural data and voxel status is scrambled (i.e. situations where there is no association between behavioural data and voxel status). To determine how likely certain test statistics are under this null hypothesis of no association between behavioural data and voxel status, the behavioural scores of the patients with and the patients without damage at a certain voxel are randomly scrambled (i.e. permuted) thousands of times, each time calculating a new test statistic. With this, a distribution of permuted test statistics is created, reflecting the probability of observing certain test statistics under the null hypothesis. Using this null distribution of permuted test statistics, the 5% threshold value can be determined, with test statistics exceeding this threshold value having a probability of less than 5% under the null hypothesis. Finally, by comparing the originally observed test statistic to

this null distribution of permuted test statistics, we can determine whether the original test statistic was extreme enough to allow rejection of the null hypothesis. In the context of voxelwise statistical lesion-behaviour mapping, this approach is extended by using the maximum test statistic (over all voxels) obtained in each permutation to create the null distribution, instead of individual voxel test statistics. As such, the 5% threshold value is not exceeded anywhere in the brain in more than 5% of the permutations, that is, permutation thresholding offers the same control of the family-wise error rate as the Bonferroni correction. Importantly, however, in situations where the individual voxels are not truly independent, permutation thresholding offers better statistical power than the Bonferroni correction. Moreover, while permutation thresholding typically focusses on the maximum test statistic obtained in each permutation (controlling the probability of observing a single false positive), this approach can also be generalised by focussing on the n -th extreme test statistic, where $n > 1$ (Mirman et al., in press, this issue). This so-called continuous permutation-based family-wise error rate correction method (controlling the probability of observing n false positives) might allow for a better balance between false positives and false negatives in typical lesion-behaviour mapping studies where the anatomical interpretation of the results rarely depends on a single voxel.

Finally, an alternative, less conservative approach to correcting for multiple comparisons is offered by *false discovery rate thresholding* (Benjamini and Hochberg, 1995; Genovese et al., 2002). Here, the goal is not to control the family-wise error rate, but to control the proportion of false positives amongst observed positives. As a consequence, a false discovery rate threshold of 5% means that up to 5% of the observed positives might be false positives. In situations where no positives are observed, false discovery rate thresholding will provide the same control of the family-wise error rate as the Bonferroni correction. However, in situations where positives are observed, false discovery rate thresholding will result in more positives surviving the correction for multiple comparisons than either the Bonferroni correction or permutation thresholding. In fact, as the amount of observed positives increases, the false discovery rate threshold decreases. This adaptiveness of false discovery rate thresholding, however, comes at the price of reduced control of the family-wise error rate (as up to 5% of the positives surviving the correction for multiple comparisons could be false positives). Moreover, in smaller samples ($n = 30-60$), false discovery rate thresholding might considerably underestimate the proportion of false positives amongst observed positives (Mirman et al., in press, this issue). In situations where control of the family-wise error rate is

paramount and/or where the test assumptions of false discovery rate correction may be violated, permutation thresholding (see above) should thus be the preferred approach to correct for multiple comparisons in lesion-behaviour mapping analyses.

5.3 Avoid dividing samples into subsamples on the basis of an a priori hypothesis

Often, we have an a priori hypothesis concerning the parts of the brain that might contribute to a certain cognitive process. Accordingly, it might seem intuitive to divide the patients and their brain lesions into different subsamples and perform separate lesion-behaviour mapping analyses for each of these subsamples. For example, based on the a priori hypothesis that action-related aspects of cognition are represented in anterior parts of the brain while perception-related aspects of cognition are located in posterior brain regions, one might divide an unselectively recruited patient sample into a subsample of patients with more anteriorly located brain damage and a subsample of patients with more posteriorly located brain damage. There are, however, several problems with this approach (see Figure 4). Firstly, patients with large lesions (for example covering both anterior and posterior parts of the brain) are difficult to categorise. This might lead to an extra category (e.g., ‘anterior & posterior’) for which no clear a priori hypothesis exists. As mentioned before (see section 2.2 above), exclusion of these patients is no solution, as this would not only result in a significant loss of valuable information, but would also create a bias towards smaller lesions and potentially milder cognitive symptoms, ultimately leading to different anatomical conclusions. Secondly, dividing a single patient sample into subsamples will anatomically bias the results of a lesion-behaviour mapping analysis into the direction of this a priori hypothesis. Dividing a patient sample into, for example, a subsample with more anterior brain damage and a subsample with more posterior brain damage, will reveal two neural correlates: one neural correlate somewhere in the more anterior regions of the brain, and one neural correlate somewhere in the more posterior regions (see Figure 4B, left side). This result is a priori expected and simply a consequence of dividing the sample into these two anatomical subsamples, regardless of the cognitive deficit displayed by the patients. The a priori hypothesis of an anterior vs. posterior dissociation of perception- vs. action-related cognitive processes will thus lead to an observation that corresponds to this hypothesis. Had we taken the same data sample of stroke patients as before, but instead divided this sample into one subsample with more ventrally located brain damage and one subsample with more dorsally located brain damage (based on the equally defensible a priori hypothesis that perception-related aspects of cognition are represented in more ventral brain areas whereas action-related

aspects of cognition are represented in more dorsal areas of the brain), we would have revealed a different result (see Figure 4B, right side). In this case, the lesion-behaviour mapping analysis of the sample with more ventral brain damage would have revealed a neural correlate somewhere in more ventrally located regions of the brain, while the lesion-behaviour mapping analysis of the subsample with more dorsal brain damage would have found a neural correlate somewhere in more dorsally located regions. The problem illustrated here, is that the results of a lesion-behaviour mapping analysis can be biased by the categorization of patients on the basis of their lesion location. That is, the categorization of an unselected patient sample on the basis of an a priori anatomical hypothesis will lead to anatomical results that correspond to this hypothesis. As such, this should be avoided.

