TECHNICAL REPORT

Discrete Dynamic Bayesian Network Analysis of fMRI Data

John Burge,[1,2] Terran Lane,[1] Hamilton Link,[1] Shibin Qiu,[1] and Vincent P. Clark[2,3,4]*

[1]*Department of Computer Science, University of New Mexico, Albuquerque*
[2]*The MIND Institute, Albuquerque, New Mexico*
[3]*Department of Psychology, University of New Mexico, Albuquerque*
[4]*Department of Neuroscience, University of New Mexico, Albuquerque*

Abstract: We examine the efficacy of using discrete Dynamic Bayesian Networks (dDBNs), a data-driven modeling technique employed in machine learning, to identify functional correlations among neuroanatomical regions of interest. Unlike many neuroimaging analysis techniques, this method is not limited by linear and/or Gaussian noise assumptions. It achieves this by modeling the time series of neuroanatomical regions as discrete, as opposed to continuous, random variables with multinomial distributions. We demonstrated this method using an fMRI dataset collected from healthy and demented elderly subjects (Buckner, et al., [2000]: J Cogn Neurosci 12:24-34) and identify correlates based on a diagnosis of dementia. The results are validated in three ways. First, the elicited correlates are shown to be robust over leave-one-out cross-validation and, via a Fourier bootstrapping method, that they were not likely due to random chance. Second, the dDBNs identified correlates that would be expected given the experimental paradigm. Third, the dDBN's ability to predict dementia is competitive with two commonly employed machine-learning classifiers: the support vector machine and the Gaussian naïve Bayesian network. We also verify that the dDBN selects correlates based on non-linear criteria. Finally, we provide a brief analysis of the correlates elicited from Buckner et al.'s data that suggests that demented elderly subjects have reduced involvement of entorhinal and occipital cortex and greater involvement of the parietal lobe and amygdala in brain activity compared with healthy elderly (as measured via functional correlations among BOLD measurements). Limitations and extensions to the dDBN method are discussed. *Hum Brain Mapp 30:122–137, 2009.* © 2007 **Wiley-Liss, Inc.**

Key words: Bayesian networks; dementia; nonlinear analysis; functional connectivity; Talairach atlas; amygdala

INTRODUCTION

Many methods for identifying correlates of activity among brain regions are based on the general linear model (GLM) [Cox, 1996; Friston, 2005; Gold, et al. 1998; Smith et al., 2004] or assume that activity among regions is linearly related with a Gaussian noise model (or both). In other words, the activity found in one region, if multiplied by some scalar value, will be similar to the activity found in other correlated regions. This is a powerful and robust approach, but risks missing fundamentally nonlinear relationships, which have been shown to be important [Lahaye et al., 2003]. For example, if two brain regions acting in

concert can strongly influence a third region, but either of the two acting alone has minimal effects, the relationship will be nonlinear and would not be correctly described by a linear model. Threshold effects are also nonlinear: linear models assume that changes in activity in one region (A) have directly proportional consequences on an affected region (B). If region A pushes B into a saturated (maximum possible intensity) state, then an increase in intensity in A will produce no change in B, violating the linearity assumption. In such cases, two regions with low correlation as measured by GLM methods may actually be significantly correlated. To detect these and other interactions among brain regions, nonlinear methods must be used to increase detection sensitivity.

In the present study, we introduce methods for detecting general nonlinear functional networks from fMRI data and demonstrate their application to a functional magnetic resonance imaging (fMRI) dataset. The dataset was obtained by Buckner et al. [2000] and provided by the fMRI Data Center [fMRIDC, 2004]. Buckner et al. examined the neurophysiological and hemodynamic correlates of healthy aging and dementia by acquiring fMRI data during a visual-motor response task. Dementia was of the Alzheimer's type and measured by the Clinical Dementia Rating [Morris, 1993]. Using GLM based analysis methods, Buckner et al. found qualitatively similar activation maps among all groups, but quantitatively differing amplitude responses in only a few select regions. Over all, they described few quantitative differences among groups.

Our approach is based on the Bayesian network (BN) framework [Pearl, 1986]. Specifically, we employ discrete dynamic Bayesian networks (dDBNs) [Murphy, 2002] that are capable of identifying general nonlinear correlates among ROIs from quantized fMRI data. We validate elicited correlates in three ways. First, we show that the elicited correlates are robust over leave-one-out cross-validation and, based on a Fourier bootstrapping confidence test [Prichard and Theiler, 1994], that they are not likely due to chance. Second, we show that the dDBNs identify correlates that would be expected given the experimental paradigm, for example, between visual and motor processing ROIs. Third, we create a BN classifier and compare its discriminative accuracy with that of two other machine-learning classifiers: support vector machines (SVMs) [Burges, 1998] and Gaussian naïve Bayesian networks (GNBNs) [Duda et al., 2001].

While the BNs are shown to be competitive with both the SVM and the GNBN, classification is not our primary focus in this article. It is used to verify that important features contained within the fMRI dataset are not obscured when using dDBN-modeling techniques on quantized data. For a more detailed discussion on classification, see [Burge and Lane 2005a,b, 2006].

Finally, we provide a brief analysis of the correlates elicited from Buckner et al.'s data that suggests there is reduced influence of entorhinal and visual cortex and an increased influence of the amygdala and parietal cortex in the neural networks of mildly demented subjects.

METHODS

Bayesian and Dynamic Bayesian Networks

For an introductory overview of Bayesian networks (BNs), we refer the reader to [Charniak, 1991], and for a detailed analysis to [Heckerman et al., 1995; Jensen, 2001; Perl, 1986]. A Bayesian network is a graphical model that compactly represents the joint probability distribution of a set of n random variables (RVs): X_1, X_2,..., X_n. The RVs are represented as nodes in the BN's directed acyclic graphical topology and correlations between RVs are represented as directed links between *child* and *parent* nodes. A child and its set of parents, Pa(X_i), is a *family*. For BNs that model discrete random variables (dBNs), the relationship between a child and its parents is parameterized by a conditional probability table (CPT) that explicitly parameterizes all possible parent and child value combinations. Figure 1 gives a simple hypothetical dBN that models the risk of having dementia.

A discrete dynamic Bayesian network (dDBN) is a specialization of a dBN that models temporal processes. Its graphical topology is divided into columns of nodes such that each column represents a time frame. Each random variable is represented by one node in each of the columns. Links are allowed to connect nodes between columns, provided the link points forward in time. Ideally, there would be one column for every time frame and links could connect nodes separated by arbitrary time steps (including nodes in the same time frame). However, such dDBNs are intractably large and require far more data and computational resources to learn than is likely to be available.

To deal with this intractability, we make the Markov order 1 and stationary assumptions [Papoulis, 1991]. This results in dDBN topologies composed of two columns of RVs. Links between the columns represent temporally consecutive correlations among RVs averaged across the entire time series. As with many modeling assumptions, there is a tradeoff between specificity and tractability. Not including isochronal links prevents correlations within the same time frame from being modeled and the stationary assumption will cause nonstationary effects that likely exist in fMRI data [Bhattacharya et al., 2006] to be averaged together. This tradeoff may be reasonable for many effects, but known egregious nonstationary effects should be dealt with via the inclusion of additional nodes. E.g., correlations among ROIs that change due to differing external stimulus can be parameterized separately via the inclusion of a Stimulus node (see next section). Further, a sensitivity analyses determining the optimal TR lag would be warranted (though was not an option in our experiments as we were using a preexisting fMRI dataset).
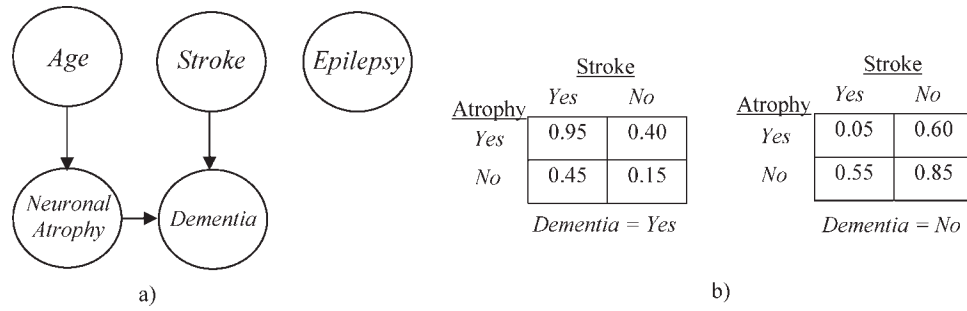
Figure 1.

(a) Hypothetical example of a Bayesian network modeling the risk of dementia. The directed links connecting Dementia with Stroke and Neuronal Atrophy indicate being demented is directly influenced by these conditions, and indirectly influenced by age. The absence of a link between Dementia and Epilepsy indicates that the presence of dementia is not influenced by the presence of epilepsy. (b) Hypothetical conditional probability table for the Dementia node. The table indicates, among other things, that if someone has neuronal atrophy and has had a stroke, there is a 0.95 probability they will be demented. There are conditional probability tables for all other nodes in the graph that are not shown.

