Modeling of activation data in the BrainMap™ database: Detection of outliers

Finn Årup Nielsen and Lars Kai Hansen

Informatics and Mathematical Modelling

Technical University of Denmark

DK-2800 Lyngby, DENMARK

(fn,lkhansen)@imm.dtu.dk

April 2, 2001

Running title: Modeling of BrainMap data.

Keywords: Meta-Analysis, Data Analysis, Estimation Techniques, Probabilistic Models, Neuroanatomy, Databases, Neural Networks (Computer).

Address for correspondence:

Finn Årup Nielsen

Informatics and Mathematical Modelling, Building 321,

Technical University of Denmark, Lyngby, Denmark

Tel: +45 4525 3921

Fax: +45 4587 2599

E-mail: fn@imm.dtu.dk

ABSTRACT

We describe a system for meta-analytical modeling of activation foci from functional neuroimaging studies. Our main vehicle is a set of density models in Talairach space capturing the distribution of activation foci in sets of experiments labeled by lobar anatomy. One important use of such density models is identification of novelty, i.e., low probability database events. We rank the novelty of the outliers and investigate the cause for 21 of the most novel, finding several outliers that are entry and transcription errors or infrequent or non-conforming terminology. We briefly discuss the use of atlases for outlier detection.

INTRODUCTION

Given the rapid accumulation of functional neuroimaging data remarkably little effort goes into mathematical and statistical meta-analyses. Notable contributions are found in Indefrey and Levelt (2000) who modeled the relation between the cognitive components of language and the associated brain anatomy of the level of gyri, using the binomial distribution, while Paus (1996) estimated the mean and the standard deviation of a set of locations in order to describe the regions corresponding to the frontal eye fields. Functional volumes modeling (FVM) proposed in Fox et al. (1997, 1999, 2001) was used to model intersubject variability of activation foci corresponding to the M1-mouth area. Multidimensional scaling was used in Lloyd (1999, 2000) for visualization of 35 PET studies based on activations in Brodmann areas. However, typical reviews and meta-analyses make little or no modeling beyond tabelization and visualization, see e.g., (Cabeza and Nyberg 2000; Farah and Aguirre 1999; Allison, Puce, and McCarthy 2000; Decety and Grèzes 1999). A review of meta-analytic approaches can be found in Fox, Parsons, and Lancaster (1998).

In this contribution we will use non-parametric modeling to identify *outliers*. Beckman and Cook (1983) distinguish between two kinds of outliers, *discordant* outliers are “any observation that appears surprising or discrepant to the investigator” and *contaminant* outliers are “any observation that is not a realization from the target distribution”. An example of discordant outliers in functional neuroimaging would be the surprising tactile processing in the occipetal lobe (Zangaladze et al. 1999; Hamilton et al. 2000). Contaminant outliers can be typographical/transcription errors.

Classic outlier detection use relatively simple parametric models, see Beckman and Cook (1983) for a review. New algorithms that are able to model more complex distributions have been promoted in the connectionistic literature (Bishop 1994; Roberts and Tarassenko 1994), — here the method is often refered to as “novelty detection”. Mixure models and kernel methods are usually employed but also self-organizing maps and so-called neural tree algorithms have been used (Ypma and Duin 1998; Martinez 1998). These models have been applied in detection of, e.g., epileptic episodes in EEG (Roberts and Tarassenko 1994; Roberts 1999) as well as industrial problems like early machine fault diagnostics (Ypma and Duin 1998).

In this contribution we will focus on kernel density methods for detection of outliers in the BrainMap database (Fox and Lancaster 1994). We are not analyzing the location-behaviour correlate: Instead we will confine ourselves to the relationship between 3D coordinates in Talairach space (Talairach and Tournoux 1988) and the anatomical labels of the locations. In this connection the contaminant outliers will be transcription errors and discordant outliers will be locations that have “surprising” anatomical labeling.