--- Figure 4 about here ---

6 Anatomical interpretation of lesion analysis results

Following a voxelwise statistical lesion-behaviour mapping analysis, we obtain a statistical map highlighting the voxels where voxel status (lesioned vs. non-lesioned) and patient behaviour are significantly related. In the case of a lesion subtraction analysis, on the other hand, we obtain a map highlighting areas of the brain where lesions are descriptively more frequent in patients with than in patients without the cognitive deficit of interest (often thresholded to isolate those percentage relative frequency difference values thought to be meaningful [with typical threshold values of 20-50%]). Anatomical interpretation then consists of describing the location of these significant or meaningful voxels, typically with the help of a brain atlas. For convenience, coordinates of peak voxels or a coordinate ranges of a cluster can be provided to describe the location of the results of the voxelwise lesion-behaviour mapping analysis or subtraction analysis. It is, however, important to realise that all voxels identified as statistically significant in a voxelwise statistical lesion-behaviour mapping analysis, or meaningful in a subtraction analysis, have the same importance, and thus should be given equal weights when interpreting the results.

Nowadays, there are many different cortical atlases to choose from. The first division that can be made is between atlases derived from single-subject data and atlases derived from multi-subject data. Whereas atlases derived from single-subject data remain popular (i.e. the

Brodmann atlas, or the AAL atlas of Tzourio-Mazoyer et al., 2002), probabilistic atlases derived from multi-subject data should be preferred, as these are able to quantify the intersubject variability in location and extent of each anatomical area. Within these multi-subject atlases, a second division can be made based on the brain characteristics used to parcellate distinct areas in different atlases. Whereas some probabilistic multi-subject atlases are based on macroscopical landmarks such as gyri and sulci (e.g. Hammers et al., 2003; Shattuck et al., 2008), others are based on histology (Zilles et al., 1997), or on functional connectivity patterns (e.g. Joliot et al., 2015). Finally, in addition to these cortical atlases, multi-subject atlases exist for white matter fiber tracts, based on either DTI fiber tracking (e.g. Zhang et al., 2010; Thiebaut de Schotten et al., 2011), or on histology (Bürgel et al., 2006). Which atlas to choose for the anatomical interpretation of the results of a lesion-behaviour mapping study is not a trivial issue. It is important to realise that different atlases might result in different anatomical interpretations of the same lesion-behaviour mapping results (de Haan and Karnath, 2017).

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (HA 5839/4-1 to BdH; KA 1258/20-1, KA 1258/23-1 to HOK). We would like to thank Christoph Sperber for helpful and inspiring discussions.

References

- Abela, E., Missimer, J., Wiest, R., Federspiel, A., Hess, C., Sturzenegger, M., Weder, B., 2012. Lesions to primary sensory and posterior parietal cortices impair recovery from hand paresis after stroke. *PloS One* 7, e31275. doi:10.1371/journal.pone.0031275
- Adams, J.H., Graham, D.I., Gennarelli, T.A., Maxwell, W.L., 1991. Diffuse axonal injury in non-missile head injury. *J. Neurol. Neurosurg. Psychiatry* 54, 481–483.
- Andersen, S.M., Rapcsak, S.Z., Beeson, P.M., 2010. Cost function masking during normalization of brains with focal lesions: still a necessity? *NeuroImage* 53, 78–84. doi:10.1016/j.neuroimage.2010.06.003
- Ashburner, J., Friston, K.J., 2005. Unified segmentation. *NeuroImage* 26, 839–851.
- Ashburner, J., Friston, K.J., 2003. Spatial normalization using basis functions, in: Frackowiak, R.S.J., Friston, K.J., Frith, C., Dolan, R., Price, C.J., Zeki, S., Ashburner, J., Penny, W.D. (Eds.), *Human Brain Function*. Academic Press, San Diego.
- Ashton, E.A., Takahashi, C., Berg, M.J., Goodman, A., Totterman, S., Ekholm, S., 2003. Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI. *J. Magn. Reson. Imaging* 17, 300–308. doi:10.1002/jmri.10258
- Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* 54, 2033–2044. doi:10.1016/j.neuroimage.2010.09.025
- Bates, E., Wilson, S.M., Saygin, A.P., Dick, F., Sereno, M.I., Knight, R.T., Dronkers, N.F., 2003. Voxel-based lesion-symptom mapping. *Nat. Neurosci.* 6, 448–450. doi:10.1038/nn1050
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300.
- Bergström, M., Ericson, K., Levander, B., Svendsen, P., Larsson, S., 1977. Variation with time of the attenuation values of intracranial hematomas. *J. Comput. Assist. Tomogr.* 1, 57–63.
- Bonato, M., Sella, F., Berteletti, I., Umiltà, C., 2012. Neuropsychology is nothing without control: A potential fallacy hidden in clinical studies. *Cortex* 48, 353–355. doi:10.1016/j.cortex.2011.06.017
- Brett, M., Leff, A.P., Rorden, C., Ashburner, J., 2001. Spatial normalization of brain images with focal lesions using cost function masking. *NeuroImage* 14, 486–500.