Figure 2 gives an example of how dDBNs can be used to model fMRI data. In the example are three ROIs: the amygdala, BA 7, and BA 40. A hypothetical time series based on a simple 1-bit *high* versus *low* quantization is given for each region. The $BA40_{t+1}$ node is shown as a child with three candidate parents. Proposing each candidate as the sole parent of the $BA40_{t+1}$ node results in a corresponding CPT describing the statistical relationship between the parent's values and subsequent children val-

ues. For example, the top CPT describes the relationship between the $Amygdala_t$ node and the $BA40_{t+1}$ node. The bottom-left entry of 0.33 indicates that when the amygdala is *high*, BA 40 has a 33% chance of being *low* in the next time step.

The least uniform CPT results when the $BA7_t$ node is the parent. Thus, BA 7 is considered the best predictor of the subsequent values of BA 40. This simple example only includes pairwise relationships, but CPTs can model more

Figure 2.

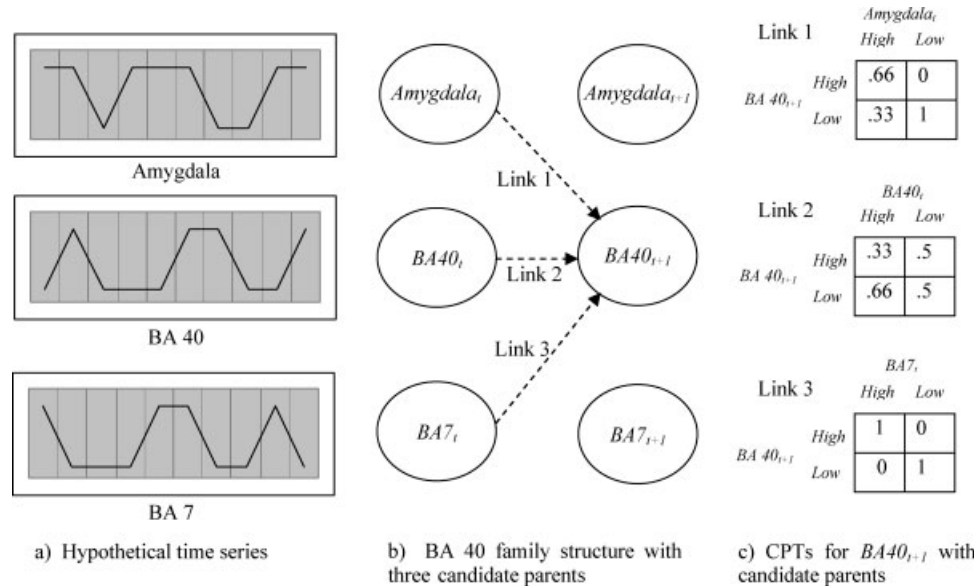Example of a dDBN modeling three ROIs. (a) Hypothetical ROI average voxel intensity time series for each of the regions. (b) dDBN for three ROIs showing three candidate parents for the $BA40_{t+1}$ node (other links would exist but are not shown for clarity). (c) The CPTs for the $BA40_{t+1}$ node corresponding to each of the three candidate parents. Notice that BA 7 is a perfect predictor of BA 40 in this example.

1. Initialize the BN with no links between t nodes and $t+1$ nodes and with the *Stimulus* node set as a parent of all other nodes.

2. Repeat for every node, X_i, in column $t+1$

3.   Repeat *numParents* times

4.   Repeat for every node, X_k, in column t

5.   Add X_k as a parent of X_i,

6.   Calculate the BDe score for X_i (i.e., measure goodness of fit)

7.   Remove X_k from $Pa(X_i)$

8.   End 4.

9.   Permanently add the parent X_k to node X_i with highest BDe score

10.   End 3.

11. End 2.

12. Empty the parent set of all but *numBestToKeep* of the highest scoring *families*

13. Return high scoring and high confidence structures

Figure 3.
dDBN structure search algorithm used to locate structures that describe the data well.

complex interactions between many simultaneous parents and the child node.

Structure Search

When a dDBN's topological structure is not known a priori, it must be searched for. The formulation of dDBN search, and BN search in general, is a well-understood problem [Heckerman et al., 1995] and known to be NP-Hard [Chickering et al., 1994]. Ideally, every possible structure would be individually constructed and the one that best modeled the data would be chosen. A function, referred to as a structure score, must be devised that is capable of indicating how well a given structure represents the relationships among variables in the data.

Several scoring functions have been proposed. These include the Akaike Information Criterion (AIC) [Akaike, 1973], Bayesian Information Criterion (BIC) [Schwarz, 1978], Bayesian Dirichlet score (BD) [Cooper and Herskovits, 1992], Bayesian Dirichlet Equivalence score (BDe) [Heckerman et al., 1995], etc. All of these scores attempt to balance the fit of a proposed model to the data with the model's complexity (a necessity given that the best models are otherwise the most complex models). AIC and BIC do this by penalizing the posterior likelihood of the model based on the number of parameters the model requires. BD and BDe take a Bayesian approach and integrate out the model's parameters (under the philosophy that all settings to the parameters are potentially correct) and use Bayesian statistics with Dirichlet priors. We use the BDe score as it is widely employed and well understood.

Like most scoring metrics, the BDe score decomposes into the aggregation of the scores of each of the BN's families. The BDe score for a family with child X_i and parents $Pa(X_i)$ given a dataset D is:

$$BDe(X_i, Pa(X_i) \mid D)$$

$$= \prod_{j=1}^{q_i} \left[\frac{\Gamma\left(N'_{i,j}\right)}{\Gamma\left(N_{i,j} + N'_{i,j}\right)} \prod_{k=1}^{r_i} \frac{\Gamma\left(N_{i,j,k} + N'_{i,j,k}\right)}{\Gamma\left(N'_{i,j,k}\right)} \right]$$

where r_i is the arity of node X_i, q_i is the number of combinations of values the nodes in $Pa(X_i)$ can take, $N_{i,j}$ is the number of times $Pa(X_i)$ is in its *j*th configuration in D and $N_{i,j,k}$ is the number of times $X_i = k$ while $Pa(X_i)$ is in its *j*th configuration in D. Unlike non-Bayesian scores, the BDe score does not incorporate any CPT probabilities, as they have been marginalized away. The N' values define the Bayesian prior information. We use uninformative priors, setting $N'_{i,j} = q_i$ and $N'_{i,j,k} = 1$.

As BDe is decomposable, the parent set for each family can be searched for independently. We employ a greedy algorithm (also referred to as a forward stepwise selection algorithm), detailed in Figure 3, to perform the structure search. For each node in column $t+1$, the algorithm finds the single best predictive parent in column t. It then iteratively adds subsequent parents that make the best predictor in conjunction with previously added parents. However, since structure search is NP-Hard, this heuristic (like all other heuristics) is not guaranteed to find the optimal solution.

The search algorithm is parameterized by two variables, *numParents* and *numBestToKeep*, which indicate the maximum number of parents each family is allowed to have and how many families are used for classification, respectively. When training dDBNs to be used in a classifier, it is important to set these RVs to values that avoid overfitting. See the *dDBN Classification Efficacy* section for more details. When training dDBNs to identify ROI correlates, *numBestToKeep* was set at its maximum value (allowing all families

to retain their structure) and *numParents* was set to the maximum value found to avoid overfitting during the classification task.

In addition to the links identified in the search, a binary Stimulus node was added to the network and set as the parent of all other nodes. This doubles the size of every CPT, allowing each node to have an independent set of parameters for each of the two possible stimuli presented to the subjects. This is important as a subject's neural response will likely be different given different stimulation and thus will gregariously violate the stationary assumption made by the dDBNs.

As with all Bayesian techniques, prior information can guide the results of the model. While calculating the BDe score for each proposed network (Step 6 of the algorithm), we employed an uninformed structure prior and an uninformed conditional probability prior. This asserts that there is no bias towards believing any particular group of ROIs are more or less likely to appear interconnected in the network (though highly connected networks are penalized). Nor is there any bias on how relationships should be parameterized.

Since the links in a BN are directed, BNs are capable of representing causal relationships. However, when the links are selected during structure search, causality between the parents and child in a family cannot always be implied. In our work, a BN's links correspond to correlations that may be causal, but are not necessarily so (further testing to exclude the possibility of confounding factors would be required).

After a structure had been selected, the parameters for the dDBN (i.e., the terms in the CPT) are commonly computed with maximum likelihood estimates. Given a fully parameterized dDBN, B, the posterior likelihood given a set of data, D, can be computed:

$$\ell(B \mid D) \propto P(B)P(D \mid B)$$

$$= P(B) \prod_{j=1}^{m} \prod_{t=1}^{T_j} \prod_{i=1}^{n} P\left(X_i = D_{i,t}^j \mid \mathrm{Pa}(X_i) = D_{Pa(X_i),t-1}^j\right)$$

where D is a set of m data points in which each data point contains a value for each of the n nodes across T_j time points, $D_{i,t}^j$ is the value of the ith node in tth time point of the jth data point, $D_{\mathrm{Pa}(X_i),t-1}^j$ is the parent configuration of the ith node in $(t-1)$th time point of the jth data point and $P(B)$ is a prior probability distribution. Given multiple Bayesian networks, one for each class of data, the Bayesian network that more accurately models a data point is said to be the one with the highest posterior likelihood.