METHODS

We downloaded the “paper”, “experiment” and “location” webpages from the BrainMap website (http://ric.uthscsa.edu/services/). Each “paper” contains one or more “experiments” and each “experiment” contains one or more “locations”. A “location” is a 3-dimensional coordinate representing an activation or deactivation focus with associated anatomical labeling. From the “location” webpages (each containing one Talairach coordinate) we obtained the values from two fields: “Coordinates in Talairach, 1988 space” and “Lobar anatomy”. The values (strings) from the “Lobar anatomy” field were tokenized using non-letter characters as separators. All words and all phrases where recorded and given their own class, e.g., the string “midline occipital lobe” generated an event in the classes “midline”, “occipital”, “lobe”, “midline occipital”, “occipital lobe” and “midline occipital lobe”.

[FIG. 1 about here.]

We downloaded 7263 location webpages. 3935 of these locations had an associated anatomical label, thus went into one or several of the word/phrase classes. There were 1231 word/phrase classes, see figure 1.

We next construct a probability density model in the three-dimensional Talairach space $\mathbf{x}$ by conditioning on the word/phrase w class: $p(\mathbf{x}|w)$. We use a relative simple estimator to model the probability density: A variation of the Specht kernel estimator (Specht 1990), where the width of the Gaussian kernel (σ^2) is optimized by leave-one-out (LOO) cross-validation. Our implementation is based on a fast one-dimensional Newton optimization of the leave-one-out cost function (negative log probability),

$$E(\sigma^2, w) = -\sum_{n=1}^{N_w} \log p_{-n}(\mathbf{x}_n|\sigma^2, w), \tag{1}$$

where $\mathbf{x}_n$ is the three-dimensional Talairach coordinate of the n'th location with label w, N_w is the number of locations labeled w, and the density based on all examples except the n'th is given by

$$p_{-n}(\mathbf{x}|\sigma^2, w) = \frac{1}{N_w - 1} \sum_{n' \neq n}^{N_w} \left(2\pi\sigma^2\right)^{-3/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{x}_{n'})^2\right). \tag{2}$$

The kernel method is flexible enough to model, e.g., a bimodel distribution which will be necessary in connection with a bilateral set of locations associated with the temporal lobes.

As the probability density estimation will be affected by outliers we use a two-stage heuristic: In the first stage we use all coordinates of the given class to obtain the probability density. In the second stage we exclude the 5%

most unlikely coordinates and estimate the probability density on the remaining 95%. If there are no outliers in the training data this procedure will introduce a small bias in the probability density (the width of the distribution will be under-estimated). This can potentially make the novelty detector more conservative, thus increasing the sensitivity to outliers, which we will accept for the present application. The bias could potentially be controlled by use of a set of carefully screened test foci.

Having established a probability density model we are able to evaluate new sets of foci. Novelty detection is implemented using the estimated density value $p(\mathbf{x})$ as test statistic. We rank the locations according to their densities with potential outliers among the low density values, see, e.g., (Hansen et al. 2000).

By using the bibliographic information from the BrainMap "paper" webpages the Entrez-PubMed service (http://www.ncbi.nlm.nih.gov/PubMed/) can be inquired. Authors, volume, first page and year of publication in an AND query were found to identify an article uniquely for those articles we investigated. There were a few entries with discrepancies in the bibliographic information between BrainMap and Entrez-PubMed that made the AND-query void.

RESULTS AND DISCUSSION

[FIG. 2 about here.]

In figure 2 we present the density formed by modeling the 294 entries labeled "cerebellum" in a cornercube environment (Rehm et al. 1998). The first level density is shown as wireframe, and the trimmed second level density is shown as a rendered surface. Note that two isolated "blobs" (that were created by isolated outlying locations in the training data) were correctly eliminated by the heuristic.

[FIG. 3 about here.]