- Brunner, E., Munzel, U., 2000. The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biom. J.* 42, 17–25. doi:10.1002/(SICI)1521-4036(200001)42:1<17::AID-BIMJ17>3.0.CO;2-U
- Büki, A., Povlishock, J.T., 2006. All roads lead to disconnection? -Traumatic axonal injury revisited. *Acta Neurochir. (Wien)* 148, 181-193; discussion 193-194. doi:10.1007/s00701-005-0674-4
- Bürgel, U., Amunts, K., Battelli, L., Mohlberg, H., Gilsbach, J.M., Zilles, K., 2006. White matter fiber tracts of the human brain: three-dimensional mapping at microscopic resolution, topography and intersubject variability. *NeuroImage* 29, 1092–1105. doi:10.1016/j.neuroimage.2005.08.040
- Burger, P.C., Heinz, E.R., Shibata, T., Kleihues, P., 1988. Topographic anatomy and CT correlations in the untreated glioblastoma multiforme. *J. Neurosurg.* 68, 698–704. doi:10.3171/jns.1988.68.5.0698
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi:10.1038/nrn3475
- Claes, A., Idema, A.J., Wesseling, P., 2007. Diffuse glioma growth: a guerilla war. *Acta Neuropathol. (Berl.)* 114, 443–458. doi:10.1007/s00401-007-0293-7
- Clas, P., Groeschel, S., Wilke, M., 2012. A semi-automatic algorithm for determining the demyelination load in metachromatic leukodystrophy. *Acad. Radiol.* 19, 26–34. doi:10.1016/j.acra.2011.09.008
- Cohen, J., 1983. The cost of dichotomization. *Appl. Psychol. Meas.* 7, 249–253.
- Cox, R.W., 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162-173.
- Cox, R.W., 2012. AFNI: What a long strange trip it's been. *NeuroImage* 62, 743-747. doi:10.1016/j.neuroimage.2011.08.056
- Crawford, J.R., Garthwaite, P.H., 2005. Testing for suspected impairments and dissociations in single-case studies in neuropsychology: evaluation of alternatives using monte carlo simulations and revised tests for dissociations. *Neuropsychology* 19, 318–331. doi:10.1037/0894-4105.19.3.318
- Crawford, J.R., Howell, D.C., 1998. Comparing an individual's test score against norms derived from small samples. *Clin. Neuropsychol. Neuropsychol. Dev. Cogn. Sect. D* 12, 482–486. doi:10.1076/clin.12.4.482.7241

- Crinion, J., Ashburner, J., Leff, A., Brett, M., Price, C., Friston, K., 2007. Spatial normalization of lesioned brains: performance evaluation and impact on fMRI analyses. *NeuroImage* 37, 866–875.
- Damasio, H., Damasio, A.R., 1989. *Lesion analysis in neuropsychology*. Oxford University Press, New York.
- de Haan, B., Clas, P., Juenger, H., Wilke, M., Karnath, H.-O., 2015. Fast semi-automated lesion demarcation in stroke. *NeuroImage Clin.* 9, 69–74. doi:10.1016/j.nicl.2015.06.013
- de Haan, B., Karnath, H.-O., 2017. “Whose atlas I use, his song I sing?” - The impact of anatomical atlases on fiber tract contributions to cognitive deficits after stroke. *NeuroImage* 163, 301-309. doi:10.1016/j.neuroimage.2017.09.051
- Fandino, J., Kollias, S.S., Wieser, H.G., Valavanis, A., Yonekawa, Y., 1999. Intraoperative validation of functional magnetic resonance imaging and cortical reorganization patterns in patients with brain tumors involving the primary motor cortex. *J. Neurosurg.* 91, 238–250. doi:10.3171/jns.1999.91.2.0238
- Gennarelli, T.A., Thibault, L.E., Adams, J.H., Graham, D.I., Thompson, C.J., Marcincin, R.P., 1982. Diffuse axonal injury and traumatic coma in the primate. *Ann. Neurol.* 12, 564–574. doi:10.1002/ana.410120611
- Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the False Discovery Rate. *NeuroImage* 15, 870–878.
- Gillebert, C.R., Humphreys, G.W., Mantini, D., 2014. Automated delineation of stroke lesions using brain CT images. *NeuroImage Clin.* 4, 540–548. doi:10.1016/j.nicl.2014.03.009
- Goebel, R., 2012. BrainVoyager – past, present, future. *NeuroImage* 62, 748-756. doi:10.1016/j.neuroimage.2012.01.083
- González, R.G., Schaefer, P.W., Buonanno, F.S., Schwamm, L.H., Budzik, R.F., Rordorf, G., Wang, B., Sorensen, A.G., Koroshetz, W.J., 1999. Diffusion-weighted MR imaging: diagnostic accuracy in patients imaged within 6 hours of stroke symptom onset. *Radiology* 210, 155–162. doi:10.1148/radiology.210.1.r99ja02155
- Griffis, J.C., Allendorfer, J.B., Szaflarski, J.P., 2016. Voxel-based gaussian naïve Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans. *J. Neurosci. Methods* 257, 97-108. doi:10.1016/j.jneumeth.2015.09.019
- Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of

- the human brain, with particular reference to the temporal lobe. *Hum. Brain Mapp.* 19, 224–247.
- Hillis, A.E., Newhart, M., Heidler, J., Barker, P.B., Herskovits, E.H., Degaonkar, M., 2005. Anatomy of spatial attention: Insights from perfusion imaging and hemispatial neglect in acute stroke. *J. Neurosci.* 25, 3161–3167.
- Hillis, A.E., Wityk, R.J., Barker, P.B., Beauchamp, N.J., Gailloud, P., Murphy, K., Cooper, O., Metter, E.J., 2002. Subcortical aphasia and neglect in acute stroke: the role of cortical hypoperfusion. *Brain* 125, 1094–1104.
- Hillis, A.E., Wityk, R.J., Tuffiash, E., Beauchamp, N.J., Jacobs, M.A., Barker, P.B., Selnes, O.A., 2001. Hypoperfusion of Wernicke’s area predicts severity of semantic deficit in acute stroke. *Ann. Neurol.* 50, 561–566.
- Holodny, A.I., Schulder, M., Ybasco, A., Liu, W.-C., 2002. Translocation of Broca’s area to the contralateral hemisphere as the result of the growth of a left inferior frontal glioma. *J. Comput. Assist. Tomogr.* 26, 941–943.
- Ingre, M., 2013. Why small low-powered studies are worse than large high-powered studies and how to protect against “trivial” findings in research: comment on Friston (2012). *NeuroImage* 81, 496–498. doi:10.1016/j.neuroimage.2013.03.030
- Inoue, K., Madhyastha, T., Rudrauf, D., Mehta, S., Grabowski, T., 2014. What affects detectability of lesion-deficit relationships in lesion studies? *NeuroImage Clin.* 6, 388–397. doi:10.1016/j.nicl.2014.10.002
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. FSL. *NeuroImage* 62, 782–790. doi:10.1016/j.neuroimage.2011.09.015
- Johnson, V.E., Stewart, W., Smith, D.H., 2013. Axonal pathology in traumatic brain injury. *Exp. Neurol.* 246, 35–43. doi:10.1016/j.expneurol.2012.01.013
- Joliot, M., Jobard, G., Naveau, M., Delcroix, N., Petit, L., Zago, L., Crivello, F., Mellet, E., Mazoyer, B., Tzourio-Mazoyer, N., 2015. AICHA: An atlas of intrinsic connectivity of homotopic areas. *J. Neurosci. Methods* 254, 46–59. doi:10.1016/j.jneumeth.2015.07.013
- Karnath, H.-O., Rennig, J., 2016. Investigating structure and function in the healthy human brain: validity of acute versus chronic lesion-symptom mapping. *Brain Struct. Funct.* doi:10.1007/s00429-016-1325-7
- Karnath, H.-O., Rennig, J., Johannsen, L., Rorden, C., 2011. The anatomy underlying acute versus chronic spatial neglect: A longitudinal study. *Brain* 134, 903–912. doi:10.1093/brain/awq355

