Proceedings of Machine Learning Research vol 73:188-199, 2017 AMBN 2017

Restricted Quasi Bayesian Networks as a Prototyping Tool for Computational Models of Individual Cortical Areas

Naoto Takahashi NAOTO.TAKAHASHI@AIST.GO.JP
Yuuji Ichisugi Y-ICHISUGI@AIST.GO.JP
Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST)
Tokyo (Japan)

Abstract

We propose *restricted quasi Bayesian networks* as an efficient prototyping tool for designing computational models of individual cortical areas of the brain. Restricted quasi Bayesian networks are simplified Bayesian networks that only distinguish probability value 0 from other values. Using our tool, it is possible to concentrate on the essential part of model design and efficiently construct prototypes. We demonstrate that restricted quasi Bayesian networks actually work well as a prototyping tool by implementing a syntactic parser for an ambiguous English sentence.

Keywords: Bayesian networks; cerebral cortical areas; prototyping tool; context free grammar.

1. Introduction

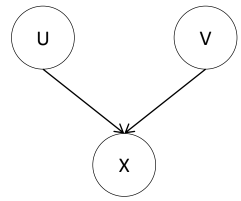
The cerebral cortex is a part of the brain. In humans, it is most deeply related with intelligence, and is divided into about 50 areas, like the visual area, the language area, and so on.

All areas in the cerebral cortex share similar anatomical structure. Thus it is reasonable to assume a common functional principle. One of the promising hypotheses is that Bayesian network (Pearl, 1988) or a kind of probabilistic graphical model is the common functional principal (Lee and Mumford, 2003; George and Hawkins, 2005; Rao, 2005; Ichisugi, 2007; Röhrbein et al., 2008; Litvak and Ullman, 2009; Chikkerur et al., 2010; Ichisugi, 2011; Hosoya, 2012; Dura-Bernal et al., 2012; Raju and Pitkow, 2016; Pitkow and Angelaki, 2017)

One way to study the mechanism of information processing in the brain is to create computational models of individual cortical areas. If we succeed in creating such models using Bayesian networks, the said hypothesis will be more reliable.

However, creating realistic models of cortical areas using fairly large-scale Bayesian networks forces us to resolve inessential problems, for example, tuning hyper parameters. Moreover, it is often difficult to trace the real cause of unsatisfying results when the created model does not behave as expected.

It is often helpful to create prototypes before creating a real model. By creating prototypes, we can estimate the hopefulness of the fundamental design of the real model that we are going to create.



		$X=1$	$X=0$
$U=1$	$V=1$	0.9	0.1
$U=1$	$V=0$	0.6	0.4
$U=0$	$V=1$	0.3	0.7
$U=0$	$V=0$	0	1

		$X=1$	$X=0$
$U=1$	$V=1$	>0	>0
$U=1$	$V=0$	>0	>0
$U=0$	$V=1$	>0	>0
$U=0$	$V=0$	0	>0

Figure 1: A conditional probability table in an ordinary Bayesian network (upper table) and that in a restricted quasi Bayesian network (lower table).

In this paper, we propose *restricted quasi Bayesian networks* as an efficient prototyping tool for creating models of cortical areas. Restricted quasi Bayesian networks are simplified Bayesian networks that only distinguish probability value 0 from other values. In other words, restricted quasi Bayesian networks cannot represent the joint probabilities of random variables, but they can represent whether a given value combination of random variables is possible or not.

Figure 1 gives an example. Suppose the variables U, V and X are all binary. If the conditional probability table of X in an ordinary Bayesian network is given by the upper table, its restricted quasi Bayesian network version is given by the lower table.

Since restricted quasi Bayesian networks do not have learning ability, conditional probability tables must be prepared by the designer. Because of their limited capabilities, restricted quasi Bayesian networks may not be applied to practical, real-world problems. However, they release us from inessential problems and allow us to concentrate on the essential part of model design. As a result of agile prototyping activities, we would find potential problems in the model, which can be extremely difficult to find in a practical system.

The rest of this paper is organized as follows. Section 2 introduces restricted Bayesian networks and explains their three node types. Section 3 further simplifies restricted Bayesian networks to define restricted *quasi* Bayesian networks and describes their properties. In Section 4, we present a syntactic parser for a small context free grammar using a restricted quasi Bayesian network, then we show in Section 5 that the presented network successfully parses an ambiguous English sentence to give all the possible parse trees. Section 6 briefly summarizes the paper and addresses our plan for the future work.

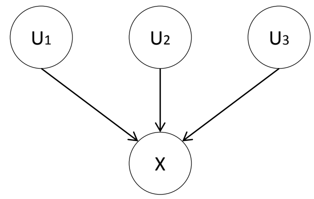


Figure 2: An OR node with three parent nodes.

2. Restricted Bayesian Networks

In an ordinary Bayesian network, the size of a conditional probability table increases exponentially against the number of parent nodes. Therefore, naively designed Bayesian networks easily cause combinatorial explosion.

