Improving and validating methods in lesion behaviour mapping

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Christoph Sperber
Geboren in Kirchheim/Teck, Deutschland

07 - 2018

Tag der mündlichen Prüfung: 18.12.2018

Dekan der Math.-Nat. Fakultät: Prof. Dr. W. Rosenstiel
Dekan der Medizinischen Fakultät: Prof. Dr. I. B. Autenrieth

1. Berichterstatter: Prof. Dr. Dr. Hans-Otto Karnath

2. Berichterstatter: Prof. Dr. Martin Giese

Prüfungskommission: Prof. Dr. Dr. Hans-Otto Karnath

Prof. Dr. Martin Giese

Prof. Dr. Hanspeter Mallot

Dr. Surjo Soekadar

Declaration:

*I hereby declare that I have produced the work entitled “*Improving and validating methods in lesion behaviour mapping*”, submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.*

Tübingen, den ....

Datum / Date Unterschrift /Signature

Abstract

The investigation of diseased brain is one of the major methods in cognitive neuroscience. This approach allows numerous insights both into human cognition and brain architecture. Most prominent is the method of lesion behaviour mapping, where inferences about functional brain architecture are drawn from focally lesioned brains. In the last 15 years, the state-of-the-art implementation of lesion behaviour mapping has been voxel-based lesion behaviour mapping, which is based on the framework of statistical parametric mapping. Recently, the validity of this method has been criticised and multivariate methods have been proposed to complement or even replace it.

In my thesis, I aim to evaluate these different methodological approaches to lesion behaviour mapping and to provide guidelines on how lesion-brain inference should be drawn. In my first empirical work, I investigate the validity of voxel-based lesion behaviour mapping. It shows that previous studies overestimated biases inherent to the method, and that validity can be improved by the use of correction factors. The second empirical work deals with a recently developed method of multivariate lesion behaviour mapping. On the one hand, I clarify how this method can be used to obtain valid lesion-brain inference. On the other hand, I show that the method is not able to overcome all limitations of voxel-based lesion behaviour mapping. In my last work, I apply multivariate lesion behaviour mapping to investigate the neural correlates of higher motor cognition. This analysis is the first to identify a brain network to underlie apraxia, a disorder of higher motor cognition, which underlines the benefits of the new multivariate approach in brain networks.

Table of contents

ABSTRACT 1

ACKNOWLEDGEMENTS 4

1 COGNITIVE NEUROSCIENCE AND DISEASED BRAINS 5

2 LESION-DEFICIT INFERENCE – FROM PAUL BROCA TO STATISTICAL PARAMETRIC MAPPING 6

2.1 Early history 6

2.2 The advent of brain imaging and first lesion-behaviour mapping studies 7

2.3 Voxel-wise Statistical Mapping 8

- 2.3.1 Lesion visualisation and delineation 9
- 2.3.2 Spatial normalisation 10
- 2.3.3 Voxel-based lesion symptom mapping 11
- 2.3.4 Voxel-based lesion behaviour mapping – a mass-univariate method 12

3 CHALLENGES AND LIMITATIONS OF THE MASS-UNIVARIATE APPROACH 13

3.1 The multiple comparison problem 13

3.2 Voxel-wise power and rarely lesioned brain voxels 16

3.3 The effect of lesion size 17

3.4 The complexity of functional brain architecture and the partial injury problem 18

3.5 The complexity of lesion anatomy 19

4 INVESTIGATING THE VALIDITY OF LESION BEHAVIOUR MAPPING METHODS 20

4.1 A simulation approach to test the validity of lesion behaviour mapping 21

4.2 The independence of statistical tests in the mass univariate approach put on trial 22

5 MULTIVARIATE LESION BEHAVIOUR MAPPING 23

6 EMPIRICAL RESEARCH QUESTIONS IN MY THESIS 26

6.1 Impact of correction factors in human brain lesion-behavior inference 26

6.2 An empirical evaluation of multivariate lesion behaviour mapping 26

6.3 The network underlying human higher-order motor control: Insights from machine learning-based lesion-behaviour mapping 27

7 FUTURE CHALLENGES AND RESEARCH DIRECTIONS IN LESION BEHAVIOUR MAPPING 27

7.1 Optimisation of (multivariate) lesion behaviour mapping 27

7.2 Translational utilisation of lesion behaviour mapping 28

REFERENCES 30

LIST OF PAPERS/MANUSCRIPTS APPENDED 37

LIST OF ALL MY PAPERS/MANUSCRIPTS THAT UTILISED, INVESTIGATED, OR REVIEWED LESION BEHAVIOUR MAPPING 112

APPENDED PAPERS/MANUSCRIPTS 38

IMPACT OF CORRECTION FACTORS IN HUMAN BRAIN LESION-BEHAVIOR INFERENCE . 39

AN EMPIRICAL EVALUATION OF MULTIVARIATE LESION BEHAVIOUR MAPPING 62

THE NETWORK UNDERLYING HUMAN HIGHER-ORDER MOTOR CONTROL: INSIGHTS FROM MACHINE LEARNING-BASED LESION-BEHAVIOUR MAPPING 87

CURRICULUM VITAE .... ERROR! BOOKMARK NOT DEFINED.

Acknowledgements

First, I would like to express my great appreciation to my advisor Prof. Hans-Otto Karnath for his supervision and guidance. All the great opportunities that he offered me were challenging, but in the end it turned out very well and it got me where I am now.

Besides my advisor, I would also like to thank the rest of my thesis committee for their guidance: Prof. Martin Giese and Prof. Hanspeter Mallot. I also would like to thank Prof. Georg Goldenberg for the cooperation on the topic of higher order motor control.

I would like to offer special thanks to the Friedrich Naumann Foundation for financial support, and even more for their ideal promotion.

I thank the members of the Section of Neuropsychology for all the team work and the great time we have had together. Special thanks go to Johannes Rennig, Bianca de Haan, Daniel Wiesen, Marc Himmelbach, Ina Baumeister, Sonja Cornelsen, Maren Prass, Gabriela Zaiser, Sophia Nestmann, and Jasmin Klopfer.
I am particularly grateful to Johannes Rennig and Bianca de Haan. Without their initial guidance, my scientific projects never would have gotten that far.
I also want to acknowledge the awesome and efficient cooperation with Daniel Wiesen.

It was a great honour to be accompanied by Hannah Tomczyk, who really isn't that bad at all.

All moose were trained by Þōrvaldr Eirikssonr, and no moose were harmed or inappropriately nuzzled while performing my research.

1 Cognitive neuroscience and diseased brains

The scientific aim of cognitive neuroscience is to understand how the human brain works. A major part of this is to map the anatomo-behavioural architecture of the brain, that is to find out which brain regions are responsible for a certain cognitive behaviour. Until the middle of the 20th century, the mapping of deceased brain has been the only available method in this field. The neural correlates of a cognitive function were inferred from patients with focal neurologic damage, who suffered from deficits in the cognitive function.

In the last decades, new methods emerged. On the one hand, imaging techniques such as electro- or magnetoencephalography, positron emission tomography, and functional magnetic resonance imaging were established in the field. One important limitation of these imaging methods is that they might not only identify regions where activity is causal for a certain cognitive function, but also regions where activity is just correlative. On the other hand, non-invasive transcranial neurostimulation methods based on electric or magnetic stimulation of the brain came up. Latter methods allow drawing the conclusion if activity of certain brain regions is *necessary* for a cognitive function. Unfortunately, transcranial neurostimulation is limited to a few experimental protocols, to a few possible stimulation loci, and effects are often weak. Here, the mapping of diseased brains comes back into play. Brains with focal damage also offer the opportunity to study causality in brain-behaviour relations (Rorden & Karnath, 2004). This is the reason why the mapping of diseased brains is still an important method in cognitive neurosciences in the 21st century.

The most common approach in mapping diseased brains involves the study of stroke patients, which was termed *lesion behaviour mapping*. My thesis deals with this method, and the greater goal of my work is to find the best way to perform lesion behaviour mapping. I investigate the validity of both established and novel methods in the field, I empirically define guidelines for novel methods, and I apply these methods to identify the networks underlying higher motor cognition.

In the synopsis of my thesis, I first give an introduction into the methods of lesion behaviour mapping in chapter 2. I outline the method's historical evolution and depict procedures that are prerequisites for state of the art lesion behaviour mapping. Last, I provide an overview of the methodological paradigm that dominated lesion behaviour mapping in the last 15 years: voxel-based lesion behaviour mapping. In chapter 3, I characterise several challenges and limitations in the field of voxel-based lesion behaviour mapping. Chapter 4 outlines an approach that is suited to investigate the impact of these limitations on the validity of lesion behaviour mapping, and which I used to investigate the different methods. Chapter 5 overviews a new method in the field: multivariate lesion behaviour mapping. This method is thought to overcome several of the limitations mentioned in the chapter before. In chapter 6, I provide a short overview on the empirical works in my thesis. Finally, in chapter 7, I picture possible future research directions in lesion behaviour mapping.

2 Lesion-deficit inference – from Paul Broca to statistical parametric mapping

2.1 Early history

Historically, studies on patients with brain damage were the first studies ever to investigate the functional anatomy of the brain. One of the most famous milestones in the field dates back to the middle of the 19th century (Broca, 1861). In 1861, the French physician and anatomist Paul Broca heard about Louis Victor Leborgne, who suffered from a loss of speech production. This patient was unable to speak any words other than the syllable “tan”. Broca showed that both Leborgne’s cognitive capabilities and his ability to understand speech were largely preserved. When Leborgne died, Broca performed an autopsy and found the left inferior frontal cortex to be damaged. Broca replicated this finding in several other patients with deficient speech production, but intact comprehension. A general conclusion of these findings was relevant in providing a paradigm for future research: cognitive functions are anatomically localised in the brain.

For more than 100 years, the methodological approach used by Paul Broca - neurological single case studies with post mortem autopsy - was almost the only available method to investigate brain-function relationships. Some lesser known exceptions were the studies by Tatsuji Inouye (see Glickstein & Whitteridge, 1987) and by Gordon Holmes (Holmes & Lister, 1916). They studied patients with non-lethal gunshot wounds to the brain in the russo-japanese war and the First World War. By examining entry and exit wounds in the skull, they were able to map the primary visual cortex with high precision. The most innovative aspect about these studies was the examination of a group of patients. Still, single case studies were the standard in the field, because anatomical information was usually only obtainable post mortem.

2.2 The advent of brain imaging and first lesion-behaviour mapping studies

In 1971, the first in vivo X-ray computed tomography scan of a human brain was carried out. Scanning and data processing, however, still took several hours, and the final image consisted of only one single low-resolution slice. Two years later, application of nuclear magnetic resonance in image generation was first described (Lauterbur, 1973) and used to obtain in vivo images in living organisms (Lauterbur, 1974). Based on these foundations, both X-ray computed tomography (CT) and magnetic resonance imaging (MRI) evolved at a tremendous pace into essential methods in many clinical fields, culminating in the Noble Prize awards of 1979 for Allan Cormack and Godfrey Hounsfield, and 2003 for Paul Lauterbur and Peter Mansfield.

The development of these imaging methods was of outstanding relevance for the diagnosis and treatment of stroke. CT allows to differentiate between ischaemic and haemorrhagic stroke, which is of vital significance in thrombolysis therapy (Freeman & Aguilar, 2012). Moreover, a wide array of more specialised CT or MRI clinical imaging protocols emerged, which allow to visualise brain perfusion, vessels, and diffusion (Jäger, 2000). Most importantly for the field of lesion behaviour mapping, structural CT or MRI that visualises the extent of stroke became available. For the first time in history, researchers were able to localize structural brain damage after stroke in vivo. This allowed researchers to perform anatomo-behavioural studies more efficiently than ever before on groups rather than single patients.

First anatomo-behavioural studies using these imaging methods qualitatively assessed brain damage. To do so, neuroradiologists – or other scientists with comparable expertise in brain anatomy and stroke imaging – visually inspected brain scans and assessed if certain areas were damaged. Alternatively, for a topographical approach, scientists manually transferred the lesion onto a template. In more detail, the lesion borders were drawn by hand on a schematic diagram of the brain, which could either be an over-simplified line drawing without any or only a few anatomical landmarks or any kind of brain template. Further analyses of these topographical data were performed qualitatively. For example, individual lesions were overlapped to identify brain areas that are often affected when a symptom is present.

This approach based on simple overlap topographies is however severely limited: brain regions often affected in patients with a symptom are not necessarily the neural correlates of the symptom, but instead areas that are simply more often affected in stroke in general (Rorden & Karnath, 2004). This issue can elegantly be visualised by computing a simple overlap of stroke patients in general, i.e. patients unselected for any symptom. In a study on 439 unselected acute right hemisphere stroke patients (Sperber & Karnath, 2016), we found overlap maxima in the centre of the territory of the middle cerebral artery, including Heschl's gyrus, insula, and putamen. Overlap maxima are thus not specific in identifying a symptom's neural correlates, but can originate from general stroke anatomy. The solution to this problem is the inclusion of control patients into the analysis (Rorden & Karnath, 2004). Control patients are stroke patients that do not suffer from the investigated symptom. The underlying rationale is that stroke in all patients follows its typical anatomy, however, only in the group of patients with the symptom the neural correlates of the symptom are damaged. Anatomo-behavioural studies should thus compare both groups. So-called lesion subtraction analysis (Rorden & Karnath, 2004) has often been used in this context. This analysis method requires normalised lesion data (see below, chapter 2.3.2). The analysis is computed for each voxel (= volumetric pixel), i.e. for each 3D imaging point, individually. For each voxel, in both groups the proportion of patients with damage in this voxel is identified. The difference between both proportional values now can indicate if a voxel is part of a symptom's neural substrate. E.g., if a voxel is damaged in 60% of patients with the symptom, but in 15% of patients without the symptom (resulting in a difference of +45%), the voxel is assumed to be part of it. On the other hand, if a voxel is damaged in 60% of patients with the symptom, and also in 60% of patients without the symptom (resulting in a difference of 0%), the voxel is likely not neural substrate of the symptom. Latter example again illustrates how simple overlap analyses can be misleading, and why control patients are required in studies on patients with brain lesions.

2.3 Voxel-wise Statistical Mapping

Lesion subtraction analysis was an innovative method that lead to many new insights on brain architecture. Yet, it is only a qualitative approach. A more or less wide range of non-zero values is always present in a lesion subtraction analysis. Whether these values are just random stochastic fluctuations or indicative for an actual brain-behaviour relation, is not obvious. This problem was overcome by the implementation of voxel-wise statistical mapping into the field of lesion behaviour mapping. Before I discuss the principles of this method, I need to introduce some pre-requisites that are commonly used for this method. Raw brain imaging obtained by CT or MR is not directly usable in voxel-wise lesion behaviour mapping. First, lesioned areas in the brain have to be identified, and second, the images have to be warped into a common space.

2.3.1 Lesion visualisation and delineation

Identification of damaged brain tissue after stroke is not a trivial task (for reviews see Provenzale et al., 2003; Merino & Warach, 2010). A first major issue is that we need to find an imaging modality that can be used to identify structurally damaged tissue. Optimal solutions, however, vary as a function of time since stroke, ranging from hyper-acute (~ first 48 hours after stroke) and acute (first 2 weeks after stroke), to chronic stroke (>3 months after stroke). When a patient with acute neurological symptoms arrives on a stroke unit, acquisition of brain imaging is a first important step in stroke diagnosis. Non-contrast CT is very sensitive to haemorrhagic stroke even in the early acute stage of stroke. On the other hand, ischemic stroke – with about 80% of stroke patients the most common stroke aetiology – often cannot be identified with acute CT in hyper-acute stroke. Furthermore, CT can fail to identify smaller stroke. Similarly, MR imaging has some limitations in acute stroke. T1-weighted MR imaging can achieve high imaging resolution, but it does not visualise acute structural damage at all. On the other hand, it is sensitive to chronic stroke. T2-weighted imaging can visualise acute stroke with a resolution that is superior to CT. In the hyper-acute stage, diffusion-weighted MR imaging can visualise the core ischemic zone, where diffusion broke down due to structural damage. Diffusion-weighted imaging, however, only provides low resolution, and can - to a minimal degree - be misleading about the extent of structural damage (Inoue et al., 2014a).

The challenge of stroke visualisation is not solved just by choosing the right imaging modality at the right time. More fundamental concerns arise in the comparison of acute versus chronic damage. In acute stroke, deficits might not only arise from structural damage, but also from diaschisis (Carrera & Tononi, 2014; Silasi & Murphy, 2014) and temporary malperfusion (Karnath et al., 2005; Zopf et al., 2012; Sebastian et al., 2014). In the chronic stage, brain architecture might be altered due to neural plasticity, i.e., the brain’s ability to reorganize its anatomo-behavioural architecture in reaction to brain damage (e.g., Chelette et al., 2013; Vaina et al., 2014; Veldema et al., 2017). This hampers the transfer of findings in a clinical study to general anatomy of the healthy brain. Furthermore, post-stroke atrophy of brain tissue can limit the usability of chronic imaging to visualise structurally damaged areas and spatial normalisation (see below, chapter 2.3.2.). The complicated topic of choosing a time point of stroke imaging and behavioural assessment for lesion behaviour mapping was already discussed by several studies, including an own review paper (Karnath & Rennig, 2017; Shahid et al., 2017; de Haan & Karnath, 2018; Sperber & Karnath, 2018). We can summarise here that lesion visualisation for lesion behaviour mapping is not a trivial task, and that no perfect solution exists.

As soon as a neuroscientist has decided to choose a certain time point after stroke (i.e. acute vs. chronic stroke) for clinical imaging consistently across all subjects, the structural lesion can be visualised using clinical imaging as illustrated above. The next step is lesion delineation, where for each patient, each voxel is identified as either damaged or intact. This can be done either manually, or by different automatic or semi-automatic algorithms (e.g. Seghier et al., 2008; de Haan et al., 2015). The result of this procedure is a binary image, the so-called lesion map.

2.3.2 Spatial normalisation

Spatial normalisation replaces the former procedure of manually transferring a lesion onto a template (see above, chapter 2.2.). Lesion subtraction analyses and voxel-based statistical analyses work on both types of data. Yet, normalisation is preferred for being an objective method, that is independent of a researcher and his anatomical expertise.

Lesion maps of different patients are not directly comparable. Brains have different shapes and sizes, and patients can lie at different positions in the scanner. Therefore, a voxel with the same coordinates in two different lesion maps in native space may belong to different brain regions. However, when comparing a voxel between two patients in an analysis, we would like both voxels to belong to the same structure, e.g. the tip of the middle temporal pole. The spatial correspondence of two lesion maps can be achieved by spatial normalisation into a common space. In this process, the individual brain scan is warped onto a template by using linear and non-linear transformations. ‘Template’ here refers to a brain image averaged from multiple real brain images and set in a well-defined coordinate system. In normalisation, transformations are applied in a way that squared intensity differences between individual brain and template are minimised. The odd intensity values in lesioned areas can be controlled for by different strategies (Brett et al., 2001; Nachev et al., 2008). The resulting normalised brain image is roughly about the same size and shape in every subject, and set in a common coordinate system. The same transformation parameters are applied to the lesion map, which can now be used in a voxel-wise group analysis.

2.3.3 Voxel-based lesion symptom mapping

In spatially normalised lesion maps, a voxel coordinate is comparable between all subjects. This allows valid application of voxel-wise statistics. Voxel-wise mapping of statistical parameters was first applied on functional data obtained by either positron emission tomography or functional MR imaging (Friston et al., 1991; Friston et al., 1995). This framework, termed ‘statistical parametric mapping’, has been the leading analysis paradigm in the analysis of neuroimaging data for years. Its vast success is likely rooted in its simplicity: in a data sample of spatially normalised images, each voxel is analysed by any statistical parametric test. The resulting statistics are remapped into three-dimensional image space. Areas where many voxels show significant signal are interpreted as regionally specific effects. Although the general rationale to apply this framework to lesion analysis has been suggested in the mid-90s (Friston et al., 1995), it has been implemented for the first time only some years later in a landmark study by Bates et al. (2003). The method was termed ‘voxel-based lesion symptom mapping’ (VLSM), and it was used to investigate stroke patient samples with continuous behavioural scores. Its exact implementation worked as following: for each voxel, the patient sample is divided into two groups – a group of patients with damage to this voxel, and a group of patients without damage to this voxel. The behavioural variable in both groups is now compared by an independent t-test, ultimately producing a map of t-statistics. These can then be assessed for their significance. If a voxel yields a significant test, with more severe symptoms in the group of patients with damage in the voxel, then damage to the voxel is thought to underlie the symptom. A statistical map can then be interpreted in reference to a brain atlas in the same space (i.e. in the same coordinate system) in order to identify brain areas that are connected to the investigated symptom.

The VLSM-framework is not restricted to the t-test, but can be used with other statistics, such as more complex general linear models, binomial, or non-parametric tests (e.g. Karnath et al., 2004; Rorden et al., 2007; Schwartz et al., 2012). General linear models are flexible and can include further variables in order to control for covariates and more complex effects. If the behavioural variable is not continuous, but binary (e.g. symptom present/symptom absent), a binomial test such as the χ^2-test can be applied. A significant extension of the VLSM was the addition of non-parametric tests (Rorden et al., 2007), because parametric tests like the t-test make requirements such as normal distribution of data and variance homogeneity, which are commonly violated in clinical data sets. Non-parametric mapping in lesion-behaviour mapping thus can provide higher statistical power.

VLSM and its extensions have been the state-of-the-art method in lesion behaviour mapping since its first application, and they are used to gain insights into the functional architecture of the brain until today. In order to not confuse the reader with the terminology used by Bates et al. (2003), I will from now on use the term voxel-based lesion behaviour mapping (VLBM), which refers not only to the mass-univariate t-test in VLSM (Bates et al., 2003), but to all mass-univariate voxel-wise lesion symptom mapping methods. This is also in line with the nomenclature in the empirical papers in my thesis.

2.3.4 Voxel-based lesion behaviour mapping – a mass-univariate method

My thesis investigated and applied methodological approaches that either extend or even replace VLBM in certain situations. In order to understand why this can improve our insights into brain architecture, we first need to focus on one aspect of VLBM: its mass-univariate character.

Theoretically, a voxel-wise test can be a multivariate test in a way that it includes – besides voxel-wise lesion status and behavioural variable – a covariate or a second target variable. Most often, however, univariate tests like the t-test are used. For clarity in nomenclature, I will from now on only refer to univariate tests in the VLBM-framework. In a VLBM analysis, thousands of univariate statistical tests are computed. Therefore, this approach has been termed a ‘massively univariate’ or ‘mass-univariate approach’. A central feature of a mass-univariate analysis is the statistical independence of tests. Each and every voxel is tested with a univariate statistical test that is independent of all other statistical tests. Imagine we are about to compute a VLBM analysis, and we pause the VLBM in the middle of the computations. Further imagine that in a brain region with a size of 1000 voxels, 999 have already been tested, and all of them were significantly related to the tested symptom. Intuitively, we would deem it very likely that the 1000th voxel will also contain significant signal. Still, VLBM will continue to test the 1000th voxel with another independent test, that is computed as if the 999 tests before never happened. We will further see that statistical independence leads to major limitations of the VLBM framework.

3 Challenges and limitations of the mass-univariate approach

The simplicity of the VLBM-framework is contrasted by complex challenges and limitations of the mass-univariate approach or even lesion behaviour mapping in general. In two review papers (Sperber & Karnath, 2018; Karnath et al., 2018), I provided comprehensive overviews on this topic. In my thesis, I want to focus in-depth on five such challenges that are faced in mass-univariate lesion behaviour mapping. These are i) the multiple comparison problem, ii) limitations of voxel-wise statistical power, especially in rarely lesioned areas, iii) lesion size as a possible confounding factor, iv) the complexity of functional brain architecture (or functional dependence of voxels), and v) the complexity of lesion anatomy (or lesion-related dependence of voxels).

3.1 The multiple comparison problem

In a mass-univariate test, each voxel is tested independently with a statistical test. For interpretation of a test statistic, the corresponding α error probability can be computed. The α error probability indicates how likely it is to obtain a false positive result, i.e. a significant result when actually no true effect is present. In the context of a t-test, an α error would mean that the test suggests a difference of means between two groups, although there is no true difference. At which α error probability level (or α-level) statistics are performed has to be decided a priori. As there is no perfect α-level defined by nature, scientists usually follow established conventions when choosing such level. In psychology, a commonly chosen α-level is $p < 0.05$, or $p < 5\%$. This means that if you perform 20 statistical tests on data that do not contain any signal (e.g. random noise), on average one of these statistical tests will yield a significant result.

In order to decide if voxels in a VLBM analysis are significantly associated with a symptom, we also have to choose an α-level. Usually, α-levels such as $p < 0.05$ or $p < 0.01$ are chosen. A major issue – termed the multiple comparison problem –

now arises, if we perform multiple tests at the same α-level. Imagine that we investigate 100000 voxels in a VLSM analysis at an α-level of $p < 0.05$. If there is actually no real connection between voxel-wise lesion damage and behavioural variable (i.e. no true positive signal), we will anyway obtain a significant signal in 5% of all statistical tests, resulting in 5000 voxels that are significantly associated with the symptom. With this problem in mind, we can easily dismiss the entire VLBM analysis as a null result if we only find 5000 significant voxels. The situation becomes much more difficult, if we find 9000 significant voxels. Likely, some true positive signal is present in the data. Still, many voxels will be false positives and you will not be able to distinguish which part of the signal are false or true positives. Luckily, there are solutions to the multiple comparison problem.

The multiple comparison problem is not specific to VLBM or mass-univariate imaging analyses, but it is present whenever multiple statistical tests are performed simultaneously. Non-surprisingly, many scientists and statisticians implemented strategies to overcome the multiple comparison problem. A well-known, and easily applicable correction is the Bonferroni correction. If n tests are performed at a global α-level of p(global), each individual statistical test is performed with an α-level of p(individual) = p(global)/n. If a global α-level of p(global) < 0.05 is chosen, that means that the probability to obtain one or more false positives across all tests is only 5%. While the Bonferroni correction very well corrects for false positive inflation in multiple tests, it is excessively conservative. In VLBM, where thousands of tests are computed, the α-level of an individual test will be vanishingly tiny, and likely no single test will ever yield a significant result. A less conservative solution to the multiple comparison problem is a correction by false discovery rate (FDR; Benjamini & Yekutieli, 2001). Contrary to Bonferroni correction, FDR does not intend to eliminate any false positive in the analysis. Instead, a researcher using FDR accepts a certain rate of false positive results in all positive results. If, for example, a FDR of $q = 0.05$ is chosen, this means that 5% of all positive findings are expected to be false positives. If we then find 9000 significant voxels after applying FDR, we know that about 450 voxels will be false positives. FDR thus offers a trade-off between the ability to find true signal and some false positive findings. To apply FDR on a set of statistical tests, only individual p-values are required, which makes FDR simple to compute. On the other hand, there are some drawbacks of FDR in the field of lesion behaviour mapping or in general (e.g. Mirman et al., 2018; Karnath et al., 2018).

Generally, FDR appears to be too lenient in several cases, but much more conservative if the true signal is only small. Yet, FDR is a popular correction method in VLBM, statistical parametric mapping, and multiple comparison situations in general.

For statistical mapping, further correction methods based on permutation testing are available. Permutation testing is a flexible and powerful approach in statistical testing. Generally, established statistical tests such as the t-test can be replaced with a permutation test. In comparison with the t-test, permutation testing does not rely on distributional assumptions, but it requires larger computational power. Another benefit of permutation testing is that tests can perform exact, i.e. truly at an α-level of p, and not only asymptotically at p. Theoretically, all t-tests in a VLSM could be replaced by permutation tests. This would, however, not help us with the multiple comparison problem. Still, each voxel would be tested at an individual α-level, and α errors would accumulate across the many performed tests.