RELATED WORK

Much of the current work in fMRI analysis uses statistical regression and hypothesis testing based upon the general linear model (GLM) [Friston, 2005] or discriminant analysis techniques (DATs) such as multivariate regression. These techniques commonly make linearity and Gaussian noise assumptions and may marginalize out time (though research has been done analyzing temporal (auto)-correlations in fMRI data as well, e.g., [Woolrich, 2001]). Within the neuroimaging community, the most commonly employed method used to identify relationships among ROIs is structural equation modeling (SEM) [Pearl, 1998; Penny et al., 2004], also known as path analysis. In the most general case, SEMs model the relationships among ROIs with structural equations of the form $X_i = f_i(\mathrm{Pa}(X_i), \varepsilon_i)$, where ROI X_i is said to be caused by the ROIs in the set $\mathrm{Pa}(X_i)$, ε_i is a stochastic error term and f_i is an arbitrary, possibly nonlinear, function relating the ROIs. In practice, f_i is chosen to be the linear combination of all ROIs in $\mathrm{Pa}(X_i)$ resulting in structural equations of the form, $X_i = \sum_{p \in \mathrm{Pa}(X_i)} \alpha_{i,p} X_p + \varepsilon_i$, where ε is Gaussianly distributed and $a_{i,p}$ are linear combination weights.

A set of structural equations implies a covariance matrix, and the optimal set of structural equations is said to be the sparsest set whose implied covariance matrix is equivalent to the covariance matrix measured from the data. As an example of using SEM to analyze fMRI data, Büchel and Friston [1997] showed significant changes in the relationship between neural systems in the dorsal visual stream caused by shifts of attention.

There is significant research utilizing BNs that are quite similar to these SEMs. For example, Hojen-Sorensen et al. [2000] used Hidden Markov Models (HMMs), a special case of BNs, to learn a model of the activity within the visual cortex from visual stimuli. Mitchell et al. [2003] classified instantaneous cognitive states of a subject such as "reading a book" or "looking at a picture" using a Gaussian naïve Bayesian network (GNBN), yet another specialized BN.

Patel et al., [2006] have introduced another completely data-driven Bayesian method. As with the BDe score we use [Heckerman et al., 1995], they measure correlations among discrete RVs via a Bayesian score that integrates out a model's probabilistic parameters (though they do not provide a closed-form solution and must sample from the model's posterior distribution—a potentially costly operation). However, their score only allows for the consideration of pairwise relationships between binary RVs, limiting them from modeling many interesting correlations (though future extensions are likely possible). In addition to modeling correlations between ROIs, they also introduce a notion of ascendancy, which allows their learned models to suggest causal relationships.

Recently, dynamic causal models (DCMs) have been introduced [Friston et al., 2003] for the purpose of modeling effective neural connectivity. Like SEMs and dDBNs, DCMs can explicitly model temporal interactions among ROIs. In addition, DCMs model hidden neurodynamics that are said to cause hemodynamic observations. (Note the different use of cause in DCMs than in SEMs.) This is done through the incorporation of hidden (or latent) neurodynamic RVs whose values must be inferred through a

forward hemodynamic model. The hidden RVs are linearly dependent on the product of their previous values and the values of external stimuli present, resulting in a bilinear relationship. In our current work, we do not explicitly model hidden neuronal activity as hidden RVs dramatically complicate the dDBN model selection process and is currently an open topic in dBN structure learning. While DCMs are not restricted solely to linear relationships, they must still make assumptions on the correlational form of continuous RVs (e.g., bilinear relationships).

Unlike DATs, SEMs, DCMs and previously employed continuous BNs, our dDBNs need not make assumptions on the functional form of RV correlations. Any conceivable correlation among the RVs in a dDBNs can be represented. dDBNs gain this powerful flexibility at the cost of precision, i.e., they discard much of the information in continuous-valued neuroimaging measurements to get the discrete values they require. A typical fMRI measurement may contain 12-32 bits of precision per voxel. This measurement must typically be quantized down to 1–3 bits for dDBN analysis.

We are aware of only one other source using dDBNs to model neural activity. Smith et al. [2006] used dDBNs to model neural flow in freely moving songbirds. Using dDBNs, they were able to identify nonlinear relationships not found by state-of-the-art linear methods[1]. While their methods are similar to ours, we apply dDBNs to a much larger domain (their structures included eight ROIs while ours includes over 130, resulting in search spaces of roughly 2^{11} structures compared to 2^{137} structures). Further, the neuroimaging data being modeled is significantly different. Whereas we model indirect BOLD response of ROIs extracted from fMRI data, they modeled direct neural responses measured by intracranial electrophysiological recording data. However, their results corroborate our own: in both cases, dDBNs were capable of finding nonlinear relationships that were not likely due to chance nor easily identified with traditional linear methods.

EXPERIMENTAL RESULTS

fMRI Experimental Design

A complete description of the experimental design can be found in Buckner et al. [2000]. Each subject participated in 60 trials. In each trial, a single or double flickering checkerboard stimulus was presented to the subject, who was instructed to respond to the stimulus with a button press. In the single exposure trials, a subject observed a solitary visual stimulus, while in the double exposure trials, two temporally separated images were observed. The stimulus is encoded in the dDBNs via a binary node, allowing the BN to be independently parameterized for both stimulus responses. During each trial, 8 volumes of

[1]They compared their methods to *partial directed coherence* techniques [Baccala and Sameshima, 2001].

T2* weighted echo-planar images with a TR of 2.68 s were obtained, resulting in a time series of 480 fMRI volumes per subject.

Data Preparation

The data used was provided after significant preprocessing had occurred. According to Buckner et al. [2000], all functional images were corrected for odd/even slice-intensity differences and intensity was normalized using a single scaling factor per run, motion correction using a rigid-body rotation and translation correction was applied within each functional run and between functional runs within each subject. Images were then interpolated and spatially normalized to conform to Talairach space. Finally, the slope and mean were subtracted from each voxel to remove effects of linear drift and also subtract signal variation due to underlying structural anatomy. The mean voxel intensity was saved and later used to scale effects to percent signal change. Further data preparation details may be found in Buckner et al. [2000].

Each volume contained $64 \times 64 \times 16$ voxels resulting in 16,384 voxels per volume and 7,864,320 voxels per subject. While limiting the analysis to intracranial voxels may reduce this volume of data, far too many voxels remain to be individually modeled with dDBNs. We abstracted from the voxel level to the ROI level as defined by the Talairach daemon [Lancaster et al., 2000]. The Talairach daemon defines a set of (approximately) 150 ROIs. Each ROI is a subset of the intracranial voxels corresponding to gray matter, e.g., the *Gray Matter* ROI is defined as the improper subset of these voxels and every other ROI is defined to be a proper subset of these voxels. Some ROIs consist of a subset of another ROI's voxels, e.g., the *Caudate Body* ROI contains a proper subset of the *Caudate* ROI's voxels.

Many of the ROIs could be broken up into left and right hemispheric constituents. While doing so would certainly increase the specificity of learned models, it also roughly doubles the number nodes in the models. This increases required computation time (roughly by an order of magnitude), but more importantly, halves the number of voxels each ROI is represented by (thus halving the amount of data available for parameter estimation). Given the small number of subjects involved in this study, we did not feel enough training data existed to warrant splitting up the ROIs. However, we are currently working on methods to improve parameter estimation to address this issue.

Figure 4 gives a flowchart for the experiment. First, the processed data provided by Buckner is obtained for each of the subjects. Then, for each 3D fMRI image, the aggregate mean voxel intensity (AMVI) is computed for each ROI as the average BOLD response of all the voxels within the ROI. Thus, each fMRI image is summarized as 150 AMVI values. The AMVI values for each ROI are chronologically ordered, resulting in a time series of 480 time points. The collection of all subjects' data is referred to as
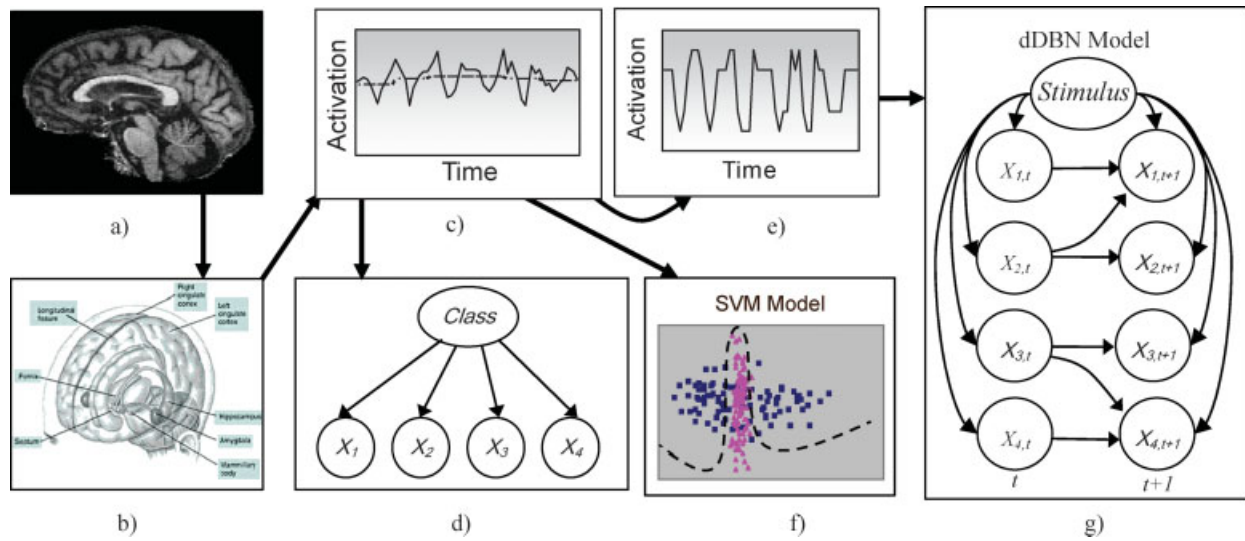
Figure 4.