Figure 3 shows the top rank outliers. The most extreme outlier is termed "SMA" and clearly has a z-coordinate that is wrong: $z = 54$ cm. This would correspond to an activation half a meter outside the brain! Such a highly abnormal location could be easily be detected, e.g., by plotting all the data in the same three-dimensional plot

A location that could not be detected using a simple measure like plotting is the seventh entry of figure 3. This location is inside the brain (see also figure 2) and is not an outlier with respect to the complete set of locations. However, when conditioning on the given label "cerebellum" it is high novelty. A systematic manual screening would require that the locations are plotted conditioned on maybe 100+ classes.

Examples where a phrase provides more information than a single word are given by the second and third entries in figure 3 — both referring to the same BrainMap location: Adding "superior" in front of "parietal" makes the location yet more unlikely.

We note that so far the analysis requires no human intervention, e.g., manual selection of the set of analyzed locations as in most current meta-analysis. To investigate the cause of the novelty it is now necessary to manually acquire, read and interpret the articles with the associated outlier locations. To speed up the investigation the context should be readily available: In figure 3 is shown hyperlink to the Brain database and — if available — a link to Entrez-PubMed and the full text article at the publisher.

[Table 1 about here.]

Table i shows a listing of the 21 most "novel" locations from the BrainMap database as well as our "manual" interpretation of the cause of the outlier. Most of the errors can be characterized as database entry errors:

The typical entry error is where the reported coordinates are given in millimeter and one of them has been interpreted as centimeter (during data entering). In table i entries 1, 3, 13 and 17 are examples of this. Errors like this are easy to resolve by reading the article and comparing it with the BrainMap entry. To resolve the cause of the novelty for other locations we contacted authors by email: The large novelty of the second entry was due to the error in the sign of the z-coordinate: $z = -51$ should have been $z = 51$ (Maurizio Corbetta, personal communication). The 15th outlier was due to a location being mixed up with an other location: The reported coordinate $-24, 42, 4$ should be either $-42, -14, 0$ or $-44, -12, 4$ (Endel Tulving, personal communication). The 16th entry was perhaps due to the x- and y-coordinate being permuted (David J. Brooks, personal communication).

Other entries are discordant outliers: Entries 4, 8, 12 and 21 from table i are all correct. In all four cases the word "lobe" produces the high novelty, and this is due to that in 71 of the 82 locations associated with "lobe" the word appear in connection with "parietal" and in 6 locations in connection with "occipital". Thus the "lobe" probability density volume is focused on these two particular lobes and locations in other lobes will have low probability density, i.e., inflated high novelty. Hence the four entries are not contaminant, but rather discordant outliers induced by the less common phrases: "temporal lobe" (entry 21), "limbic lobe" (entry 8), "subgyral frontal lobe" (entry 4) and "medial frontal lobe" (entry 12).

The 19th entry has a high novelty due to the word "area": The word appear 131 times in our data but only 5 times in connection with "Wernicke's area". The outlier entry is almost 2 centimeters below the AC-PC plane while the four other locations are around 1 centimeter above the AC-PC plane with two on the right and two on the left. The reported Wernicke's area location is a discordant outlier because it is located in the inferior/middle temporal gyrus while Wernicke's area is usually located more superior[1].

[FIG. 4 about here.]

[FIG. 5 about here.]

