Published in final edited form as:

Neuroimage. 2011 July 15; 57(2): 331–333. doi:10.1016/j.neuroimage.2010.11.012.

Exploring functional relations between brain regions from fMRI meta-analysis data: Comments on Ramsey, Spirtes, and Glymour

Jane Neumann^{a,b}, Robert Turner^a, Peter T. Fox^c, and Gabriele Lohmann^a

^a Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

^b Leipzig University Medical Center, IFB Adipositas Diseases, Leipzig, Germany

^c Research Imaging Center, University of Texas Health Science Center, San Antonio, Texas, USA

Abstract

In this paper we address Ramsey et al.'s critical assessment of our method for learning partially directed graphs from meta-analysis imaging data (Neumann et al., 2010). We argue that our method provides valid and interpretable results when applied to data representing a single experimental paradigm. Simulations further suggest that, despite theoretical limitations, the application of our method to mixed probability distributions yields reliable results with error rates at acceptable levels. Finally, we discuss the nature of meta-analysis data and the notion of causality in the context of functional neuroimaging.

Ramsey, Spirtes, and Glymour address the problem of learning causal relations from data representing a mixture of probability distributions. This is an important problem which, no doubt, deserves careful consideration in the context of functional neuroimaging. However, the critical discussion of our recent paper (Neumann et al., 2010) which the authors offer in light of this problem is based, in substantial parts, on a misconception and misrepresentation of our goals, the scope of our method, and the nature of meta-analysis data. In the following we wish to clarify some of these issues. Most importantly, we will argue that our method provides valid and interpretable results when applied to meta-analysis data representing a single or very closely related experimental paradigms for which a common underlying cognitive network can be assumed. In addition, simulations suggest that the application of our method to data from mixed probability distributions yields reliable results with error rates in the range of typical acceptance levels, despite the theoretical limitations discussed by Ramsey and colleagues.

Learning partially directed networks

Contrary to the assumption of Ramsey et al., our work *does not* aim to provide a general method for the combination of data from mixed probability distributions. Instead, we

Corresponding Author: Dr. Jane Neumann, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, D-04103 Leipzig, Germany, Phone: +49 (0)341 9940 2621, Fax: +49 (0)341 9940 2448, neumann@cbs.mpg.de.

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

attempt to explore possible functional relations between brain regions from co-activation patterns of these regions observed in independently performed imaging experiments.

The main building block of our method is an algorithm for learning the structure of directed acyclic graphs (DAG) from observational data. Learning the structure of a DAG is difficult, being limited by the size and quality of the available data. In a noisy and small data set, the probability distribution represented by the data will not perfectly match the true distribution we wish to learn. No structure learning technique, including the procedures suggested by Ramsey et al., is able to learn the corresponding Markov equivalence class with complete reliability¹. Thus, when we apply structure learning to real world data, we will always have to accept possible false positive dependencies or undetected relationships in our results; see e.g. Koller and Friedman (2009, Chapt. 18) for a detailed discussion of this issue.

In order to keep such errors at an acceptable level, we proposed to apply structure learning repeatedly, rather than using it as a ‘one-shot’ learning mechanism, and to assess the reproducibility of the results. This yields a graph containing directed and undirected edges repeatedly discovered from different subsets of the available data. Ramsey and colleagues are correct in their claim that this graph does not necessarily represent a Markov equivalence class. In fact, we made no such claim in our paper. Although in every repetition of the procedure a Markov equivalence class is obtained, the final graph combining all reliably learned edges does not preserve Markov properties.

In our view, however, the results remain interpretable. For example, the directed links from ‘parent’ to ‘child’ nodes in a DAG represent the fact that the random variable associated with the child is conditioned on the random variables of the parents². In our context, this translates into the observation that activation of one region (child) statistically depends on the activation of others (parents) in the experiments investigated. Thus, if we can assume, from additional evidence, some interaction or influence between parents and child, it is more likely that activation of the parents has a direct or indirect effect on activation of the child than vice versa.

Importantly, the edges obtained by our procedure are determined with high reproducibility and are not discovered only as part of a Markov equivalence class in a ‘one-shot’ learning scheme. They are thus less likely to represent only spurious connections than any edge obtained from the single application of structure learning to noisy data.

Observations as described above enable us to derive, from observational data and in an exploratory fashion, hypotheses about probable dependencies between brain regions in a given context. These hypotheses can, for example, serve as input for confirmatory methods. Such methods require strong hypotheses about probable connectivity patterns, since testing all theoretically possible network configurations is impossible due to computational constraints. In our view, a mechanism for exploring possible functional connectivities between brain regions from observational data is of high scientific value.