To solve the multiple comparison problem in VLBM with permutation tests, a more sophisticated approach has to be chosen (Nichols & Holmes, 2002; Nichols & Hayasaka, 2003). The permutation test somehow has to consider a variable that is derived not on voxel level, but that originates from the whole brain. One solution is to consider the maximum statistic. Like described above, e.g. t-tests are performed for each voxel individually on the real data. This will provide a statistical map of t-statistics. Next, several thousands of permutations of the behavioural data are also analysed with VLSM. Each of these analyses on random data will provide a maximum t-statistic, i.e. the highest t-value found across the whole statistical map. This will tell us which maximum t-statistics can be expected by chance. At an α-level of $p < 0.05$, we can now identify the maximum t-statistic that is yielded while only 5% of all analyses yield higher maximum statistics. Using this t-value, the original statistical map can be thresholded, and all voxels above this t-value are considered to be associated with the symptom. Another permutation approach in VLSM uses the maximum cluster size instead of maximum statistics. With an analogue approach, it is investigated what cluster sizes of significant voxels above an a priori α-level can be expected by chance. Then, in the VLSM on real data, all clusters are deemed significant that have a size that is larger than the threshold obtained in the permutation analysis. Permutation tests in VLSM are computationally demanding, and some drawbacks are known (Mirman et al., 2018). On the other hand, they are thought to provide a more appropriate correction than the lenient FDR correction.

3.2 Voxel-wise power and rarely lesioned brain voxels

If statistical parametric mapping using the t-test is applied on functional data, each voxel can be tested with the same groups. E.g., if two conditions are compared voxel-wise based on data of 30 subjects, each voxel will be subject to a t-test that compares 30 data points versus 30 data points. This (purely fictional) example for functional data will appear different if we now instead look on lesion data in a VLSM. Here, groups are defined by who has a lesion in the voxel and who has no lesion there. Regions across the brain are differently susceptible to stroke, and voxel-wise lesion frequencies vary (Sperber & Karnath, 2016). Which patient groups are compared per voxel and size of the groups thus varies across the brain. Furthermore, lesions are rare in many brain areas. As a consequence, the group of patients with a lesion in a voxel is very small (up to non-existent) in many voxels. A two-samples t-test that compares a large group of patients without lesion in a voxel with a group of zero patients with a lesion cannot be computed. As soon as latter group includes at least one patient, a t-test statistic can mathematically be computed, but it is still obvious non-sense. It becomes interesting as soon as we look at cases, where the lesion-group has two or more patients. It is difficult to define a cutoff at which a group is large enough to be validly tested with a t-test. A post-hoc voxel-wise power analysis can be helpful here (Kimberg et al., 2007). Such power analysis will also show that voxel-wise power considerably varies across voxels, and that power is lowest in areas that are rarely affected (Kimberg et al., 2007).

In summary, we have the following problem in nearly all lesion data samples: for many voxels in the brain statistical power is too low for proper voxel-wise analysis, and in more extreme cases statistical tests might be either not be computable at all, or might provide odd results. In order to account for these problems, scientists apply a criterion for minimum lesion affection in a lesion analysis. This is often verbalised in a paper’s method section as following: ‘We only tested voxels with at least x lesions’ (see, e.g., Goldenberg and Randerath, 2015; Mirman et al., 2015; Tarhan et al., 2015; Timpert et al., 2015; Watson and Buxbaum, 2015). This simply excludes all these potentially problematic voxels from the analysis. In turn, VLBM analyses are never truly a whole brain analysis, what usually is not explicitly communicated in studies.

3.3 The effect of lesion size

Lesion size is a major confounding factor in any anatomo-behavioural study that investigates brain damage. This is not limited to voxel-wise or mass univariate analyses. Most neurological symptoms correlate with lesion size. The reason is not that lesion size itself induces a symptom, but that larger lesions are more likely to affect a critical brain region. The resulting general problem can be grasped intuitively: imagine a researcher in the 19th century that aims to identify the neural correlates of a symptom. He is able to post-mortem dissect the brains of two patients with the symptom, and finds that one patient suffered from a full stroke of the middle cerebral artery territory, and the other suffered from a small stroke to the temporo-parietal junction. Obviously, investigation of latter patient with a smaller lesion tells us much more about the neural correlates of the symptom. Large lesions, on the other hand, do not provide us with spatially specific information. Unfortunately, there is even more to this problem in VLBM. Average size of a lesion in a voxel differs across the brain, i.e. there are regions that are on average affected more frequently by larger lesions than other regions. With lesion size being highly correlated with most behavioural variables, patients with more severe symptoms will also have larger lesions. Thus, a VLBM analysis might not only map the neural correlates of a symptom, but also regions that are typically affected by larger lesions.

Several approaches exist to overcome this issue. First, one might restrict a lesion analysis to only small lesions (Price et al., 2017). While this strategy indeed controls for the issues with lesion size – and partially also for other issues outlined in chapter 3.5 – it only aggravates issues related to statistical power. When we only investigate smaller lesions, lesion frequencies per voxel will be lower, and problems related to statistical power, as described in chapter 3.2, will be more prominent. In addition, such strategy would require application of more stringent exclusion criteria. This might not be optimal, because clinical samples are already difficult to acquire with less strict exclusion criteria. Another approach is to control for the effect of lesion size in the lesion analysis. The most common approach here is to control the behavioural variable for variance explained by lesion size (Karnath et al., 2004; Schwartz et al., 2012). This can be done by regression methods that can identify the variance in the behavioural variable explained by lesion size. In the actual lesion behaviour mapping analysis, only the residuals of this regression, i.e. the variance not explained by lesion size, is mapped.

Such stringent control of lesion size should theoretically overcome the problems in VLBM mentioned above. Yet, the approach was seen as controversial. First, lesion size often highly correlates with behavioural symptoms. Therefore, the regression approach removes a lot of variance from the behavioural data. The remaining variance might then be too low to find positive signal in VLBM analysis. In other words, control for lesion size comes with decreased statistical power. A second problem was controversially discussed (Karnath & Smith, 2014; Nachev, 2015; Xu et al., 2018). The average lesion size varies across the brain (see Sperber & Karnath, 2016), thus some brain areas are typically affected by larger than average lesions. If the neural correlates of a symptom are located in a brain area that is typically affected by larger lesions, then the control for lesion size might be unfairly penalised. It has been stated that this problem "will inevitably confound the anatomical inference" (Nachev, 2015). And indeed, looking at this problem from a more moderate and nuanced perspective, such biases are probable at least in some cases of VLBM analyses. If, however, such biases predominate, or if a control for lesion size does more good than harm, has never been investigated before.

3.4 The complexity of functional brain architecture and the partial injury problem

In the chapter above I elaborated on the mass-univariate nature of voxel-based lesion behaviour mapping and the independence of statistical tests. In two ways, the assumption of independence in VLBM seems to be unfitting to investigate the brain. The first is related to cognitive brain architecture. If we perform an independent statistical test on a single voxel – like in VLBM – we implicitly act as if damage to this single voxel was underlying the investigated symptom. However, lesions with the size of a single voxel (e.g. 1x1x1mm^3) will most likely go clinically unnoticed. This will also happen in a voxel that was found to be associated with a symptom in a VLBM analysis. The reason is that the single voxel alone does not underlie the investigated cognitive function. Instead, neurons in many voxels have to work together to create the neural substrate of cognitive abilities. This could be a larger cluster of neurons that together form a brain region, or multiple brain regions in a brain network. In other words, neurons in a voxel work dependently with neurons in other voxels.

The independence of tests in VLBM leads to the so-called 'partial injury problem' (Kinkingnéhun et al., 2007; Rorden et al., 2009). A cognitive module that is larger than a voxel can be damaged only partially by a lesion, i.e. some voxels of this module can be damaged and some of them remain intact. In this situation the VLBM analysis will suffer from lower power, and the analysis might fail to identify the brain module/brain network in parts or in whole. For an in-depth explanation of the partial injury problem, I'd like to refer the reader to the introduction and figure 1 of the second project of my thesis, '*An empirical evaluation of multivariate lesion behaviour mapping'*.

3.5 The complexity of lesion anatomy

The second way in which the independence of voxel-wise statistical tests seems to be unfitting to investigate the brain is related to lesion anatomy. The cerebrum is supplied by three major brain arteries, the anterior, the middle, and the posterior cerebral artery. These arteries each supply a territory in the brain with only small overlap. These territories are located the same across humans with some variance (van der Zwan et al., 1993; Tatu et al., 2012; Neumann et al., 2016). Likewise, branches of these major arteries are – with few exceptions, especially for the anterior cerebral artery – located similarly across individuals. Thus, each branch of a major artery typically supplies a certain brain region. Further, branches of the brain arteries are differently susceptible to stroke, leading to typical locations of lesions (see Caviness et al., 2002; Sperber & Karnath, 2016). Therefore, both after ischemia and haemorrhage, brain lesions follow typical patterns along the vasculature. This fact is well illustrated by Lee at al., 2009, who show MR images of typical posterior cerebral artery stroke loci with reference to the occluded branches of the posterior cerebral artery.

Following these typical patterns, lesions often damage voxels collaterally. In other words, damage to two voxels is not independent. But how does this affect the results of VLBM? Imagine that the neural correlates of a cognitive function are organised in a small brain area A. Placed next to this brain region A is another brain region B that is not related to the cognitive function. Further imagine that both areas are supplied by the same branch of a cerebral artery. Thus, whenever the branch is occluded in stroke, both brain regions will be affected at once. Therefore, if a post-stroke cognitive symptom is present after damage to area A, area B will often be damaged as well. Likewise, if area B is damaged after stroke, most often area A will also be damaged, and the symptom is present. A VLBM analysis investigating the symptom will now likely not only correctly identify area A as neural substrate of the function, but also area B. In this example, we see how the dependence of voxels in relation to stroke anatomy – from here one referred to as ‘anatomical dependence’ – can lead to wrong results in lesion behaviour mapping.

4 Investigating the validity of lesion behaviour mapping methods

In the previous chapter, I introduced a quintet of challenges and limitations present in mass-univariate lesion behaviour mapping. I discussed these points from a purely theoretical perspective, with some fictional examples. Such theoretical perspective, however, might not be sufficient in finding the optimal way to perform lesion behaviour mapping. While the multiple comparison problem does not leave much opportunity for objections, the other problems are more controversial. Especially for the last two issues – functional and anatomical dependence of voxels – there is a wide range of possible conclusions available. On the one extreme, we might now admit that some possible inaccuracies exist in lesion behaviour mapping, that we however deem to be too small or irrelevant when performing a study. On the other extreme, we might feel compelled to discard the mass-univariate entirely and abandon all findings of previous studies using this method (Mah et al., 2014; Nachev, 2015; Xu et al., 2018).

In my opinion, these problems are too complex to solve them from a purely theoretical perspective. If we aim to find out if these issues considerably affect the validity of VLBM, or if we want to find out which strategy to investigate lesion-behaviour inference is superior, we need some validation method. Unfortunately, it is not trivial to investigate the validity of lesion behaviour mapping. Ideally, the results of a lesion behaviour mapping analysis would be compared with some ground truth, i.e. a cognitive module that is anatomically well known and thus represents a gold standard. A valid analysis of the related cognitive function should then identify the ground truth. However, the knowledge we have of brain architecture is largely based on lesion behaviour mapping. Using our knowledge of the brain to define a ground truth thus is a circular error. In a review paper (Sperber & Karnath, 2018), I addressed this problem at greater length, and proposed several strategies to investigate the validity of lesion behaviour mapping methods. In my dissertation, one of these approaches plays the most important role: simulation studies.

4.1 A simulation approach to test the validity of lesion behaviour mapping

We can investigate the validity of lesion behaviour mapping with simulations. These simulations require a sample of real lesions, which are processed into binary, normalised lesion maps. The behavioural scores required for lesion behaviour mapping, however, are no real data. Instead, these are simulated data. The central idea is that we arbitrarily choose the anatomical ground truth underlying a cognitive function. For example, we decide that the inferior parietal gyrus – as defined by any brain atlas – is the neural correlate of a fictional symptom. Next, we compute a score for the behavioural symptom for each lesion. For example, we could decide that each lesion that has at least 10% of all voxels in the inferior parietal gyrus damaged is associated with the presence of a binary symptom. If we instead desire a continuous variable, we could choose an algorithm that computes a behavioural score from a lesion's damage to the inferior parietal gyrus. A simple, straight-forward algorithm is a linear relation between damage to the inferior parietal gyrus and the behavioural score. In its simplest version, the behavioural score is equivalent to the damage in the region. For example, a lesion that affects 25% of all voxels in the inferior parietal gyrus would be associated with a behavioural score of 25; the maximum obtainable score would then be 100, indicating maximal symptom severity.

There are several advantages of simulation studies. First, and most importantly, we have exact knowledge about the ground truth, i.e. the anatomical correlates of a symptom. For this very reason, we can use simulations to validate lesion behaviour mapping. Second, an infinite amount of ground truth regions or simulation algorithms can be chosen. This allows us to perform large group studies with complex designs. Third, simulations provide us with a tool to investigate all the issues that I introduced in chapter 3. By using real lesions in simulations, we will likely find the usual effects of lesion size, and damage between voxels is dependent. Further, we can choose ground truths that consist of multiple regions, thus adding functional dependence between voxels. A major limitation of simulations is limited ecological validity – simulation algorithms are likely over-simplified, with real lesion-behaviour relations being more complex in several aspects. Still, such simulations provide us with a powerful opportunity to validate lesion behaviour mapping. Consequently, several studies utilised simulation approaches to compare different approaches to lesion behaviour mapping (Rorden et al., 2009; Mah et al., 2014; Inoue et al., 2014b; Zhang et al., 2014; Pustina et al., 2018; DeMarco & Turkeltaub, 2018).

4.2 The independence of statistical tests in the mass univariate approach put on trial

Some studies used simulations to investigate if functional or anatomical dependence between voxels affects lesion behaviour mapping. The first study came from the Nachev group (Mah et al., 2014), and was – so far – the most influential one. This study aimed to investigate if and how much VLBM is affected by functional or anatomical dependence. In a first experiment, they investigated simulations that were only based on damage to one single voxel. Thus, there was no functional dependence of voxels present, but due to the use of real lesions, anatomical dependence was. They found that maps of statistically significant voxels in a VLBM were not centred on the simulation's ground truth voxel, but shifted by on average 16mm towards the centre of the vascular territory. In a second experiment, they investigated what happens if functional independence comes into play. They based simulations on two ground truth regions, in a way that damage to either region could lead to the simulated symptom. Doing so they showed that VLBM can yield results that can be grossly misplaced. The authors concluded that the only possible consequence was nothing less than the abandonment of mass-univariate lesion behaviour mapping in its entirety. The same position was hold in subsequent review papers by the Nachev group (Nachev, 2015; Xu et al., 2018). In parallel, the study by Inoue et al. (2014) investigated the same questions, with only slightly different simulations. Their findings mirrored the ones by Mah et al. (2014), thus strengthening their quality by a first replication.

The issues with functional independence were further disseminated in the studies by Zhang et al. (2014) and Pustina et al. (2018). They also investigated simulation ground truths consisting of multiple regions, however with a more elaborated design. Both studies found that VLBM often fails to identify all ground truth regions, and instead only correctly identifies some of them. These findings are in line with the partial injury problem (see chapter 3.4).

To sum up, some studies have shown that the assumption of independence in mass-univariate lesion behaviour mapping indeed leads to errors. This was found to be the case for both functional and anatomical independence of voxels. Although I do not fully agree with the rigorous criticism expressed by Nachev and colleagues, I think that these issues in VLBM are a clear limitation. From personal experience, I especially see problems in identifying brain networks, which is hampered by the partial injury problem. As examples where this issue is highly relevant, I see apraxia of pantomime (see the third project in my thesis, '*The network underlying human higher-order motor control: Insights from machine learning-based lesion-behaviour mapping')* or spatial neglect (see Karnath & Rorden, 2012). For both symptoms, VLBM studies found markedly heterogeneous results. As outlined in my review (Sperber & Karnath, 2018), this heterogeneity might have originated from the partial injury problem. VLBM analyses might have identified only parts of the underlying brain networks, and the parts found in each study might have varied due to random sampling effects or due to smaller methodological differences.

5 Multivariate lesion behaviour mapping

The findings by Mah et al. (2014) lead to vigorous discussions on the validity of mass-univariate lesion behaviour mapping, and the flames were fanned by a parallel development: the advent of multivariate lesion behaviour mapping (MLBM).

The central idea behind MLBM is to compute (statistical) models not for each voxel individually, but for multiple voxels or regions at once. Theoretically, any multivariate statistical method that can model a dependent variable (in VLBM: the behavioural score) based on multiple independent variables (in VLBM: voxel-wise lesion status) could be considered here. Methods such as multiple regression or n-way ANOVAs (for n ≥ 2), however, are limited in lesion behaviour mapping, as they are not suited to compute models based on enormous numbers of independent variables. Such methods are especially problematic if the number of observations (i.e. the number of subjects) is smaller than the number of independent variables. Thus, previous implementations of MLBM utilised more complex ways to model data. Such can be machine learning algorithms such as a support vector machine (SVM). SVMs model a binary variable based on a large number of variables (for more information see Vapnik, 1995; Hastie et al., 2008). As machine learning algorithms play an important role in my thesis, we require some technical terms: the dependent variables in machine learning are often referred to as input variables or *features*. The predicted variable, or *target variable*, can be a binary or a continuous variable. When the target variable is binary (e.g. symptoms present vs. symptom not present), the algorithms are used to perform a *classification*. When it is continuous, the algorithms are used to perform a *regression*.

The first study that performed MLBM used a SVM to classify the presence of spatial neglect (Smith et al., 2013). Aim of the study was to find a SVM that can classify patients only by using anatomical data. Such SVM model could provide new insights into lesion-deficit inference. As features, Smith et al. computed the proportion of damage to a priori chosen regions of interest. These regions of interest were taken from brain atlases. In so-called cross-validation, the performance of SVMs was assessed. It was investigated if combinations of either two or three features (i.e. damage to two or three regions of interest) provide better models than SVMs using less features, that is either one or two regions. This strategy points at a major challenge: the relevance of a single feature in SVM is difficult to access. A viable strategy is to compute a SVM on a set of features, and then remove one feature. If model performance significantly decreases, the one feature was important. The problem in MLBM is that with many features, like dozens of regions of interest or even thousands of voxels, the amount of possible feature combinations explodes. Therefore, the study by Smith et al. was restricted to a maximum of three features.

All in all, there are several disadvantages to this approach. First, it is limited to a priori chosen regions of interest. Second, although strictly speaking being multivariate, the approach is limited to a small number of features. Third, being a classification, the full variance of continuously measured behaviour cannot be investigated. These problems are shared with another multivariate approach based on game theory (Toba et al., 2017).

The next study that used MLBM was the study by Mah et al. (2014), who did not only identify problems with VLBM (see chapters 3.4 and 3.5), but additionally suggested that MLBM might be the only way to obtain valid lesion-brain inference. However, they also faced the challenge of interpreting the relevance of features in SVM. They performed a SVM on voxel level. To do so, they included the damage status of each voxel (damaged vs. not damaged) as features. With such large amount of features, a direct comparison between SVM models is not an option. Instead, they assessed the feature weights. In a SVM, each single feature is weighted in order to generate the model. This feature weighting is not informative on its own, and it does not tell us if a feature significantly contributes to a model. However, feature weights can be ranked. Mah et al. aimed to directly compare VLBM with MLBM. In the VLBM, they found a certain amount of significant voxels. Then, in the SVM, they selected the same amount of voxels, and chose the voxels that had the highest feature weight. This approach allowed a limited comparison of univariate versus multivariate lesion analysis, yet it is not a practically usable approach. Importantly, Mah et al. postulated that MLBM is able to overcome the problems related to both functional and anatomical dependence of voxels. Their simulations, however, only pointed at better performance of MLBM in regard to functional dependence (see chapter 3.4), but not anatomical dependence (see chapter 3.5). Still, the authors heavily emphasized the importance of MLBM to overcome the problems related to anatomical dependence, and maintained this position in two review papers (Nachev, 2015; Xu et al., 2018).

Only shortly after the study by Mah et al. (2014), Zhang et al. (2014) published the next study that implemented MLBM. This study finally was able to overcome the problem with interpreting features. As in the MLBM approach by Mah et al., Zhang et al. included information from individual voxels as features. Instead of a SVM, they employed a support vector regression (SVR). This is an algorithm that extents SVM so that continuous target variables can be included. The major innovation was a permutation approach to test the contribution of each voxel to the SVR model. As in Mah et al., the model was computed and feature weights were assessed. Next, the same procedure was performed for a large amount of random permutations of the behavioural data. Latter analyses shows what feature weights can be expected with random data. Like commonly done in permutation testing, statistical thresholds can be inferred from these analyses to assess statistical significance of feature weights in the analysis of real data. The resulting topography is a voxel-wise map of statistical significance.

This approach, termed support vector regression based lesion symptom mapping (SVR-LSM), has several advantages. First, it is able to model continuous target variables. Second, it includes individual voxels as features, and thus it does not depend on the quality of any a priori parcellation of the brain. Furthermore, the modelling process can include an immense amount of voxels, even up to a whole brain analysis. Last but not least, the method underwent an elaborated validation by simulations. These simulations have shown that SVR-LSM outperforms VLBM in identifying complex functional modules like brain networks. In other words, SVR-LSM is able to capture the functional dependence of voxels. Likely due to these benefits, SVR-LSM was quickly adopted in the field (Mirman et al., 2015b; Fama et al., 2017; Griffis et al., 2017; Ghaleh et al., 2018; DeMarco & Turkeltaub, 2018), and I adopted this method while working on my thesis to investigate the neural correlates of spatial neglect (Wiesen et al., submitted), the line bisection error, and apraxia. The latter is part of the present thesis.

Note that in recent years, more attempts to establish multivariate lesion-brain inference were made (e.g. Yourganov et al., 2015; Toba et al., 2017; Pustina et al., 2018). These approaches were not investigated or used in my thesis. Therefore, I will not further discuss them. This shall not hide the fact these methods also have potential to complement lesion-deficit inference.

6 Empirical research questions in my thesis

6.1 Impact of correction factors in human brain lesion-behavior inference

The first empirical work takes a second look at the validity of VLBM. It closely follows the work by Mah et al. (2014), who found systematic errors in VLBM due to functional and anatomical dependence of voxels. In my work, I challenge the severe criticism on VLBM by putting a focus on two correction factors in VLBM: correction for lesion size and restriction of the analysis to voxels with sufficient power. Both factors were neglected in the study by Mah et al. Using simulations, my study shows that this has inflated errors in VLBM, and that correction for lesion size is generally beneficial in VLBM.

6.2 An empirical evaluation of multivariate lesion behaviour mapping

The second empirical work examines the SVR-LSM approach. Although previous studies have shown the superiority of SVR-LSM in detecting functionally dependent brain modules, many open questions remained. A major theoretically relevant question is, if SVR-LSM not only captures functional dependence between voxels, but also anatomical dependence. Several authors advertised multivariate lesion behaviour mapping as only valid way to analyse lesion-deficit inference because it captures the anatomical dependence of voxels. However, it was never properly investigated before if MLBM indeed is able to do so. In my thesis, I use simulations to show that this in fact not the case – SVR-LSM still suffers from systematic biases due to lesion anatomy. Furthermore, I investigate the practically relevant questions if SVR-LSM requires a correction for multiple comparisons, and what sample sizes are required in SVR-LSM.

6.3 The network underlying human higher-order motor control: Insights from machine learning-based lesion-behaviour mapping

The third empirical work finally applies SVR-LSM to investigate the neural correlates of real behaviour. Here, I investigate the apraxia of pantomime. Previous studies investigated apraxia with VLBM, and their results were heterogeneous. A possible explanation might be that the actual neural correlates of apraxia are a complex brain network, where VLBM is limited due to the partial injury problem. In my thesis, I showed that SVR-LSM identifies multiple brain regions to underlie apraxia. All these brain regions were found by previous VLBM studies, but often in isolation. This suggests that SVR-LSM is a valuable tool when brain functions organised in networks are investigated.

7 Future challenges and research directions in lesion behaviour mapping

A large amount of methodological innovations in the field of lesion-deficit inference came up in the last years. Besides MLBM, several methods based on magnetic resonance imaging also allow insights into lesion-deficit inference. Among those are fMRI, resting-state fMRI, diffusion tensor imaging, and perfusion imaging. In a review paper, I discussed how these methods can be used in neurological patients to investigate the functional anatomy of the brain (Karnath et al., 2018). For new research directions in the near future, it suggests itself to utilise these methods to study brain anatomy. Currently, I apply MLBM to gain new insights into the neural correlates of several neurological deficits. Furthermore, I would like to deepen the understanding of anatomical networks underlying praxis skills by using diffusion tensor imaging. I believe this method is better suited to identify involved white matter tracts than lesion behaviour mapping methods. In this case, lesion behaviour mapping and fibre tracking nicely complement each other.

Additionally, there are major challenges that require methodological refinement. These methodologically oriented research directions also logically follow the works in my thesis: the optimisation of MLBM and its clinical application.

7.1 Optimisation of (multivariate) lesion behaviour mapping

With the help of the simulation approach, it was shown that MLBM is superior to VLBM in some regards. Still, MLBM is far from being perfect. One main problem is that hyperparameter optimisation in SVR-LSM should not solely aim at maximising model fit. Rather, a trade-off between model fit and parameter generalisability should be aimed at (Rasmussen et al., 2012; Zhang et al., 2014). Unfortunately, clear guidelines on hyperparameter optimisation in SVR-LSM are not existent. Another main problem is the power to detect positive signal. The study on MLBM in apraxia in my thesis, but also several previous studies of other authors using SVR-LSM found only weak signal. Here, like for VLBM (Rorden et al., 2009), strategies to improve statistical power are required. In addition, the problem with anatomical dependence still exists. I believe that there is no way to remove this bias entirely. Yet, some ways exist that might reduce it, and future studies should investigate their potential.

The SVR-LSM approach is not the only method in MLBM. Other methods based on game theory (Toba et al., 2017) or sparse canonical correlations (Pustina et al., 2018) were recently developed. Especially the latter was thoroughly evaluated and promises high potential in lesion behaviour mapping. Nobody yet compared all these MLBM methods, and it is not known which of these methods is least susceptible to biases due to anatomy, which is best suited to identify brain networks, and which is most valid with smaller sample sizes.

7.2 Translational utilisation of lesion behaviour mapping

I am deeply interested in understanding the brain, and lesion-deficit inference plays a role in doing so. However, sometimes I ask myself what my work in basic research is good for, and why society should fund my research. Even worse, family or friends might want to know what purpose my research has. Luckily, there are potential clinical applications of lesion behaviour mapping.

In MLBM, machine learning algorithms are used to model behavioural scores. How well the algorithms can predict the behavioural score from anatomical data only plays a minor role here. Prediction of scores, however, bears translational applicability when we find algorithms that can predict chronic behaviour based on acute brain imaging. Such algorithms could predict, e.g., the upper limb motor impairment six months after stroke based on brain imaging acquired two days after stroke onset. If such algorithms work with high prediction accuracy, we could provide important information for planning patient care and guide rehabilitation. If such algorithms point at high potential for recovery, rehabilitation efforts could be increased, or otherwise, with no potential for recovery, compensation strategies could be trained.

Some prediction algorithms already were established recently (Rondina et al., 2016). There are, however, several challenges left unsolved. In the moment, I work on prediction algorithms for motor impairments. A study by Rondina et al. (2016) suggested that feature selection is a major factor for well-performing algorithms. Here, approaches adopted from MLBM could provide means to perform feature selection in a data-driven way.

Another is to find out what further variables – besides voxel-wise damage – underlie pathological behaviour. In MLBM, we only want to investigate the role of voxel-wise damage in the modelling process, and maximisation of prediction accuracy is not an aim. However, for translational use, the algorithms are supposed to be maximised for prediction. As factors such as age, pre-stroke cognitive status, co-morbidity, education and so on might play a role in inducing a symptom, we need find out which additional variables have to be included into the models for a good prediction.

References

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. *Nature Neuroscience*, *6*(5), 448–450. https://doi.org/10.1038/nn1050

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat., 29*, 1165-1188.