A possible alternative scheme for detecting of novelty in the BrainMap database would be to use anatomical atlases: Figure 4 shows the Talairach cerebellum from a triangulation of manually digitized points on the surface using the nuages program (Geiger 1993). Many of the locations lie outside the Talairach cerebellum. Some of these should presumably not be called outliers. Other anatomical atlases are labeled probability volumes: The ICBM atlas (Evans, Collins, and Holmes 1996) being a prominent example. In this atlas the cerebellum has been identified and each voxel is given a probability for being "cerebellum", see figure 5. It should be noted that the volume is a *probability* volume $P(w|\mathbf{x})$, rather than *density* volume $p(\mathbf{x}|w)$, and thus cannot be directly compared with the activation focus densities. Evaluating the BrainMap location in this model we find that some of the locations have zero probability of being "cerebellum": $P(w = \text{"cerebellum"}|\mathbf{x}) = 0$. Again these are presumably not outliers. The locations in figure 5 have been transformed by the inverse operation of Matthew Brett's nonlinear transformation (Brett 1999). It is possible that more complex spatial transformation such as a three-dimensional warps produce slightly better fit between the location and the probability volume, but probably not enough to encapsulate all of the variation in the coordinate labeling. Some variation might be attributable to the anatomical reference volume and the software used in the spatial normalization. The BrainMap database does not fully record this information.

Detection of outliers from their anatomical labeling as carried out here is relevant for database cleaning, while it might be of less neuroscientific interest on its own. A more interesting opportunity lies in the modeling of the relationship between coordinate/anatomical labelings and the cognitive components. This is, however, complicated by the fact that BrainMap — and the typical neuroscientific brain mapping article — does not tabelize the interpreted cognitive components for each individual location. In BrainMap the cognitive component ("behavioral domain") is associated with the anatomical information on the level of the experiment and seldom does an experiment involve only a single reported location or cluster of locations, though see, e.g., (Tervaniemi, Medvedev, Alho, Pakhomov,

[1] The Wernicke area is not distinctly defined: Reber (1995) defines Wernicke's areas as "a loosely circumscribed cortical area in the temporal region of the dominant hemisphere of the brain". Other definitions are "left posterior temporoparietal cortex" (Price et al. 1999), "temporal-occipital region" (Atkinson et al. 1990, page 344) and "superior temporal" (Fox et al. 2001).

Roudas, van Zuijen, and Näätänen 2000) where two experiments with automatic auditory processing, phonetic and musical, respectively, result in a two individual locations or clusters of locations.

Apart from novelty detection our density volumes could be used to automatically label coordinates in the style of the Talairach deamon, see, e.g., Lancaster et al. (1997, 2000).

The kernel density modeling approach applied to coordinate/anatomical labelings in relation to cognitive components would be contrary to one of the recommendations listed by (Fox, Parsons, and Lancaster 1998): It would *not address a specific hypothesis* but rather *generate hypotheses*, — hypotheses that might be non-trivial and surprising.

CONCLUSION

We have described a meta-analysis scheme for activation foci in functional neuroimages. Our approach is based on probability density modeling using a fully automatic non-parametric kernel approach. Based on data from the BrainMap database we constructed a model of the relation between anatomical labels (words and phrases) and corresponding focus location, enabling outlier detection by ranking foci novelty according to the density value. Among 21 of the most novel outliers investigated we found both discordant (infrequent or non-conforming terminology) and contaminant (e.g. transcription errors) outliers.

To our knowledge our system is the first to combine simple text analysis with spatial modeling. It can potentially assist the neuroscientists in quality control of neuroimaging data and could be part of a database entry program.

Acknowledgments

We thank Research Imaging Center, University of Texas Health Science Center at San Antonio for access to the BrainMap database. Furthermore we thank Endel Tulving, Maurizio Corbetta, David J. Brooks, Harri Jenkins and Bennett Shaywitz for answering our emails. An early version of this paper was sent to the NIPS conference (and not accepted). We thank the NIPS reviewers for some useful comments. This paper has been supported by the Danish research councils through "THOR center for Neuroinformatics" and "Interdisciplinary Neuroresearch" and the American NIH "Human Brain Project" grant R01 DA09246 and P20 MH57180. At last we thank Daniela Balslev for useful discussions.

REFERENCES

Allison T, Puce A, McCarthy G (2000). Social perception from visual cues: role of the STS region. Trends in Cognitive Sciences 4:267–278.