- Karnath, H.-O., Rorden, C., 2012. The anatomy of spatial neglect. *Neuropsychologia* 50, 1010–1017.
- Karnath, H.-O., Sperber, C., Rorden, C., in press. Mapping human brain lesions and their functional consequences. *NeuroImage*.
- Karnath, H.-O., Steinbach, J.P., 2011. Do brain tumours allow valid conclusions on the localisation of human brain functions?--Objections. *Cortex* 47, 1004–1006. doi:10.1016/j.cortex.2010.08.006
- Karnath, H.-O., Zopf, R., Johannsen, L., Berger, M.F., Nagele, T., Klose, U., 2005. Normalized perfusion MRI to identify common areas of dysfunction: patients with basal ganglia neglect. *Brain* 128, 2462–2469. doi:10.1093/brain/awh629
- Kimberg, D.Y., Coslett, H.B., Schwartz, M.F., 2007. Power in voxel-based lesion-symptom mapping. *J. Cogn. Neurosci.* 19, 1067–1080.
- Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage* 46, 786–802.
- Koenig, M., Kraus, M., Theek, C., Klotz, E., Gehlen, W., Heuser, L., 2001. Quantitative assessment of the ischemic brain by means of perfusion-related parameters derived from perfusion CT. *Stroke* 32, 431–437.
- Lansberg, M.G., Thijs, V.N., O'Brien, M.W., Ali, J.O., de Crespigny, A.J., Tong, D.C., Moseley, M.E., Albers, G.W., 2001. Evolution of apparent diffusion coefficient, diffusion-weighted, and T2-weighted signal intensity of acute stroke. *AJNR Am. J. Neuroradiol.* 22, 637–644.
- Levine, B., Kovacevic, N., Nica, E.I., Schwartz, M.L., Gao, F., Black, S.E., 2013. Quantified MRI and cognition in TBI with diffuse and focal damage. *NeuroImage Clin.* 2, 534–541. doi:10.1016/j.nicl.2013.03.015
- Liebermeister, C., 1877. Über Wahrscheinlichkeitsrechnung in Anwendung auf therapeutische Statistik. *Samml. Klin. Vorträge Inn. Med.* No 31-64 110, 935–962.
- Mah, Y.-H., Husain, M., Rees, G., Nachev, P., 2014a. Human brain lesion-deficit inference remapped. *Brain* 137, 2522–2531. doi:10.1093/brain/awu164
- Mah, Y.-H., Jager, R., Kennard, C., Husain, M., Nachev, P., 2014b. A new method for automated high-dimensional lesion segmentation evaluated in vascular injury and

- applied to the human occipital lobe. *Cortex* 56, 51–63.
doi:10.1016/j.cortex.2012.12.008
- Mayer, T.E., Hamann, G.F., Baranczyk, J., Rosengarten, B., Klotz, E., Wiesmann, M., Missler, U., Schulte-Altedorneburg, G., Brueckmann, H.J., 2000. Dynamic CT perfusion imaging of acute stroke. *AJNR Am. J. Neuroradiol.* 21, 1441–1449.
- Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the human brain: Theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). *NeuroImage* 2, 89-101.
- Mazziotta, J.C., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., Mazoyer, B., 2001a. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1293-1322.
- Mazziotta, J.C., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Feidler, J., Smith, K., Boomsma, D., Hulshoff Pol, H., Cannon, T., Kawashima, R., Mazoyer, B., 2001b. A four-dimensional probabilistic atlas of the human brain. *J. Am. Med. Inform. Assoc.* 8, 401-430.
- McKnight, T.R., von dem Bussche, M.H., Vigneron, D.B., Lu, Y., Berger, M.S., McDermott, M.W., Dillon, W.P., Graves, E.E., Pirzkall, A., Nelson, S.J., 2002. Histopathological validation of a three-dimensional magnetic resonance spectroscopy index as a predictor of tumor presence. *J. Neurosurg.* 97, 794–802.
doi:10.3171/jns.2002.97.4.0794
- Meyer, P.T., Sturz, L., Sabri, O., Schreckenberger, M., Spetzger, U., Setani, K.S., Kaiser, H.-J., Buell, U., 2003. Preoperative motor system brain mapping using positron emission tomography and statistical parametric mapping: hints on cortical reorganisation. *J. Neurol. Neurosurg. Psychiatry* 74, 471–478.
- Mirman, D., Landrigan, J.-F., Kokolis, S., Verillo, S., Ferrara, C., Pustina, D., in press, this issue. Corrections for multiple comparisons in voxel-based lesion-symptom mapping. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2017.08.025