Brett, M., Leff, A. P., Rorden, C., & Ashburner, J. (2001). Spatial Normalization of Brain Images with Focal Lesions Using Cost Function Masking. *NeuroImage*, *14*(2), 486–500. https://doi.org/10.1006/nimg.2001.0845

Broca, P. P. (1861). Remarks on the seat of the faculty of articulated language, following an observation of aphemia. *Bulletin de la Société Anatomique, 6,* 330–357. Transl. Green DC.

Carrera, E., & Tononi, G. (2014). Diaschisis: Past, present, future. *Brain*, *137*(9), 2408–2422. https://doi.org/10.1093/brain/awu101

Caviness, V. S., Makris, N., Montinaro, E., Sahin, N. T., Bates, J. F., Schwamm, L., … Kennedy, D. N. (2002). Anatomy of Stroke, Part I: An MRI-Based Topographic and Volumetric System of Analysis. *Stroke*, *33*(11), 2549–2556. https://doi.org/10.1161/01.STR.0000036083.90045.08

Chelette, K. C., Carrico, C., Nichols, L., & Sawaki, L. (2013). Long-term cortical reorganization following stroke in a single subject with severe motor impairment. *NeuroRehabilitation*, *33*(3), 385–389. https://doi.org/10.3233/NRE-130968

de Haan, B., Clas, P., Juenger, H., Wilke, M., & Karnath, H.-O. (2015). Fast semi-automated lesion demarcation in stroke. *NeuroImage. Clinical*, *9*, 69–74. https://doi.org/10.1016/j.nicl.2015.06.013

DeMarco, A. T., & Turkeltaub, P. E. (2018). A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion-symptom mapping. *Human Brain Mapping*, *21*(May), 2461–2467. https://doi.org/10.1002/hbm.24289

Fama, M. E., Hayward, W., Snider, S. F., Friedman, R. B., & Turkeltaub, P. E. (2017). Subjective experience of inner speech in aphasia: Preliminary behavioral relationships and neural correlates. *Brain and Language*, *164*, 32–42. https://doi.org/10.1016/j.bandl.2016.09.009

Freeman, W. D., & Aguilar, M. I. (2012). Intracranial hemorrhage: Diagnosis and management. *Neurologic Clinics*, *30*(1), 212–240. https://doi.org/10.1016/j.ncl.2011.09.002

Ghaleh, M., Skipper-Kallal, L. M., Xing, S., Lacey, E., DeWitt, I., DeMarco, A., & Turkeltaub, P. (2018). Phonotactic processing deficit following left-hemisphere stroke. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *99*, 346–357. https://doi.org/10.1016/j.cortex.2017.12.010

Glickstein, M., & Whitteridge, D. (1987). Tatsuji Inouye and the mapping of the visual fields on the human cerebral cortex. *Trends in Neurosciences*, *10*(9), 350–353. https://doi.org/10.1016/0166-2236(87)90066-X

Goldenberg, G., & Randerath, J. (2015). Shared neural substrates of apraxia and aphasia. *Neuropsychologia*, *75*, 40–49. https://doi.org/10.1016/j.neuropsychologia.2015.05.017

Hastie, T., Tibshirani, R., & Friedmann, J. (2008). The Elements of Statistical Learning. 2nd Edition. Berlin: Springer.

Holmes, G., & Lister, W. T. (1916). Disturbances of vision from cerebral lesions, with special reference to the cortical representation of the macula. Brain, 39, 34-73.

Griffis, J. C., Nenert, R., Allendorfer, J. B., & Szaflarski, J. P. (2017). Damage to white matter bottlenecks contributes to language impairments after left hemispheric stroke. *NeuroImage: Clinical*, *14*, 552–565. https://doi.org/10.1016/j.nicl.2017.02.019

Inoue, K., Madhyastha, T., Rudrauf, D., Mehta, S., & Grabowski, T. (2014). What affects detectability of lesion–deficit relationships in lesion studies? *NeuroImage: Clinical*, *6*, 388–397. https://doi.org/10.1016/j.nicl.2014.10.002

Inoue, M., Mlynash, M., Christensen, S., Wheeler, H. M., Straka, M., Tipirneni, A., … Albers, G. W. (2014). Early diffusion-weighted imaging reversal after endovascular reperfusion is typically transient in patients imaged 3 to 6 hours after onset. *Stroke*, *45*(4), 1024–1028. https://doi.org/10.1161/STROKEAHA.113.002135

Jäger, H. R. (2000). Diagnosis of stroke with advanced CT and MR imaging. *British Medical Bulletin*, *56*(2), 318–333.

Karnath, H. O., Zopf, R., Johannsen, L., Berger, M. F., N??gele, T., & Klose, U. (2005). Normalized perfusion MRI to identify common areas of dysfunction: Patients with basal ganglia neglect. *Brain*, *128*(10), 2462–2469. https://doi.org/10.1093/brain/awh629

Karnath, H.-O., Fruhmann Berger, M., Küker, W., & Rorden, C. (2004). The anatomy of spatial neglect based on voxelwise statistical analysis: a study of 140 patients. *Cerebral Cortex (New York, N.Y.: 1991)*, *14*(10), 1164–1172. https://doi.org/10.1093/cercor/bhh076

Karnath, H.-O., & Rennig, J. (2017). Investigating structure and function in the healthy human brain: validity of acute versus chronic lesion-symptom mapping.

Brain Structure & Function, *222*(5), 2059–2070. https://doi.org/10.1007/s00429-016-1325-7

Karnath, H.-O., & Rorden, C. (2012). The anatomy of spatial neglect. *Neuropsychologia*, *50*(6), 1010–1017. https://doi.org/10.1016/j.neuropsychologia.2011.06.027

Karnath, H.-O., & Smith, D. V. (2014). The next step in modern brain lesion analysis: multivariate pattern analysis. *Brain: A Journal of Neurology*, *137*(Pt 9), 2405–2407. https://doi.org/10.1093/brain/awu180

Karnath, H.-O., Sperber, C., & Rorden, C. (2018). Mapping human brain lesions and their functional consequences. *NeuroImage*, *165*(May 2017), 180–189. https://doi.org/10.1016/j.neuroimage.2017.10.028

Kimberg, D. Y., Coslett, H. B., & Schwartz, M. F. (2007). Power in Voxel-based lesion-symptom mapping. *Journal of Cognitive Neuroscience*, *19*(7), 1067–1080. https://doi.org/10.1162/jocn.2007.19.7.1067

Kinkingnéhun, S., Volle, E., Pélégrini-Issac, M., Golmard, J. L., Lehéricy, S., du Boisguéheneuc, F., … Dubois, B. (2007). A novel approach to clinical-radiological correlations: Anatomo-Clinical Overlapping Maps (AnaCOM): Method and validation. *NeuroImage*, *37*(4), 1237–1249. https://doi.org/10.1016/j.neuroimage.2007.06.027

Lauterbur, P. C. (1973). Image formation by induced local interactions. Examples employing nuclear magnetic resonance. *Nature (London, United Kingdom)*, *242*, 190–191. https://doi.org/10.1038/242190a0

Lauterbur, P. C. (1974). Magnetic resonance zeugmatography. *Pure and Applied Chemistry*, *40*(1–2), 149–157. https://doi.org/10.1351/pac197440010149

Mah, Y.-H., Husain, M., Rees, G., & Nachev, P. (2014). Human brain lesion-deficit inference remapped. *Brain: A Journal of Neurology*, *137*(Pt 9), 2522–2531. https://doi.org/10.1093/brain/awu164

Merino, J. G., & Warach, S. (2010). Imaging of acute stroke. *Nature Reviews Neurology*, *6*(10), 560–571. https://doi.org/10.1038/nrneurol.2010.129

Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O. K., Coslett, H. B., & Schwartz, M. F. (2015). Neural organization of spoken language revealed by lesion-symptom mapping. *Nature Communications*, *6*, 6762. https://doi.org/10.1038/ncomms7762

Mirman, D., Landrigan, J.-F., Kokolis, S., Verillo, S., Ferrara, C., & Pustina, D. (2018). Corrections for multiple comparisons in voxel-based lesion-symptom mapping. *Neuropsychologia*, *115*(December 2016), 112–123. https://doi.org/10.1016/j.neuropsychologia.2017.08.025

Mirman, D., Zhang, Y., Wang, Z., Coslett, H. B., & Schwartz, M. F. (2015). The ins and outs of meaning: Behavioral and neuroanatomical dissociation of semantically-driven word retrieval and multimodal semantic recognition in aphasia. *Neuropsychologia*, *76*(3), 208–219. https://doi.org/10.1016/j.neuropsychologia.2015.02.014

Nachev, P. (2015). The first step in modern lesion-deficit analysis. *Brain: A Journal of Neurology*, *138*(Pt 6), e354. https://doi.org/10.1093/brain/awu275

Nachev, P., Coulthard, E., Jäger, H. R., Kennard, C., & Husain, M. (2008). Enantiomorphic normalization of focally lesioned brains. *NeuroImage*, *39*(3), 1215–1226. https://doi.org/10.1016/j.neuroimage.2007.10.002

Neumann, J. O., Giese, H., Nagel, A. M., Biller, A., Unterberg, A., & Meinzer, H. P. (2016). MR Angiography at 7T to Visualize Cerebrovascular Territories. *Journal of Neuroimaging*, *26*(5), 519–524. https://doi.org/10.1111/jon.12348

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25. https://doi.org/10.1002/hbm.1058

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, *12*(5), 419–446. https://doi.org/10.1191/0962280203sm341ra

Price, C. J., Hope, T. M., & Seghier, M. L. (2017). Ten problems and solutions when predicting individual outcome from lesion site after stroke. *NeuroImage*, *145*(Pt B), 200–208. https://doi.org/10.1016/j.neuroimage.2016.08.006

Provenzale, J. M., Jahan, R., Naidich, T. P., & Fox, A. J. (2003). Assessment of the patient with hyperacute stroke: imaging and therapy. *Radiology*, *229*(2), 347–359. https://doi.org/10.1148/radiol.2292020402

Pustina, D., Avants, B., Faseyitan, O. K., Medaglia, J. D., & Coslett, H. B. (2018). Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. *Neuropsychologia*, *115*(August), 154–166. https://doi.org/10.1016/j.neuropsychologia.2017.08.027

Rasmussen, P. M., Hansen, L. K., Madsen, K. H., Churchill, N. W., & Strother, S. C. (2012). Model sparsity and brain pattern interpretation of classification models in neuroimaging. *Pattern Recognition*, *45*(6), 2085–2100. https://doi.org/10.1016/j.patcog.2011.09.011

Rondina, J. M., Filippone, M., Girolami, M., & Ward, N. S. (2016). Decoding post-stroke motor function from structural brain imaging. *NeuroImage. Clinical*, *12*, 372–380. https://doi.org/10.1016/j.nicl.2016.07.014

Rorden, C., Fridriksson, J., & Karnath, H.-O. (2009). An evaluation of traditional and novel tools for lesion behavior mapping. *NeuroImage*, *44*(4), 1355–1362. https://doi.org/10.1016/j.neuroimage.2008.09.031

Rorden, C., & Karnath, H.-O. (2004). Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nature Reviews. Neuroscience*, *5*(10), 813–819. https://doi.org/10.1038/nrn1521

Rorden, C., Karnath, H.-O., & Bonilha, L. (2007). Improving lesion-symptom mapping. *Journal of Cognitive Neuroscience*, *19*(7), 1081–1088. https://doi.org/10.1162/jocn.2007.19.7.1081

Schwartz, M. F., Faseyitan, O., Kim, J., & Coslett, H. B. (2012). The dorsal stream contribution to phonological retrieval in object naming. *Brain*, *135*(12), 3799–3814. https://doi.org/10.1093/brain/aws300

Sebastian, R., Schein, M. G., Davis, C., Gomez, Y., Newhart, M., Oishi, K., & Hillis, A. E. (2014). Aphasia or Neglect after Thalamic Stroke: The Various Ways They may be Related to Cortical Hypoperfusion. *Frontiers in Neurology*, *5*(November), 1–8. https://doi.org/10.3389/fneur.2014.00231

Seghier, M., Ramlackhansingh, A., & Crinion, J. (2008). Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage*, *41*, 1253–1266. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811908002553

Shahid, H., Sebastian, R., Schnur, T. T., Hanayik, T., Wright, A., Tippett, D. C., … Hillis, A. E. (2017). Important considerations in lesion-symptom mapping: Illustrations from studies of word comprehension. *Human Brain Mapping*, *38*(6), 2990–3000. https://doi.org/10.1002/hbm.23567

Silasi, G., & Murphy, T. H. (2014). Stroke and the connectome: How connectivity guides therapeutic intervention. *Neuron*, *83*(6), 1354–1368. https://doi.org/10.1016/j.neuron.2014.08.052

Smith, D. V, Clithero, J. a, Rorden, C., & Karnath, H.-O. (2013). Decoding the anatomical network of spatial attention. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(4), 1518–1523. https://doi.org/10.1073/pnas.1210126110

Sperber, C., & Karnath, H.-O. (2018). On the validity of lesion-behaviour mapping methods. *Neuropsychologia*, *115*(July 2017), 17–24. https://doi.org/10.1016/j.neuropsychologia.2017.07.035

Sperber, C., & Karnath, H.-O. (2016). Topography of acute stroke in a sample of 439 right brain damaged patients. *NeuroImage. Clinical*, *10*, 124–128. https://doi.org/10.1016/j.nicl.2015.11.012

Tarhan, L. Y., Watson, C. E., & Buxbaum, L. J. (2015). Shared and Distinct Neuroanatomic Regions Critical for Tool-related Action Production and Recognition: Evidence from 131 Left-hemisphere Stroke Patients. *Journal of Cognitive Neuroscience*, *27*(12), 2491–2511. https://doi.org/10.1162/jocn_a_00876

Tatu, L., Moulin, T., Vuillier, F., & Bogousslavsky, J. (2012). Arterial territories of the human brain. *Frontiers of Neurology and Neuroscience*, *30*, 99–110. https://doi.org/10.1159/000333602

Timpert, D. C., Weiss, P. H., Vossel, S., Dovern, A., & Fink, G. R. (2015). Apraxia and spatial inattention dissociate in left hemisphere stroke. *Cortex*, *71*, 349–358. https://doi.org/10.1016/j.cortex.2015.07.023

Toba, M. N., Zavaglia, M., Rastelli, F., Valabrégue, R., Pradat-Diehl, P., Valero-Cabré, A., & Hilgetag, C. C. (2017). Game theoretical mapping of causal interactions underlying visuo-spatial attention in the human brain based on stroke lesions. *Human Brain Mapping*, *3471*(November 2016), 3454–3471. https://doi.org/10.1002/hbm.23601

Van der Zwan, A., Hillen, B., Tulleken, C.A., Dujovny, M. (1993). A quantitative investigation of the variability of the major cerebral arterial territories. *Stroke, 24*, 1951-9.

Vapnik, V. N. (1995). The nature of statistical learning theory. Berlin: Springer.

Vaina, L. M., Soloviev, S., Calabro, F. J., Buonanno, F., Passingham, R., & Cowey, A. (2014). Reorganization of retinotopic maps after occipital lobe infarction. *Journal of Cognitive Neuroscience*, *26*(6), 1266–1282. https://doi.org/10.1162/jocn_a_00538

Veldema, J., Bösl, K., & Nowak, D. A. (2017). Motor Recovery of the Affected Hand in Subacute Stroke Correlates with Changes of Contralesional Cortical Hand Motor Representation. *Neural Plasticity*, *2017*, 6171903. https://doi.org/10.1155/2017/6171903

Watson, C. E., & Buxbaum, L. J. (2015). A distributed network critical for selecting among tool-directed actions. *Cortex*, *65*, 65–82.

Wiesen, D., Sperber, C., Yourganov, G., Rorden, C., & Karnath. H. O. (submitted). The perisylvian network of spatial neglect: insights from machine learning-based lesion-behaviour mapping.

Xu, T., Jha, A., & Nachev, P. (2018). The dimensionalities of lesion-deficit mapping. *Neuropsychologia*, *115*(May), 134–141. https://doi.org/10.1016/j.neuropsychologia.2017.09.007

Yourganov, G., Smith, K. G., Fridriksson, J., & Rorden, C. (2015). Predicting aphasia type from brain damage measured with structural MRI. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *73*, 203–215. https://doi.org/10.1016/j.cortex.2015.09.005

Zhang, Y., Kimberg, D. Y., Coslett, H. B., Schwartz, M. F., & Wang, Z. (2014). Multivariate lesion-symptom mapping using support vector regression. *Human Brain Mapping*, *5876*, 5861–5876. https://doi.org/10.1002/hbm.22590

Zopf, R., Klose, U., & Karnath, H. O. (2012). Evaluation of methods for detecting perfusion abnormalities after stroke in dysfunctional brain regions. *Brain Structure and Function*, *217*(2), 667–675. https://doi.org/10.1007/s00429-011-0363-4

List of papers/manuscripts appended

Sperber C, Karnath H-O. 2017. Impact of correction factors in human brain lesion-behavior inference. ***Hum Brain Mapp***. 38:1692–1701.

Sperber C, Wiesen D, Karnath H-O. *An empirical evaluation of multivariate lesion behaviour mapping.*

Sperber C, Wiesen D, Goldenberg G, Karnath H-O. *The network underlying human higher-order motor control: Insights from machine learning-based lesion-behaviour mapping.*

Appended papers/manuscripts

Impact of correction factors in human brain lesion-behavior inference

Christoph Sperber[1]& Hans-Otto Karnath[1,2]

[1]Centre 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

Abstract

Statistical voxel-based lesion-behavior mapping (VLBM) in neurological patients with brain lesions is frequently used to examine the relationship between structure and function of the healthy human brain. Only recently, two simulation studies noted reduced anatomical validity of this method, observing the results of VLBM to be systematically misplaced by about 16 mm. However, both simulation studies differed from VLBM analyses of real data in that they lacked the proper use of two correction factors: lesion size and ‘sufficient lesion affection’. In simulation experiments on a sample of 274 real stroke patients we found that the use of these two correction factors reduced misplacement markedly compared to uncorrected VLBM. Apparently, the misplacement is due to physiological effects of brain lesion anatomy. Voxel-wise topographies of collateral damage in the real data were generated and used to compute a metric for the inter-voxel relation of brain damage. ‘Anatomical bias’ vectors that were solely calculated from these inter-voxel relations in the patients’ real anatomical data, successfully predicted the VLBM misplacement. The latter has the potential to help in the development of new VLBM methods that provide even higher anatomical validity than currently available by the proper use of correction factors.

Introduction

To identify critical brain regions representing cognitive functions in the human brain early neuroscience had to rely on posthumous autopsy of individual brain damage (Broca, 1861; Wernicke, 1874). Today, modern imaging methods in combination with new statistical procedures allow to infer lesion-behaviour relationship at a group level. Voxel-based lesion-behavior mapping (VLBM) techniques with either parametric (Bates et *al.*, 2003) or non-parametric (Rorden et *al.*, 2007) statistics is frequently used for this purpose (overview cf. Table 1 in Karnath & Rennig, 2016). The central aspect of this inferential method is the attempt to control for regions that are not critical for the behavioral deficit under consideration; i.e. they aim to rule out regions of the brain that are simply vulnerable to damage and thus commonly damaged in stroke patients. The statistical procedure allowed numerous new insights and replaced the simple lesion overlap strategy, which included marked anatomical biases (cf. Rorden & Karnath, 2004).

One technical assumption of the VLBM method is statistical independence of all voxels, i.e. that the lesion status of a voxel is treated independently of the lesion status of adjacent voxels. In reality, however, the anatomy of stroke follows typical patterns that are defined by the vascular trees (Phan et *al.*, 2005; Lee et *al.*,2009; Sperber & Karnath, 2015). Two recent studies thus have assessed the localization accuracy of the VLBM method (Inoue et *al.*, 2014; Mah et *al.*, 2014). Both studies used a simulation approach based on large neurological patient samples with brain damage. They observed a bias within the lesion-deficit maps, displacing inferred critical regions from their true anatomical locations by about 16 mm towards areas of greater general lesion affection. Mah et *al.* (2014) speculated that "the pattern of mislocalization across the brain will depend on the complex interaction between the multivariate lesion distribution and brain functional architecture". They suggested to use novel machine learning techniques – such as multivariate pattern analysis (Smith et *al.*, 2013; Mah et *al.*, 2014; Zhang et *al.*, 2014) – that employ high-dimensional inference to accurately describe the true locus. Multivariate pattern analysis indeed appears to be an enrichment of modern lesion analysis to train and then test predictive models based on the pattern of damage to multiple regions (Karnath & Smith, 2014). However, this does not necessarily need to rule out the value of VLBM for certain scientific approaches per se.

In fact, the two previous simulation studies (Inoue et *al.*, 2014; Mah et *al.*, 2014) computed the VLBM analyses without the proper use of two commonly used correction factors, which might have led to underestimation of anatomical accuracy. Despite of the very large sample size included by Mah et *al.* (2014) the authors did not control for lesion size. However, for most behavioral deficits lesion size − independent from lesion location − is the best predictor for severity of the behavioral deficit; larger lesions are more likely to affect critical anatomical structures (Karnath et *al.*, 2004). If a sufficiently large dataset is available, VLBM studies of real data sets thus control this effect, typically by regressing out lesion size from the behavioral scores. The simulation study by Inuoe et *al.* (2014) indeed corrected for lesion size. Surprisingly, they found VLBM with a correction for lesion size to produce a larger bias than without correction. However, the study by Inuoe et *al.* (2014) was based on a lesion sample very different from the typical stroke samples used in VLBM studies of real data sets. The authors did not only include patients with stroke but also with other etiologies, such as e.g. encephalitis or surgical resections. It appears as if the proportion of non-stroke patients was very high in that the lesion overlay with frontal and fronto-temporal maxima markedly differed from the typical topography of unselected strokes with a maximum of overlap in the center of the territory of the middle cerebral artery (Phan et *al.*, 2005; Mah et *al.*, 2014; Sperber & Karnath, 2015). Thus, it remains to be tested in which way a VLBM study based on only stroke etiology is modified by a correction for lesion size.

A further discrepancy between the simulation study by Inuoe et al. (2014) and VLBM studies of real data sets is that the latter typically restrict statistical analysis to voxels that are affected by a certain proportion of lesions. This restriction to only voxels with 'sufficient lesion affection' prevents that results are biased by brain regions that are only rarely affected by stroke and thus do not carry sufficient information. In contrast to this common practice, the simulation study by Inuoe et *al.* (2014) did not control for this factor. In the study of Mah et *al.* (2014), 'sufficient lesion affection' was controlled with a criterion of n = 4, equivalent to 0.7% of the total sample. Real VLBM studies usually apply such correction in the range of $5 \leq n \leq 10$, equivalent to roughly 5%-10% of the whole sample (e.g. Goldenberg & Randerath, 2015; Mirman et al., 2015a; Tarhan et al., 2015; Timpert et al., 2015; Watson & Buxbaum, 2015).

Taken together, it remains an open question whether or not a VLBM bias occurs under the proper control for lesion size and for ‘sufficient lesion affection’ in a stroke patient sample. If indeed a considerable misplacement remains, it would be interesting to find out the origin of this bias. Mah et *al.* (2014) speculated that such bias might originate from systematic ‘parasitic’ voxel-voxel relations of collateral brain damage in the general anatomy of stroke and the lesion-deficit relation itself, which inevitably stays a black box in real settings. To clarify this question, we aimed to quantify the inter-voxel relations and experimentally test if these alone are able to predict the size of possible VLBM misplacement.

Methods

Patients with acute first unilateral, right hemispheric stroke admitted to the Centre of Neurology at the University of Tübingen were recruited. Patients with diffuse, bilateral, or cerebellar lesions, with tumors, marked anatomical distortion due to intracerebral hemorrhage, or patients without obvious lesion in MRI or spiral CT were excluded. A sample of 274 patients (mean age = 61.2 years; SD = 13.5) was recruited. Of these patients 233 had an infarct and 41 a hemorrhage. Patients or their relatives gave informed consent to participate in our study, which was performed according to the ethical standards laid down in the 1964 Declaration of Helsinki.

Brain lesions were demonstrated by MRI in 144 cases and by spiral CT in 130 cases. On average, imaging was acquired 4.5 days (SD = 7.4 days) after stroke onset. Binary lesion maps were created by manual delineation of lesion boundaries on axial slices of the patient’s individual scan using MRIcron (www.mccauslandcenter.sc.edu/mricro/mricron). For patients who underwent MR scanning, diffusion-weighted imaging (DWI) in the hyper acute stage until 48 hours after stroke onset and T_2-weighted fluid attenuated inversion recovery (FLAIR) imaging in later stages after stroke onset were used to delineate the lesions. If available, these scans were co-registered with a high-resolution T_1-weighted structural scan using SPM8 (www.fil.ion.ucl.ac.uk/spm). Brain scans were warped into MNI space with 1x1x1 mm^3 resolution by using SPM8 spatial normalization algorithms and the Clinical Toolbox (Rorden et *al.*, 2012), which provides age-specific templates both for MRI and CT scan normalization. Delineation of lesion borders and quality of normalization were verified by consensus of two experienced investigators.

Experiment 1: The spatial bias of VLBM in a realistic analysis setting

To investigate the performance of VLBM, two previous studies (Inuoe et *al*., 2014; Mah et *al*., 2014) used simulated 'behavioral' scores instead of the patients' real behavior to avoid circular reasoning. A priori, a so called 'truth model' which was thought to be the neural substrate of the simulated behavior was selected. The 'truth model' was defined by a brain region taken from a brain atlas (Inoue et *al.*, 2014; Mah et *al.*, 2014) or was even as simple as a single voxel (Mah et *al.*, 2014). Subsequently, an algorithm to compute continuous simulated 'behavioral' scores from damage to the truth model brain regions was implemented. Based on this algorithm 'behavioral' scores were calculated as a function of damage to these brain regions. As a final step, these truth model brain regions were compared to the voxel-wise, three-dimensional statistical map that was obtained in a VLBM analysis. The present experiment used a simulation procedure analogous to these previous simulation procedures. The aim was to test the impact of additional control for lesion size and for 'sufficient lesion affection'.

Simulation of 'behavioral' scores

To define our truth model, we chose the Automatic Anatomic Labelling atlas (AAL) (Tzourio-Mazoyer et *al.*, 2002) distributed with MRIcron, providing 45 right hemisphere cortical and subcortical regions. Since the AAL atlas is slightly larger (few voxels at the borders) than the templates used for normalization, overlaying voxels were manually removed from the AAL. In line with the two previous simulation studies (Inoue et *al.*, 2014; Mah et *al.*, 2014), we chose a simple algorithm to compute continuous simulated scores from damage to the truth model. This strategy appears convincing since (a) no realistic mathematic model of lesion-score relationship exists and (b) a simple simulation model should be affected by a bias genuine to VLBM the same way as a complex model. The simplest model to compute continuous scores based on damage to brain areas is a linear model, i.e. a model that computes 'behavioral' scores *s* as a linear function of the proportion of the damage to a truth model area *x*: $s(x) = a * x$.

The 'behavioral' scores were set between 0 (no deficit) and 100 (maximal deficit). For example, a patient without any damage to a given area received a simulated 'behavioral' score of 0 and a patient with 27% damage of all voxels in this area received a score of 27. For each of the 45 AAL regions we performed three simulation runs, each with randomly drawn samples of 100 lesions, resulting in 135 simulations per condition. Limiting sample size to 100 lesions allowed us to draw conclusions for real VLBM settings. The simulation was implemented using custom scripts in MATLAB 2009 and the 'Tools for NIFTI and ANALYZE image' toolkit (http://www.mathworks.com/matlabcentral/fileexchange/8797-tools-for-nifti-and-analyze-image).