Atkinson RL, Atkinson RC, Smith EE, Bem DJ, Hilgard ER (1990). Introduction to Psychology (Tenth ed.). San Diego: Harcourt Brace Jovanovich.

Becker JT, Mintun MA, Diehl DJ, Dobkin J, Martidis A, Madoff DC, Dekosky ST (1994). Functional neuroanatomy of verbal free recall: A replication study. Human Brain Mapping 1:284–292.

Beckman RJ, Cook RD (1983). Outlier.........s. Technometrics 25:119–149. Discussions 150–161.

Bishop CM (1994). Novelty detection and neural network validation. IEE Proceedings — Vision, Image and Signal Processing 141:217–222. Document No. 19941330.

Brett M (1999). The MNI brain and the Talairach atlas. Internet: http://www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html.

Buckner RL, Raichle ME, Miezin FM, Petersen SE (1996). Functional anatomic studies of memory retrieval for auditory words and visual pictures. The Journal of Neuroscience 16:6219–6235.

Cabeza R, Nyberg L (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. Journal of Cognitive Neuroscience 12:1–47.

Corbetta M, Miezin FM, Shulman GL, Petersen SE (1993). A PET study of visuospatial attention. The Journal of Neuroscience 13:1202–1226.

Decety J, Grèzes J (1999). Neural mechanisms subserving the perception of human action. Trends in Cognitive Sciences 3:172–178.

Evans AC, Collins DL, Holmes CJ (1996). Automatic 3D regional MRI segmentation and statistical probability anatomy maps. In Myers R, Cunningham VJ, Bailey DL, and Jones T, editors. Quantification of Brain Function Using PET, P 123–130. Academic Press, Inc.

Farah MJ, Aguirre GK (1999). Imaging visual recognition: PET and fMRI studies of the functional anatomy of human visual recognition. Trends in Cognitive Sciences 3:179–186.

Fox PT, Huang A, Parsons LM, Xiong JH, Zamarippa F, Rainey L, Lancaster JL (2001). Location-probability profiles for the mouth region of human primary motor-sensory cortex: Model validation. NeuroImage

13:196–209.

Fox PT, Huang AY, Parsons LM, Xiong JH, Rainey L, Lancaster JL (1999). Functional volumes modeling: Scaling for group size in averaged images. Human Brain Mapping 8:143–150. Special Issue: Proceedings of the BrainMap '98 Workshop.

Fox PT, Lancaster JL (1994). Neuroscience on the net. Science 266:994–996.

Fox PT, Lancaster JL, editors (2000). Sixth International Conference on Functional Mapping of the Human Brain. NeuroImage, Volume 11. Academic Press.

Fox PT, Lancaster JL, Parson LM, Xiong JH, Zarnarripa F (1997). Functional volumes modeling: Theory and preliminary assessment. Human Brain Mapping 5:306–311.

Fox PT, Parsons LM, Lancaster JL (1998). Beyond the single study: functional/location metanalysis in cognitive neuroimaging. Current Opinion in Neurobiology 8:178–187.

Geiger B (1993). Three-dimensional modeling of human organs and its application to diagnosis and surgical planning. Technical Report 2105, Institut National de Recherche en Informatique et Automatique, 06902 Sophia Antipolis, France.

George MS, Ketter TA, Parekh BA, Gill DS, Huggins T, Marangell L, Pazaglia PJ, Post R (1994). Spatial ability in affective illness: differences in regional brain activation during a spatial matching task ($H_2^{15}O$ PET). Neuropsychiatry, Neuropsychology, and Behavioral Neurology 7:143–153.

Grafton ST, Woods RP, Mazziotta JC (1993). Within-arm somatotopy in human motor areas determined by positron emission tomography imaging of cerebral blood flow. Experimental Brain Research 95:172–176.

Hamilton R, Keenan JP, Catala M, Pascual-Leone A (2000). Alexia for braille following bilateral occipital stroke in an early blind woman. NeuroReport 11:237–240.