- Mohr, J.P., Biller, J., Hilal, S.K., Yuh, W.T., Tatemichi, T.K., Hedges, S., Tali, E., Nguyen, H., Mun, I., Adams, H.P., 1995. Magnetic resonance versus computed tomographic imaging in acute stroke. *Stroke* 26, 807–812.
- Motta, M., Ramadan, A., Hillis, A.E., Gottesman, R.F., Leigh, R., 2014. Diffusion-perfusion mismatch: an opportunity for improvement in cortical function. *Front. Neurol.* 5, 280. doi:10.3389/fneur.2014.00280
- Nachev, P., Coulthard, E., Jäger, H.R., Kennard, C., Husain, M., 2008. Enantiomorphic normalization of focally lesioned brains. *NeuroImage* 39, 1215–1226. doi:10.1016/j.neuroimage.2007.10.002
- Neumann-Haefelin, T., Wittsack, H.J., Wenserski, F., Siebler, M., Seitz, R.J., Mödder, U., Freund, H.J., 1999. Diffusion- and perfusion-weighted MRI. The DWI/PWI mismatch region in acute stroke. *Stroke* 30, 1591–1597.
- Ojemann, J.G., Miller, J.W., Silbergeld, D.L., 1996. Preserved function in brain invaded by tumor. *Neurosurgery* 39, 253-258; discussion 258-259.
- Ostrom, Q.T., Gittleman, H., Xu, J., Kromer, C., Wolinsky, Y., Kruchko, C., Barnholtz-Sloan, J.S., 2016. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2009–2013. *Neuro-Oncol.* 18, v1–v75. doi:10.1093/neuonc/nov207
- Pallud, J., Capelle, L., Taillandier, L., Badoual, M., Duffau, H., Mandonnet, E., 2013. The silent phase of diffuse low-grade gliomas. Is it when we missed the action? *Acta Neurochir. (Wien)* 155, 2237–2242. doi:10.1007/s00701-013-1886-7
- Povlishock, J.T., 1993. Pathobiology of traumatically induced axonal injury in animals and man. *Ann. Emerg. Med.* 22, 980–986.
- Povlishock, J.T., Katz, D.I., 2005. Update of neuropathology and neurological recovery after traumatic brain injury. *J. Head Trauma Rehabil.* 20, 76–94.
- Pustina, D., Avants, B., Faseyitan, O.K., Medaglia, J.D., Coslett, H.B., in press, this issue. Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2017.08.027
- Pustina, D., Coslett, H. B., Turkeltaub, P. E., Tustison, N., Schwartz, M. F., Avants, B., 2016. Automated segmentation of chronic stroke lesions using LINDA: Lesion identification with neighborhood data analysis. *Hum. Brain Mapp.* 37, 1405-1421. doi:10.1002/hbm.23110

- Ricci, P.E., Burdette, J.H., Elster, A.D., Reboussin, D.M., 1999. A comparison of fast spin-echo, fluid-attenuated inversion-recovery, and diffusion-weighted MR imaging in the first 10 days after cerebral infarction. *AJNR Am. J. Neuroradiol.* 20, 1535–1542.
- Rondina, J.M., Filippone, M., Girolami, M., Ward, N.S., 2016. Decoding post-stroke motor function from structural brain imaging. *NeuroImage Clin.* 12, 372–380. doi:10.1016/j.nicl.2016.07.014
- Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., Karnath, H.-O., 2012. Age-specific CT and MRI templates for spatial normalization. *NeuroImage* 61, 957–965. doi:10.1016/j.neuroimage.2012.03.020
- Rorden, C., Brett, M., 2000. Stereotaxic display of brain lesions. *Behav. Neurol.* 12, 191–200.
- Rorden, C., Karnath, H.-O., 2004. Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nat. Rev. Neurosci.* 5, 813–819. doi:10.1038/nrn1521
- Rorden, C., Karnath, H.-O., Bonilha, L., 2007. Improving lesion-symptom mapping. *J. Cogn. Neurosci.* 19, 1081–1088. doi:10.1162/jocn.2007.19.7.1081
- Rudrauf, D., Mehta, S., Grabowski, T. J., 2008. Disconnection's renaissance takes shape: Formal incorporation in group-level lesion studies. *Cortex* 44, 1084–1096. doi:10.1016/j.cortex.2008.05.005
- Schaefer, P.W., Hunter, G.J., He, J., Hamberg, L.M., Sorensen, A.G., Schwamm, L.H., Koroshetz, W.J., Gonzalez, R.G., 2002. Predicting cerebral ischemic infarct volume with diffusion and perfusion MR imaging. *AJNR Am. J. Neuroradiol.* 23, 1785–1794.
- Scherer, H.J., 1940. The forms of growth in gliomas and their practical significance. *Brain* 63, 1–35. doi:10.1093/brain/63.1.1
- Schiffbauer, H., Ferrari, P., Rowley, H.A., Berger, M.S., Roberts, T.P., 2001. Functional activity within brain tumors: a magnetic source imaging study. *Neurosurgery* 49, 1313–1320; discussion 1320–1321.
- Schlaug, G., Benfield, A., Baird, A.E., Siewert, B., Lövblad, K.O., Parker, R.A., Edelman, R.R., Warach, S., 1999. The ischemic penumbra: operationally defined by diffusion and perfusion MRI. *Neurology* 53, 1528–1537.
- Shahid, H., Sebastian, R., Schnur, T.T., Hanayik, T., Wright, A., Tippett, D.C., Fridriksson, J., Rorden, C., Hillis, A.E., 2017. Important considerations in lesion-symptom mapping: Illustrations from studies of word comprehension. *Hum. Brain Mapp.* 38, 2990–3000. doi:10.1002/hbm.23567

- Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W., 2008. Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage* 39, 1064–1080.
- Shaw, K., Brennan, N., Woo, K., Zhang, Z., Young, R., Peck, K.K., Holodny, A., 2016. Infiltration of the basal ganglia by brain tumors is associated with the development of co-dominant language function on fMRI. *Brain Lang.* 155–156, 44–48. doi:10.1016/j.bandl.2016.04.002
- Skirboll, S.S., Ojemann, G.A., Berger, M.S., Lettich, E., Winn, H.R., 1996. Functional cortex and subcortical white matter located within gliomas. *Neurosurgery* 38, 678-684; discussion 684-685.
- Smith, D.V., Clithero, J.A., Rorden, C., Karnath, H.-O., 2013. Decoding the anatomical network of spatial attention. *Proc. Natl. Acad. Sci. U. S. A.* 110, 1518–1523. doi:10.1073/pnas.1210126110
- Soares, J.M., Magalhães, R., Moreira, P.S., Sousa, A., Ganz, E., Sampaio, A., Alves, V., Marques, P., Sousa, N., 2016. A hitchhiker's guide to functional magnetic resonance imaging. *Front. Neurosci.* 10, 515. doi:10.3389/fnins.2016.00515
- Soares, J.M., Marques, P., Alves, V., Sousa, N., 2013. A hitchhiker's guide to diffusion tensor imaging. *Front. Neurosci.* 7, 31. doi:10.3389/fnins.2013.00031
- Sperber, C., Karnath, H.-O., 2016. Topography of acute stroke in a sample of 439 right brain damaged patients. *NeuroImage Clin.* 10, 124–128. doi:10.1016/j.nicl.2015.11.012
- Sperber, C., Karnath, H.-O., 2017. Impact of correction factors in human brain lesion-behavior inference. *Hum. Brain Mapp.* 38, 1692–1701. doi:10.1002/hbm.23490
- Sperber, C., Karnath, H.-O., in press, this issue. On the validity of lesion-behaviour mapping methods. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2017.07.035
- Su, E., Bell, M., 2016. Diffuse Axonal Injury, in: Laskowitz, D., Grant, G. (Eds.), *Translational Research in Traumatic Brain Injury*, *Frontiers in Neuroscience*. CRC Press/Taylor and Francis Group, Boca Raton (FL).
- Swanson, K., Alvord, E.C., Murray, J.D., 2004. Dynamics of a model for brain tumors reveals a small window for therapeutic intervention. *Discrete Contin. Dyn. Syst. - Ser. B* 4, 289–295. doi:10.3934/dcdsb.2004.4.289
- Swanson, K.R., Bridge, C., Murray, J.D., Alvord, E.C., 2003. Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion. *J. Neurol. Sci.* 216, 1–10.

- Taniguchi, M., Kato, A., Ninomiya, H., Hirata, M., Cheyne, D., Robinson, S.E., Maruno, M., Saitoh, Y., Kishima, H., Yoshimine, T., 2004. Cerebral motor control in patients with gliomas around the central sulcus studied with spatially filtered magnetoencephalography. *J. Neurol. Neurosurg. Psychiatry* 75, 466–471.
- Thiebaut de Schotten, M., Ffytche, D.H., Bizzi, A., Dell'Acqua, F., Allin, M., Walshe, M., Murray, R., Williams, S.C., Murphy, D.G.M., Catani, M., 2011. Atlasing location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. *NeuroImage* 54, 49–59. doi:10.1016/j.neuroimage.2010.07.055
- Thiel, A., Herholz, K., Koyuncu, A., Ghaemi, M., Kracht, L.W., Habedank, B., Heiss, W.D., 2001. Plasticity of language networks in patients with brain tumors: a positron emission tomography activation study. *Ann. Neurol.* 50, 620–629.
- Ticini, L.F., de Haan, B., Klose, U., Nagele, T., Karnath, H.-O., 2010. The role of temporoparietal cortex in subcortical visual extinction. *J. Cogn. Neurosci.* 22, 2141–2150.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15, 273–289.
- von Kummer, R., Bourquain, H., Bastianello, S., Bozzao, L., Manelfe, C., Meier, D., Hacke, W., 2001. Early prediction of irreversible brain damage after ischemic stroke at CT. *Radiology* 219, 95–100. doi:10.1148/radiology.219.1.r01ap0695
- Warach, S., Chien, D., Li, W., Ronthal, M., Edelman, R.R., 1992. Fast magnetic resonance diffusion-weighted imaging of acute human stroke. *Neurology* 42, 1717–1723.
- Wilke, M., de Haan, B., Juenger, H., Karnath, H.O., 2011. Manual, semi-automated, and automated delineation of chronic brain lesions: A comparison of methods. *NeuroImage* 56, 2038–2046. doi:10.1016/j.neuroimage.2011.04.014
- Wilke, M., Holland, S. K., Altaye, M., Gaser, C., 2008. Template-O-Matic: A toolbox for creating customized pediatric templates. *NeuroImage* 41, 903–913. doi:10.1016/j.neuroimage.2008.02.056
- Wilson, S. M., 2016. Lesion-symptom mapping in the study of spoken language understanding. *Lang. Cogn. Neurosci.* 32, 891–899. doi:10.1080/23273798.2016.1248984
- Winkler, A.M., Kochunov, P., Glahn, D.C. FLAIR Templates. Available at <http://glahngroup.org> or <http://brainder.org>.