Experimental design flowchart. (**a**) Raw data is first acquired. (**b**) The data is then grouped into ROIs with the Talairach daemon. (**c**) The aggregate mean voxel intensity (AMVI) is computed for each ROI over time. (**d, f**) The Gaussian naïve Bayesian network and support vector machine models are learned from the AMVI time series. (**e**) The AMVI time series are quantized (HighLow quantization shown). (**g**) The dynamic Bayesian network model is learned from the quantized data. The Stimulus node is set as a parent to all other nodes and incorporates experimental stimulus into the model. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

dataset $D = \{D_1, D_2, \ldots, D_{26}\}$ where D_i comprises the 150 ROI times series for the ith subject. Each D_i can be thought of as a table, with the columns being ROIs, the rows being time frames, and each cell giving the AMVI of a single Talairach region within a single fMRI image. D is then divided into two datasets, D_h and D_d, containing data for the healthy and demented subjects.

At this point, both SVM and GNBN analysis can be applied to the data. However, dDBNs require discrete data. We have found that a 2-bit quantization works best for Buckner et al.'s data. A 1-bit quantization discarded too much of the original signal and a 3-bit (or higher) quantization required too many parameters to be learned given the amount of training data available (as the number of bits increases, the amount of model parameters increases polynomially with degree equal to the maximum number of parents in any one family).

To quantize the data, the time series are divided into windows for each stimulus exposure presented to the subject. Each exposure lasted eight fMRI image acquisitions, and thus each window was composed of eight time points. For each window, a trend is calculated as the mean of the AMVI values within the window. This trend is the average BOLD response for a ROI across an entire stimulus exposure. The trend values are then subtracted from the entire time series, resulting in a new time series with zero mean (consequently, low-frequency drifts in the data are removed).

Let V_{cont} be a value in the zero-mean time series and V_{max} and V_{min} be the maximum and minimum values seen in the zero-mean time series. V_{disc}, the discrete value for V_{cont}, is then computed as follows. $V_{disc} =$ very low if $V_{cont} < V_{min}/2$, $V_{disc} =$ low if $V_{cont} < 0$, $V_{disc} =$ high if $V_{cont} < V_{max}/2$ and $V_{disc} =$ very high otherwise[2]. These discrete states, {very low, low, high, very high} correspond to the ROI's intensity at a time point relative to a small temporal window of time. Figure 4e illustrates the quantization for the AMVI time series seen in Figure 4c.

Two dDBNs are then learned: dDBN$_h$ and dDBN$_d$, obtained by training on D_h and D_d. To classify a previously unseen data point, D_{test}, the posterior likelihood of each of the learned models given D_{test} is calculated. The class of the data point is set to the class corresponding to the model with the highest posterior likelihood.

Comparison of dDBN Structure Between Groups

An important subtlety in interpreting dDBN structures learned from data is that the non-existence of a link does not necessarily indicate that the link's parent and child are statistically independent. The proper interpretation is that there are other links, corresponding to different relation-

---

[2]Other quantizations were attempted, such as using Fourier low pass filters or sliding windows instead of fixed windows or different mid-point boundaries corresponding to standard deviations, etc. All were found to be reasonable.
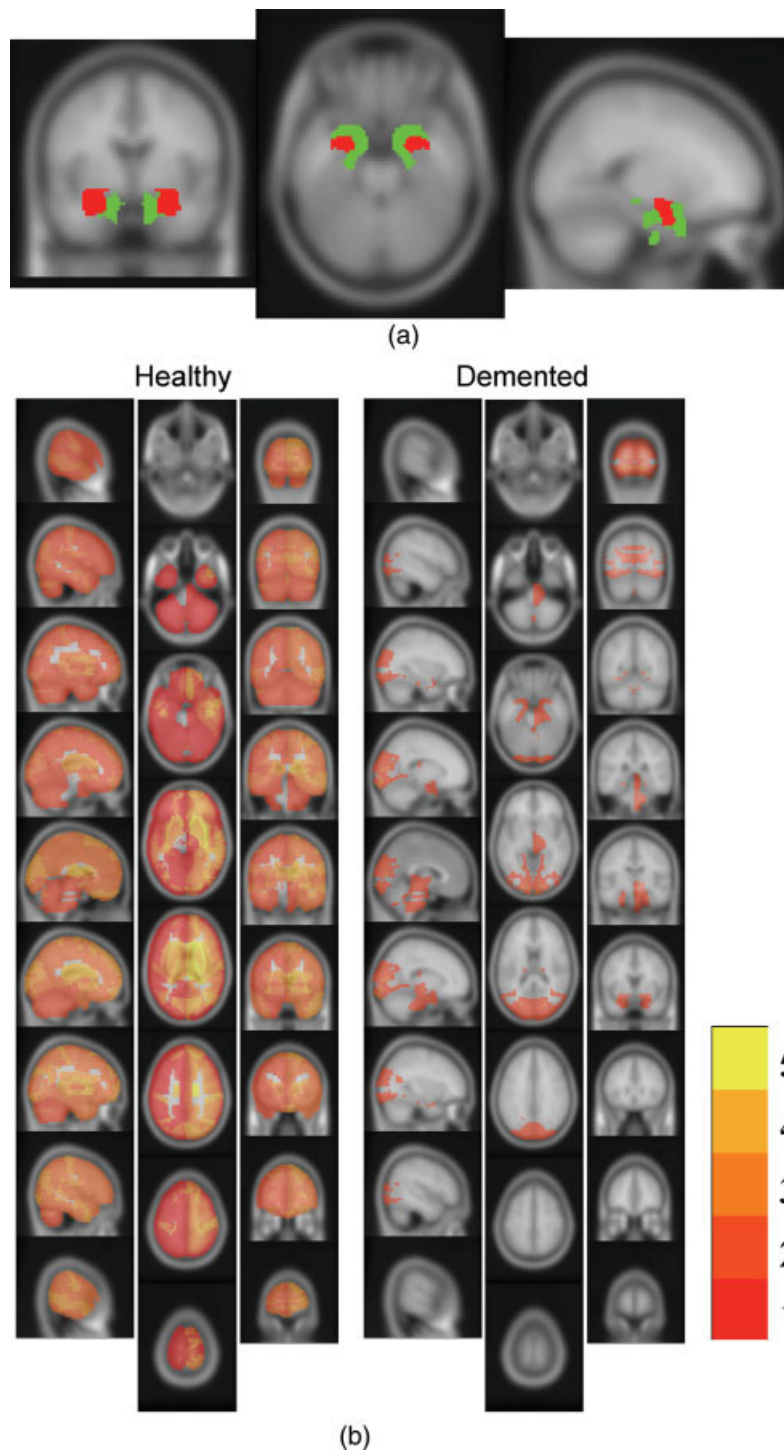
(a)

(b)

Figure 5.