¹This was confirmed by our first simulations. From several hundred co-activation patterns, structure learning *sometimes but not always* successfully learned the correct Markov equivalence class. Importantly, this result was *not* an artefact of our particular learning algorithm, as an exhaustive search over all theoretically possible DAGs with three and four nodes yielded identical results. Unfortunately, an exhaustive search is only computationally feasible for small networks due to the super-exponential growth of theoretically possible network structures with increasing network size.

²One can not, of course, interpret individual edges in a graph in isolation; this holds true for both DAGs and the graphs representing our learning results.

The nature of meta-analysis data

Contrary to the assumption of Ramsey et al., in our method we *do not* combine fMRI time series across subjects. Instead, in a meta-analysis we combine lists of coordinates in MNI or Talairach space that encode co-activation patterns obtained in previous imaging experiments of groups of subjects.

Each such list represents the spatial distribution of activation maxima across cortical and subcortical regions of the brain. It is important to make this distinction, as fMRI time series are not directly comparable across subjects, and simply applying structure learning to a pool of such time series would be fatal. In contrast, imaging studies employing the same experimental task are generally assumed, and in fact commonly observed, to yield comparable activation patterns³. In fact, most meta-analyses select experiments based on their common cognitive demand. Co-activation patterns from such experiments are thus assumed to reflect a single joint probability distribution. This is an assumption⁴ that we need to make for most comparisons across imaging experiments and in any attempt to replicate imaging results. Under this assumption, learning the structure of a DAG from co-activation patterns is a valid approach.

Some differences in the activation patterns will of course exist across experiments, arising from differences in scanning protocols and stimulus presentation, data modelling techniques, statistical tests, etc. Our observational data will thus represent the joint probability distribution in a noisy manner. This noise can be reduced, though not fully eliminated, prior to structure learning by meta-analysis techniques such as Activation Likelihood Estimation (Turkeltaub et al., 2002) and replicator dynamics (Neumann et al., 2005). These methods facilitate the extraction of co-activation patterns that are frequently observed across studies, thereby reducing spurious results in the individual studies and outliers. In our paper, we suggested to apply such methods as preprocessing steps to structure learning.

Learning from mixed probability distributions

As discussed in our paper, we expect our method to provide very reliable results when applied to data covering a single experimental paradigm⁵. This was strongly supported by the results from our first simulations, not discussed by Ramsey et al., where test data was sampled from a single joint probability distribution. However, we also pointed out that statistical dependencies between brain regions will almost certainly differ to some degree across different experimental paradigms. Here, the assumption of a single joint probability distribution might not hold, and we face the limitations on structure learning formulated in Yule's theorem (Yule, 1903).

A graph learned from mixed distributions might not fully represent all conditional independence relations present in the observed system. In particular, learning might result in novel dependencies, referred to as 'false positives' in Ramsey et al.'s simulations, that are absent in the individual joint distributions. The precise relationship between graphs resulting

³There are special cases where this assumption is violated, for example, in tasks where subjects are likely to employ different problem solving strategies, or when comparing groups of healthy subjects and patients. Here, special care needs to be taken in data analysis and interpretation.

⁴Validating this assumption in a specific case is hard, as assessing the true properties of the joint distribution would require some way of 'unmixing' it. While Ramsey et al. discuss possible approaches to 'unmixing' probability distributions, they acknowledge that none of the currently available methods is feasible for fMRI data.

⁵In our paper, the need to apply our method to real world data representing several (rather than one) experimental paradigms arose from the limited size of data sets available for individual paradigms. This problem will continuously be alleviated given the exponentially growing number of imaging studies performed each year. This will shortly allow for the application of our method to data representing individual cognitive tasks.

from mixed and individual distributions is hard to assess. Nevertheless, our and Ramsey et al.'s own simulations support the observation that applying structure learning to mixed distributions provides results with error rates at acceptable levels.

Our own simulations show that in the case of unequal mixing proportions, structure learning is likely to produce the graph that is more frequently represented in the input data set. In practical applications, an unequal mixing proportion will more often than not be the case. A post-hoc assessment of the input data (after preprocessing) can provide additional information on the possible dominance of a paradigm and its associated cognitive processes in the input data set⁶. Such additional assessment can reach from counting and categorizing the experimental paradigms that gave rise to the individual co-activation patterns to a detailed investigation of the functional heterogeneity of a specific brain network or region as described, for example, by Laird et al. (2009) and Robinson et al. (2010).

In the more extensive simulations presented by Ramsey et al. only a single novel dependency (false positive adjacency) was obtained on average from mixed data. More interestingly, the probability of falsely obtaining a directed dependency from two mixed distributions was below 5%⁷, which in the context of empirical data analysis is in the range of typical acceptance levels⁸. We would therefore argue that the application of our method to data sets where the assumption of a single joint probability distribution might not be guaranteed is not unreasonable, provided the results are carefully interpreted as what they are: possible statistical dependencies between the activation of brain regions arising from an exploratory analysis.