Comparison of the statistical VLBM map and truth model

In order to reduce computational demands, we chose a mass-univariate t-test with false discovery rate (FDR) correction, implemented in Nii-Stat software (www.nitrc.org/projects/niistat/). This test is commonly used in modern VLBM studies and requires only minimal computational power. All statistical analyses were tested for a $p = .05$ level. One of our hypotheses was that limiting the analysis to only voxels with sufficient 'general lesion affection' should improve the performance of VLBM. In accordance with a widely accepted criterion, we defined the threshold for 'sufficient affection' as 5% of the whole sample. We contrasted the effect of data restriction to voxels with 'sufficient affection' to the procedure without this restriction by setting Nii-Stat to only test voxels at least damaged in $n = 5$ patients (equal to 5%) versus to test all voxels at least damaged in $n = 1$ patient. In particular, the latter condition has been used in the simulation study by Inuoe et al. (2014). However, if voxels with less than 5% lesion affection were still included in our simulation as a part of truth model brain regions − while excluding the same from the analysis − this would a priori cause inability of any lesion analysis method to identify the truth model. Therefore, we not only applied the 5%-criterion in the analysis as a correction factor, but we also introduced a further condition were we applied the 5%-criterion in the analysis *as well as* the simulation. For this condition, we simulated scores based on an alternative set of truth model brain regions that only covered aforementioned voxels. In detail, we created a modified version of the AAL by simply removing all voxels that did not fulfill the 5%-criterion. Seven regions were eliminated completely (supplementary motor area, medial superior frontal gyrus, orbital part of middle frontal gyrus, anterior cingulum, middle cingulum, posterior cingulum, paracentral lobule). This modified AAL offered a second, alternative set of truth models that considered 'sufficient lesion affection' already in the simulation.

Our second hypothesis was that controlling for lesion size improves VLBM performance. We thus carried out a VLBM analysis on each subsample once without controlling for lesion size and once with controlling for lesion size. To implement a control for lesion size we used the built-in default procedure of Nii-Stat: before the mass-univariate test is computed, lesion size is linearly regressed on the behavioral scores. Following this regression, we used the residuals for the actual VLBM analysis.

As dependent variables we included the same measurement of spatial misplacement defined by the centers of mass of truth model and statistical map as used in the two previous simulation studies by Mah et *al.* (2014) and Inoue et *al.* (2014). In detail, for each simulation step an a priori truth model region and a statistical map were available. For both these three-dimensional binary images the centre of mass was calculated and the Euclidean distance was measured. Additionally, we calculated 'sensitivity' (true positive rate: hits/(hits+false negatives)) and 'precision' (positive predictive value: hits/(hits+false positives)). The advantage of these parameters is that they do not rely on correct rejections, as these might be inflated due to the size of the image bounding box. In fact, the study by Inoue et *al.* (2014) found this parameter to be close to ceiling level across all groups.

Results

The sample of 274 lesions covered nearly the total right hemisphere and 770556 voxels were damaged in at least one patient (Fig. 1A). A majority of lesions lay in the territory of the middle cerebral artery with a centre of affection around putamen and insula. The topography closely resembled the one on stroke patients provided in the supplementary material in Mah et *al.* (2014). Of these 770556 voxels 81.4% were covered by at least 5% of all lesions (equivalent to 14 lesions) (Fig. 1B). Our two hypotheses were tested in a 3x2 design, with factors 'control for sufficient lesion affection' (not controlled with $n = 1$ criterion; controlled in the analysis only with $n = 5$ criterion; controlled in analysis *and* simulation with $n = 5$ criterion) and 'control for lesion size' (controlled; not controlled). Over 97% of all VLBM analyses yielded significant results and were included into the final analysis. As both factors were only partially paired, a repeated measure ANOVA could not be computed. Therefore, as often done in this situation (e.g., Samawi & Vogel, 2013), we here calculated and report results of an independent ANOVA.

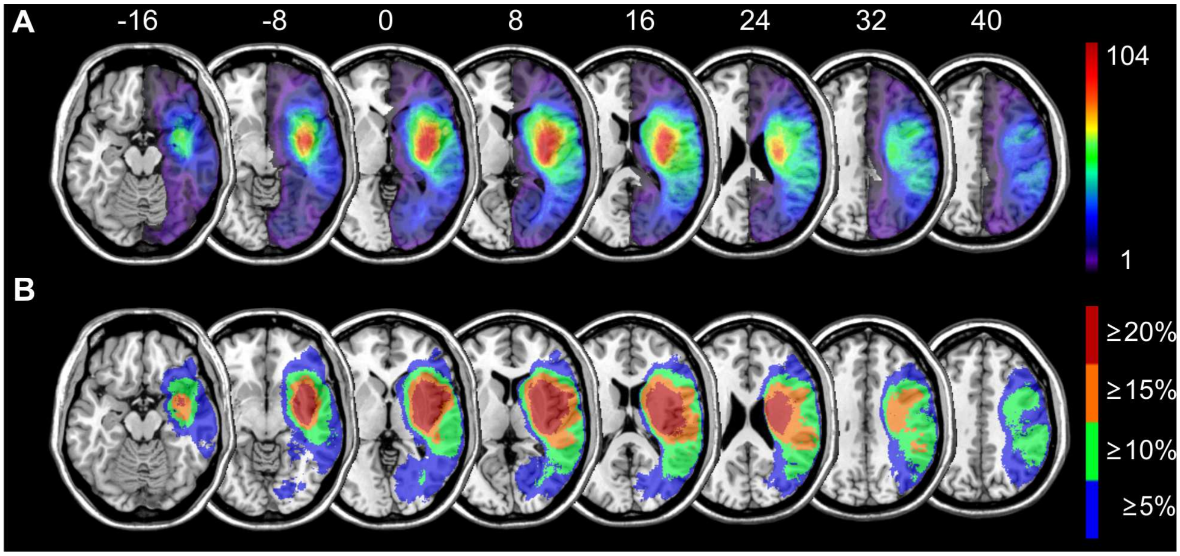


Figure 1: Topography of brain lesions

Lesion topography for all 274 patients with (A) continuous color scaling and (B) with an alternative step-wise color scaling to show all voxels that were damaged in at least x % of all patients. Numbers above the slices indicate z-coordinate in MNI space.

In addition, we performed a repeated measures ANOVA only using available paired data; all significant results of the repeated measures ANOVA turned out to be significant again and thus are not reported here. In case of significant effects, Bonferroni-corrected post-hoc tests were calculated. Averaged over all groups, the misplacement was 18.6 mm (*SD* = 11.3 mm). The ANOVA revealed that misplacement was affected by 'control for sufficient lesion affection' (*F*(2,739) = 14.02; *p* < .001) and 'control for lesion size' (*F*(1,739) = 88.73; *p* < .001) (Fig. 2A). Both factors did not interact (*F*(2,739) = 1.55; *p* = .21). Post-hoc tests showed that misplacement was lower if VLBM analyses were controlled for lesion size and for 'sufficient lesion affection' both in the analysis and simulation. Under these conditions, misplacement was reduced to 11.5 mm (*SD* = 6.3 mm). Sensitivity was generally high (*Sens.* = .73; *SD* = .33) (Fig. 2B). Factor 'control for sufficient lesion affection' had a significant impact on sensitivity (*F*(2,739) = 130.88; *p*< .001) with the *n* = 5 criterion in both simulation and analysis outperforming the other groups. Factor 'control for lesion size' neither affected sensitivity as a main effect (*F*(1,739) = 1.21; *p* = .27) nor in an interaction with 'control for sufficient lesion affection' (*F*(2,739) = 0.67; *p* = .51). 'Precision' was generally very low (*prec.* = 0.12; *SD* =

.11) and was affected both by ‘control for lesion size’ ($F(1,739) = 186.84$; $p < .001$) and ‘control for sufficient lesion affection’ ($F(2,739) = 8.09$; $p < .001$) (Fig. 2C). Again, the interaction was not significant ($F(2,739) = 0.73$; $p = .48$). ‘Control for lesion size’ improved ‘precision’; post-hoc tests revealed that ‘control for sufficient lesion affection’ with the $n = 5$ criterion in both simulation and analysis was inferior to the general $n = 1$ criterion. These two groups did not significantly differ from the condition with the $n = 5$ criterion in the analysis only.

The simulated behavioral scores correlated with lesion size both in the condition with the full AAL simulation (average correlation r = .43; SD = .21) and with the modified AAL simulation (for ‘sufficient lesion affection’; average correlation r =.47; SD = .22). This is in the range of behavior-lesion size correlations in real patient data, that may range from low, non-significant correlations to high correlations of r = .7 (e.g., Kertesz & Ferro, 1984; Brott et al., 1988; Wittmann et al., 2004). The average peak t-values of statistical maps were t = 8.51 (SD = 0.72) for all simulations and t = 7.99 (SD = 0.62) for simulations both controlled for lesion size and ‘sufficient lesion affection’ (see Fig. 3 for example t-maps). Thus, our simulation-based peak t-values were in the high upper range of peak t-values in real VLBM studies (e.g., Verdon et al., 2010).

Discussion

Simulation experiment 1 revealed that misplacement of VLBM results can be minimized by the use of lesion size as a covariate and the exclusion of voxels with low lesion affection. Under these conditions, the misplacement could be reduced by 48% compared to uncorrected VSLM, adding up to only 11.5 mm (Fig. 2a). The following experiment should clarify whether this bias is due to natural stroke anatomy determined by the vascular architecture and systematically biased inter-voxel relations of collateral damage, i.e. if it represents an ‘anatomical bias’.

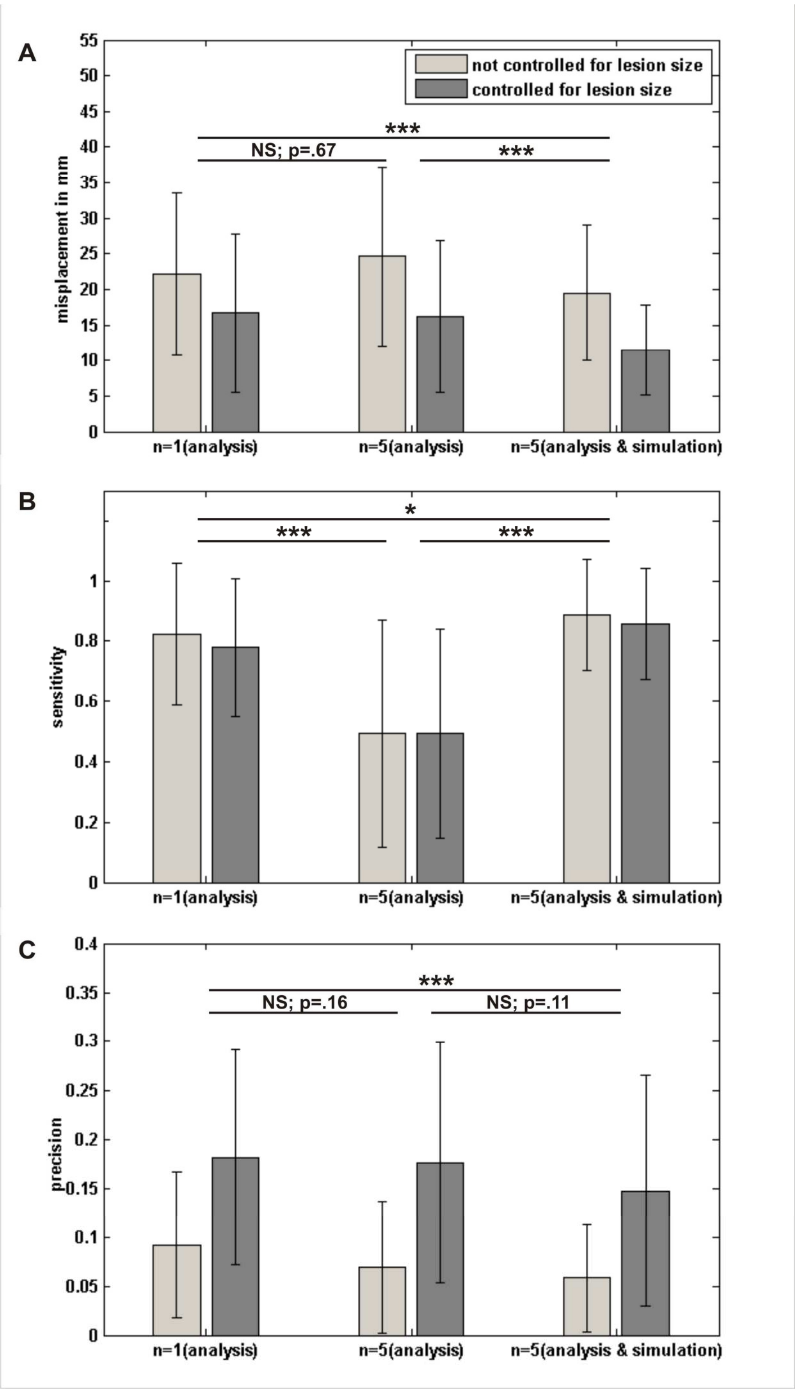


Figure 2: Effects of factors 'control for lesion size' and 'control for sufficient lesion affection' in VLBM analysis

Results of the 3x2 ANOVA conducted in experiment 1 addressing the effects of factors 'control for lesion size' and 'control for sufficient lesion affection' in VLBM for (A) misplacement, (B) sensitivity, and (C) precision. Error bars represent standard deviation. Asterisks indicate significance in post-hoc tests on the effects of 'sufficient lesion affection' control (*p < .05, ***p < .001).

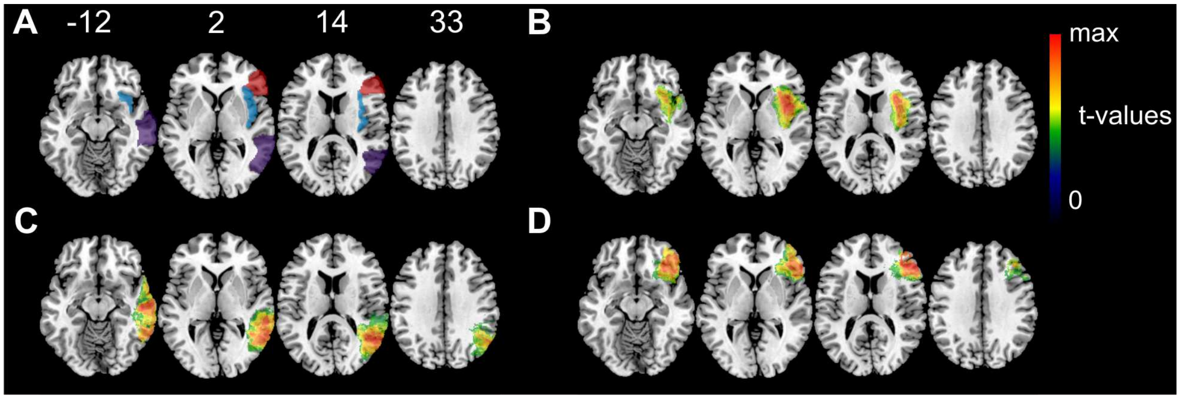


Figure 3: Example t-maps

For three regions of interest example t-maps from experiment 1 are shown. All maps originate from the condition with both control for lesion size and 'control for sufficient lesion affection' in simulation and analysis. (A) three regions of interest taken from the AAL atlas: insula (blue), middle temporal gyrus (purple), and inferior frontal gyrus, triangular (red) (B) VLBM results for the insula with t(max) = 7.35 (C) VLBM results for the middle temporal gyrus with t(max) = 7.91 (D) VLBM results for the inferior frontal gyrus, triangular with t(max) = 8.04. Color coding in B-D indicate t-values thresholded to only show voxels with significant t-values p < .05. Numbers above the slices indicate z-coordinate in MNI space.

Experiment 2: Effects of stroke anatomy on VLBM

Mah et *al.* (2014) computed voxel-wise VLBM misplacement vectors, i.e. vectors based on the misplacement of statistical VLBM results compared to a truth model region/voxel, indicating that such results were systematically biased. In contrast, we here aimed to calculate voxel-wise vectors based on the patients' anatomical data, i.e. on the data before any statistical analyses were applied. Therefore, we generated voxel-wise topographies of collateral damage in the real data and used them to compute a metric for the inter-voxel relation of brain damage. If indeed inter-voxel relations should be the cause for the misplacement in lesion mapping, the VLBM-misplacement vectors observed by Mah et *al.* (2014) should be reliably predictable by our 'anatomical bias' vectors based on anatomy.

A voxel-wise vector for 'anatomical bias'

For each simulation run, first a voxel ('truth model voxel') was chosen. To define the target point of an anatomical bias vector, i.e. a centre of anatomical affection, we identified all lesions that included this truth model voxel to create a 'voxel overlay'

(Fig. 4A). This topography already offers information on the anatomy of stroke that affects the truth model voxel. However, it neglects lesions that do not include the truth model voxel. Therefore, we calculated an element-wise division of the voxel overlay divided by the overlay of the whole 274 patient sample (Fig. 1A) to produce topographies of inter-voxel relation. This results in a single topography for each chosen truth voxel individually (Fig. 4B). The values in this topography indicate how many lesions that lie in any voxel also include the truth voxel. The proportional values vary between 0 (0% of all lesions in this voxels also contain the truth model voxel) and 1 (100% of all lesions in this voxel also contain the truth model voxel). For example, if in the topography for a certain truth model voxel any voxel contains the value 0.27, this means that 27% of all lesions in this voxel also damaged the truth model voxel. To prevent a high impact of voxels that are generally rarely affected by stroke and to stay close to the study by Mah et *al.* (2014), we limited this analysis to voxels that were damaged in at least 4 patients. The centre of mass of this topography was identified (Fig. 4B) and used to define a vector of 'anatomical bias' (purple vector in Fig. 4D). Due to high computational demands, this analysis was not carried out for the whole brain, but for 100 randomly chosen voxels.

A voxel-wise vector for misplacement

The creation of a voxel-wise VLBM misplacement vector was implemented analogous to the study by Mah et *al.* (2014). Given a truth model voxel, a binary 'behavioral' score was simulated. If the lesion of a patient also included damage to the truth model voxel, the patient received a behavioral score of '1' (present deficit); else he received a '0' (no deficit). These 'behavioral' scores were used in a lesion analysis on the whole 274 patients data sample in Nii-Stat, using the Liebermeister test (Rorden et *al.*, 2007) and FDR correction. Only voxels damaged in at least 4 patients were investigated. This analysis of simulation data yielded a binary statistical map for each truth model voxel (Fig. 4C). The misplacement was defined as the vector from the truth model voxel to the centre of mass of this statistical map (green vector in Fig. 4D).

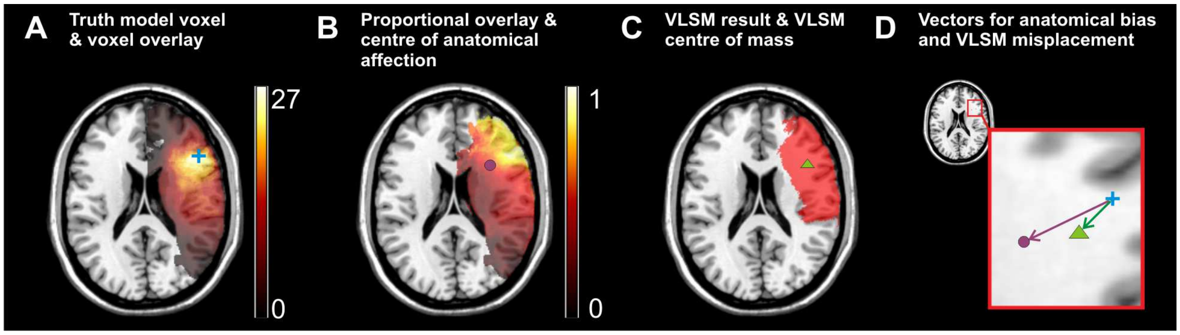
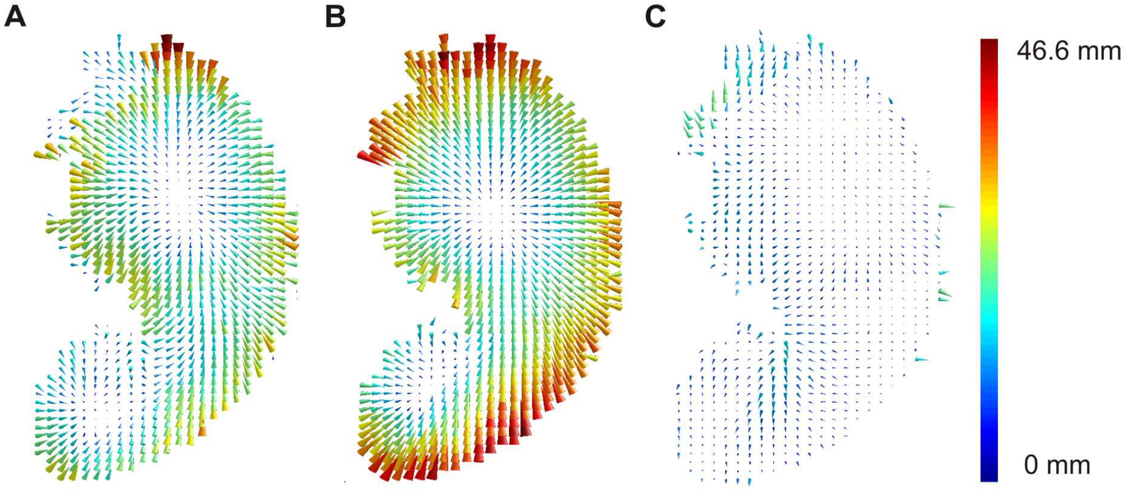

A
Truth model voxel & voxel overlay
27
0
B
Proportional overlay & centre of anatomical affection
1
0
C
VLSM result & VLSM centre of mass
D
Vectors for anatomical bias and VLSM misplacement

Figure 4: Example for the computation of voxel-wise anatomical bias and VLBM misplacement

For one exemplary truth model voxel (MNI coordinates x=47, y=24, z=20), the procedures in experiment 2 are illustrated. (A) All lesions that include the chosen truth model voxel (blue cross) are identified to create a 'voxel overlay'. (B) For each voxel damaged in at least 4 patients the 'voxel overlay' is element-wisely divided by the total overlay of all 274 patients (see Fig. 1A) to produce a topography of inter-voxel relation. The centre of mass of the resulting topography (purple circle) offers a voxel-wise centre of anatomical affection. (C) The truth model voxel is used to simulate a binary 'behavioral' deficit. A lesion analysis computes a statistical map (red area) and the centre of mass of this map (green triangle) provides the centre of VLBM results. (D) The previously defined coordinates and the truth model voxel (blue) are used to define a vector of 'anatomical bias' (purple arrow) and a vector of misplacement (green arrow). All illustrations are shown on slice z=20. Note that for the present figure the resulting centers are projected back to same z-slice for illustration purposes.

Results

The 100 randomly chosen voxels were damaged in at least 13 and maximally 96 of all 274 patients (mean = 32.7; *SD* = 18.1). For each of these voxels we calculated the VLBM misplacement vectors (Fig. 5A) and anatomical misplacement vectors (Fig. 5B). On average the anatomical misplacement vector was 25.7 mm (*SD* = 7.7mm) long. By using a FDR correction at *p* = .05, the average VLBM misplacement vector for the same voxels was 18.3 mm (*SD* = 6.1 mm) long and thus significantly smaller than the anatomical misplacement vectors (*t*(99) = 16,66; *p*< .001). As the length of the VLBM misplacement vectors depended solely on false alarms – and thus on how conservative a test is – we ran a second simulation on the same voxels, but with a FDR correction at p = .01. For this more conservative test, the misplacement was 16.3 mm (*SD* = 5.5 mm) and significantly lower than with the less conservative test (*t*(99) = 15,98; *p* < .001). The length of vectors for VLBM misplacement and for anatomical bias correlated highly both for FDR correction at *p* = .05 (Pearson´s *R* = .82; p < .001) and *p* = .01 (*R* = .76; p < .001). To measure directional similarity, we computed the cosine similarity that ranges between 1 if two vectors have the same direction and -1 if they point into the opposite direction. If two vectors are exactly orthogonal, cosine similarity is 0. Cosine similarity was $cos(\theta)$ = .91 (*SD* = 0.12) for p = .05 and $cos(\theta)$ = .88 (*SD* = 0.15) for p = .01. Thus, although anatomical vectors were significantly larger than misplacement vectors, both sets of vectors appeared to be highly similar.



[Figure 5 near here]

Figure 5: Vector maps for 'anatomical bias', VLBM misplacement, and corrected VLBM misplacement

The vector graphics visualize the results of experiment 2 exemplarily for slice z=17. (A) Vector map for the misplacement of statistical VLBM results at p=.05. (B) Vector map for 'anatomical bias'. Voxel-wise vectors here were based on the inter-voxel relation in the anatomical data, i.e. on the data before any statistical analyses were applied. (C) Vector map for 'corrected misplacement vectors' using the minimization factor k=0.6495. For illustration purposes, the length of the vectors does not show the real vector length, but is scaled using the same factor in all graphics. Color-coding indicates the length of the vectors in mm.

Considering the similarity of the vectors and the assumption, that an 'anatomical bias' is the reason for a VLBM misplacement, one should be able to predict VLBM misplacement with the 'anatomical bias' and thus correct the VLBM misplacement. Therefore, we computed the position vector of the VLBM centre of mass and subtracted the 'anatomical bias' vector, i.e. we corrected the VLBM centre by the information provided from inter-voxel relation of brain damage in the anatomical data. The distance between the truth model voxel and this new corrected centre of

VLBM results was expressed as ‘corrected misplacement vectors’. On average the corrected misplacement vector was 11.2 mm (*SD* = 3.2 mm) long for *p* = .05 and 13.1 mm (*SD* = 4.0 mm) long for *p* = .01. Given the different sets of misplacement vectors for varying p-levels and the larger vectors for anatomical misplacement, we expected lower ‘corrected misplacement vectors’ for more optimal correction with vectors individualized for the chosen p-level. Therefore, for both p-levels, we looked at every pair of misplacement vector $\overrightarrow{\mathrm{m}}$ and anatomical bias vector $\overrightarrow{\mathrm{ab}}$ and searched via minimization function for a factor k for which the corrected misplacement $c = |\vec{m} - k * \overrightarrow{ab}|$ was minimal. For a significance level of *p* = .05, the corrected misplacement was minimized by an average factor of *k* = 0.6495. This minimization factor was applied to the ‘corrected misplacement vectors’ (Fig. 5C). On average, these ‘corrected misplacement vectors’ had a length of 6.8 mm (*SD* = 2.9 mm), which was a significant improvement compared to the uncorrected misplacement (*t*(99) = 19,48; *p*< .001). For a significance level of *p* = .01 we found *k* = 0.5655 to be the average optimal factor that significantly reduced uncorrected misplacement to 7.0 mm (*SD* = 3.0 mm) (*t*(99) = 17,38; *p*< .001. Cosine similarity between the original misplacement vector and this corrected misplacement vector was *cos(θ)* = .34 (*SD* = 0.42) for *p* = .05 and *cos(θ)* = .40 (*SD* = 0.41) for *p* = .01.

Discussion

Experiment 2 tested if VLBM misplacement can be predicted by its underlying stroke anatomy. In fact, we revealed that the ‘anatomical bias’ based on the inter-voxel relation affected the VLBM results. In other words, measurable aspects of stroke anatomy indeed appear to be the source of VLBM misplacement. The VLBM-misplacement vectors observed by Mah et *al.* (2014) thus can be reliably predicted by our ‘anatomical bias’ vectors based on anatomy.

General discussion

The concept of a correction for lesion size by linear regression has recently been criticized on a theoretical level (Nachev, 2015): as lesion size varies with anatomical location, it was argued that the correction would confound the anatomical interference and could even amplify the misplacement of VLBM results. In contrast to this assumption, we here observed in a large sample of stroke patients that lesion size in fact has a significant impact on VLBM accuracy. A closer look at the inter-voxel relation explains this effect: larger lesions inflate the number of ‘parasitic’ inter-voxel relations over long distance (see Fig. 4A/4B) and thus enlarge the bias in VLBM. Beyond, the present simulation demonstrated that the VLBM misplacement is reduced by controlling for rarely affected brain areas (‘control for sufficient lesion affection’). In combination, the use of factors ‘correction for lesion size’ and ‘sufficient lesion affection’ markedly reduced the misplacement of VLBM results compared to uncorrected VSLM. The two variables reduced VLBM misplacement in an additive manner, i.e. both correction factors independently improved VLBM accuracy.