(**a**) Shows location of amygdala and entorhinal ROIs in red and green, respectively. All ROIs are plotted onto selected coronal, axial and sagittal slices of mean normalized T1 images averaged over 305 volunteers. (**b**) Shows children of entorhinal ROIs for healthy elderly subjects (left) and demented elderly (right). Color of individual voxels is determined by the number of ROIs which include that voxel (each voxel can contributes to five ROIs as defined by the Talairach daemon), as shown in the legend at lower right of the figure. Note reduction of influence of entorhinal cortex in demented elderly. (**c**) Shows children of amygdala for healthy elderly (left) and demented elderly (right). Note increase in influence of amygdala in demented elderly.
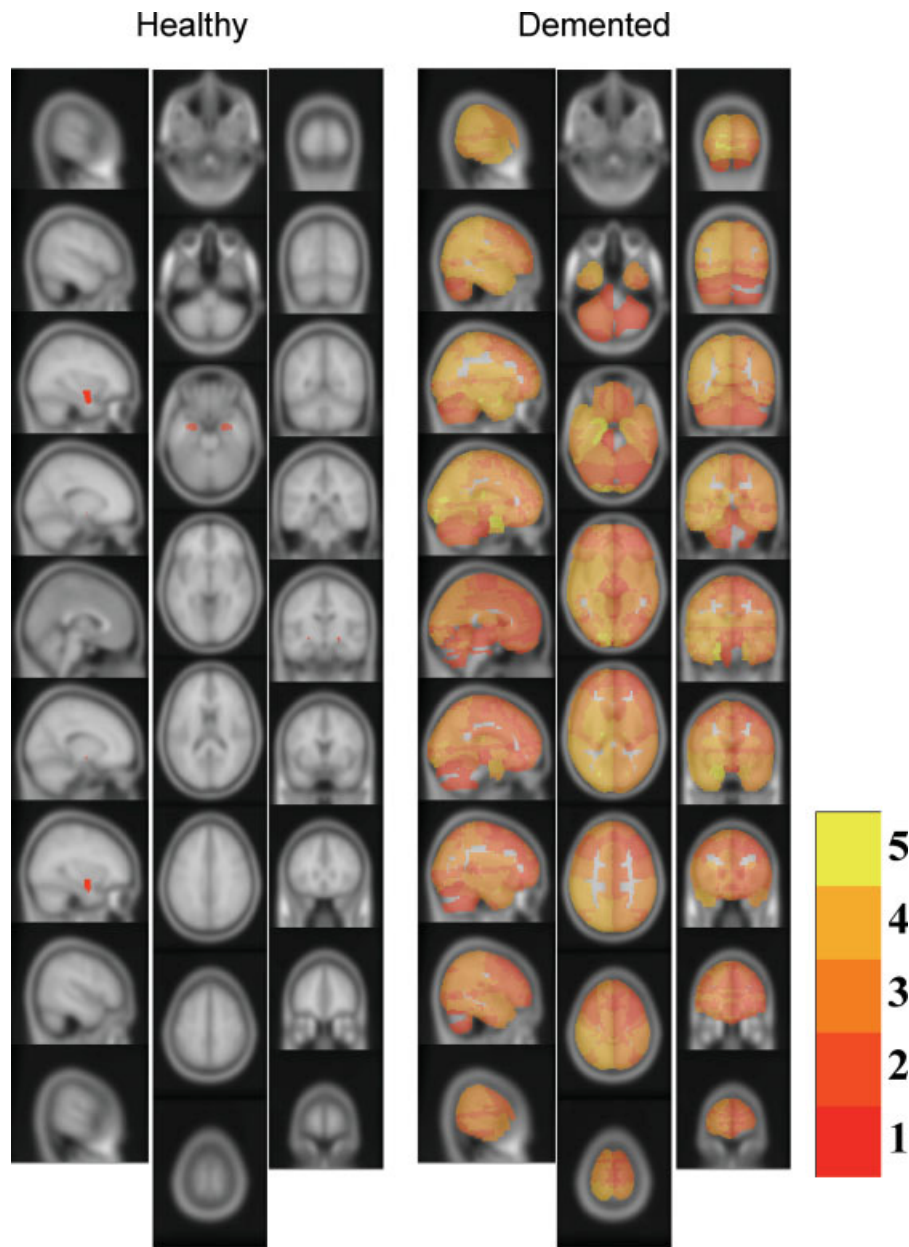
(c)

Figure 5.
(Continued)

ships that are stronger (as measured by the BDe score). As a corollary, finding a link in class *A* but not in class *B* does not necessarily indicate that the link's corresponding relationship is weaker in class *B* (though, that is often the case). For instance, the amygdala is found as a parent of the right cerebrum in demented subjects, but not in healthy subjects. This alone only indicates that regions other than the amygdala had greater correlations with the right cerebrum in healthy subjects, i.e., the relative importance of the amygdala was greater in demented subjects than in healthy subjects. To determine if the absolute strength of a link is greater in one class, its strength must be explicitly measured in both classes. The correlational strength between the amygdala and right cerebrum was found to increase by 4% in demented subjects, so not only was this relationship relatively stronger in demented sub-

TABLE I. The five highest scoring families found by the dDBN structure search for the classification of healthy elderly (left) and demented elderly (right) datasets

	Healthy elderly					Demented elderly			
Structure	Parents	Robust	z and P	Jux. Inc	Structure	Parents	Robust	z and P	Jux. Inc
Gray matter	Cingulate gyrus	86%	5.31	4%	BA 40	Amygdala	100%	2.27	10%
	Rectal gyrus	80%	10^{-8}	0.21		Uvula of Vermis	100%	0.01	0.002
	Cuneus	73%				Parietal Lobe	81%		
	BA 34	20%				Lateral Dorsal Nucleus	81%		
Left cerebrum	Cuneus	80%	4.79	0%	Inferior	Amygdala	90%	3.04	10%
	Cingulate gyrus	80%	10^{-7}		Parietal	Uvula of Vermis	90%	.001	0.002
	BA 28	46%			Lobule	Inferior Parietal Lobule	72%		
	BA 13	13%				Inf. Occipital Gyrus	36%		
BA 4	Pyramis	100%	4.14	3%	BA 7	Amygdala	90%	3.93	17%
	Occipital Lobe	100%	10^{-5}	0.16		Inferior Parietal Lobule	81%	10^{-4}	0.0007
	BA 4	100%				Declive	81%		
	Rectal gyrus	93%				Inf. Semi Lunar Lobule	63%		
BA 6	Cerebellar tonsil	100%	2.22	4%	Right	BA 31	90%	3.34	6%
	Dentate	66%	.013	0.14	Cerebrum	Amygdala	81%	10^{-3}	0.05
	Paracentral lobule	26%				Insula	72%		
	Gray matter	13%				Inf. Semi Lunar Lobule	36%		
Right cerebrum	Right cerebrum	66%	4.17	0%	Parietal	Amygdala	81%	2.40	10%
	Cuneus	66%	10^{-5}		Lobe	Parietal Lobe	45%	.008	0.005
	Dentate	53%				Lat. Dorsal Nucleus	18%		
	Uncus	33%				Right Cerebellum	9%		

Structure—the child ROI in the family listed in order of BDe score. Parents—the parents found for the child ROI listed in order of cross-validation robustness. Robust—the number of times the parent was found in the family during leave one out cross-validation tests. z and P—The confidence score, using the Fourier bootstrap method, for the family (upper) and the corresponding P value it represents (lower). Jux. Inc.—(Top). The increase in likelihood of the family given its own data compared to the likelihood of the family given the opposite class's data. (Bottom) The significance level at which an increase occurred.

jects, but it was absolutely stronger as well. Unless specified otherwise, when we indicate that a correlation is stronger in one class it will be because it was both relatively and absolutely stronger.

We will summarize the results found on three types of dDBN families: those with the highest scores in each study group, those that were likely to be involved in the cognitive operations associated with the visual-motor paradigm used by Buckner et al. [2000], and those that were likely to be affected by the neurodegeneration associated with dementia.

High-Scoring Families

Table I lists the top five scoring families for the healthy and demented dDBNs. These families are significant as they represent the strongest correlations found in the data. Further, all of the correlations are found to be stronger going forward in time than backward in time, suggesting causal relationships (though further testing would be required to ascertain causality).

For the healthy elderly, the children of three of the top five families covered large areas of the brain (gray matter, left and right cerebrum). The most significant parents of these regions included visual areas (the cuneus and occipital cortex), as well as a number of limbic regions (the cingulate gyrus, BA 34 and BA 28, which together comprise

entorhinal cortex). Other children were the primary motor (BA 4) and premotor (BA 6) areas involved in executing motor responses to the visual stimuli. The behavior of BA 4 was significantly correlated with that of the occipital lobe, which is consistent with the use of visual-motor response networks for the performance of this task. Indeed, 14 regions show the transfer of visual information to the rest of the cerebrum with the cuneus as a parent, including the average over all of gray matter and a number of subcortical regions including the striatum. The absolute strength of several of these correlations was found to be diminished in demented patients (Jux. Inc. column in Table I), though due to the relatively small number of datasets, the significance of these reductions could only be made at the $P < 0.1$–0.2 range.

By contrast, four of the top five children in the families obtained from demented data specifically include portions of the parietal lobe (the parietal lobe, inferior parietal lobe, BA 40 and BA 7), whereas none of the top healthy families do. Further, all of these relationships were significantly stronger in demented patients than in healthy patients ($P < 0.05$). This suggests that parietal regions may become more involved in simple visual-motor tasks in early stages of dementia, as temporal brain regions become compromised. In addition, most of these regions do not have a high degree of functional connectivity with posterior visual areas, but instead show functional connectivity with a variety of parietal, cerebellar, and sub-cortical regions, as

well as the amygdala. Other findings regarding the amygdala are discussed in more detail below.

The families identified in the healthy dDBN included many regions involved in visual, motor attentional and memory processing, many of which were identified by Buckner et al. [2000] using GLM based analyses. The cingulate gyrus was found as a parent for the average overall of gray matter, and for a variety of individual brain regions. This is consistent with the role of the cingulate gyrus in response selection, attention and coordination of function across brain regions. The entorhinal cortex (comprised of BA 28 and 34) was also found to be a parent of a large number of other brain regions in the healthy elderly. The entorhinal cortex provides the main source of input to the hippocampus, and is essential for forming declarative memories [Suzuki and Amaral, 1994].