Causality

Contrary to the assumption of Ramsey et al., we *do not* reject causal representations in fMRI applications. However, in our context, one can not infer causal relations from directed connections in the obtained graph.

In order to understand the conditions under which relations can be interpreted as causal, we need to make a clear distinction between associations and causality: "An associational concept is any relationship that can be defined in terms of a joint distribution of observed variables, and a causal concept is any relationship that cannot be defined from the distribution alone" (Pearl, 2009, p.99). Thus, by definition, knowing the joint distribution of observed variables is not sufficient for making causal statements, but requires additional knowledge or assumptions; see also Lindquist and Sobel (this issue) for a discussion of this problem.

Whether or not results from fMRI applications represent causal relationships or mere associations strongly depends on the particular method used for assessing them. Network analysis techniques such as DCM employ causal concepts. However, in these approaches causal relations are not discovered or learned, rather, they are presupposed in an *a priori* model and tested against the available data. In other words, causality does not arise from the data, it is assumed. Likewise, methods for causal discovery, including those mentioned by

⁶Such detailed analysis was beyond the scope of our original paper which was aimed at introducing the technique rather than discovering particular relations between brain regions. Consequently, we refrained from any strong interpretation of our result graphs.

⁷A graph with 8 nodes can contain 28 edges with 3 possible directionalities each. For the extreme cases of 2 completely non-overlapping and 2 identical input graphs in a mixture, a maximum of 76 and 68 possible false positive directionalities can be obtained, resp.

⁸In our view counting the number of false negatives and true positives the way suggested in the authors' simulations is misleading. The union of two DAGs will more often than not result in a graph which is no longer a DAG. By definition, this graph can not fully be recovered by structure learning for DAGs and there are no means of determining which of the edges should belong to any 'correct' result graph. Consequently, true positives and false negatives can not be determined.

Ramsey and colleagues, need to rely on assumptions such as causal faithfulness. As the authors state, such search procedures are only consistent estimators of causal relations “subject to sampling and distribution assumptions”. Without these assumptions⁹, causal relations cannot be inferred from data. To quote Pearl (2009, p99) again: “one cannot substantiate causal claims from associations alone, even at the population level—behind every causal conclusion there must lie some causal assumption that is not testable in observational studies.”

Causal assumptions necessary for learning causal relations between brain regions are hard to validate. They are violated, for example, by the presence of hidden common causes or causal feedback. Neither of these can be ruled out in our context, given that after preprocessing, network learning is restricted to regions of interest and real brain networks most certainly contain feedback connections. Consequently, without any additional knowledge, we can not interpret the connections discovered by DAG structure learning from fMRI meta-analysis data as causal, but as statistical dependencies pointing to functional relations between brain regions. This does not mean, however, that the relations discovered by our method are meaningless – see above.

Acknowledgments

We wish to thank Nihat Ay, Max Planck Institute for Mathematics in the Sciences, Leipzig, for very helpful discussions on graphical models. This work was supported by NIH Research Grant R01 MH74457 (PI: Peter T. Fox) and the Federal Ministry of Education and Research (BMBF), Germany (FKZ 01EO001).

References

- Koller, D.; Friedman, N. Probabilistic Graphical Models: Principles and Techniques. MIT Press; 2009.
- Laird AR, Eickhoff SB, Li K, Robin DA, Glahn DC, Fox PT. Investigating the Functional Heterogeneity of the Default Mode Network Using Coordinate-Based Meta-Analytic Modeling. *Journal of Neuroscience*. 2009; 29(46):14496–14505. [PubMed: 19923283]
- Neumann J, Fox PT, Turner R, Lohmann G. Learning partially directed functional networks from meta-analysis imaging data. *NeuroImage*. 2010; 49(2):1372–1384. [PubMed: 19815079]
- Neumann J, Lohmann G, Derrfuss J, von Cramon DY. The meta-analysis of functional imaging data using replicator dynamics. *Human Brain Mapping*. 2005; 25(1):165–173. [PubMed: 15846812]
- Pearl, J. Causality. Cambridge University Press; 2000.
- Pearl J. Causal inference in statistics: An overview. *Statistics Surveys*. 2009; 3:96–146.
- Robinson JL, Laird AR, Glahn DC, Lovullo WR, TFP. Metaanalytic connectivity modeling: Delineating the functional connectivity of the human amygdala. *Human Brain Mapping*. 2010; 31(2):173–184. [PubMed: 19603407]
- Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA. Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*. 2002; 16:765–780. [PubMed: 12169260]
- Yule GU. Notes on the theory of association of attributes in statistics. *Biometrika*. 1903; 2(2):121–134.

⁹See e.g. Koller and Friedman (2009, p1040) and Pearl (2000, Chap. 2/3) for a detailed discussion of the key assumptions necessary for causal network learning.