The correction factors ‘lesion size’ and ‘sufficient lesion affection’ also increased variables ‘sensitivity’ and ‘precision’. Variable ‘sensitivity’ was very high in general, thus the actual anatomical correlate of a simulated behavior was correctly identified together with a high number of false alarms that were spatially oriented in the direction of the misplacement. The operationalisation of ‘misplacement’ used in the present as well as the two previous simulation studies (Inuoe et al., 2014; Mah et al., 2014) thus could be criticized, as the simple Euclidean distance between two centers of mass omits such information and can result from an infinite number of different configurations that can differ in sensitivity, precision etc. This problem is underlined by the fact that many VLBM studies provided results that were not located primarily in subcortical structures but rather at cortical grey matter regions (e.g., Karnath et *al.*, 2004; Kalénine et *al.*, 2010; Karnath et *al.*, 2011; Manuel et *al.*, 2013; Mirman et *al.*, 2015), although a pure misplacement effect should shift cortical structures towards the centre of the vascular territories.

Although the control for ‘sufficient lesion affection’ improved performance of VLBM analyses, it is important to note that this method at the same time limits VLBM analyses. In the literature, VLBM studies usually provide a simple overlay topography of all lesions and display results on a template for the whole brain. The fact that such studies actually did not test parts of the brain is often not referred to explicitly. However, VLBM analyses self-evidently do not provide any information about brain areas that are not tested, i.e. that fall below the criterion for ‘sufficient lesion affection’. Therefore, we included the non-realistic experimental condition with control for ‘sufficient lesion affection’ in analysis and simulation into experiment 1. This condition simulated behavioral scores only based on areas that were above the criterion for ‘sufficient lesion affection’. With this condition, experiment 1 has shown that control for ‘sufficient lesion affection’ improves performance of VLBM within the area of tested voxels. At the same time, the condition with control for ‘sufficient lesion affection in the analysis only’ has shown that this correction also impairs VLBM if related to the whole brain. While misplacement was not significantly affected, sensitivity was decreased. This is not surprising, as positive signals could not be identified in voxels that were not tested and misses thus were inflated. To conclude, limiting VLBM for ‘sufficient lesion affection’ trades in spatial extent of the analysis (i.e. less voxels are tested) for a more valid VLBM performance in voxels that are tested. This conclusion can be transferred to real VLBM studies. Contrary to the present condition with ‘sufficient lesion affection in simulation and analysis’, in real VLBM studies brain regions relevant to behavior might also lie in brain areas that are not tested, i.e. that are removed from the analysis due to correction for ‘sufficient lesion affection’. Such areas thus should be considered as a black box that still could contribute to behavior. Following this principle, the condition with 'sufficient lesion affection in simulation and analysis' in our present study is transferrable to real VLBM studies.

The misplacement of statistical VLBM maps apparently is due to physiological effects of brain lesion anatomy. Lesion anatomy here includes the lesion-deficit relationship as well as the inter-voxel relation. While the lesion-deficit relationship describes the relationship between the lesion of a certain region and its behavioral consequences, inter-voxel relation is the voxel-wise topographies of collateral damage. In a simple simulation setting that was comparable to the simple simulation settings in the two previous studies (Inuoe et al., 2014; Mah et al., 2014), we successfully corrected the VLBM misplacement by ‘anatomical bias’ vectors, solely calculated from inter-voxel relations in the patients’ real anatomical data. However, in the present as well as in the two previous simulation studies (Inuoe et al., 2014; Mah et al., 2014) the lesion-deficit relationship only played an intermediate role as it was used to compute simulated ‘behavioral’ scores – based on a stroke anatomy with systematically biased inter-voxel relations. A systematic bias in the lesion-deficit relationship itself (e.g., higher impact of subcortical voxels inside a single truth model region on simulated behavioral scores) was not introduced. Thus, the biased inter-voxel relation alone was the reason for VLBM misplacement. Furthermore, magnitude of VLBM misplacement was affected by the VLBM’s p-level, however the correlation between misplacement and ‘anatomical bias’ was high in both tested p-values.

In VLBM studies of real data sets it is unlikely that systematically biased lesion-deficit relations itself generally contribute to the VLBM misplacement. The black box of lesion-deficit relations rather plays a mediating role, as it determines the severity of a deficit based on lesions that suffer from biased inter-voxel relations. Given that the inter-voxel relation data is the main source of VLBM misplacement, magnitude and direction of VLBM misplacement could be estimated and new correction algorithms that even further improve validity of VLBM results are imaginable. A possibility for such prospective correction could be an anatomical parcellation atlas that incorporates the anatomy of stroke and the underlying inter-voxel relation – at the cost of data resolution compared to a voxel-wise analysis. Also, the development of retrospective correction algorithms is imaginable. By using a valid anatomical reference sample such algorithms could be applied post-hoc on previous VLBM studies. As experiment 2 was based on a simple simulation model and 'anatomical bias' depended on VLSM parameters, more complex algorithms will be required for such purpose. Possible candidates for such corrections that are able to identify spurious results vs. true results are, e.g., voxel-wise vectorial algorithms or a seed-based approach that directly uses the inter-voxel relations (analogous to, e.g., resting state analyses [Fox & Raichle, 2009]).

To conclude, the misplacement bias in VLBM results is in fact much smaller if appropriate correction factors are used. Although such correction might be biased by the variability of lesion size across the brain (Nachev, 2015), positive effects obviously prevail. The misplacement appears to be due to physiological effects of brain lesion anatomy. The latter has the potential to help in the development of new VLBM methods providing even higher validity than currently available by the proper use of correction factors.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (KA 1258/23-1). Christoph Sperber was supported by the Friedrich Naumann Foundation. The authors would like to thank David V. Smith for his valuable and stimulating comments on the manuscript. We also thank the staff of the Division of Neuropsychology at Tübingen University who helped to collect and pre-process the imaging data; in particular, Bianca de Haan, Johannes Rennig, Urszula Mihulowicz, Julia Suchan, and Dongyun Li.

References

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF (2003): Voxel-based lesion-symptom mapping. Nat. Neurosci. 6:448–50.

Broca P (1861): Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). Bulletin de la Société Anatomique 6: 330–57.

Brott T, Marler JR, Olinger CP, Adams HP, Tomsick T, Barsan WG, Biller J, Eberle R, Hertzberg V, Walker M (1989): Measurements of acute cerebral infarction: lesion size by computed tomography. Stroke 20(7):871-5.

Fox MD, Raichle ME (2007): Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci. 8:700–711.

Goldenberg G, Randerath J (2015): Shared neural substrates of apraxia and aphasia. Neuropsychologia 75:40–49.

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.

Kalénine S, Buxbaum LJ, Coslett HB (2010): Critical brain regions for action recognition: Lesion symptom mapping in left hemisphere stroke. Brain 133:3269–3280.

Karnath H-O, Fruhmann Berger M, Küker W, Rorden C (2004): The anatomy of spatial neglect based on voxelwise statistical analysis: a study of 140 patients. Cereb. Cortex 14:1164–72.

Karnath H-O, Rennig J, Johannsen L, Rorden C (2011): The anatomy underlying acute versus chronic spatial neglect: a longitudinal study. Brain 134:903–12.

Karnath H-O, Smith DV (2014): The next step in modern brain lesion analysis: multivariate pattern analysis. Brain 137:2405–7.

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

Kertesz A., Ferro JM (1984): Lesion size and location in ideomotor apraxia. Brain 10, 921–33.

Lee E, Kang D-W, Kwon SU, Kim JS (2009): Posterior cerebral artery infarction: diffusion-weighted MRI analysis of 205 patients. Cerebrovasc. Dis. 28:298–305.

Mah Y-H, Husain M, Rees G, Nachev P (2014): Human brain lesion-deficit inference remapped. Brain 137, 2522-31.

Manuel AL, Radman N, Mesot D, Chouiter L, Clarke S, Annoni J-M, Spierer L (2013): Inter- and intrahemispheric dissociations in ideomotor apraxia: a large-scale lesion-symptom mapping study in subacute brain-damaged patients. Cereb. Cortex 23:2781–9.

Mirman D, Chen Q, Zhang Y, Wang Z, Faseyitan OK, Coslett HB, Schwartz MF (2015): Neural organization of spoken language revealed by lesion–symptom mapping. Nat. Commun. 6:6762.

Nachev P (2015): The first step in modern lesion-deficit analysis. Brain 138:e354.

Phan TG, Donnan G a, Wright PM, Reutens DC (2005): A digital map of middle cerebral artery infarcts associated with middle cerebral artery trunk and branch occlusion. Stroke 36:986–91.

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. 5:813–9.

Rorden C, Karnath H-O, Bonilha L (2007): Improving lesion-symptom mapping. J. Cogn. Neurosci. 19:1081–8.

Rorden C, Bonilha L, Fridriksson J, Bender B, Karnath HO (2012): Age-specific CT and MRI templates for spatial normalization. Neuroimage 61:957–965.

Samawi, H. M., & Vogel, R (2013): Notes on two sample tests for partially correlated (paired) data. Journal of Applied Statistics 41(1), 109–117.

Smith D V, Clithero JA, Rorden C, Karnath H-O (2013): Decoding the anatomical network of spatial attention. Proc. Natl. Acad. Sci. 110:1518–23.

Sperber C, Karnath H-O (2015): Topography of acute stroke in a sample of 439 right brain damaged patients. NeuroImage Clin. 10:124–128.

Tarhan LY, Watson CE, Buxbaum LJ (2015): Shared and Distinct Neuroanatomic Regions Critical for Tool-related Action Production and Recognition: Evidence from 131 Left-hemisphere Stroke Patients. J. Cogn. Neurosci. 27:2491–511.

Timpert DC, Weiss PH, Vossel S, Dovern A, Fink GR (2015): Apraxia and spatial inattention dissociate in left hemisphere stroke. Cortex 71:349–358.

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.

Verdon V, Schwartz S, Lovblad KO, Hauert CA, & Vuilleumier P (2010): Neuroanatomy of hemispatial neglect and its functional components: a study using voxel-based lesion-symptom mapping. Brain. 133, 880–94.

Watson CE, Buxbaum LJ (2015): A distributed network critical for selecting among tool-directed actions. Cortex. 65:65–82.

Wernicke C (1874): Der aphasische Symptomencomplex. Breslau: Cohn und Weigart.

Wittmann M, Burtscher A, Fries W, von Steinbüchel N (2004): Effects of brain-lesion size and location on temporal-order judgment. Neuroreport 15(15):2401-2405.

Zhang Y, Kimberg DY, Coslett HB, Schwartz MF, Wang Z (2014): Multivariate lesion-symptom mapping using support vector regression. Hum. Brain Mapp. 5876:5861–5876.

An empirical evaluation of multivariate lesion behaviour mapping

Christoph Sperber[1], Daniel Wiesen[1], & Hans-Otto Karnath[1,2]

[1]Centre 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

Abstract

Multivariate lesion behaviour mapping based on machine learning algorithms has recently been suggested to complement the methods of anatomo-behavioural approaches in cognitive neuroscience. Several studies applied and validated support vector regression-based lesion symptom mapping (SVR-LSM) to map anatomo-behavioural relations. However, this promising method, as well as the multivariate approach per se, still bears many open questions. By using large lesion samples in three simulation experiments, the present study empirically tested the validity of several methodological aspects. We found that i) correction for multiple comparisons is required in the current implementation of SVR-LSM, ii) that sample sizes of at least 100 to 120 subjects are required to model voxel-wise lesion location in SVR-LSM, and iii) that SVR-LSM is susceptible to misplacement of statistical topographies along the brain's vasculature to a similar extent as mass-univariate analyses.

1 Introduction

Studies on patients with focal brain lesions are a main source of our knowledge on the anatomo-behavioural architecture of the brain (Rorden & Karnath, 2004). For statistical analysis of lesion anatomy, different approaches of voxel-based lesion behaviour mapping (VLBM) have been implemented (Bates et al., 2003; Rorden et al., 2007). The main idea behind VLBM is to test each brain voxel individually if damage to the voxel is associated with a certain behavioural measure. As this is performed by computing a univariate test at each voxel, VLBM has also been termed a ‘mass-univariate’ approach.

While over a hundred studies have utilised VLBM so far (see Karnath & Rennig, 2017), the mass-univariate approach − like all modern neuroimaging techniques − has limitations (Karnath et al., 2018). The central problem of mass-univariate analyses is the fact that multiple univariate tests are per se independent. The assumption of independence, however, does not appear to be appropriate in investigating lesion-deficit relations. First, brain functions are not organised in single voxels, but in larger anatomical modules or networks. Second, stroke lesions do not damage the brain in a voxel-wise, independent manner, but – due to vasculature – systematically with typical patterns of collateral damage.

Previous studies have addressed these issues empirically. Simulation studies have shown that VLBM might fail to identify cognitive modules organised in a network (Mah et al., 2014; Zhang et al., 2014; Pustina et al., 2018). The underlying problem has been termed the ‘partial injury problem’ (Kinkingnéhun et al., 2007; Rorden et al., 2009; but see also Pustina et al (2018) for possible issues besides the ‘partial injury problem’). This problem appears in VLBM if a cognitive module is only partially injured by lesions, e.g. if only parts of a brain network are damaged. Statistical power can then be reduced, and VLBM might fail to identify the cognitive module in parts or in whole (see Fig. 1 for an illustration).

Other simulation studies have identified a misplacement of VLBM results towards the centres of the arterial territories (Inoue et al., 2014; Mah et al., 2014). This bias originates from systematic collateral damage between voxels, i.e. from high correlation/dependence of lesion status between voxels (Sperber & Karnath, 2017).

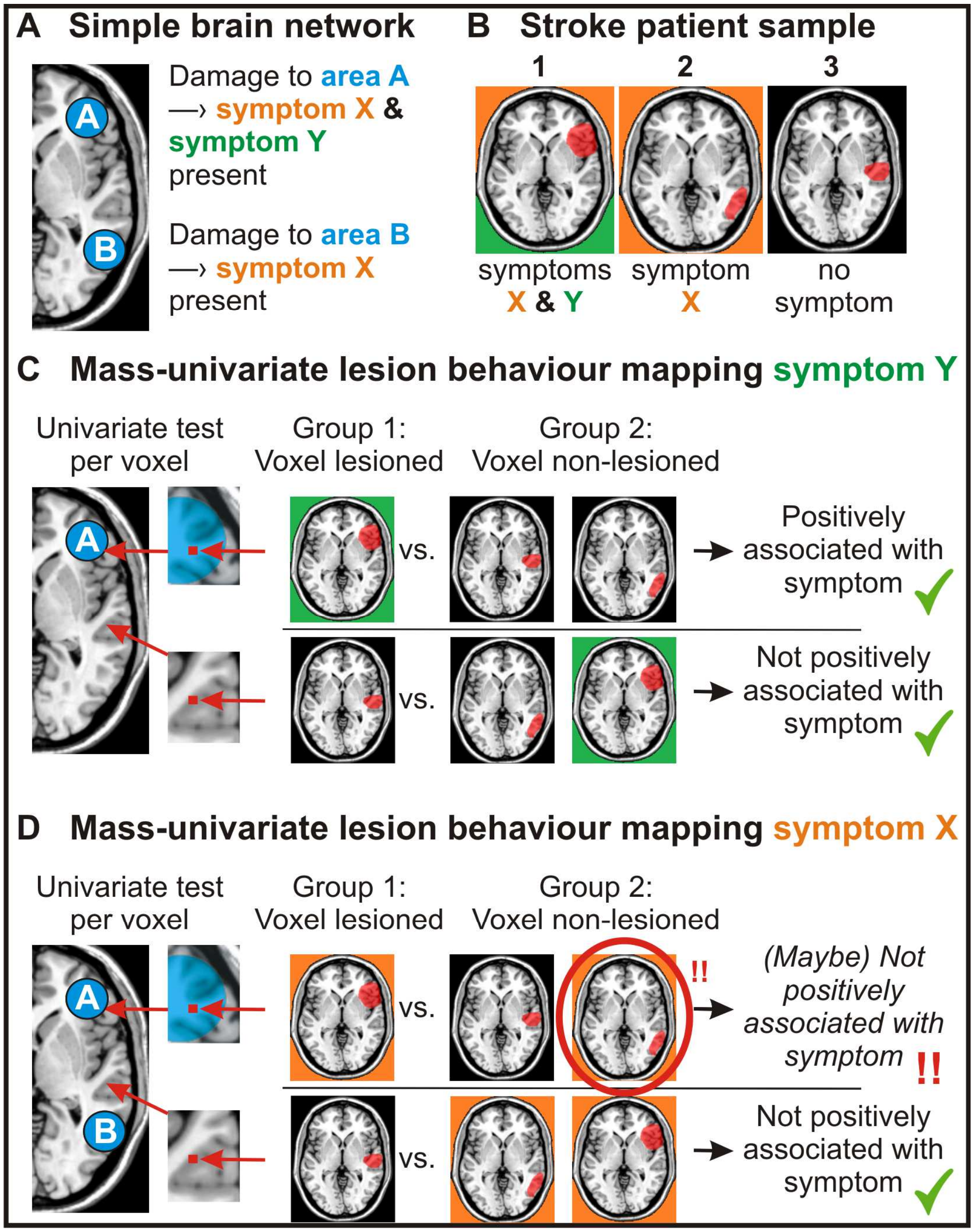


Figure 1: The 'partial injury problem'

Illustration of the 'partial injury problem' in mass-univariate lesion behaviour mapping. **(A)** A simple fictional brain network consisting of two nodes. Damage to either node causes the same symptom X, while only damage to area A induces symptom Y. **(B)** A stroke sample of three patients. Note that the neural correlates of symptom X are *partially injured* in patients 1&2 **(C)** Mass-univariate lesion behaviour mapping of symptom Y shown for two example voxels. Following the mass-univariate VLBM approach, for each voxel patients with damage to this voxel (Group 1) are statistically tested against patients without damage to this voxel (Group 2). Voxels are considered to be associated with a symptom if Group 1 is significantly associated with a more severe symptom. For symptom Y, where damage to the brain module is either complete or not present at all, VLBM results will be correct. **(D)** Mass-univariate lesion behaviour mapping of symptom X. Here, statistical power is decreased because patients with present symptoms due to lesions in other voxels (red circle) serve as counter-examples. This can reduce the ability of mass-univariate analyses to correctly identify brain networks or large neuroanatomical modules in a whole.

To overcome these issues of mass-univariate analyses, multivariate lesion behaviour mapping (MLBM) has been suggested (Smith et al., 2013; Mah et al., 2014; Zhang et al., 2014; review in Karnath et al., 2018). In MLBM, behaviour is modelled in one single model based on the lesion status of multiple voxels or regions of interest. This can be achieved by using machine learning algorithms such as support vector machines, including support vector regression (SVR; Vapnik, 1995). Several simulation studies have shown that MLBM is indeed superior to VLBM in detecting brain networks (Mah et al., 2014; Zhang et al., 2014; Pustina et al., 2018).

While it seems that MLBM is able to overcome the partial injury problem, it has not been investigated yet, how much MLBM is susceptible to misplacement due to collateral damage between voxels. A recent study found that misplacement in multivariate analyses is low compared to some VLBM approaches (Pustina et al., 2018). However, it was not investigated if the remaining misplacement occurs spatially random or if it still occurs systematically along the brain's vasculature.

Another open question concerns the sample sizes required for MLBM. Multivariate models naturally contain a large number of variables. Therefore, MLBM might require much larger sample sizes for parameter estimation than VLBM. A recent study investigated the performance of VLBM and MLBM at different sample sizes, and MLBM was found to be equal or even superior to VLBM also with smaller sample sizes (Pustina et al., 2018). But still, it is not known how many subjects are required to obtain a 'good' multivariate model.

A third issue of discussion relates to the way statistical inference is computed in MLBM. Until now, the most often used multivariate method is based on support vector regression (SVR-LSM; Zhang et al., 2014, Mirman et al., 2015b; Fama et al., 2017; Griffis et al., 2017; Ghaleh et al., 2018; Wiesen et al., submitted). SVR-LSM has several advantages over other multivariate methods: the analysis can be performed voxel-wise on a whole brain-level, and continuous behavioural variables can be modelled. The groundwork of these advantages was a novel way to determine voxel-wise statistical significance. In short, SVR generates a β-parameter for each input variable (i.e. for each voxel in SVR-LSM). Contribution of β-parameters to the multivariate model is then statistically tested by permutation testing (Zhang et al., 2014). However, there is a dissent on the practically highly relevant question if correction for multiple comparisons as an additional step in SVR-LSM is required. Fama et al. (2017) argued that "because SVR-LSM considers all voxels simultaneously in a single regression model, correction for multiple comparisons is not required". Further, Gaonkar et al., (2013) postulated that the "interdependence [of the parameters in a SVR model] has the potential to alleviate multiple comparisons problems" when used to assess voxel-wise significance in multivariate imaging analyses. On the other hand, other studies performed SVR-LSM but corrected for multiple comparisons (Griffis et al., 2017; Ghaleh et al., 2018).

The present paper aimed to scrutinise the SVR-LSM method and answer the questions outlined above by empirical means. By using simulations, we investigated three questions: i) Is a correction for multiple comparisons required in SVR-LSM? ii) What sample size is required in MLBM? iii) Does MLBM suffer from a misplacement of results towards the centres of the brain's vascular territories?

2 General Methods

Imaging data of patients with first acute unilateral right stroke admitted to the Centre of Neurology at Tübingen University Hospital were used. Only patients with a clearly demarcated, non-diffuse lesion visible in structural imaging were included. Patients or their relatives consented to the scientific use of their data. The study has been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Structural brain images acquired as part of clinical protocols by either CT or MRI were used for lesion mapping. If both imaging modalities were available, MRI was preferred. In patients where MR scans were available, we used diffusion-weighted imaging (DWI) if the images were acquired within 48 h after stroke onset or T2-weighted fluid attenuated inversion recovery (FLAIR) images for later scans. Lesions were manually delineated on transversal slices of the individual scan using MRIcron (www.mccauslandcenter.sc.edu/mricro/mricron). Scans were then warped to 1x1x1mm^3 MNI space (Collins et al., 1994) using SPM8 (www.fil.ion.ucl.ac.uk/spm) and Clinical Toolbox (Rorden et al., 2012).

Multivariate lesion behaviour mapping by SVR was performed using MATLAB 2017b, libSVM (Chang and Lin, 2011), and a publicly available collection of scripts for SVR-LSM from the study by Zhang et al. (2014). Lesion maps and behavioural data were processed to fit the input data structure of libSVM and an epsilon-SVR with radial basis function kernel was computed. The resulting β-parameters were then remapped into three-dimensional MNI space. Voxel-wise statistical significance level in this parameter map was determined by permutation testing. Using this approach, data are permuted several thousands of times and the resulting pseudo-behaviour data are used to generate SVR-β-maps. Finally, voxel-wise significance is determined by comparison of pseudo-behaviour β-maps and the β-map obtained from real behavioural data. In the present study 1000 permutations were used. Zhang and co-workers have also shown that a control for the effect of lesion size is required in SVR-LSM. Therefore, the binary lesion images were first vectorised and normalised to have a unit norm, which serves as a direct total lesion volume control (dTLVC). Derivation of statistical maps via permutation testing, kernel choice, pre-processing of behavioural variables via normalisation, back-projection of data into three-dimensional space, and control of lesion size were performed with scripting provided by Zhang et al. (2014). We consistently set hyperparameters $C = 30$ and $\gamma = 4$, which have proven to perform well in a previous study (Wiesen et al., submitted). Note that hyperparameter selection via cross validation is an important step in SVR-LSM, and has potential to maximize model quality. However, its computational demands are high, and no clear criteria are available for parameter choice yet (see Zhang et al., 2014). All further analyses were performed using MATLAB and SPSS 19; all statistical tests were computed at $p < .05$. Voxel-based lesion behaviour mapping was performed using NiiStat (https://www.nitrc.org/plugins/mwiki/index.php/niistat:MainPage).

3 Experiment 1: Correction for multiple comparisons

3.1 Methods

Experiment 1 aimed to clarify if a correction for multiple comparisons is required in SVR-LSM as implemented by Zhang et al (2014). Therefore, we evaluated false positive rates in lesion-behaviour samples without any actual underlying positive signal. To obtain such samples, we utilised a sample of 203 right brain damaged patients with normalised lesion maps recruited in a recent study by Wiesen et al. (submitted). Wiesen and co-workers found large significant clusters to underlie spatial neglect, using SVR-LSM both with and without correction for multiple comparisons via false discovery rate (FDR). In the present study, we permuted 20 times a continuous measure of spatial neglect behaviour (Rorden & Karnath, 2010) in this sample of 203 right brain damaged patients. The resulting samples thus did not contain any true signal. Note that we also wanted to exclude the possibility that control for lesion size affects the amount of false alarms. Therefore, SVR-LSM was performed with these 20 samples both with and without correction for lesion size (see above section '2 Methods': 'direct total lesion volume control'). Only voxels damaged in at least 10 patients were included in the modelling process. Dependent variable was the rate of voxels with false positive signal at p-levels .05 and .01.

3.2 Results

The analyses were performed for 349512 voxels. A large amount of false positive findings was observed in all conditions. If statistical significance is determined for each voxel independently – and independent of the fact that only a single multivariate model is computed – each analysis should yield p*349512 false positive voxels. Two-tailed one-sample t-tests showed that the number of false positive voxels did neither significantly differ from these expected values for statistical parameter thresholding at $p < .05$ (for both t-tests $p > .95$; see Fig. 2) nor at $p < .01$ (for both t-tests $p > .32$). Also, paired t-tests did not find any difference between false positive rates in analyses with versus analyses without control for lesion size (both $p > .98$). Moreover, applying FDR correction with $q = .05$ to the results, none of the 40 performed analyses (20 with and 20 without control for lesion size) yielded positive results.

3.3 Discussion

At a p-level of $p<.05$, we found that 5% out of all tested voxels contained false positive signal (and correspondingly 1% of all voxels at $p<.01$). This was not affected by direct total lesion volume control. The current implementation of SVR-LSM thus poses the same challenge as VLBM: analyses find large amounts of false positive signal, and statistical maps have to be controlled for multiple comparisons. Note that this conclusion could also have been made by examining the underlying algorithms used in SVR-LSM. SVR includes a large amount of variables (here: voxels) into a single model.

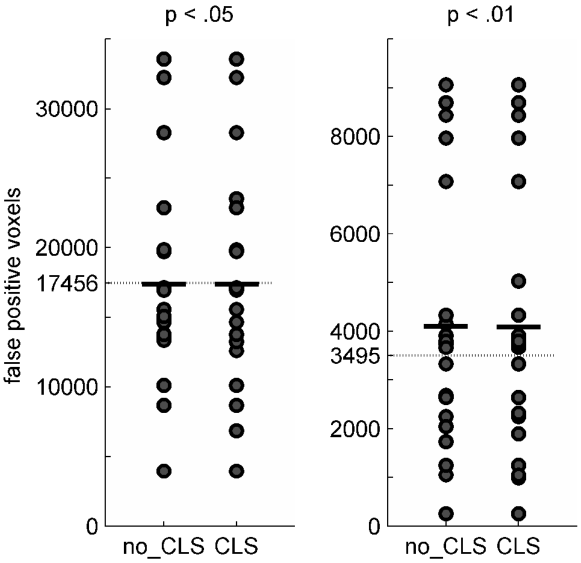


Figure 2: False positive rates in SVR-LSM

False positive rates of SVR-LSM in 20 simulation data sets that contain no true positive signal. Results are shown for multivariate lesion behaviour mapping with permutation-based statistical thresholding of parameter maps (see Zhang et al., 2014) at p < .05 and p < .01, and both with and without control for lesion size (CLS). Dotted lines indicate expected amount of false positives if statistical significance is indeed determined for each voxel independently; bold bars indicate mean values.

Generally, a comparison between two SVR models does not require a correction for multiple comparisons, although many variables are included in both models. However, permutation testing assesses statistical significance for each voxel individually. Therefore, the correction for multiple comparisons is required. From such theoretical perspective, the present empirical investigation is tautological. On the other hand, given the dissent in the field (see above section '1 Introduction'), the empirical approach employed in the present study provides a clear answer.