Families That Differ Between Healthy and Demented Elderly

Figure 5 illustrates the children sets of two regions found to differ significantly between the healthy elderly and demented elderly: the entorhinal cortex and the amygdala[3]. The influence of BA28 was reduced in the demented subjects, where it was classified as a parent for only 5 ROIs as opposed to 35 ROIs in healthy subjects. The entorhinal cortex is one of the most greatly affected by the Alzheimer's disease process, where degeneration begins long before behavioral symptoms are fully evident. Indeed, entorhinal degeneration has been found in postmortem studies of individuals with no overt signs of dementia, as young as 20 years old [Braak and Braak, 1995], and the degenerative process is well advanced by the earliest stages of AD that can be diagnosed based upon deficits in verbal memory [Gomezisla et al.,1996]. Analysis of dDBNs from the healthy young subjects showed BA 28 as a parent of the hippocampus, but this was not found in either elderly group, suggesting that the influence of entorhinal cortex on the hippocampus is reduced before behavioral signs of dementia are detected. Further, these entorhinal cortex results are robust across leave-one-out cross-validation. It was found as a parent in 40 healthy dDBN families, on average, across all 26 iterations of leave-one-out cross-validation, but in only 10 demented dDBN families on average. The influence of the cuneus was also reduced in demented relative to healthy subjects, where it was not a parent for any demented dDBN families, but was found as a parent in 14 healthy dDBN families.

The most prominent network differences between healthy and demented elderly was the increased number of networks that involved the amygdala in demented subjects. The amygdala was found to be a parent for all five of the top families associated with dementia, but was not a parent for any top families in the healthy elderly subjects (Table I). Further, the strength of these families was reduced when the same relationships were examined using data from healthy elderly, indicating that these correlations were stronger in demented subjects. In the demented population, over half of the 150 regions have the amygdala as a parent variable in the DBN. This is in contrast to healthy families, where the amygdala is a parent only of itself. Like the results from entorhinal cortex, the amygdala results were robust across leave-one-out cross-validation. It was found as a parent in 85 families within the demented dDBN on average, in all 26 cross-validation iterations, but in only 1.4 families on average within the healthy dDBN.

The stronger correlations found here between the amygdala with many other brain regions in demented vs. healthy families suggest that the amygdala may exert more influence over the function of other brain regions in dementia. This finding agrees with a number of previous behavioral and imaging findings in dementia. A predominant behavioral symptom of dementia is agitation, which includes a high state of anxiety and motor restlessness. The amygdala is central to the perception and expression of anxiety in humans and other primates [Sander et al., 2003] especially in children [Amaral et al., 2003], with a diminution of its influence with normal aging [Bauman et al., 2004]. The present findings suggest that the agitation associated with AD may result from a greater influence of the amygdala. These results are also in agreement with other imaging studies where evidence of increased influence of the amygdala was found in demented subjects, including Grady et al. [2001] and Rosenbaum et al. [2004]. Taken together, these results provide mounting evidence that the influence of the amygdala over activity in other brain regions is increased in dementia. This might result from the brains of demented subjects being "rewired" in dementia, or the neurodegenerative process may enhance the influence of brain circuits already present. Alternatively, changes in the apparent influence of the amygdala could be due to differences in functional connectivity, without changes in effective or anatomical connectivity. The question of whether dementia is associated with differences in effective connectivity of the amygdala could be examined using methods that provide more precise temporal information, such as EEG and MEG. Differences in anatomical connectivity of the amygdala might be examined using DTI. In conclusion, further exploration of the differences in functional and anatomical brain organization associated with dementia, and methods of analysis such as the dDBN method described here, will lead to a better understanding of this disorder, which may lead in turn to improved methods of diagnosis, prevention and treatment.

---

[3]Voxels in Figure 5 are highlighted based on how many times they occur in ROIs found as children to the amygdala or entorhinal cortex. Highlighting voxels based on their correlational strength, or by how commonly they occur across populations, is not possible due to dDBN structure search complexities.

STRUCTURAL VALIDATION

Fourier Bootstrapping Confidence Testing

To gauge the likelihood that elicited correlates occurred due to chance, we employ a Fourier bootstrapping method [Prichard and Theiler 1994] that creates a series of surrogate ROI time series containing the same linear auto- and cross-correlation functions present in the actual data (up to the first and second order moments-nonlinear correlations are not preserved). BDe scores for families elicited on the real fMRI datasets are calculated on the surrogate datasets to indicate how strong these correlations are in the surrogate data. A confidence measure, in the form of a z score, is calculated for each family as the distance in standard deviations of the BDe score on the true data from the mean BDe score on the surrogate data: $\mathrm{Conf} = (x - \mu_{\mathrm{rand}})/\sigma_{\mathrm{rand}}$, where x is the BDe score of the family trained on the true dataset, μ_{rand} and σ_{rand} are the mean and standard deviation of the BDe score on the surrogate datasets (see Fig. 6).

A high confidence indicates that the magnitude of the BDe score given the real data is not likely due to chance. Since the BDe score measures how strongly dependent a BN's family's child is on its parents, a high confidence would also indicate that the dependence among the family's ROIs was not likely due to chance as well. However, since the surrogate data contains all of the first and second order moments of the real data, a high confidence actually indicates that only the higher order dependencies among RVs (third moments and above) identified in a DBN family are not likely due to chance. Thus, families with high confidences correspond to relationships among RVs in the data that are not likely explained by linear correlations or by random chance.

If DBNs were restricted to modeling relationships between only the first and second moments (as many neuroimaging techniques are), this could not be used to validate the results. Fourier methods simply cannot destroy relationships among distal ROIs while simultaneously preserving proximal relationships. Wavelet-based methods can help alleviate this limitation, e.g. [Breakspear et al., 2004], but were not required in our validation as the Fourier method was successful in demonstrating confidence in elicited families.

Twenty surrogate datasets were generated to calculate the confidence scores, which were then converted to p values. A confidence score of approximately 1.65 corresponds to $P < 0.05$; a score of $\sim$2.3 corresponds to $P < 0.01$. Roughly 71% of the families identified in the DBN had confidence scores greater than 1.65. Further, the top five healthy and demented families listed in Table I all had confidence scores of at least 2.2, with some scores as high as 5.3. Overall, the confidence measure indicates that many (but not all) of the correlates identified in the dDBNs, including all of the highest scoring correlates, were not likely due to chance.

In calculating the confidence, two mathematical assumptions are made. (1) Surrogate BDe scores are distributed
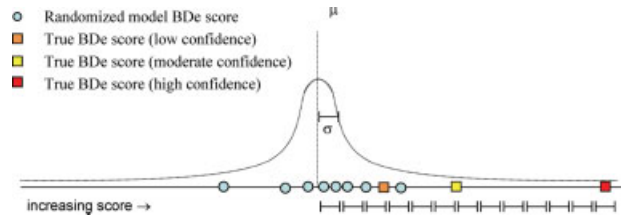

Figure 6.
Visualization of the family confidence measure. The confidence of each of the families in a dDBN trained on a true dataset vary in their reliability. Computing the BDe score of families trained on surrogate datasets yields a distribution (assumed to be Gaussian with mean m and variance s) of BDe scores on datasets with autocorrelations that are identical to the original dataset. The larger the distance of a family's BDe score from m, the more likely the relationships identified in the family did not occur by chance. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

normally. There are approximately 40,000 BDe scores generated in these tests. We've analyzed their distribution, which other than a light tail on the negative side (which does not affect our confidence score), fits a normal well. (2) The estimate of the normal's standard deviation can be performed with 20 samples. We have computed the 95% confidence intervals for the σ's and used the higher bound to calculate all of the z scores, resulting in conservative estimates.

Structural Robustness

For the structures learned by the dDBN to be meaningful, they must be robust to small changes in the training data. We analyzed the structures learned for both the healthy and demented dDBNs during the leave-one-out cross-validation iterations used in the classification task. Table I lists the percentage of iterations in which each parent ROI was selected for each of the top five scoring dDBN families. Typically, each family contained two or three parents that were significantly robust. In several families, all four parents were found to be robust, as was the case with the healthy dDBN's BA 4 family and the demented dDBN's BA 40 family. Further, some parents were particularly robust in all the families they were found in. Most notably, the amygdala was selected as a parent in 89% of the iterations (on average) for each of the top five demented families.

Only the structural results for the top five families are listed, but most of the other families in the dDBNs displayed similar robustness characteristics. Overall, while there were some parents that only show up sporadically, such as *Gray Matter* in the healthy BA 6 family, most of the resulting dDBN family structures were robust.

Nonlinear Components

A primary advantage of using dDBNs is their ability to model complex nonlinear relationships. However, the dDBNs can also model linear relationships. If the results of a dDBN analyses only found linear relationships, dDBNs would make a poor modeling choice due to their need to quantize continuous data. The Fourier bootstrapping method demonstrated that nonlinear effects were being modeled by the DBN (otherwise BDe scores on the surrogate datasets would not have been diminished), however, we've also explicitly tested the linear correlation among ROIs identified by the top DBN families.