Hansen LK, Sigurdsson S, Kolenda T, Nielsen FÅ, Kjems U, Larsen J (2000). Modeling text with generalizable Gaussian mixtures. In Proceedings of ICASSP'2000. Turkey.

Howard D, Patterson K, Wise R, Brown WD, Friston KJ, Weiller C, Frackowiak RSJ (1992). The cortical localization of the lexicons. Brain 115:1769–1782.

Imaizumi S, Mori K, Kiritani S, Kawashima R, Sugiura M, Fukuda H, Itoh K, Kato T, Nakamura A, Hatano K, Kojima S, Nakamura K (1997). Vocal identification of speaker and emotion activates different brain regions. Neuroreport 8:2809–2812.

Indefrey P, Levelt WJM (2000). The neural correlates of language production. In Gazzaniga M, editor. The Cognitive Neurosciences (2nd ed.)., P 845–865. Cambridge, MA: MIT Press.

Kosslyn SM, Alpert NM, Thompson WL, Chabris CF, Rauch SL, Anderson AK (1994). Identifying objects seen from different viewpoints, a PET investigation. Brain 117:1055–1071.

Lancaster J, Kochunov P, Woldorff M, Liotti M, Parsons M, Rainey L, Nikerson D, Fox PT (2000). Automatic Talairach labels for functional activation sites. See Fox and Lancaster (2000), P S483.

Lancaster JL, Rainey L, Summerlin J, Freitas CS, Fox PT, Evans AC, Toga AW, Mazziotta JC (1997). Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. Human Brain Mapping 5:238–242.

Leblanc R, Meyer E, Bub D, Zatorre RJ, Evans AC (1992). Language localization with activation positron emission tomography scanning. Neurosurgery 31:369–373.

Lloyd D (1999). Terra cognita: From functional neuroimaging to the map of the mind. http://www.trincoll.edu/˜dlloyd/terra.html. (Submitted to Brain and Mind).

Lloyd D (2000). Multivariate meta-analysis of studies in the brainmap archive. See Fox and Lancaster (2000), P S911.

Martinez D (1998). Neural tree density estimation for novelty detection. IEEE Transactions on Neural Networks 9:330–338.

O'Sullivan BT, Roland PE, Kawashima R (1994). A PET study of somatosensory discrimination in man. Microgeometry versus macrogeometry. European Journal of Neuroscience 6:137–148.

Owen AM, Milner B, Petrides M, Evans AC (1996). Memory for object features versus memory for object location: A positron-emission tomography study of encoding and retrieval processes. Proceedings of the National Academy of Sciences USA 93:9212–9217.

Pardo JV, Raichle ME, Fox PT (1991). Localization of a human system for sustained attention by positron emission tomography. Nature 349:61–63.

Paus T (1996). Location and function of the human frontal eye field. Neurophychologia 34:475–483.

Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. Nature 331:585–589.

Price CJ, Veltman DJ, Ashburner J, Josephs O, Friston KJ (1999). The critical relationship between the timing of stimulus presentation and data acquisition in blocked designs with fMRI. NeuroImage 10:36–44.

Reber AS (1995). The Penguin Dictionary of Psychology. Penguin.

Rehm K, Lakshminaryan K, Frutiger SA, Schaper KA, Sumners D, Strother SC, Anderson JR, Rottenberg DA (1998). A symbolic environment for visualizing activated foci in functional neuroimaging datasets. Medical Image Analysis 2:215–226.

Roberts S, Tarassenko L (1994). A probabilistic resource allocating network for novelty detection. Neural Computation 6:270–284.

Roberts SJ (1999). Novelty detection using extreme value statistics. IEE Proceedings-Vision, Image and Signal Processing 146:124–129.