4 Experiment 2: Sample size required for multivariate models

4.1 Methods

Experiment 2 investigated what sample size is required to perform valid SVR-LSM. Six samples with simulated 'behavioural' data based on damage to either single or multiple areas were investigated. To obtain simulated data, 283 real normalised lesion maps of patients with right hemisphere damage were used. Most patients of this sample were also part of a recent simulation study (Sperber & Karnath, 2017). As ground truth, regions in the Automatic Anatomic Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) were chosen. Patients' individual 'behavioural' scores were then computed as a linear function of the individual lesion's damage to the atlas region. This was a continuous score that ranged from 0 (no damage to the region) to 1 (damage to 100% of all voxels in the atlas region). For multi-region models, the score was computed based on the region with most damage. Following simulation regions were chosen: i) insula ii) middle frontal gyrus iii) inferior parietal lobule iv) inferior frontal gyrus triangular + supramarginal gyrus v) caudatum + middle temporal gyrus vi) inferior frontal gyrus triangular + supramarginal gyrus + middle temporal gyrus.

Different sample sizes up to 140 patients were investigated in steps of 20, i.e. seven different sample sizes (20, 40, 60, 80, 100, 120, 140), and it was investigated what sample size is required to obtain a valid SVR model. This leads to the non-trivial question how to assess if a multivariate model is good. The model should fit the data; however, a good fit does not imply that a model is good, as it can suffer from over-fitting. Rather, a good model should also provide high generalisability. Therefore, we primarily assessed the correspondence of SVR-LSM maps, i.e. the final permutation-thresholded β-maps, and the reproducibility of SVR model β-parameters (see Rasmussen et al., 2012) between distinct samples at different sample sizes. Further, we assessed the prediction accuracy via cross-validation. To investigate model performance at sample size of n lesions, 2*n lesions were randomly drawn and assigned to two exclusively disjunct samples of n lesions each. For all analyses in Experiment 2, only voxels damaged in at least 10 patients in the 2*n sample were tested. This ensured that two paired analyses were always based on the same voxels. Next, a SVR was computed to model behavioural scores based on the status of all voxels. From the SVR models and corresponding β-maps, reproducibility of β-maps and prediction accuracy were assessed. To obtain the variable 'reproducibility of β-maps', the correlation between β-parameters in both β-maps was computed. Note that β-weights of individual voxels only provide limited interpretability, as they only indirectly relate to the behavioural scores. Yet, reproducibility of β-weights can be interpreted in the context of generalisability with caution (as, e.g., in Rasmussen et al., 2012; Zhang et al., 2014), especially as all comparisons in the present study were based on the same behavioural variable and the same hyperparameters. Second, 'prediction accuracy' was assessed by applying the model obtained in the first sample for a prediction in the second sample. Then, the correlation between predicted and true behavioural values was computed. The procedure of drawing 2*n lesions and randomly assigning them to two equally large groups was repeated 50 times for each data point. The correspondence of actual SVR-LSM maps (i.e. the final p-maps obtained from β-maps via permutation testing) was assessed by also drawing 2*n lesions and computing SVR-LSM maps independently for both samples. Then, correspondence of both maps was assessed by i) comparing both FDR-corrected, thresholded maps and computing the Dice Index, which provides a measure of similarity of two binary spatial images between 0 (no spatial overlap) to 1 (maximal spatial overlap), and ii) assessing 'reproducibility of p-maps' by computing the correlation between p-values in both maps. In order to save computational resources, the latter procedure was repeated five times for each data point.

4.2 Results

Plotting the data course of the investigated variables across sample sizes (Fig. 3) revealed several noticeable features: first, the data course of the variables qualitatively differed between reproducibility of both β- and p-maps and Dice index on the one hand and prediction accuracy on the other hand. Independent t-tests (see supplementary) revealed that reproducibility of β-maps significantly increased for all variables with each increase in sample size, except for the step from 80 to 100 patients in the simulation based on the inferior frontal gyrus. Increment-wise improvements, however, decreased rapidly with increments for larger sample sizes, and the plotted curves suggest that model performance asymptotically approaches a limiting value. Non-surprisingly, Dice indices and reproducibility of p-maps qualitatively followed a similar trend. In contrast, prediction accuracy was not significantly improved by increases in sample size above 100 subjects, while for some simulation regions prediction accuracy already peaked with sample sizes of 40 subjects. Second, standard deviations of reproducibility and prediction accuracy descriptively decreased rapidly with increasing sample sizes. Third, among simulations based on single regions (insula, middle frontal gyrus, and inferior parietal lobule), model performance differed across most data points. Independent t-tests (see supplementary material) confirmed that regarding reproducibility of β-maps, for all sample sizes, SVR performed best for simulations on the insula, and worst for simulations on the middle frontal gyrus.

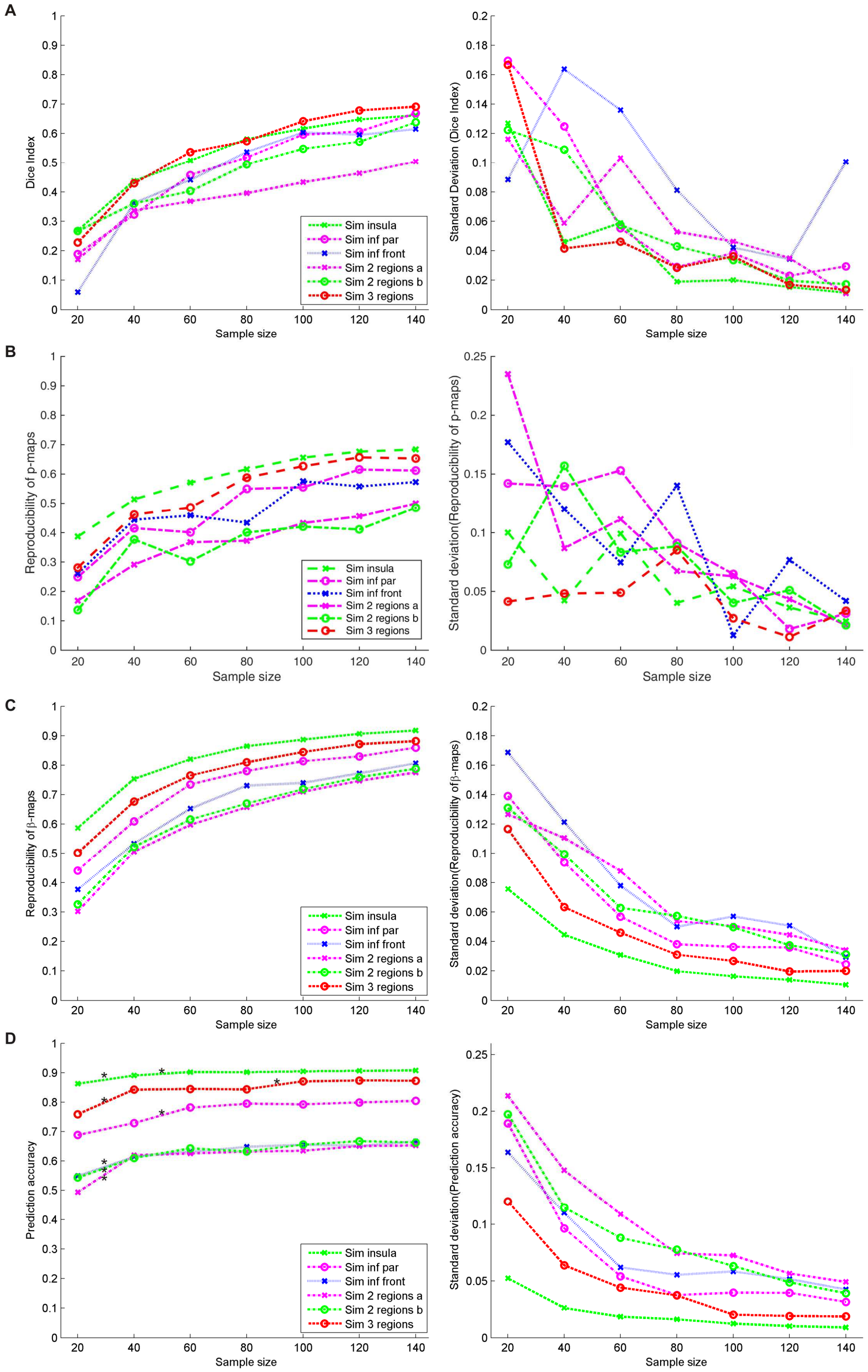

A
Dice Index
Standard Deviation (Dice Index)
B
Reproducibility of p-maps
Standard deviation(Reproducibility of p-maps)
C
Reproducibility of β-maps
Standard deviation(Reproducibility of β-maps)
D
Prediction accuracy
Standard deviation(Prediction accuracy)
Sample size
20
40
60
80
100
120
140
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.02
0.04
0.06
0.08
0.12
0.14
0.16
0.18
0.05
0.15
0.25
Sim insula
Sim inf par
Sim inf front
Sim 2 regions a
Sim 2 regions b
Sim 3 regions
*

Figure 3: Model performance across different sample sizes

(A) Dice indices (left panel) and its standard deviation (right panel) at different sample sizes. Note that each data point here is based on only 5 iterations. Therefore, standard deviations are larger than for variables in panels C&D. **(B)** Reproducibility of p-maps and its standard deviation at different sample sizes. Each data point is based on 5 iterations of the experimental procedure. **(C)** Reproducibility of β-maps and its standard deviation at different sample sizes. Each data point is based on 50 iterations of the experimental procedure. All 20-subject increments except for one (see text) were connected to significant increases in reproducibility according to independent sample t-tests. **(D)** Prediction accuracy and its standard deviation at different sample sizes. Each data point is based on 50 iterations of the experimental procedure. For the left panel of Fig 3C asterisks indicate significant changes for a 20-subject increment.

4.3 Discussion

Improvements in the reproducibility of β-maps across all sample sizes were found, while small samples appeared to profit the most from increases in size. Increases in size based on already large samples were still significant; however, they only provided smaller benefits. A very similar trend was observed for Dice indices and reproducibility of p-maps. On the other hand, prediction accuracy was relatively stable across sample sizes and did not further improve by increases in sample size above 100 subjects. To conclude, MLBM by SVR-LSM seems to require large samples to provide a model that maximizes the use of anatomical information for parameter estimation. Optimal sample sizes appear to be larger than 140 subjects, whereas one can doubt the usefulness of increases beyond ~ 100 to 120 subjects; nominal gains in reproducibility beyond these sizes are very small. However, if SVR is not used for a parametrical mapping as in SVR-LSM, but for prediction of clinical outcome, performance already peaks with smaller sample sizes with about 40 to 100 subjects. Furthermore, with larger sample sizes standard deviations were low for both reproducibility and prediction accuracy, suggesting that model performance is quite stable for defined sample sizes. In other words, given a certain (larger) sample size, model parameters were equally good (or bad) across iterations.

5 Experiment 3: Spatial bias of statistical results

5.1 Methods

Experiment 3 aimed to clarify if MLBM was suffering from a misplacement of results towards the centres of the arterial territories. We thus largely copied the simulation approach to investigate misplacement of topographical results in VLBM that was used by two previous studies (Mah et al., 2014; Sperber & Karnath, 2017). Since such a simulation approach is based on highly artificial simulations, it is not perfectly transferable to real data. Yet, its simplicity fits well empirical questions that aim to assess general principles in lesion behaviour mapping (for further discussion see Mah et al., 2014). As both previous studies using this simulation approach clearly visualised misplacement on axial MNI slice z = 17, we limited the analysis to this slice. Simulations were run using a real right hemisphere lesion sample of 283 patients. In short, 464 equally distributed voxels on slice z = 17 were selected. For each voxel and patient, it was determined if the lesion includes the voxel. Contrary to previous studies, however, simulated 'behavioural' scores were not binary, but continuous to allow application of support vector *regression*. To do so, random normally distributed values were drawn. If a lesion included the simulation voxel, values were drawn from a distribution with $\mu = 1.4$ and $\sigma = 0.4$, else from a distribution with $\mu = 0.4$ and $\sigma = 0.4$. Any resulting negative scores were set to zero.

Resulting 'behavioural' data of 464 simulations were then used in VLBM and SVR-LSM. Only voxels damaged in at least 5 patients were considered, and all resulting topographies were corrected for multiple comparisons by FDR correction at $p < .05$. Misplacement was then defined as a vector ranging from the voxel the simulation was based on to the centre of mass of the thresholded, binary statistical topography. As correction for lesion size is a crucial factor in anatomical misplacement (Sperber & Karnath, 2017), VLBM and SVR-LSM were performed with two different strategies to control for lesion size. First, SVR-LSM was performed using dTLVC (Zhang et al., 2014); second, we applied an approach that controls lesion size via regression both on the behavioural variable and the anatomical data (deMarco & Turkeltaub, 2018). To perform the latter type of correction, we integrated publicly available MATLAB scripts by deMarco & Turkeltaub (https://github.com/atdemarco/svrlsmgui) into our custom-modified scripts by Zhang et al. (2014).

5.2 Results

For mass-univariate voxel-based lesion behaviour mapping, a misplacement of topographical results by 13.5 mm (SD = 5.8 mm; median = 13.5 mm) for analyses without control for lesion size, and by 8.7mm (SD = 4.4 mm; median = 8.2 mm) for analyses with control for lesion size by nuisance regression was found. As shown in a previous study (Sperber & Karnath, 2017), control for lesion size significantly reduced misplacement (t(887) = 13.89, p < .001). Further, a vector visualisation (Fig. 4 A and B) revealed that misplacement was systematically oriented towards the centres of the middle and posterior arterial territories. These findings thus replicated previous studies that investigated misplacement of topographical results in VLBM (Mah et al., 2014; Sperber & Karnath, 2017). Peak Z-standardised statistics in the statistical topographies were Z = 8.2 (SD = 0.7) for VLBM without control for lesion size and Z = 7.7 (SD = 0.7) for VLBM with control for lesion size.

SVR-LSM topographies with control for lesion size by dTLVC were misplaced by 11.4 mm (SD = 5.0 mm; median = 11.0 mm). This misplacement was smaller than in VLBM without control for lesion size (t(828) = 5.54, p < .001), but larger than misplacement in VLBM with control for lesion size (t(815) = 8.17, p < .001). SVR-LSM with control for lesion size by regression both on anatomy and behaviour was misplaced by 12.5 mm (SD = 7.7 mm; median = 11.2 mm), which was larger than SVR-LSM with control for lesion size by dTLVC (t(841) = 2.35; p < .05), but still smaller than uncorrected VLBM (t(913) = 2.26; p < .05). Visual inspection of the data revealed that this difference originated from few outliers; the median misplacement obtained from both methods was roughly the same.

In order to compare directionality of vectors between conditions, cosine similarity was assessed. In a comparison of two groups of vectors, average cosine similarity will be 1 if all vector pairs show the same directionality, and 0 if directionality between vector pairs is entirely random. Cosine similarity of misplacement vectors in VLBM without control for lesion size and SVR-LSM with control for lesion size by dTLVC was 0.80 (SD = 0.29), which was significantly larger than zero (t(378) = 54.29, p < .001). For SVR-LSM with control for lesion size by regression and VLBM without control for lesion size, cosine similarity was 0.78 (SD = 0.32), which was also larger than zero (t(378) = 48.31; p < .001. Thus, directionality of misplacement vectors in the univariate and the mass-univariate analyses were similar. Correspondingly, the vector visualisation of misplacement in

SVR-LSM (Fig. 4C) also appeared to follow the vasculature.

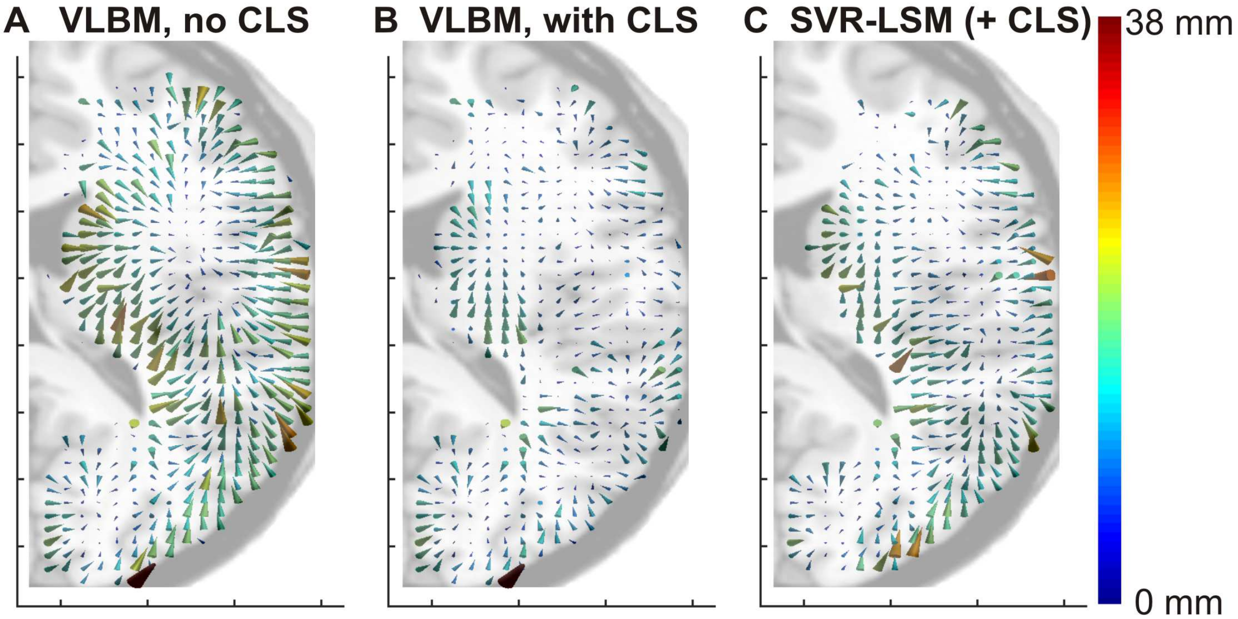


Figure 4: Misplacement of topographical results in VLBM and SVR-LSM

Vector maps for spatial misplacement in VLBM and SVR-LSM. Each vector shows the misplacement ranging from the 'ground truth' voxel to the centre of mass of the binary topographical map. All results are based on analyses controlled with FDR correction at $p < .05$. All analyses were limited to MNI slice $z = 17$. **(A)** Vector map for voxel-based lesion behaviour mapping. **(B)** Vector map for voxel-based lesion behaviour mapping with control for lesion size (CLS) via nuisance regression. **(C)** Vector map for multivariate lesion behaviour mapping using support vector regression, including a control for lesion size via dTLVC as suggested by Zhang et al. (2014).

5.3 Discussion

The centres of mass of statistical topographies in SVR-LSM were misplaced in a similar spatial direction as in VLBM; they were oriented towards the centres of the middle and posterior arterial territories. The magnitude of this replacement was between the magnitude of misplacement in VLBM without and VLBM with control for lesion size. Thus, SVR-LSM is not less susceptible to misplacement compared to VLBM. Different approaches to control for lesion size – by using dTLVC (Zhang et al., 2014) or via regression both on the behavioural variable and the anatomical data (deMarco & Turkeltaub, 2018) − did not eliminate the misplacement in SVR-LSM. To conclude, multivariate analysis in lesion behaviour mapping does not per se account for the complexity of lesion anatomy, and inter-voxel correlations can negatively affect results.

6 General Discussion

Recently, we outlined that the validity of lesion behaviour mapping methods can be tested empirically with different approaches (Sperber & Karnath, 2018). In the present study, we did so with SVR-LSM, a novel promising method with the potential to overcome some of the shortcomings of mass-univariate lesion behaviour mapping. We found that i) correction for multiple comparisons is required in SVR-LSM, ii) that sample sizes above ~ 100 to 120 subjects are required to model voxel-wise lesion location in the context of SVR-LSM, and iii) that SVR-LSM is susceptible to misplacement of statistical topographies along the vasculature in the same way as mass-univariate analyses.

Our results resolve the controversy on multiple comparisons in SVR-LSM (Zhang et al., 2014; Mirman et al., 2015; Fama et al., 2017; Griffis et al., 2017). They show that SVR-LSM requires a correction for multiple comparisons. This can be a correction by false discovery rate (FDR; Benjamini & Yekutieli, 2001) as carried out in the present study. However, future empirical studies are required to find the best solution to the multiple comparisons problem in SVR-LSM. For univariate analyses, several alternative solutions to the multiple comparison problem have been proposed (e.g., Nichols & Hayasaka, 2003; Rorden et al, 2007; Karnath et al., 2018; Mirman et al., 2018). Correction by FDR is easy to apply and computationally fast, and therefore well fits in a large scale simulation study. However, it has several shortcomings, e.g. if samples are of small size or only contain low signal (Karnath et al., 2018; Mirman et al., 2018). Note that for a proper application of FDR on single real behavioural data sets, larger numbers of permutations should be used than in the present study. The present approach simply has given way to computational limitations.

The present findings further give an answer to the question whether valid parameter estimation for multivariate analyses generally requires large data sets. Some researchers postulated that multivariate lesion behaviour mapping depends on large-scale multi-centre studies, which are able to provide such large samples (Mah et al., 2014; Xu et al., 2018). Findings in the present study partially support this assumption. Multivariate modelling based on voxel-wise lesion information in SVR consistently improved its generalisability with increases in sample size even up to 140 subjects. On the other hand, improvements beyond ~100 subjects were very small and appeared to approach a limiting value. This closely resembles findings on multivariate modelling of fMRI data (Churchill et al., 2014). The authors also observed that increases in sample size led to a plateau in regards to prediction accuracy already with small samples, while reproducibility of model parameters improved if already large samples were increased. Under the perspective of practicability, our data suggest that sample sizes of about 100 to 120 patients appear to be a good trade-off between model quality and feasibility regarding data input. It is of practical relevance to find out exactly how many patients are required to map a certain function. Cross validation, which is anyway required for hyper-parameter optimisation in SVR-LSM (see Zhang et al., 2014), can provide insights into model quality, however only smaller sub-samples of the total sample can be compared. Thus, with some limitations, cross validation could indicate if a sample was adequate in the modelling process.

However, caution should be advised if real symptoms are investigated. For real behavioural variables, using an adequate sample size for SVR-LSM would not imply that correlations close to r = 1.0 should be expected. First, anatomical information in real symptoms (compared to the present simulation samples) can vary. Critical brain regions can be organised with different complexity, as spatial normalisation of brain scans is noisy, or as inter-individual differences in brain anatomy exists. Second, a multitude of factors besides structural lesion information can affect post-stroke behaviour, such as age, inter-individual anatomical differences, time post stroke, or pre-stroke cognitive status (for review Price et al., 2017). Therefore, model generalisability of a SVR β-map can only be as good as behaviour can be explained by structural lesion information, e.g. voxel-wise lesion status. This makes it difficult to evaluate model performance. For example, imagine a SVR model based on a real data sample. If this model offers a cross-validation reproducibility of r = 0.4, one can hardly tell if this model already optimally includes anatomical voxel-wise information. Furthermore, differences in model performance across single-region simulations observed in the present study point at a role of lesion coverage. For simulations based on regions with higher lesion coverage (cf. Sperber & Karnath, 2016), experiment 2 has shown that models with higher lesion coverage perform better. Thus, SVR-LSM might not perform equally well for all regions, with worse performance if critical brain regions are covered by fewer lesions. Therefore, researchers that apply SVR-LSM should take care that their sample contains a considerable amount of cases in the pathological range, i.e. cases in which the critical brain region is damaged. Alternatively, investigation of behaviour that is only rarely pathological will require larger samples. Future studies on large samples of real data are required for further insights into optimal sample sizes in MLBM. Given that real data are more complex than simulated data, the present study provides a lower boundary for required sample sizes. Requirements for real data samples might be larger. Such future studies could also compare different approaches of MLBM in respect to required sample sizes (e.g. Yourganov et al., 2015; Pustina et al., 2018). Experiment 2 in the present study only investigated SVR-LSM and should not be generalised to MLBM in general.

In the discussion of spatial misplacement inherent to mass-univariate analyses (Mah et al., 2014), it was noted that the main reason to use multivariate instead of mass-univariate analyses was the complex architecture of lesions which leads to the misplacement of statistical results (Nachev, 2015; Xu et al., 2018). In contrast, the present study suggests that multivariate lesion behaviour mapping is susceptible to misplacement of statistical topographies to the same extent as mass-univariate VLSM analyses. A simple thought experiment illustrates why these findings in fact are not surprising: Imagine a sample of 100 lesions that is used in a lesion behaviour mapping study. As commonly found in lesion samples, many voxel pairs are damaged in exactly the same lesions, so-called ‘unique patches‘ (Pustina et al., 2018). In other words, there are many voxel pairs for which typical lesion anatomy leads to a perfect inter-voxel correlation of damage. Further imagine that for one of these perfectly correlated voxel pairs, one voxel belongs to a cognitive module which induces a cognitive symptom when damaged, and the other voxel does not belong to the cognitive module. In this case, we do not see any possibility that a lesion analysis - be it univariate or multivariate - could correctly identify only one voxel to belong to the cognitive module, but not the other without using any a priori information. As the present study suggests, this problem persists for multivariate analyses even if inter-voxel correlations are not perfect, but still high.

The problem of misplacement will not be alleviated by analysing data on region level rather than on voxel level, as suggested by Nachev (2015). Lesional damage between neighbouring regions will likely correlate, and results will also be misplaced. Nevertheless, the remaining misplacement bias in MLBM results − as well as in VLSM results – does not reach such levels as originally assumed in the study by Mah et al. (2014). Moreover, the quantification of misplacement as implemented in both the present and previous studies has limitations (cf. Sperber & Karnath, 2017; Pustina et al., 2018). First, a simple vector based on the centre of mass of a topographical map omits a lot of information contained in three-dimensional topographies (see Sperber & Karnath [2017] and Pustina et al. [2018] for more elaborated approaches based on more complex simulations). Second, such vectors can only point towards subcortical regions, and not towards areas outside the brain. In other words, the direction of possible biases is already predefined. This limitation might account for parts of the misplacement. However, misplacement vectors also clearly follow the arterial territories (see, e.g., Fig. 2 of the present article, or Mah et al. [2014]), what indicates that lesion anatomy is a central factor in the generation of the misplacement. Another, more general limitation is the ecological validity of simulation studies. While simulations provide a powerful tool to test the validity of lesion-behaviour mapping methods (Sperber & Karnath, in press; Xu et al., 2018), it is not known how well findings in simulations can be transferred to analyses in real data. In the present and in a previous study (Sperber & Karnath, 2017), we found peak statistics in VLBM to be very high compared to lesion studies on real data. This hints at an over-proportionally high underlying positive signal, which was present although we introduced random noise in experiment 3. The high signal, in turn, leads to more positive findings in any lesion analysis. Likely, this will induce an overestimation of misplacement in an analysis where all positive findings – except for one voxel – are false alarms. Indeed, misplacement in a simulation study has also been found to vary between p-levels of VLBM (Sperber & Karnath, 2017), with lower misplacement for more conservative p-levels, i.e. less false alarms. To conclude, as in the two previous studies using the same ‘artificial’ simulation approach (Mah et al., 2014; Sperber & Karnath, 2017), the present experiment 3 does not show that SVR-LSM topographies based on real data are misplaced by exactly ‘x’ mm (11.4mm in the present experiment), but rather that lesion anatomy generally is a biasing factor in SVR-LSM, similar as in VLBM.