- Wu, O., Cloonan, L., Mocking, S.J.T., Bouts, M.J.R.J., Copen, W.A., Cougo-Pinto, P.T., Fitzpatrick, K., Kanakis, A., Schaefer, P.W., Rosand, J., Furie, K.L., Rost, N.S., 2015. Role of Acute Lesion Topography in Initial Ischemic Stroke Severity and Long-Term Functional Outcomes. *Stroke* 46, 2438–2444. doi:10.1161/STROKEAHA.115.009643
- Wunderlich, G., Knorr, U., Herzog, H., Kiwit, J.C., Freund, H.J., Seitz, R.J., 1998. Precentral glioma location determines the displacement of cortical hand representation. *Neurosurgery* 42, 18-26; discussion 26-27.
- Zhang, Y., Kimberg, D.Y., Coslett, H.B., Schwartz, M.F., Wang, Z., 2014. Multivariate lesion-symptom mapping using support vector regression. *Hum. Brain Mapp.* 35, 5861–5876. doi:10.1002/hbm.22590
- Yourganov, G., Smith, K.G., Fridriksson, J., Rorden, C., 2015. Predicting aphasia type from brain damage measured with structural MRI. *Cortex* 73, 203-215. doi:10.1016/j.cortex.2015.09.005
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G, Ho, S., Gee, J. C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage* 31, 1116-1128. doi:10.1016/j.neuroimage.2006.01.015
- Zhang, Y., Zhang, J., Oishi, K., Faria, A.V., Jiang, H., Li, X., Akhter, K., Rosa-Neto, P., Pike, G.B., Evans, A., Toga, A.W., Woods, R., Mazziotta, J.C., Miller, M.I., van Zijl, P.C.M., Mori, S., 2010. Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *NeuroImage* 52, 1289–1301. doi:10.1016/j.neuroimage.2010.05.049
- Zilles, K., Schleicher, A., Langemann, C., Amunts, K., Morosan, P., Palomero-Gallagher, N., Schormann, T., Mohlberg, H., Bürgel, U., Steinmetz, H., Schlaug, G., Roland, P.E., 1997. Quantitative analysis of sulci in the human cerebral cortex: Development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture. *Hum. Brain Mapp.* 5, 218–221.
- Zopf, R., Fruhmann Berger, M., Klose, U., Karnath, H.-O., 2009. Perfusion imaging of the right perisylvian neural network in acute spatial neglect. *Front. Hum. Neurosci.* 3, 15. doi:10.3389/neuro.09.015.2009
- Zopf, R., Klose, U., Karnath, H.-O., 2012. Evaluation of methods for detecting perfusion abnormalities after stroke in dysfunctional brain regions. *Brain Struct. Funct.* 217, 667–675. doi:10.1007/s00429-011-0363-4

Figure legends

Figure 1: Number of publications per year that used the lesion method between the years 1995 and 2015. This result was obtained by running a Pubmed (<https://www.ncbi.nlm.nih.gov/pubmed>) literature search with the search term „(‘lesion analysis’ OR ‘lesion mapping’ OR ‘VLSM’ OR ‘VLBM’) AND brain“, followed by a manual exclusion of non-empirical articles (i.e. methodological articles, review articles, etc). Note the clear and steady increase in number of publications that used the lesion method per year over the last 10 years.

Figure 2: Illustration of the typical lesion study pipeline. When performing a lesion study, the researcher has to decide which patients to select, how to assess the patient’s lesion location and behavioural status, how to spatially normalise the patient’s brain image and lesion map, how to perform the voxelwise (statistical) comparisons over all patients, and finally, how to anatomically interpret the obtained results.

Figure 3: Illustration of the thought experiment described in the manuscript text. Both lesions depicted *on the top* equally affect the brain area crucially related to a certain cognitive function of interest. The diagram *on the bottom*, reflects the rate of spontaneous recovery over time of this function. The difference in measured behavioural deficit scores of these patients (‘54’ vs. ‘27’) does not result from differences in the relevance of the individual lesions for the cognitive function of interest. Instead, this difference is the consequence of including the two patients at different time points following stroke-onset: one patient is recruited and behaviourally tested in the acute phase after stroke (and thus shows a max. behavioural deficit score of ‘54’) while the other patient is first seen and tested in the intermediate/chronic stroke phase, following partial recovery of the behavioural deficit (from an initial behavioural deficit score of ‘54’ down to a behavioural deficit score of ‘27’).

Figure 4: The consequences of dividing an unselectively recruited patient sample into subsamples on the basis of an a priori anatomical hypothesis. A: Sketch where one and the same stroke patient sample (middle image, simple lesion overlap in yellow) is divided into either an anterior and posterior subsample (left image), or a ventral and dorsal subsample (right image), in order to perform separate lesion-behaviour mapping analyses in each of these subsamples. This division of a single patient sample into subsamples will generate

results that match the a priori hypothesis. B: Illustration of this effect in an unselectively recruited sample of 20 stroke patients (patients taken from Sperber and Karnath [2016]). One and the same patient sample (top middle image, simple lesion overlap) is divided into either an anterior and posterior subsample (left images), or a ventral and dorsal subsample (right images). Whereas dividing the entire patient sample into an anterior and posterior subsample results in an area of maximum lesion overlap at a z-coordinate of 6 and 24 respectively, dividing the same patient sample into a ventral and dorsal subsample results in an area of maximum lesion overlap at a z-coordinate of -12 and 33 respectively. Moreover, when dividing the entire patient sample into an anterior and posterior subsample, 7 patients (i.e. 35% of the sample) could not be categorised. When dividing the entire patient sample into a ventral and dorsal subsample, 11 patients (i.e. 55% of the sample) could not be categorised. The number of overlapping lesions is illustrated by colour, from violet (n=1) to red (n=maximum lesion overlap). The numbers at the bottom of the Figure indicate MNI z-coordinates. Images are in neurological orientation.