Schlösser R, Hutchinson M, Joseffer S, Rusinek H, Saarimaki A, Stevenson J, Dewey SL, Brodie JD (1998). Functional magnetic resonance imaging of human brain activity in a verbal fluency task. Journal of Neurology, Neurosurgery, and Psychiatry 64:492–495.

Shaywitz BA, Pugh KR, Constable RT, Shaywitz SE, Bronen RA, Fulgright RK, Shankweiler DP, Katz L, Fletcher JM, Skudlarski P, Gore JC (1995). Localization of semantic processing using functional magnetic resonance imaging. Human Brain Mapping 2:149–158.

Specht DF (1990). Probabilistic neural networks. Neural Networks 3:109–118.

Talairach J, Tournoux P (1988). Co-planar Stereotaxic Atlas of the Human Brain. New York: Thieme Medical Publisher Inc.

Tervaniemi M, Medvedev SV, Alho K, Pakhomov SV, Roudas MS, van Zuijen TL, Näätänen R (2000). Laterilized automatic auditory processing of phonetic versus musical information: A PET study. Human Brain Mapping 10:74–79.

Tulving E, Kapur S, Markowitsch HJ, Craik FIM, Habib R, Houle S (1994). Neuroanatomical correlates of retrieval in episodic memory: Auditory sentence recognition. Proceedings of the National Academy of Sciences of the USA 91:2012–2015.

Wills AJ, Jenkins IH, Thompson PD, Findley LJ, Brooks DJ (1994). Red nuclear and cerebellar but no olivary activation associated with essential tremor: a positron emission tomographic study. Annals of Neurology 36:636–642.

Ypma A, Duin RPW (1998). Novelty detection using self-organizing maps. In Kasabov N, Kozma R, Ko K, O'Shea R, Coghill G, and Gedeon T, editors. Progress in Connectionist-Based Information Systems. Proceedings of the 1997, International Conference on Neural Information Processing and Intelligent Information Systems (ICONIP97), Volume 2, Singapore, P 1322–1325. IEE: Springer Verlag.

Zangaladze A, Epstein CM, Grafton ST, Sathian K (1999). Involvement of visual cortex in tactile discrimination of orientation. Nature 401:587–590.

Fig 1 The 1231 classes sorted according to frequency. The most occurring words are “gyrus”, “cortex” and “frontal”. The most occurring phrase is “frontal gyrus” as the 12th most frequent class. There are approximately 600 classes with more than two examples.

Fig. 2 Probability density estimate of the “cerebellum” class in Talairach space in a corner cube environment. The wireframe-model is the first stage probability density estimation where all the locations are included and the polygon model is the second stage probability density estimate where the 5% most extreme are excluded. Note that two isolated “blobs” created by isolated, outlying locations were eliminated going from the first to the second level density. This figure as well as figures 4 and 5 are made with the Brede Matlab toolbox available at http://hendrix.imm.dtu.dk/software/brede.

Fig. 3 An automatically generated list of those locations estimated to have the highest novelty. “Paper”, “Exp.” and “Loc.” correspond to the identifiers used in the BrainMap database. X, y, z and “Lobar anatomy” are the associated fields in the database with the coordinates in centimeter and the “loglikehood” is our novelty measure. The “Full text” column indicates whenever it is possible to extract a link from the Entrez-PubMed to the electronic full text of the articles.

Fig. 4 Surface of the cerebellum from the Talairach Atlas with the “cerebellum” locations. The inferior part of the cerebellum is not in the atlas, thus not in the visualization. The contour shadows are the convex hull of the digitized Talairach cerebellum.

Fig. 5 The ICBM cerebellum with all the cerebellum locations from BrainMap. The locations have been transformed by the inverse operation of Matthew Brett’s nonlinear transformation (see text).

Table i BrainMap outliers. The entries are ordered according to novelty. The second column indicates the paper, experiment and location identifier of the BrainMap database. The third to fifth column are x, y and z with the "reported" coordinates from BrainMap (*not* the corrected "Talairach 1988" coordinates).