The present study also bears implications for translational uses of multivariate modelling based on structural lesion data. As we discussed elsewhere (Karnath et al., 2018), these methods have potential to be used in long-term prediction of post-stroke outcome, e.g. in guiding rehabilitation measures. A recent study has shown that prediction of hemiparesis based on structural imaging can be performed with high accuracy using voxel-wise lesion information (Rondina et al., 2016). Experiment 2 in the present study contributes by showing that multivariate models maximize the use of structural lesion data already with small samples. Prediction accuracy was already relatively high even with our smallest investigated sample size of 20 patients, and did hardly improve with further increases beyond 40 to 80 subjects. However, contrary to SVR-LSM, the ultimate aim of prediction algorithms is to maximize prediction accuracy. Therefore, profound knowledge of non-anatomical variables that affect post-stroke behaviour is required (for review Price et al., 2017). Such variables can easily be included into SVR. However, it is not known yet if this requires larger sample sizes. Further, strategies for anatomical feature selection could improve prediction accuracies (see, e.g., Rondina et al., 2016). Importantly, when only prediction of behaviour is desired, both the multiple comparison problem (experiment 1) and anatomical biases (experiment 3) are not relevant.

To conclude, the present study could clarify some of the open and, in part, controversially debated questions related to SVR-LSM. Multivariate lesion behaviour mapping does not appear to resolve all methodological issues in the field of lesion-behaviour inference. Nevertheless, this new and promising approach to lesion analysis appears to supplement traditional mass-univariate analysis methods, in particular if larger patient samples are accessible.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (KA 1258/23-1). Christoph Sperber was supported by the Friedrich Naumann Foundation. Daniel Wiesen was supported by the Luxembourg National Research Fund (FNR/11601161). The authors acknowledge support by the state of Baden-Württemberg through bwHPC.

References

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. *Nature Neuroscience*, *6*(5), 448–450. https://doi.org/10.1038/nn1050

Benjamini, Y., Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29, 1165-1188.

Chang, C.C., Lin, C.J. (2011). LIBSVM: a library for support vector machines. ACM *Trans Intell Syst Technol*,. 2, 1–27.

Churchill, N. W., Yourganov, G., & Strother, S. C. (2014). Comparing within-subject classification and regularization methods in fMRI for large and small sample sizes. *Human Brain Mapping*, *35*(9), 4499–4517. https://doi.org/10.1002/hbm.22490

Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C. (1994). Automatic 3D intersubject regis- tration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.*, 18, 192–205.

DeMarco, A. T., & Turkeltaub, P. E. (2018). A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion-symptom mapping. *Human Brain Mapping*, *21*(May), 2461–2467. https://doi.org/10.1002/hbm.24289

Fama, M. E., Hayward, W., Snider, S. F., Friedman, R. B., & Turkeltaub, P. E. (2017). Subjective experience of inner speech in aphasia: Preliminary behavioral relationships and neural correlates. *Brain and Language*, *164*, 32–42. https://doi.org/10.1016/j.bandl.2016.09.009

Gaonkar, B., Sotiras, A., & Davatzikos, C. (2013). Deriving statistical significance maps for support vector regression using medical imaging data. *Int Workshop Pattern Recognit Neuroimaging*, 13–16. https://doi.org/10.1016/j.surg.2006.10.010.Use

Ghaleh, M., Skipper-Kallal, L. M., Xing, S., Lacey, E., DeWitt, I., DeMarco, A., & Turkeltaub, P. (2018). Phonotactic processing deficit following left-hemisphere stroke. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *99*, 346–357. https://doi.org/10.1016/j.cortex.2017.12.010

Griffis, J. C., Nenert, R., Allendorfer, J. B., & Szaflarski, J. P. (2017). Damage to white matter bottlenecks contributes to language impairments after left hemispheric stroke. *NeuroImage: Clinical*, *14*, 552–565. https://doi.org/10.1016/j.nicl.2017.02.019

Inoue, K., Madhyastha, T., Rudrauf, D., Mehta, S., & Grabowski, T. (2014). What affects detectability of lesion–deficit relationships in lesion studies? *NeuroImage: Clinical*, *6*, 388–397. https://doi.org/10.1016/j.nicl.2014.10.002

Karnath, H.-O., & Rennig, J. (2017). Investigating structure and function in the healthy human brain: validity of acute versus chronic lesion-symptom mapping. *Brain Structure & Function*, *222*(5), 2059–2070. https://doi.org/10.1007/s00429-016-1325-7

Karnath, H.-O., Sperber, C., & Rorden, C. (2018). Mapping human brain lesions and their functional consequences. *NeuroImage*, *165*(May 2017), 180–189. https://doi.org/10.1016/j.neuroimage.2017.10.028

Kinkingnéhun, S., Volle, E., Pélégrini-Issac, M., Golmard, J. L., Lehéricy, S., du Boisguéheneuc, F., … Dubois, B. (2007). A novel approach to clinical-radiological correlations: Anatomo-Clinical Overlapping Maps (AnaCOM): Method and validation. *NeuroImage*, *37*(4), 1237–1249. https://doi.org/10.1016/j.neuroimage.2007.06.027

Mah, Y.-H., Husain, M., Rees, G., & Nachev, P. (2014). Human brain lesion-deficit inference remapped. *Brain: A Journal of Neurology*, *137*(Pt 9), 2522–2531. https://doi.org/10.1093/brain/awu164

Mirman, D., Landrigan, J.-F., Kokolis, S., Verillo, S., Ferrara, C., & Pustina, D. (2018). Corrections for multiple comparisons in voxel-based lesion-symptom mapping. *Neuropsychologia*, *115*(December 2016), 112–123. https://doi.org/10.1016/j.neuropsychologia.2017.08.025

Mirman, D., Zhang, Y., Wang, Z., Coslett, H. B., & Schwartz, M. F. (2015). The ins and outs of meaning: Behavioral and neuroanatomical dissociation of semantically-driven word retrieval and multimodal semantic recognition in aphasia. *Neuropsychologia*, *76*(3), 208–219. https://doi.org/10.1016/j.neuropsychologia.2015.02.014

Nachev, P. (2015). The first step in modern lesion-deficit analysis. *Brain: A Journal of Neurology*, *138*(Pt 6), e354. https://doi.org/10.1093/brain/awu275

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, *12*(5), 419–446. https://doi.org/10.1191/0962280203sm341ra

Price, C. J., Hope, T. M., & Seghier, M. L. (2017). Ten problems and solutions when predicting individual outcome from lesion site after stroke. *NeuroImage*, *145*(Pt B), 200–208. https://doi.org/10.1016/j.neuroimage.2016.08.006

Pustina, D., Avants, B., Faseyitan, O. K., Medaglia, J. D., & Coslett, H. B. (2018). Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. *Neuropsychologia*, *115*(August), 154–166. https://doi.org/10.1016/j.neuropsychologia.2017.08.027

Rasmussen, P. M., Hansen, L. K., Madsen, K. H., Churchill, N. W., & Strother, S. C. (2012). Model sparsity and brain pattern interpretation of classification models in neuroimaging. *Pattern Recognition*, *45*(6), 2085–2100. https://doi.org/10.1016/j.patcog.2011.09.011

Rondina, J. M., Filippone, M., Girolami, M., & Ward, N. S. (2016). Decoding post-stroke motor function from structural brain imaging. *NeuroImage. Clinical*, *12*, 372–380. https://doi.org/10.1016/j.nicl.2016.07.014

Rorden, C., Fridriksson, J., & Karnath, H.-O. (2009). An evaluation of traditional and novel tools for lesion behavior mapping. *NeuroImage*, *44*(4), 1355–1362. https://doi.org/10.1016/j.neuroimage.2008.09.031

Rorden, C., & Karnath, H.-O. (2004). Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nature Reviews. Neuroscience*, *5*(10), 813–819. https://doi.org/10.1038/nrn1521

Rorden, C., Karnath, H.-O., & Bonilha, L. (2007). Improving lesion-symptom mapping. *Journal of Cognitive Neuroscience*, *19*(7), 1081–1088. https://doi.org/10.1162/jocn.2007.19.7.1081

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H.-O. (2012). Age-specific CT and MRI templates for spatial normalization. *NeuroImage*, *61*(4), 957–965. https://doi.org/10.1016/j.neuroimage.2012.03.020

Rorden, C., & Karnath, H. O. (2010). A simple measure of neglect severity. *Neuropsychologia*, *48*(9), 2758–2763. https://doi.org/10.1016/j.neuropsychologia.2010.04.018

Smith, D. V, Clithero, J. a, Rorden, C., & Karnath, H.-O. (2013). Decoding the anatomical network of spatial attention. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(4), 1518–1523. https://doi.org/10.1073/pnas.1210126110

Sperber, C., & Karnath, H.-O. (2016). Topography of acute stroke in a sample of 439 right brain damaged patients. *NeuroImage. Clinical*, *10*, 124–128. https://doi.org/10.1016/j.nicl.2015.11.012

Sperber, C., & Karnath, H.-O. (2017). Impact of correction factors in human brain lesion-behavior inference. *Human Brain Mapping*, *38*(3), 1692–1701. https://doi.org/10.1002/hbm.23490

Sperber, C., & Karnath, H.-O. (2018). On the validity of lesion-behaviour mapping methods. *Neuropsychologia*, *115*(May), 17–24. https://doi.org/10.1016/j.neuropsychologia.2017.07.035

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer, NY, USA.

Wiesen, D., Sperber, C., Yourganov, G., Rorden, C., Karnath, H.-O. (submitted). The perisylvian network of spatial neglect: insights from machine learning-based lesion-behavior mapping.

Xu, T., Jha, A., & Nachev, P. (2018). The dimensionalities of lesion-deficit mapping. *Neuropsychologia*, *115*(May), 134–141. https://doi.org/10.1016/j.neuropsychologia.2017.09.007

Yourganov, G., Smith, K. G., Fridriksson, J., & Rorden, C. (2015). Predicting aphasia type from brain damage measured with structural MRI. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *73*, 203–215. https://doi.org/10.1016/j.cortex.2015.09.005

Zhang, Y., Kimberg, D. Y., Coslett, H. B., Schwartz, M. F., & Wang, Z. (2014). Multivariate lesion-symptom mapping using support vector regression. *Human Brain Mapping*, *5876*, 5861–5876. https://doi.org/10.1002/hbm.22590

The network underlying human higher-order motor control: Insights from machine learning-based lesion-behaviour mapping

Christoph Sperber[1], Daniel Wiesen[1], Georg Goldenberg[3], Hans-Otto Karnath[1,2]

[1]Centre 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

[3] Neurological Department, Technical University Munich, Munich, Germany

Abstract

Neurological patients with apraxia of pantomime provide us with a unique opportunity to study the neural correlates of high-order motor function. Previous studies using lesion-behaviour mapping methods led to inconsistent anatomical results, reporting various lesion locations to induce this symptom. We hypothesised that the inconsistencies might arise from limitations of mass-univariate lesion-behaviour mapping approaches if our ability to pantomime the use of objects is organised in a brain network. Thus, we here investigated apraxia of pantomime by using multivariate lesion behaviour mapping based on support vector regression in a sample of 130 left-hemisphere stroke patients. Indeed, this method identified a common network to underlie high-order motor control, including the inferior parietal lobule, posterior parts of superior and middle temporal cortex, insula, as well as a periventricular frontal white matter bottleneck, adjacent to inferior frontal gyrus. Further, long association fibres were affected, such as the superior longitudinal fascicle, inferior occipito-frontal fascicle, and superior occipito-frontal fascicle. The resulting topography integrates findings of different previous studies. The findings thus not only underline the benefits of multivariate lesion-behaviour mapping in brain networks, but also pacify a longstanding discussion on the anatomy of human higher-order motor control.

Introduction

Following brain damage primarily of the left hemisphere, patients can suffer from high-order motor disorders, not caused by primary motor or sensory deficits (Heilman & Rothi, 1993). These disorders are summarised under the umbrella term 'apraxia' (Wheaton & Hallett, 2007; Goldenberg, 2011) and can include, for example, our ability to imitate or execute gestures (De Renzi et al., 1980; Goldenberg, 1996; Rumiati et al., 2009; Mengotti et al., 2015), to perform motor imagery (Ochipa et al., 1997; Buxbaum et al., 2005), or to mechanically reason (Goldenberg & Hagmann, 1998; Baumard et al., 2014). A prominent disorder in this field is apraxia of pantomime (hereinafter simply referred to as 'apraxia') in which patients fail to pantomime the use of a common tool as if they hold the tool in hand, while they are typically able to use the real tool with less or no errors (De Renzi et al., 1982; Goldenberg & Hagmann, 1998; Wada et al., 1999; Lausberg et al., 2003; Goldenberg et al., 2004; Laimgruber et al., 2005; Sperber et al., 2018).

Multiple studies investigated neurological patients to uncover the neural correlates of apraxia (for a review see Niessen et al., 2014). Their results, however, were inconsistent. Most frequently, apraxia was associated with lesions to the inferior parietal lobe or adjacent parietal regions (Halsband et al., 2001; Buxbaum et al., 2003, 2005; Weiss et al., 2008; Hoeren et al., 2014; Goldenberg & Randerath, 2015). On the other hand, several studies found lesions in the inferior frontal gyrus to induce apraxia (Goldenberg et al., 2007; Manuel et al., 2013; Weiss et al., 2016). Besides, regions such as the insula (Goldenberg et al., 2007; Hermsdörfer et al., 2013; Hoeren et al., 2014), premotor and precentral areas (Weiss et al., 2016), and the middle temporal gyrus (Manuel et al., 2013) were also reported to be critical. Interestingly, results in most studies were limited to only a few of these areas.

There are many possible methodological reasons for these inconsistencies (for a review see Sperber & Karnath, in press). A potential source for heterogeneous results could be the general analysis approach of the above mentioned studies. While they used different analysis techniques – such as voxel-based lesion behaviour mapping (VLBM; Manuel et al., 2013; Hoeren et al., 2014; Goldenberg & Randerath, 2015; Weiss et al., 2016), subtraction plots (Goldenberg et al., 2007; Weiss et al., 2008; Hermsdörfer et al., 2013), or region-of-interest analyses (Halsband et al., 2001) – all studies followed a univariate approach. Univariate methods such as mass-univariate VLBM, however, can fail to identify neural correlates of pathological behaviour if behaviour is organised in larger modules or networks (Rorden et al., 2009; Mah et al., 2014; Zhang et al., 2014). The previous anatomo-behavioural studies on apraxia thus might have been unable to gain full insight to the neural correlates of human higher-order motor control, and instead only identified single components of a possible network.

In recent years, multivariate lesion-behaviour mapping methods based on machine learning have been developed (Smith et al., 2013; Mah et al., 2014; Zhang et al., 2014; Yourganov et al., 2015). Multivariate lesion-behaviour mapping methods can include multiple variables – e.g., the lesion status of multiple voxels or regions of interest – into one single model. Simulation studies have revealed that such methods perform better than traditional VLBM tools if damage to multiple brain regions can induce a particular symptom (Mah et al., 2014; Zhang et al., 2014; Pustina et al., in press). Thus, in theory, multivariate analyses might resolve the inconsistencies in the field of apraxia if our assumption on a possible network underlying neuropsychological deficits in human higher-order motor control is correct. To test our hypothesis, we reanalysed a large sample of left hemisphere stroke patients using a multivariate lesion behaviour mapping method.

Methods

Subjects

We retrospectively analysed data of 130 left brain damaged patients (mean age = 56.5 ± 12.3 years; range 26-83) that had been admitted to the Neuropsychological Department of the Bogenhausen Hospital in Munich (for demographic and clinical data see Table 1). The patients were investigated in two previous studies (Goldenberg et al., 2007; Goldenberg & Randerath, 2015). All patients had a first ever left hemisphere stroke at least three weeks before the examination. Neuropsychological examination and imaging were part of clinical protocols at the Bogenhausen Hospital. Patients consented to the scientific re-use of their data; the study has been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

	Pantomime normal	Pantomime defective
Age (years)	55.4 (11.4)	57.7 (13.1)
Ischemia/haemorrhage	54/14	49/13
Lesion size (mm³)	79.2 (58.4)	111.1 (71.2)
Time since lesion (weeks)	14.2 (18.6)	14.9 (17.9)
Aphasia classification	6 none/ 12 global/ 9 broca/ 16 wernicke/ 10 amnestic/ 15 other	0 none/ 27 global/ 4 broca/ 12 wernicke/ 8 amnestic/ 11 other
Hemiparesis present (%)	44.1	61.3
Imitation finger (test score)	17.8 (2.9)	14.6 (4.4)
Imitation finger defective (%)	32.4	59.7
Imitation hand (test score)	16.8 (3.8)	14.4 (4.6)
Imitation hand defective (%)	44.9	67.7
Pantomime (test score)	49.2 (3.0)	29.6 (10.6)

Table 1: Demographic and clinical data

Demographic and clinical data of all 130 patients. Pantomime was tested using a 20-item test (Goldenberg et al., 2003, 2007); imitation of hand and finger gestures was assessed using a 10-item test each (Goldenberg, 1996). Patients were considered to have defective pantomime when they scored below a cutoff of 45 points. This cutoff was defined with a sample of 49 healthy control subjects (Goldenberg et al., 2007). Maximum score obtainable in the pantomime task was 55 points (cutoff < 45) and 20 points in the imitation tasks (finger posture imitation cutoff < 17; hand posture imitation cutoff < 18). Aphasia was classified according to the Aachen Aphasia Test. Numbers in parentheses indicate standard deviations.

Neuropsychological examination

Pantomime of tool use was assessed with a 20-item test (Goldenberg et al., 2003, 2007). Patients were asked to imitate the use of common tools as if they hold the actual tool in hand. For each item the examiner named an action and the corresponding tool and simultaneously showed a picture of the object (e.g., ‘How do you brush your teeth with a tooth brush?’ while showing a picture of a tooth brush). The picture was removed before the patient initiated the task. Patients with hemiparesis were instructed to use the left, ipsilesional hand to perform the task. To ensure that patients did understand the task, practice items were performed and patients that did not understand the task (e.g., when a patient drew the outlines of the tool on the table instead of performing pantomime) were excluded. For each of the 20 items one point was scored for the correct grip and finger posture and a maximum of one to three points for aspects such as movement amplitude, trajectories, or hand position in relation to the own body. Maximum score was 55 points. The inter-rater reliability was previously found to be very satisfying both for the number of correct features per item (kappa = 0.61) and for the total test score (kappa = 0.94; Goldenberg et al., 2003). Furthermore, all patients were tested for aphasia with the Aachen Aphasia Test (Huber et al., 1983) and for apractic deficits in imitation of postures with the fingers or the hand (Goldenberg, 1996).

Imaging and lesion mapping

Structural imaging was acquired either by MRI (n = 118) or CT (n = 12) on average 14.6 weeks (SD 18.2) after stroke-onset. The interval between imaging and neuropsychological examination was maximally three weeks. Lesions were manually mapped on transversal slices of the T1-weighted ‘ch2’ template scan from the Montreal Neurological Institute using MRIcro software (Rorden & Brett, 2000; http://people.cas.sc.edu/rorden/mricro/index.html). The ‘ch2’ template is oriented to fit Talairach space (Talairach & Tournoux, 1998) and is distributed with the MRIcro software. Lesions were mapped on a fixed set of twelve slices with z-coordinates -40, -32, -24, -16, -8, 0, 8, 16, 24, 32, 40, and 50 by using the closest matching or identical transversal slice found in the imaging. A topography of all lesions is shown in Figure 1A. To obtain an estimate for lesion size for the demographic data (Table 1) and the supplementary analysis, lesions were interpolated by converting each individual slice into a volume of 8mm thickness.

Multivariate lesion behaviour mapping

Multivariate lesion behaviour mapping (MLBM) was performed using support vector regression (SVR; Vapnik, 1995; Drucker et al., 1996), which is a multiple regression method based on machine learning. This method is an extension of support vector machines (Cortes and Vapnik, 1995). SVR is able to model continuous variables and has been successfully implemented in SVR-based lesion symptom mapping (SVR-LSM) to map lesion-behaviour relationships with high resolution on a whole-brain voxel-level (Zhang et al., 2014; Mirman et al., 2015b; Fama et al., 2017; Griffis et al., 2017). Further, SVR-LSM can be used with a control for the effect of lesion size by normalisation of each subject’s lesion data vector (Zhang et al., 2014).

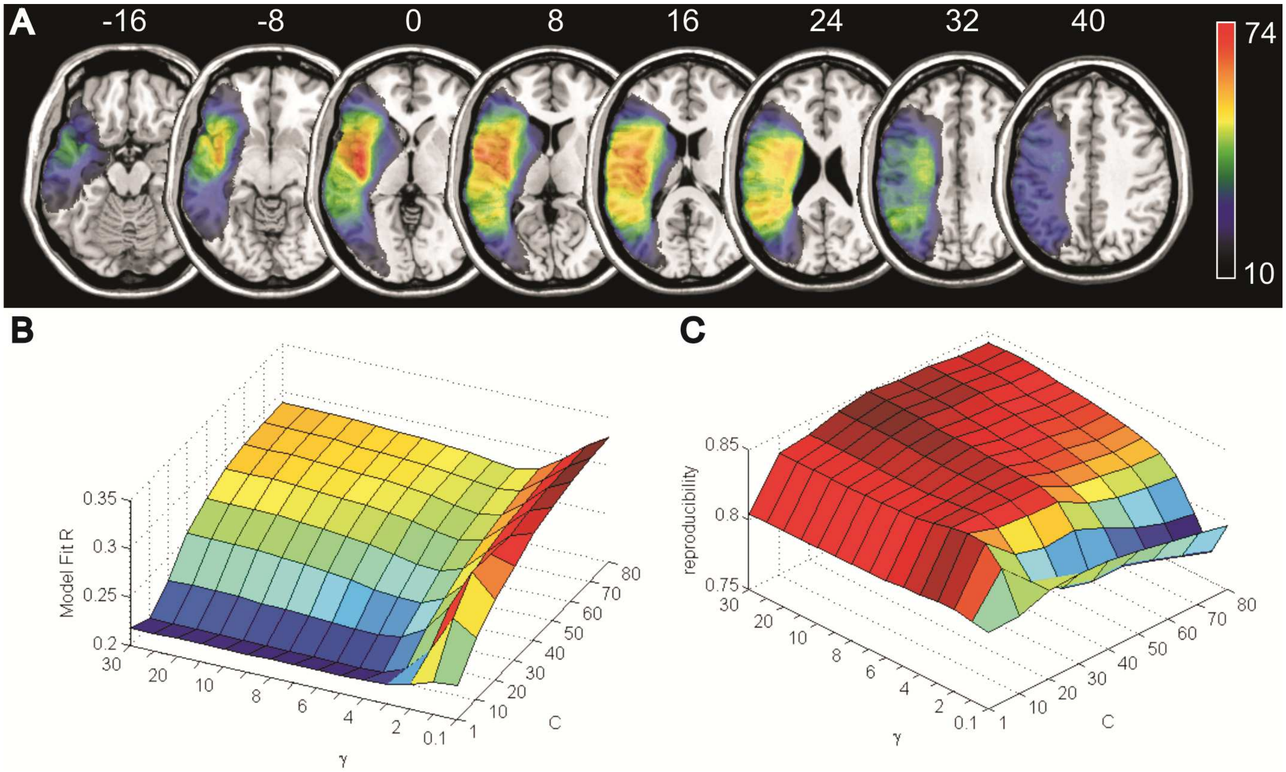


Figure 1: Topography of brain lesions and hyper-parameter optimisation for the SVR-MLBM

(A) Lesion topography of all 130 brain lesions with colour-coding that depicts the number of overlaying lesions per voxel. Only voxels affected in at least ten patients, i.e. voxels included in the multivariate analysis, are shown. Numbers above the slices indicate z-coordinates in MNI space. (B) Results of the hyper-parameter optimisation by grid search for C and γ. **(B)** Model Fit R and **(C)** reproducibility (see Rasmussen et al., 2012; Zhang et al., 2014) are plotted for the a priori set of chosen parameters.

Most importantly, the MLBM method based on SVR was validated in a set of simulation studies for simple brain networks (Zhang et al., 2014). Thus, at least in such artificial situations, voxel-wise SVR was empirically proven to be able to identify critical brain regions assembled in networks.

The analysis was performed with MATLAB 2016a and libSVM (Chang and Lin, 2011). We modified a publicly available collection of scripts (https://github.com/yongsheng-zhang/SVR-LSM) used in the study by Zhang et al. (2014) and adopted algorithms for control for lesion size and for the derivation of a topography from SVR parameters. The detailed methodological process and theoretical background is described in Zhang et al. (2014). In short, binary lesion images were vectorised and then normalised to have a unit norm. This procedure provides a control for the effect of lesion size. Lesion data and behavioural data were then further processed to fit the required data format of the libSVM toolbox. Using default libSVM options, an epsilon-SVR with radial basis function kernel was performed. The β-parameters obtained from the SVR were then remapped onto a three-dimensional brain topography. To assess statistical significance, a permutation approach has been chosen. By randomisation of behavioural scores, a large number of SVR-β-maps were generated and voxel-wise significance of β-parameters could be derived. The topography was computed using permutation testing with 4000 permutations at $p < 0.05$, and only voxels with at least ten lesions were tested. To obtain an optimal model, we performed an optimisation for hyper-parameters C and γ via grid search. The range of investigated parameters was chosen as in the study by Zhang et al. (2014): C = 1, 10, 20, 30, 40, 50, 60, 70, 80, and γ = 0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30. Using a five-fold cross-validation, we evaluated both model fit and reproducibility of each parameter set (see Rasmussen et al., 2012; Zhang et al., 2014). Resulting topographies were interpreted according to the AAL atlas (Tzourio-Mazoyer et al., 2002) for grey matter regions and to the probabilistic cytoarchitectonic fibre tract atlas (Bürgel et al., 2006) for white matter structures. For the white matter atlas, overlay of the thresholded statistical map with the probabilistic map at $p \geq 0.3$ was identified.

Results

The grid search showed that model fit and reproducibility generally were diametrical (Fig. 1B and C). The ideal model should provide high model fit while maximising reproducibility (Rasmussen et al., 2012). We chose C = 30 and γ = 4, as this set of hyper-parameters provided both a comparatively satisfying model fit (r = .24) and high reproducibility (reproducibility = .84).

The permutation-thresholded topography (Fig. 2) revealed a network of left frontal, temporal, parietal, and subcortical regions underlying apraxia. Table 2 lists affected grey and white matter structures. Large areas of significant voxels were found in the inferior parietal lobule, including angular and supramarginal gyri. In the frontal lobe, we found the largest cluster in the periventricular white matter and a few significant voxels in the orbital part of the inferior frontal lobe. Furthermore, larger significant clusters were found in superior and middle temporal gyrus, insula, precentral gyrus, caudatum, amygdala, and hippocampus. Significant clusters also included white matter structures, including superior longitudinal fascicle, inferior occipito-frontal fascicle, and superior occipito-frontal fascicle.