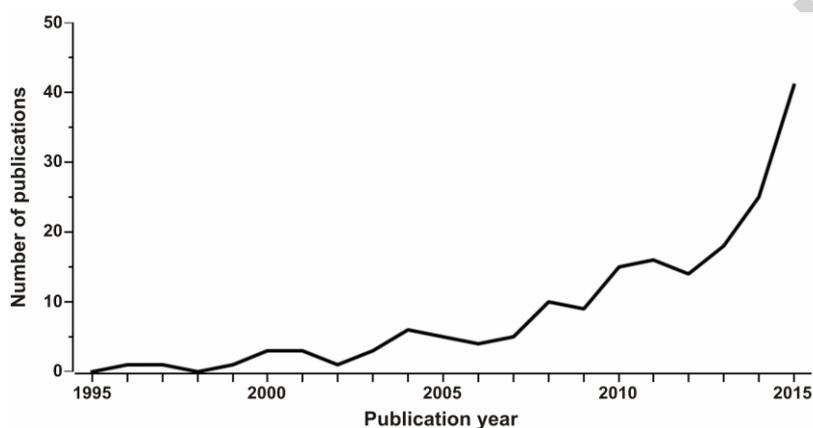


Fig. 1

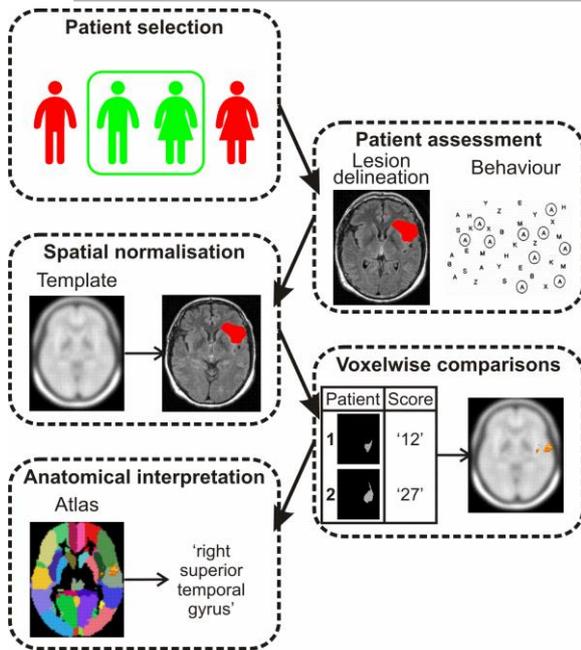


Fig. 2

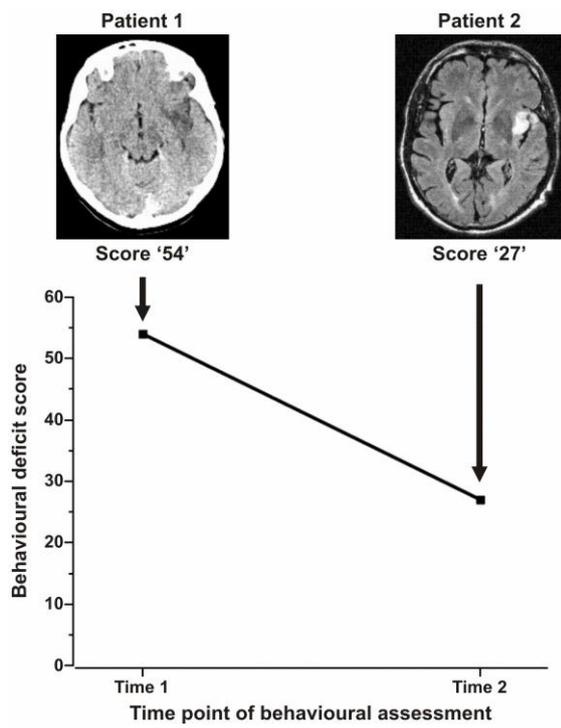


Fig. 3

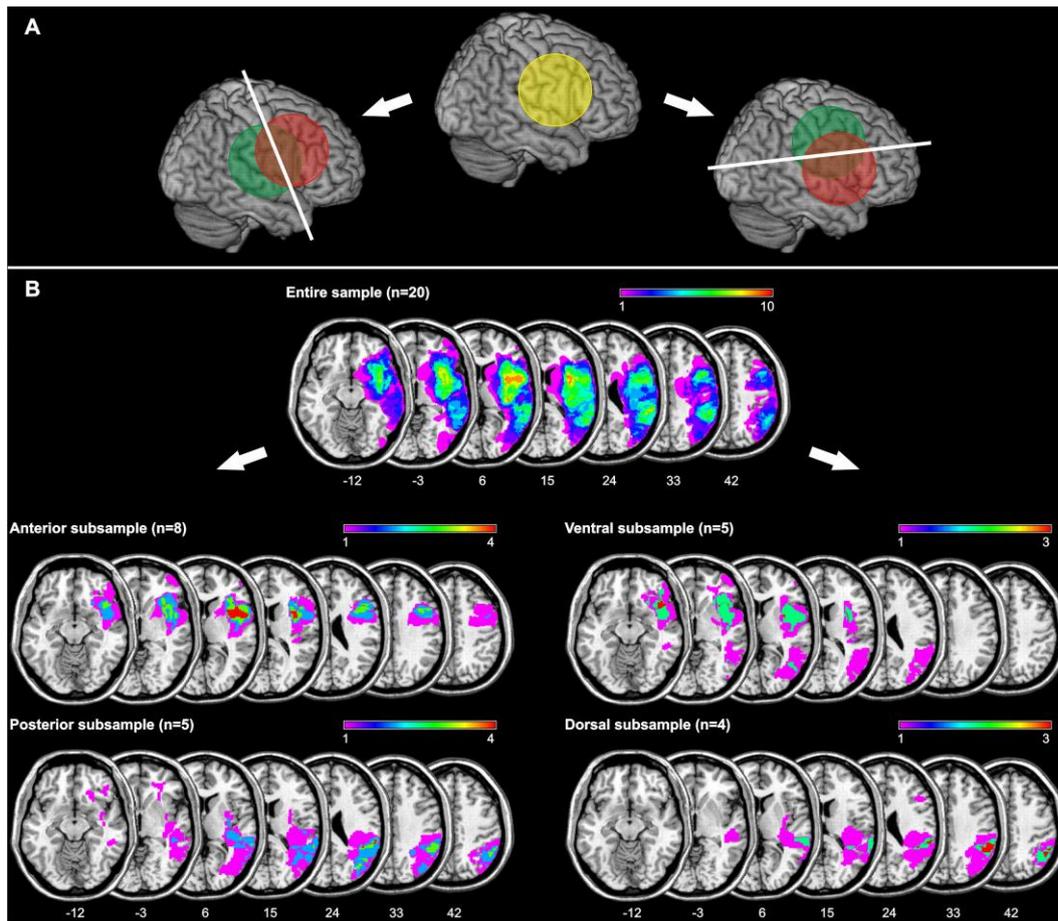


Fig. 4

- The lesion method is an influential and popular method to study human brain function
- But virtually no papers or books exist to assist scientists interested in this method
- We here provide a hitchhiker's guide with practical guidelines and references