No.	BrainMap	x	y	z	BrainMap label	Comment	Reference
1	267, 2, 1	−5	7	540	SMA	Millimeter and centimeter for z-coordinate confused during BrainMap entry	(Buckner et al. 1996, table 4, entry 1)
2	29, 10, 8	48	−23	−51	Lateral superior parietal	Resolved: Transcription mistake.	(Corbetta et al. 1993, table 5)
3	141, 1, 10	35	150	28	Dorsolateral prefrontal	Millimeter and centimeter for y-coordinate confused during BrainMap entry	(Kosslyn et al. 1994, tabel 2, entry 10)
4	249, 1, 59	−31.8	48.1	2.2	Subgyral frontal lobe	Correct	S. K. Brannan, 1997, Unpublished
5	280, 1, 9	24	−70	−24	Dorsal parietal cortex	Is labeled “Right cerebellum” in the article	(Schlösser et al. 1998, table 1, entry 9)
6	4, 2, 7	−6	42	−8	Cerebellum — superior anterior	Not possible to find the foci in the article.	(Petersen et al. 1988)
7	280, 1, 7	38	24	−8	Dorsolateral parietal	Is labeled “Right orbitofrontal cortex” in the article	(Schlösser et al. 1998, table 1, entry 7)
8	249,1,29	−2	26	16	Limbic Lobe	Correct	S. K. Brannan, 1997, Unpublished
9	277, 3, 3	−50	−42	−14	Inferior frontal gyrus, posterior	Is labeled “inferior temporal gyrus posterior (area 37)” in the article	(Owen et al. 1996, table 2, entry 3)
10	115, 2, 5	−38	54	0	Middle temporal gyrus	Not resolved.	(Shaywitz et al. 1995, page 155)
11	19,2,17	24	−47	38	Frontal	Not resolved	(Pardo et al. 1991, Table 1a, entry 17)
12	47,4,1	−36	32	28	Medial frontal lobe	Correct	(George et al. 1994)
13	65, 2, 23	57	26	45	Anterior cingulate	Millimeter and centimeter for x-coordinate confused during BrainMap entry.	(O’Sullivan et al. 1994, table 4, entry 10)
14	52, 1, 2	36	−46	36	Inferior frontal gyrus	Probably misunderstanding of the text during entry. The foci is around supramarginal gyrus and the denoted “BA40”.	(Becker et al. 1994, page 287)
15	61, 1, 12	−24	42	4	Temporal/insular	Resolved: Transcription mistake.	(Tulving et al. 1994, table 1)
16	130,5,8	−38	−8	4	cingulate	Perhaps a transcription error with the x- and y-coordinate being permuted	(Wills et al. 1994, table 5, entry 14)
17	48,2,3	80	−56	−16	anterior cerebellum	Millimeter and centimeter for x-coordinate confused during BrainMap entry	(Grafton et al. 1993, table 1, entry 18)
18	273,1,6	43	−14	15	parietal-occipital junction	Not resolved	(Imaizumi et al. 1997, table 1, entry 6)
19	89,1,8	−58	−37	−17	Wernicke’s area	Correct, though labeled “Lt inferior temporal gyrus; Lt middle temporal gyrus (Wernicke’s area)” in the article	(Leblanc et al. 1992, table 1, entry 8)
20	29,8,5	−37	−93	−8	Lingual/fusiform	Perhaps correct	(Corbetta et al. 1993, table 6)
21	26,3,4	40	−74	4	medial occipital gyrus/temporal lobe	Correct. Labeled “*middle* occipital gyrus...” in the article	(Howard et al. 1992, page 1776)

Table i. BrainMap outliers. The entries are ordered according to novelty. The second column indicates the paper, experiment and location identifier of the BrainMap database. The third to fifth column are x, y and z with the “reported” coordinates from BrainMap (*not* the corrected “Talairach 1988” coordinates).