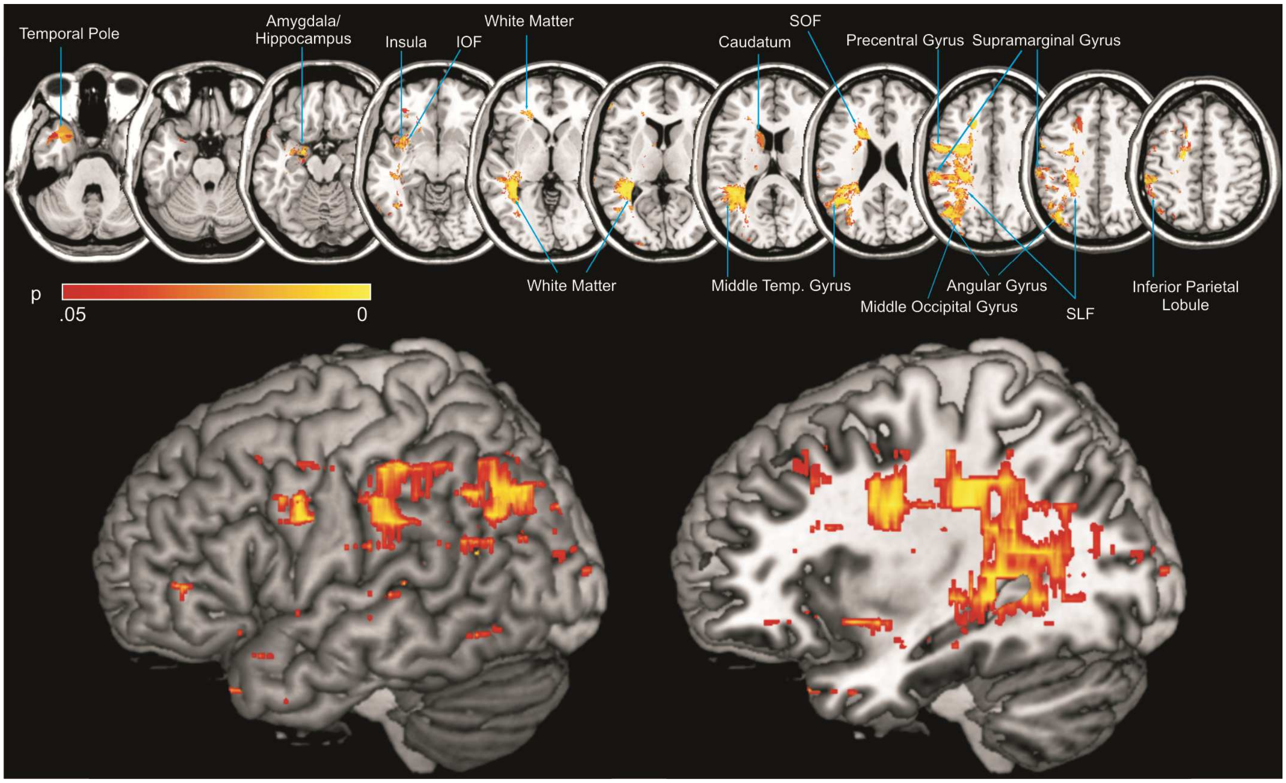


Figure 2: Results of the multivariate lesion-behaviour mapping

Permutation-thresholded β-map of SVR-MLBM on apraxia scores ($p < 0.05$), illustrating the anatomical regions significantly associated with apraxia of pantomime. Significance clusters were interpreted according to the AAL atlas (Tzourio-Mazoyer et al., 2002) for grey matter regions and to the probabilistic cytoarchitectonic fibre tract atlas (Bürgel et al., 2006) for white matter structures. Lower panels show three-dimensional renderings of the same map using the 3D-interpolation algorithm provided by MRIcron (http://people.cas.sc.edu/rorden/mricron/index.html; 8mm search depth in the left panel and 16mm in the right). Abbreviations: SLF – superior longitudinal fascicle; SOF – superior occipitofrontal fascicle; IOF – inferior occipitofrontal fascicle.

Grey matter structure	Percent affected	White matter structure	Percent affected
Angular gyrus	43.5	Sup. longitudinal fasc.	48.5
Amygdala	30.8	Inf. occ.-frontal fasc.	23.6
Supramarginal gyrus	27.9	Sup. occ.-frontal fasc.	20.5
Middle temporal pole	20.2	Corticospinal tract	18.8
Inf. parietal lobe	13.7	Optic radiation	13.1
Precentral gyrus	12.4	Uncinate fascicle	12.6
Caudatum	11.8	Callosal body	10.0
Middle occipital gyrus	10.5	Acoustic radiation	4.5
Sup. temporal pole	10.5		
Hippocampus	8.4		
Middle temporal gyrus	7.2		
Supp. motor area	5.2		
Superior temporal gyrus	4.4		
Superior frontal lobe	4.0		
Insula	3.9		
Inf. frontal lobe/orbital	2.2		
Postcentral gyrus	2.2		

Table 2: Topological grey and white matter analysis

Topological analysis of grey and white matter structures covered by the significant statistical map (see also Fig. 2). For grey matter structures, left hemispheric regions taken from the AAL Atlas (Tzourio-Mazoyer et al., 2002) with at least 2% affection are reported. For white matter structures, ROIs in the probabilistic histological atlas (Bürgel et al., 2006) were defined at a probability of p ≥ .3 to obtain binary maps. Of these binary maps, only left hemisphere parts were considered (MNI-coordinate x < 91). Further, for both grey and white matter atlas ROIs only z-slices that were part of the statistical analysis were considered.

The finding of the largest frontal cluster in the periventricular frontal white matter, but not in inferior frontal gyrus, was surprising given that several previous studies found that the inferior frontal gyrus was associated with apraxia. We therefore took a closer look at the frontal significant cluster. Recent studies discussed the role of periventricular frontal white matter lesions in aphasia and found that damage to a white matter bottleneck increased aphasic disturbances (Mirman et al., 2015a; Griffis et al., 2017). To find out if damage to this white matter bottleneck could also underlie apraxia, we compared the frontal cluster in our topography with an atlas-based reconstruction of the bottleneck. Analogue to previous studies (Mirman et al., 2015a; Griffis et al., 2017), the bottleneck was reconstructed using the ICBM DTI-81 atlas (Mori et al., 2008; Oishi et al., 2008). Probabilistic maps were thresholded at a probability of p ≥ 0.3. Indeed, the cluster affected the white matter bottleneck, consisting of inferior occipito-frontal fascicle, uncinate fascicle, and thalamic projection fibres connected to inferior and middle frontal gyrus (Fig. 3).

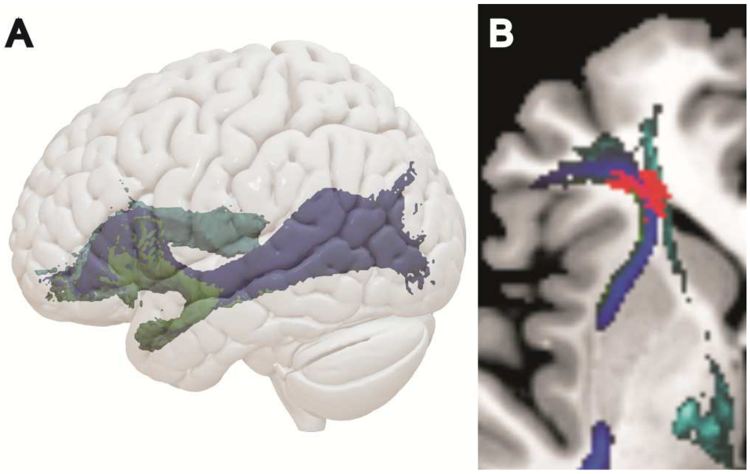


Figure 3: Frontal white matter bottleneck

Significant cluster in frontal white matter in relation to frontal white matter fibres. Analogue to previous studies (Mirman et al., 2015a; Griffis et al., 2017), white matter fibre tracts were reconstructed using probabilistic maps taken from the ICBM DTI-81 atlas (Mori et al., 2008; Oishi et al., 2008) and thresholded at p ≥ 0.3. **(A)** 3D-atlas reconstruction of relevant fibres consisting of inferior occipito-frontal fascicle (blue), uncinate fascicle (green), and thalamic projection fibres connected to inferior and middle frontal gyrus (cyan). **(B)** 2D-plot of the white matter bottleneck on MNI slice z = 0 and the significant cluster obtained in the MLBM analysis (red).

Discussion

In a large sample of 130 left brain damaged patients, we investigated apraxia of pantomime by using multivariate lesion behaviour mapping. A main advantage of MLBM is that the role of different brain regions can be investigated in combination. Indeed, we were able to confirm the hypothesis of a network underlying human high-order motor control, including left frontal, temporal, parietal, and subcortical regions. Different parts of this network have been observed in previous studies in isolation. This has contributed to the inconsistent reports and conclusions on the neural representation of apraxia. In fact, the present multivariate analysis uncovered that these regions belong to a common network underlying high-order motor control.

The most likely reason for the inability of previous mass-univariate approaches to identify the whole network could be the 'partial injury problem'. This is a methodological problem inherent to several lesion analysis methods including

VLBM (Rorden et al., 2009; for review Karnath et al., 2018). As example, imagine a simple brain network consisting of two distinct areas A and B. If damage to either area A or area B can induce the same symptom, patients showing the deficit may exist that have damage to area A, but not area B, and vice versa. Such patients will be used as counter examples in the voxel-wise statistical tests and statistical power to detect the neural correlates of the symptom is therefore reduced. Thus, given that a complex brain network underlies apraxia, mass-univariate methods can fail to identify the brain network in a whole. Previous mass-univariate studies therefore did not provide 'wrong' results, but simply were unable to identify all critical brain regions involved at once. Hence, MLBM appears to be a beneficial innovation in research on distributed networks, including apraxia of pantomime.

The SVR-LSM analysis did not associate apraxia with lesions to inferior frontal gyrus, but to adjacent white matter. At first glance, this result was surprising, because several previous studies found the inferior frontal gyrus itself to underlie apraxia (Goldenberg et al., 2007; Manuel et al., 2013; Weiss et al., 2016). A post-hoc analysis, however, shed further light on this finding. Recent studies discussed the role of periventricular frontal white matter lesions in aphasia and found that damage to a white matter bottleneck increased aphasic disturbances (Mirman et al., 2015a; Griffis et al., 2017). Apraxia is closely linked to aphasia (e.g., Kertesz & Hooper, 1982; Goldenberg & Randerath, 2015; Weiss et al., 2016) and, in fact, we found the same periventricular frontal white matter bottleneck affected in apraxia, consisting of inferior occipito-frontal fascicle, uncinate fascicle, and thalamic projection fibres connected to inferior and middle frontal gyrus. Interestingly, this periventricular bottleneck region was also included in previous studies that found the inferior frontal gyrus to underlie apraxia by using VLBM (Manuel et al., 2013) or subtraction analysis (Goldenberg et al., 2007). However, one VLBM study found only frontal cortical regions, but not white matter areas to underlie apraxia (Weiss et al., 2016). Furthermore, recent studies that associated aphasia with damage to white matter bottlenecks (Mirman et al., 2015a; Griffis et al., 2017) also identified a posterior periventricular white matter bottleneck. Large significant clusters in this area were also found by our study, which indicates that this bottleneck as well might play a role in apraxia. In general, the present analysis pointed out a prominent role of white matter damage in apraxia. Besides the white matter bottlenecks, several other white matter structures, including the superior longitudinal fascicle, the inferior occipito-frontal fascicle, and the superior occipito-frontal fascicle, were found significant. These white matter fibres constitute a perisylvan network that connects frontal, temporal, and parietal brain regions, and represent a crucial part of the human language network (Catani et al., 2005; Catani & Mesulam, 2008; Turken & Dronkers, 2011). This once more underlines the close relation of language and praxis processes (e.g., Kertesz & Hooper, 1982; Goldenberg & Randerath, 2015; Weiss et al., 2016).

The role of white matter damage found in the present study is also in line with other previous studies. For example, Kertesz and Ferro (1984) reported that smaller lesions that induce apraxia are predominantly found in the periventricular white matter. Also, a combined fMRI-DTI study identified a fronto-temporo-parietal network to underlie the ability to pantomime, which includes frontal white matter fibres (Vry et al., 2013). The finding of a network underlying apraxia is convincing since the ability to pantomime object use is a complex task that requires a wide range of cognitive and motor abilities. Several cognitive models of praxis skills have been proposed in line with findings in apractic patients (e.g., Barbieri & de Renzi, 1988; Cubelli et al., 2000; Bartolo et al., 2003; Johnson-Frey, 2004; Frey, 2008; Jax et al., 2014; Goldenberg, 2017). Although there is no consensus on the cognitive model underlying pantomime (Goldenberg, 2017), the different models generally assume paths along multiple cognitive processes that lie between phonological or visual analysis of the input stimuli (e.g., the word 'tooth brush' or a picture of a tooth brush) and the motor response. For example, the most classical cognitive model of apraxia assumes that gestures such as pantomime are conceptualised, converted into a motor programme, and then executed (e.g., Liepmann, 1908; Barbieri & De Renzi, 1988; Jax et al., 2014). Given such complex cognitive models, a disruption of different cognitive functions could induce deficits in pantomime. These deficits may also show different characteristics with differently affected cognitive subcomponents. Accordingly, errors in apraxia can qualitatively differ and dissociate (e.g., Buxbaum, 2001; Halsband et al., 2001; Goldenberg, 2011; Manuel et al., 2013). Thus, previous studies so far might have mapped different aspects of pathological behaviour functions at once. Indeed, it has been shown that the neural correlates of different apractic error types can dissociate (Manuel et al., 2013).

The findings in the present study are not only able to reconcile several discrepancies within the lesion-behaviour mapping literature in apraxia, but also discrepancies between lesion-behaviour mapping studies and fMRI studies. As in lesion studies, fMRI experiments that investigated pantomime also found several different left hemisphere regions to be involved (for reviews see Johnson-Frey, 2004; Lewis, 2006; Niessen et al., 2014; see also Lausberg et al., 2015; Vry et al., 2015; Martin et al., 2016; Chen et al., 2017). Parietal regions, including the intraparietal sulcus, inferior parietal lobe, and/or superior parietal lobe, were found activated in nearly all studies during pantomime (Niessen et al., 2014). Beyond, activation in middle and inferior frontal gyrus, inferior, middle and superior temporal lobe, inferior occipital gyrus, precentral gyrus, and insula were reported in some of these studies (Lewis, 2006; Niessen et al., 2014; Lausberg et al., 2015; Martin et al. 2016). The present finding suggesting a complex network to underlie higher-order motor control thus is in line with these observations derived from healthy subjects.

Albeit the results of the present study resolve some of the inconsistencies in the field, open questions also remain. For example, the present analysis found critical voxels in hippocampus, amygdala, and the temporal pole. It is possible that this finding represents an artefact, as patients with damage to the hippocampus consistently show larger lesions that also affect temporal regions (cf. Goldenberg & Randerath, 2015). Indeed, in the present sample, lesions with high affection of the hippocampus or amygdala i) seemed to be very large and ii) regularly included other regions that were found to be critical for apraxia in the main analysis, including temporal regions, parietal regions, or white matter fibres (see supplementary data). Thus, it seems that multivariate analyses appear to be prone to errors if damage systematically co-occurs between two or more regions, i.e. if statistical independence of damage to voxels/brain regions is violated. A first study recently has investigated this issue (Pustina et al., in press). They found that multivariate analyses are superior to mass-univariate analyses with respect to this bias, but still not perfect (for further discussion Karnath et al., 2018). Another still debated issue in MLBM is correction for multiple comparisons (see e.g., Zhang et al., 2014; Mirman et al., 2015b; Fama et al., 2017). The initial modelling procedure in MLBM only computes a single SVR model, thus correction for multiple comparisons is not necessary. However, it is unclear if correction for multiple comparisons is necessary when the thresholded topography is derived from model parameters (Zhang et al., 2014). Some studies explicitly did not control for multiple comparisons (Zhang et al., 2014; Fama et al., 2017), while others successfully used correction by false discovery rate (FDR; Mirman et al., 2015b; Griffis et al., 2017) or cluster-based permutation testing

(Mirman et al., 2015b). Both correction methods, however, have their flaws (see Mirman et al., in press for cluster-size permutation; see Karnath et al., 2018 for critical discussion of FDR correction). When FDR with q = .05 was used in the present analysis, no voxels survived correction. The low signal in our data might be related to generally low model fit (with high reproducibility on the hand; see Rasmussen et al., 2012 for discussion), or with on average large lesions in our sample combined with correction for lesion size. In general, the use of multivariate methods in lesion-behaviour mapping only emerged recently (Smith et al., 2013); the method thus still needs further elaboration and optimisation. Also, the question remains how the network underlying apraxia of pantomime is involved in other deficits of higher-order motor skills, such as apraxia of imitation or apraxia of real tool use. The present study as well as a first study that investigated apraxia of imitation using a multivariate region-pair approach (Achilles et al., 2017) suggest that multivariate lesion-behaviour mapping may also be able to improve our understanding of these symptoms. It may deepen our knowledge on how brain regions in the network and their cognitive functions work together, and how the ability to pantomime relates to other higher-order motor skills such as imitation of gestures or real tool use. This could allow us to take the next steps in understanding human motor cognition both from a neuroscientist's and a clinician's perspective.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (KA1258/15-1), by the Friedrich Naumann Foundation (to C.S.), and the Luxembourg National Research Fund (FNR/11601161 to D.W.).

Achilles EIS, Weiss PH, Fink GR, Binder E, Price CJ, Hope TMH. 2017. Using multi-level Bayesian lesion-symptom mapping to probe the body-part-specificity of gesture imitation skills. Neuroimage. 161: 94–103.

Barbieri C, De Renzi E. 1988. The executive and ideational components of apraxia. Cortex. 24:535–544.

Bartolo A, Cubelli R, Della Sala S, Drei S. 2003. Pantomimes are special gestures which rely on working memory. Brain Cogn. 53(3):483-494.

Baumard J, Osiurak F, Lesourd M, Gall D Le. 2014. Tool use disorders after left brain damage. Front Psychol. 5:1-12.

Bürgel U, Amunts K, Battelli L, Mohlberg H, Gilsbach JM, 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.

Buxbaum LJ. 2001. Ideomotor apraxia: a call to action. Neurocase. 7(6):445-458.

Buxbaum LJ, Sirigu A, Schwartz MF, Klatzky R. 2003. Cognitive representations of hand posture in ideomotor apraxia. Neuropsychologia. 41(8):1091-1113.

Buxbaum LJ, Johnson-Frey SH, Bartlett-Williams M. 2005. Deficient internal models for planning hand-object interactions in apraxia. Neuropsychologia. 43(6):917-929.

Catani M, Jones DK, Ffytche DH. 2005. Perisylvian language networks of the human brain. Ann. Neurol. 57: 8–16.

Catani M, Mesulam M. 2008. The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. Cortex. 44: 953–61.

Chang CC, Lin CJ. 2011. LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol. 2:27:1–27:27.

Chen Q, Garcea FE, Jacobs RA, Mahon BZ. 2017. Abstract Representations of Object-Directed Action in the Left Inferior Parietal Lobule. Cereb Cortex. 1-13.

Cortes C, Vapnik V. 1995. Support-vector networks. Mach Learn. 20:273–297.

Cubelli R, Marchetti C, Boscolo G, Della Sala S. 2000. Cognition in action: testing a model of limb apraxia. Brain Cogn. 44(2):144-165.

De Renzi E, Motti F, Nichelli P. 1980. Imitating gestures: A quantitative approach to ideomotor apraxia. Arch Neurol. 37(1):6-10.

De Renzi E, Faglioni P, Sorgato P. 1982. Modality-specific and supramodal mechanisms of apraxia. Brain. 105:301-312.

Drucker H, Burges CJC, Kaufman L, Burges CJC, Kaufman L, Smola A, Vapnik V. 1996. Support vector regression machines. Adv Neural Inf Process Syst (NIPS). 9:155–161.

Fama ME, Hayward W, Snider SF, Friedman RB, Turkeltaub PE. 2017. Subjective experience of inner speech in aphasia: Preliminary behavioral relationships and neural correlates. Brain Lang. 164:32–42.

Frey SH. 2008. Tool use, communicative gesture and cerebral asymmetries in the modern human brain. Philos Trans R Soc Lond B Biol Sci. 363(1499):1951-1957.

Goldenberg G. 2011. Apraxien. Göttingen: Hogrefe.

Goldenberg G. 1996. Defective imitation of gestures in patients with damage in the left or right hemispheres. J Neurol Neurosurg Psychiatry. 61(2):176-180.

Goldenberg G, Hentze S, Hermsdörfer J. 2004. The effect of tactile feedback on pantomime of tool use in apraxia. Neurology. 63(10):1863-1867.

Goldenberg G, Hagmann S. 1998. Tool use and mechanical problem solving in apraxia. Neuropsychologia. 36(7):581-589.

Goldenberg G. 2017. Facets of Pantomime. J Int Neuropsychol Soc. 23(2):121-127.

Goldenberg G, Hartmann K, Schlott I. 2003. Defective pantomime of object use in left brain damage: Apraxia or asymbolia? Neuropsychologia. 41(12):1565-1573.

Goldenberg G, Hermsdörfer J, Glindemann R, Rorden C, Karnath H-O. 2007. Pantomime of tool use depends on integrity of left inferior frontal cortex. Cereb Cortex. 17(12):2769-2776.

Goldenberg G, Randerath J. 2015. Shared neural substrates of apraxia and aphasia. Neuropsychologia. 75:40-49.

Griffis JC, Nenert R, Allendorfer JB, Szaflarski JP. 2017. Damage to white matter bottlenecks contributes to language impairments after left hemispheric stroke. NeuroImage Clin. 14: 552–565.

Halsband U, Schmitt J, Weyers M, Binkofski F, Grützner G, Freund HJ. 2001. Recognition and imitation of pantomimed motor acts after unilateral parietal and premotor lesions: A perspective on apraxia. Neuropsychologia. 39(2):200-216.

Heilman KM, Rothi LJG. 1993. Apraxia. In: Clinical Neuropsychology. New York, Oxford: Oxford University Press.

Hermsdörfer J, Li Y, Randerath J, Roby-Brami A, Goldenberg G. 2013. Tool use kinematics across different modes of execution. Implications for action representation and apraxia. Cortex. 49(1):184-199.

Hoeren M, Kümmerer D, Bormann T, Beume L, Ludwig VM, Vry MS, et al. 2014. Neural bases of imitation and pantomime in acute stroke patients: distinct streams for praxis. Brain. 137: 2796–2810.

Huber W, Poeck K, Weniger D, Willmes K. Aachener Aphasie Test. Goettingen: Hogrefe; 1983.

Jax SA, Rosa-Leyra DL, Buxbaum LJ. 2014. Conceptual- and production-related predictors of pantomimed tool use deficits in apraxia. Neuropsychologia. 62(2):194-201.

Johnson-Frey SH. 2004. The neural bases of complex tool use in humans. Trends Cogn Sci. 8(2):71-78.

Karnath H-O, Sperber C, Rorden C. 2018. Mapping human brain lesions and their functional consequences. Neuroimage. 165:180-189.

Kertesz A, Hooper P. 1982. Praxis and language: The extent and variety of apraxia in aphasia. Neuropsychologia. 20(3),275–286.

Kertesz A, Ferro JM. 1984. Lesion size and location in ideomotor apraxia. Brain. 107: 921–33.

Laimgruber K, Goldenberg G, Hermsdörfer J. 2005. Manual and hemispheric asymmetries in the execution of actual and pantomimed prehension. Neuropsychologia. 43(5):682-692.

Lausberg H, Cruz RF, Kita S, Zaidel E, Ptito A. 2003. Pantomime to visual presentation of objects: Left hand dyspraxia in patients with complete callosotomy. Brain. 126(2):343-360.

Lausberg H, Kazzer P, Heekeren HR, Wartenburger I. 2015. Pantomiming tool use with an imaginary tool in hand as compared to demonstration with tool in hand specifically modulates the left middle and superior temporal gyri. Cortex. 71:1-14.

Lewis JW. 2006. Cortical networks related to human use of tools. Neuroscientist. 12(3):211-231.

Liepmann H. 1908. Drei Aufsätze aus dem Apraxiegebiet. Berlin: Karger.

Mah Y-H, Husain M, Rees G, Nachev P. 2014. Human brain lesion-deficit inference remapped. Brain. 137(Pt 9):2522-2531.

Manuel AL, Radman N, Mesot D, Chouiter, L, Clarke, S, Annoni, J-M, Spierer, L. 2013. Inter- and intrahemispheric dissociations in ideomotor apraxia: a large-scale lesion-symptom mapping study in subacute brain-damaged patients. Cereb Cortex. 23(12):2781-2789.

Martin M, Nitschke K, Beume L, Dressing, A, Bühler, LE, Ludwig, VM, et al. 2016. Brain activity underlying tool-related and imitative skills after major left hemisphere stroke. Brain. 139(Pt 5):1497-1516.

Mengotti P, Ripamonti E, Pesavento V, Rumiati RI. 2015. Anatomical and spatial matching in imitation: Evidence from left and right brain-damaged patients. Neuropsychologia. 79:1-16.

Mirman D, Chen Q, Zhang Y, Wang Z, Faseyitan OK, Coslett HB, et al. 2015a. Neural organization of spoken language revealed by lesion-symptom mapping. Nat. Commun. 6: 6762.

Mirman D, Zhang Y, Wang Z, Coslett HB, Schwartz MF. 2015b. The ins and outs of meaning: Behavioral and neuroanatomical dissociation of semantically-driven word retrieval and multimodal semantic recognition in aphasia. Neuropsychologia. 76:208–219.

Mirman D, Landrigan J-F, Kokolis S, Verillo S, Ferrara C, Pustina D. in press. Corrections for multiple comparisons in voxel-based lesion-symptom mapping. Neuropsychologia.

Mori S, Oishi K, Jiang H, Jiang L, Li X, Akhter K, et al. 2008. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. Neuroimage. 40: 570–82.

Niessen E, Fink GR, Weiss PH. 2014. Apraxia, pantomime and the parietal cortex. NeuroImage Clin. 5:42-52.

Ochipa C, Rapcsak SZ, Maher LM, Rothi LJ, Bowers D, Heilman KM. 1997. Selective deficit of praxis imagery in ideomotor apraxia. Neurology. 49(2):474-480.

Oishi K, Zilles K, Amunts K, Faria A, Jiang H, Li X, et al. 2008. Human brain white matter atlas: identification and assignment of common anatomical structures in superficial white matter. Neuroimage. 43: 447–57.

Pramstaller PP, Marsden CD. 1996. The basal ganglia and apraxia. Brain. 119: 319–340.

Pustina D, Avants B, Faseyitan O, Medaglia J, Coslett HB. In press. Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. Neuropsychologia.

Rasmussen PM, Hansen LK, Madsen KH, Churchill NW, Strother SC. 2012. Model sparsity and brain pattern interpretation of classification models in neuroimaging. Pattern Recognit. 45:2085–2100.

Rorden C, Brett M. 2000. Stereotaxic display of brain lesions. Behav Neurol. 12:191–200.

Rorden C, Fridriksson J, Karnath H-O. 2009. An evaluation of traditional and novel tools for lesion behavior mapping. Neuroimage. 44(4):1355-1362.

Rumiati RI, Carmo JC, Corradi-Dell'Acqua C. 2009. Neuropsychological perspectives on the mechanisms of imitation. Philos Trans R Soc Lond B Biol Sci. 364:2337-2347.

Smith DV, Clithero JA, Rorden C, Karnath HO. 2013. Decoding the anatomical network of spatial attention. Proc Natl Acad Sci U S A. 110(4):1518-1523.

Sperber C, Karnath HO. In Press. On the validity of lesion-behaviour mapping methods. Neuropsychologia.

Sperber C, Christensen A, Ilg W, Giese MA, Karnath HO. In Press. Apraxia of object-related action does not depend on visual feedback. Cortex.

Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system - an approach to cerebral imaging. New York: Thieme.

Turken AU, Dronkers NF. 2011. The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. Front. Syst. Neurosci. 5: 1.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello, F, Etard, O, Delcroix, N, et al. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15(1):273-289.

Vapnik VN. 1995. The nature of statistical learning theory. New York: Springer.

Vry M-S, Tritschler LC, Hamzei F, Rijntjes M, Kaller CP, Hoeren M, et al. 2015. The ventral fiber pathway for pantomime of object use. Neuroimage. 106: 252–63.

Wada Y, Nakagawa Y, Nishikawa T, Aso, N, Inokawa, M, Kashiwagi, A, et al. 1999. Role of somatosensory feedback from tools in realizing movements by patients with ideomotor apraxia. Eur Neurol. 41(2):73-78.

Weiss PH, Rahbari NN, Hesse MD, Fink GR. 2008. Deficient sequencing of pantomimes in apraxia. Neurology. 70(11):834-840.

Weiss PH, Ubben SD, Kaesberg S, Kalbe E, Kessler J, Liebig T, et al. 2016. Where language meets meaningful action: a combined behavior and lesion analysis of aphasia and apraxia. Brain Struct. Funct. 221: 563–76.

Wheaton LA, Hallett M. 2007. Ideomotor apraxia: a review. J Neurol Sci. 260(1-2):1-10.

Yourganov G, Smith KG, Fridriksson J, Rorden C. 2015. Predicting aphasia type from brain damage measured with structural MRI. Cortex. 73:203-215.

Zhang Y, Kimberg DY, Coslett HB, Schwartz MF, Wang Z. 2014. Multivariate lesion-symptom mapping using support vector regression. Hum Brain Mapp. 5876:5861-5876.