A linear regressor was fit to each family with the child node treated as a dependent RV and the parent nodes treated as independent RVs. The linear regressor was fit by minimizing the sum of squared residuals. The strength of linear fit was measured via the coefficient of determination,

$$r = \sum_{i=1}^{n} \left(Y_i' - \langle Y \rangle \right)^2 \bigg/ \sum_{i=1}^{n} \left(Y_i - \langle Y \rangle \right)^2$$

where Y is a set of points in $\Re^5$ corresponding to the AMVI values of the dependent and independent RVs, n is the total number of rows in all of the demented (or healthy) datasets, Y_i' is the linear estimate of Y and $\langle Y \rangle$ is the mean value of the dependent RV.

We found that there was significant discourse between the ordering of dDBN families based on their BDe score and the ordering based on strength of their linear fit. For example, the fifth highest healthy family, the Right Cerebrum, had a higher r than the other four highest healthy families and the highest-ranking demented family, BA 40, had a lower r-value than three of the other top five demented families. Overall, the top ranking families were not ordered based on the strength of their linear fit as would be expected if DBNs were simply modeling linear correlations.

We also compared the linear fit of the top 10 families with the linear fit found in a large sample (200,000) of families with randomly generated structures. Approximately 5% of these families had stronger linear fits (but much smaller BDe scores) than the average linear fit of the top 10 families identified by the dDBN structure search. Linear-based methods would favor these families over those selected by the dDBN and thus would likely have a hard time identifying many of the relationships the dDBN search found.

dDBN Classification Efficacy

To validate that the dDBN structure search method is capable of identifying meaningful relationships among ROIs, we test the classification efficacy of learned dDBNs with that of two other commonly employed machine learning classifiers, support vector machines (SVMs) [Burges, 1998] and Gaussian naïve Bayesian networks (GNBNs)
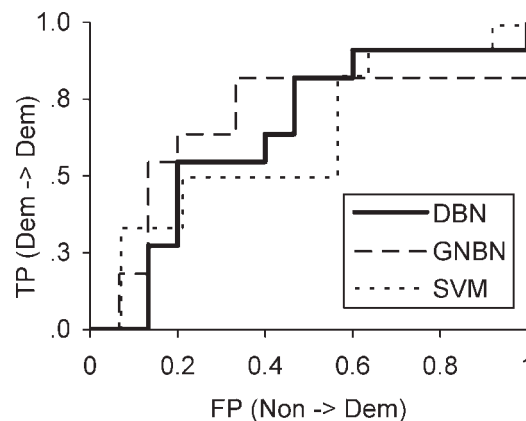

Figure 7.

ROC curves for classifying demented vs. healthy subjects. The *x*-axis is the percent of false positives (FPs) (healthy subjects classified as demented subjects) and the *y*-axis is the percent of true positives (TPs) (demented subjects classified as demented subjects). No algorithm dominated although the SVM was continually beaten by at least one other algorithm.

[Duda et al., 2001]. See Appendix A and B for experimental details on SVMs and GNBNs.

For the dDBNs, classification efficacy is determined using a two stage cross-validation process. The inner cross-validation stage trains the *numBestToKeep* and *numParents* parameters. The outer cross-validation stage learns a BN with the best *numBestToKeep* and *numParents* found in the inner cross-validation stage and classifies left out subject data. In other words, the testing data is not used in any way to learn the structure, parameters or hyper-parameters used in the model that performs the classification on the testing data.

The dDBNs correctly classified 18 of the 26 subjects, yielding an accuracy of just under 70%. The GNBN and SVM correctly classified 17 subjects, yielding 65% accuracy. With only a single subject difference separating the accuracies of the three classifiers, the difference is not significant, although, the ROC curve given in Figure 7 indicates that the classification behavior (with regards to true positives and false negatives) between the dDBN, GNBN and SVM was different. None of the three classifiers completely dominated each other and over some portion of the curve, each classifier performed better or worse than one of the other classifiers. The GNBN performed better than the dDBN and SVM when false positives are weighed roughly equally to true positives. As penalties for false positives are given less weight (towards the right of the ROC curve), the dDBN and SVM outperformed the GNBN. AUC values for the dDBN, GNBN, and SVM classifiers were 0.64, 0.67, and 0.61, respectively.

There has been other work on the classification of dementia based on neuroimaging data. Using multivariate linear regression on positron emission tomography (PET)

data, Azari et al. examined the neurophysiological effects of dementia in elderly subjects [Azari et al., 1993]. They correctly classified subjects with mild or moderate dementia from healthy subjects with 87% accuracy. The subjects examined in Buckner et al.'s study suffered only from very mild or mild dementia as opposed to the subjects with more severe dementia examined by Azari et al. Although Buckner et al. did not attempt to ascertain classification accuracies using the GLM, the fact that they found few functional differences between groups suggests classification would be a difficult task. The comparison between the present results and those of Azari et al. is further complicated by the different imaging techniques used (fMRI vs. PET), which vary in anatomical coverage, temporal sampling and hemodynamic properties being imaged.

CONCLUSIONS AND FUTURE WORK

To examine the changes in nonlinear functional connectivity among neuroanatomical regions associated with early stages of dementia, we used discrete dynamic Bayesian networks (dDBNs) to model the neuroanatomical relationships present in both healthy subjects and subjects with mild or very mild dementia from data collected by Buckner et al. [2000]. Unlike most methods used to model neuroimaging data, dDBNs use discrete RVs, as opposed to continuous RVs, which allows dDBNs to model arbitrarily non-linear relationships but also requires quantization of the continuously-valued fMRI measurements.

The dDBN identified a variety of neural networks in data obtained from healthy elderly subjects that would be expected for a visual-motor task, including visual-motor networks, and networks involved in motivation, memory, attentional/integrative processing and emotion. Differences in networks found in data obtained from demented subjects included reduced influence of entorhinal cortex on other brain regions, as would be expected given the large degree of neurodegeneration in this brain area with Alzheimer's disease, the primary cause of dementia in elderly adults. An increased influence of the amygdala, in both absolute and relative strength, on many other brain regions was also found. This was not wholly unexpected, given the increased anxiety and agitation found in demented subjects, as well as the results of previous imaging studies that have found increased linear correlations between the amygdala and other brain regions [Grady et al., 2001; Rosenbaum et al., 2004]. However, dDBN was able to identify these correlations with greater sensitivity by examining both linear and non-linear relationships. Taken together, these results suggest a significant change in functional brain networks in early stages of dementia.

To validate our results, we showed that most of the relationships identified by the dDBN had high confidence, in the form of a z score, that classifiers based on dDBNs were capable of classifying competitively with other commonly employed machine learning classifiers, and that the dDBNs

identified relationships that would be expected given the underlying experimental paradigm employed by Buckner et al. We also showed that the dDBN selected structures based on nonlinear criteria and that they were robust to leave one out cross-validation.

Our application of dDBNs to the analysis of neuroimaging data presented here provides a single example of the usefulness of the discrete Bayesian Network framework, but many variations and extensions to our work are possible. Improvements to the structure search heuristic, such as application of the *Optimal Reinsertion* algorithm of Moore and Wong [2003] or application of hierarchical structure search methodologies [Burge and Lane 2006], are being investigated. For purposes of classification, there are structure scoring functions that are better suited than the BDe metric we used, such as the class-conditional likelihood metric [Grossman and Domingos 2005] or the approximate conditional likelihood metric [Burge and Lane 2005a]. Inclusion of hidden RVs into the dDBN topology, such as is the case in the hemodynamic forward model in DCMs [Friston et al., 2003], may be quite useful and is an open area of research in machine learning.

Certainly, the lack of such a hemodynamic model may be obscuring important correlates, given the 3-s delay between fMRI images. One solution is to add additional delays into the dDBN analysis so that correlations that take longer than 3-s to develop may be captured. This would improve model sensitivity, although it significantly complicates structure search. Another approach would be to disregard time altogether and use static Bayesian networks to model correlations among ROIs occurring at the same time, though explicitly modeling time in fMRI models has been shown to increase model sensitivity [Lahaye et al., 2003]. A model could also incorporate both static and temporal links, though this can complicate model selection. In all, there is a tremendous amount of research on dDBNs that could be performed in order to offer a different approach for neuroimaging analysis, a field currently dominated by continuous RV analysis methods.

ACKNOWLEDGMENTS

REFERENCES

Akaike H (1973): Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Second International Symposium on Information Theory. Budapest: Academiai Kiado. pp 267–281.

Amaral DG, Bauman MD, Schumann CM (2003): The amygdala and autism: Implications from non-human primate studies. Genes Brain Behavior 2:295–302.

Azari NP, Pettigrew KD, Schapiro MB, Haxby JV, Grady CL, Pietrini P, Salerno JA, Heston LL, Rapoport SI, Horwitz B (1993). Early detection of Alzheimer's disease: A statistical approach

using positron emission tomographic data. J Cerebral Blood Flow Metab 13:438–447.

Baccala LA, Sameshima K (2001): Partial directed coherence: A new concept in neural structure determination. Biol Cybern 84:463–474.

Bauman MD, Lavenex P, Mason WA, Capitanio JP, Amaral DG (2004): The development of social behavior following neonatal amygdala lesions in rhesus monkeys. J Cogn Neurosci 16:1388–1411.

Bhattacharya S, Ho MR, Purkayastha S (2006): A Bayesian approach to modeling dynamic effective connectivity with fMRI data. Neuroimage 30:794–812.

Braak H, Braak E (1995): Staging of Alzheimers-disease-related neurofibrillary changes. Neurobiol Aging 6:271–278.

Breakspear M, Brammer MJ, Bullmore ET, Das P, Williams LM (2004): Spatiotemporal wavelet resampling for functional neuroimaging data. Hum Brain Mapp 23:1–25.

Büchel C, Friston KJ (1997): Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modeling and fMRI. Cerebral Cortex 7:768–778.

Buckner RL, Snyder A, Sanders A, Marcus R, Morris J (2000): Functional brain imaging of young, nondemented, and demented older adults. J Cogn Neurosci 12:24–34.

Burge J, Lane T (2005a): Learning class-discriminative dynamic Bayesian networks. Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany.

Burge J, Lane T (2005b): Comprehensibility of generative versus class discriminative dynamic Bayesian multinets. Human Comprehensible Machine Learning Workshop. AAAI, Pittsburg, PA.

Burge J, Lane T (2006): Bayesian network structure search on hierarchically related random variables. Proceedings of the 17th European Conference on Machine Learning, Berlin, Germany.

Burges C (1998): A Tutorial on Support Vector Machines for Pattern Recognition. Boston: Kluwer Academic Publishers.

Charniak E (1991): Bayesian networks without tears. AI Mag 12:4.

Chickering D, Geiger D, Heckerman D (1994): Learning Bayesian networks is NP-Hard. Technical Report MSR-TR-94-17. Microsoft.

Cooper G, Herskovits E (1992): A Bayesian method for the induction of probabilistic networks from data. Machine Learning 9:309–347.

Cox RW (1996): AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Duda RO, Hart PE, Stork DG (2001): Pattern Classification, 2nd ed. New York: Wiley Interscience. ISBN 0-471-05669-3.

fMRIDC (2004): The fMRI Data Center. Dartmouth. Available at http://lx50.fmridc.org/f/fmridc.

Friston KJ (2005): Models of brain function in neuroimaging. Ann Rev Psychol 56:57–87.

Friston KJ, Harrison L, Penny W (2003): Dynamic causal modeling. NeuroImage 19:1273–1302.

Fung G, Mangasarian OL (2001): Proximal support vector machine classifiers. Proceedings of the Seventh ACM SIGKDD. San Francisco. pp 77–86.

Gold S, Christian B, Arndt S, Zeien G, Cizadlo T, Johnson DL, Flaum M, Andreasen NC (1998): Functional MRI statistical software packages: A comparative analysis. Hum Brain Mapp 6:73–84.

Gomezisla T, Price JL, McKeel DW, Morris JC, Growdon JH, Hyman BT (1996): Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. J Neurosci 16:4491–4500.

Grady CL, Furey ML, Pietrini P, Horwitz B, Rapoport SL (2001): Altered brain functional connectivity and impaired short-term memory in Alzheimer's disease. Brain 124:739–756.

Grossman D, Domingos P (2004): Learning Bayesian network classifiers by maximizing conditional likelihood. International Conference on Machine Learning, Alberta, Canada. pp 361–368.

Heckerman D, Geiger D, Chickering DM (1995), Learning Bayesian networks: The combination of knowledge and statistical data. Machine Learning 20:197–243.

Hojen-Sorensen P, Hansen L, Rasmussen C (2000): Bayesian modeling of fMRI time series. In: Solla SA, Leen TK, Muller KR. editors. Advances in Neural Information Processing Systems. Cambridge: MIT Press.

Jensen FV (2001): Bayesian Networks and Decision Graphs. New York: Springer-Verlag.

Lahaye P, Poline J, Flandin G, Dodel S, Garnero L (2003): Functional connectivity: Studying nonlinear, delayed interactions between BOLD signals. Neuroimage 20:962–974.

Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT (2000): Automated Talairach Atlas labels for functional brain mapping. Hum Brain Mapp 10:120–131.

Mitchell T, Hutchinson R, Just M, Niculescu RS, Pereira F, Wang X (2003): Classifying Instantaneous Cognitive States from fMRI Data. American Medical Informatics Association Annual Symposium. October 2003.

Moore A, Wong W (2003): Optimal Reinsertion: A New Search Operator for Accelerated and More Accurate Bayesian Network Structure Search Learning. Proceedings of the Twentieth International Conference on Machine Learning. Arlington, VA: AUAI Press. pp 552–559.

Morris JC (1993): Clinical dementia rating. Neurology 43:2412–2414.

Murphy K (2002): Dynamic Bayesian Networks: Representation, Inference and Learning. PhD dissertation, Berkeley, University of California, Computer Science Division.

Papoulis A (1991): Probability, Random Variables, and Stochastic Processes, 3rd ed. New York: McGraw-Hill.

Patel R, Bowman F, Rilling J (2006): A Bayesian approach to determining connectivity of the human brain. Hum Brain Mapp 27:267–276.

Pearl J (1986): Fusion, propagation and structuring in belief networks. Artif Intell 29:241–288.

Pearl J (1998): Graphs, causality, and structural equation models. Sociologic Methods Res 27:226–284.

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004), Modeling functional integration: A comparison of structural equation and dynamic causal models. Neuroimage 23 (Suppl 1):S264–S274.

Prichard D, Theiler J (1994), Generating surrogate data for time series with several simultaneously measured variables. Phys Rev Lett 14:951–954.

Rosenbaum RS, Furey ML, Horwitz B, Grady CL (2004): Altered communication between emotion-related brain regions supports short-term memory in Alzheimer's disease. Soc Neurosci Abstr 203:8.

Sander D, Grafman J, Zalla T (2003): The human amygdala: An evolved system for relevance detection. Rev Neurosci 14:303–316.

Schwarz G (1978): Estimating the dimension of a model. Ann Stat 6:461–464.

Smith VA, Yu J, Smulders TV, Hartemink AJ, Jarvis ED (2006): Computational inference of neural information flow networks. PLoS Comput Biol 6:1436–1449.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy R, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004): Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 (Suppl 1):208–219.

Suzuki WA, Amaral DG (1994): Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. J Neurosci 14:1856–1877.

Veropoulos K, Campbell C, Cristianini N (1999) Controlling the sensitivity of support vector machines. Proceedings of the International Joint Conference on Artificial Intelligence, Workshop ML3, Stockholm, Sweden, pp 55–60.

Woolrich MW, Ripley BD, Brady M, Smith SM (2001): Temporal autocorrelation in univariate linear modeling of FMRI Data. Neuroimage 14:1370–1386.

APPENDIX A

Gaussian Naïve Bayesian Network fMRI Analysis

The GNBN in our classification experiments contained a single binary class node and 300 observable nodes. The 300 nodes were divided into 150 goups of two nodes; with one group for each Talairach ROI. The two nodes in each region were Trial1Var$^{(i)}$ and Trial2Var$^{(i)}$ where (i) indicates the ith anatomical region. The Gaussian distributions in each of the nodes were set as the maximum likelihood estimate (MLE) distributions for either the means or the variances of the AMVI values across all of a subject's fMRI images. The variances for the Trial1 nodes were computed only on fMRI images scanned while the subject performed a trial with a single visual stimulus. Similarly, the variances for the Trial2 nodes were computed on images taken when a subject performed a trial with two visual stimuli. To increase classification performance, only a subset of the nodes in the GNBN are used for classification. This size of this subset is selected empirically with cross validation. Classification of a testing data point was performed via a standard BN likelihood ratio test.

APPENDIX B

Support Vector Machine fMRI Analysis

Two datasets were created for analysis with the SVMs, D_{raw} and D_{tal}. D_{tal} contains the AMVI values the dDBN used for classification. This resulted in each subject being represented by 480 data points, each with ~150 dimensions, one per Talairach ROI. D_{raw} uses the raw fMRI voxel information as dimensions for a single data point. In this dataset, each subject is represented by 480 data points with each data point having 65,536 dimensions (one dimension per voxel, one data point per fMRI image and 480 images per subject).

Ideally, extracranial voxels would be removed and classification performed solely on intracranial voxels. However, extracranial voxels in one subject may be intracranial voxels in another, complicating the process of training solely on intracranial voxels. Given the vastly decreased intensities of extracranial voxels compared to intracranial voxels (2–3 orders of magnitude), the inclusion of extracranial voxels in the SVM classification should not cause significant performance degradation.

An SVM is trained on each dataset with both linear and Gaussian kernels. As SVMs are sensitive to having an imbalance in the number of representatives for each class, we modify the diagonal elements of the kernel matrices as suggested by [Veropoulos et al., 1999]. We used the Proximal SVM [Fung and Mangasarian, 2001] to perform the classification.