Humans, on the other hand, often resolve complex, real-world problems in real time. If the cerebral cortex is a kind of Bayesian networks, it must be a special type of Bayesian networks that do not cause combinatorial explosion.

With this working hypothesis, we are looking for a way to limit the number of parameters in Bayesian networks that we use in the models of cortical areas. We first restrict the nodes in Bayesian networks to the following three types to reduce the size of conditional probability tables: OR nodes, gate nodes and exclusive nodes. All nodes are binary random variables and take 0 or 1 as values. We call the resulted network *restricted Bayesian network*.

All these three types of nodes can be transformed into different network structures, to which we can apply the method of Heckerman (1993) to accelerate inference. When a node has n parents, we can transform that part into a network of the depth $O(n)$ with at most constant number of parents. Using the transformed network, we can execute one step of approximate inference and learning at $O(n)$.

2.1 OR Nodes

The OR nodes in restricted Bayesian networks follow the noisy-OR model (Pearl, 1988).

Suppose binary random variables $U_i (1 \leq i \leq n)$ and X take 0 or 1 as values. If we define w_i as

$$w_i = P(X = 1|U_1 = 0, \ldots, U_{i-1} = 0, U_i = 1, U_{i+1} = 0, \ldots, U_n = 0), \tag{1}$$

then the conditional probability $P(X = 1|U_1, \ldots, U_n)$ under the noisy-OR model is given as the following.

$$P(X = 1|U_1, \ldots, U_n) = 1 - \prod_{i=1}^{n} (1 - w_i)^{U_i} \tag{2}$$

We depict an OR gate as in Figure 2.

2.2 Gate Nodes

A gate node has two types of parent: one or more controller and an upstream. Let $U_1, \ldots, U_{n-1}$ be the controllers of a gate node X, and U_n be the upstream of X. Us-

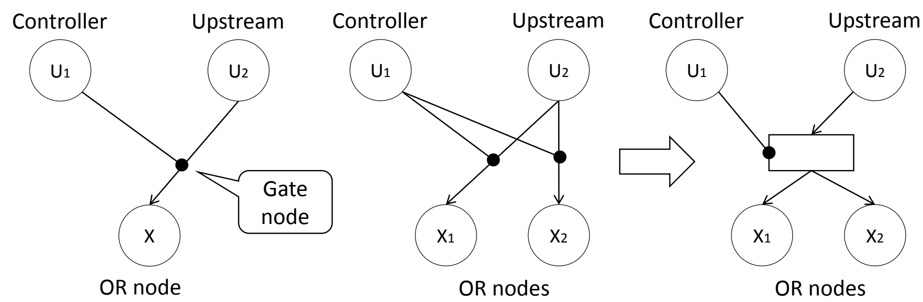


Figure 3: (Left) A gate node (the small black dot) along with its upstream and a controller. (Middle) Two gate nodes sharing the same upstream and the controller. (Right) A shorthand of the two gate nodes that are shown in the middle part.

ing the definition of w_i in Equation 1, the conditional probability $P(X = 1|U_1, \ldots, U_n)$ is defined as follows.

$$P(X = 1|U_1, \ldots, U_n) = (1 - w_n)^{1-U_n} \prod_{i=1}^{n-1} (1 - w_i)^{U_i} \tag{3}$$

Gate nodes are qualitatively similar to the noisy-OR model, but their definitions of $P(X = 0)$ and $P(X = 1)$ are exchanged and the values of the controllers are inverted.

In this paper, we use gate nodes that have only one controller. We depict a gate node as a small black dot as in the left part of Figure 3. When multiple gate nodes share the same upstream and the same controller (as in the middle part of Figure 3), they are depicted as in the right part of Figure 3.

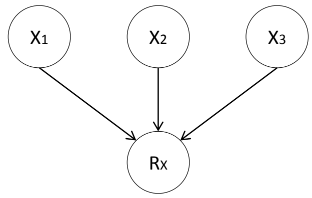
2.3 Exclusive Nodes

An exclusive node has two or more parent nodes and takes value 1 if and only if at most one parent node takes 1 (Figure 4). Exclusive nodes have no adjustable parameters. By assigning 1 to an exclusive node as its observed value, it is possible to limit the number of its active parents at most one.

Exclusive nodes are used to represent multi-valued variables using a set of binary variables. Suppose a 4-valued variable X takes values v_0, v_1, v_2 and v_3. X can be represented by three binary variables X_1, X_2, and X_3, and one exclusive variable R_X as their child. We can interpret, for example, $X_1 = X_2 = X_3 = 0$ as $X = v_0$, $X_1 = 1$ and $X_2 = X_3 = 0$ as $X = v_1$, $X_2 = 1$ and $X_1 = X_3 = 0$ as $X = v_2$, and $X_3 = 1$ and $X_1 = X_2 = 0$ as $X = v_3$. See Figure 5.

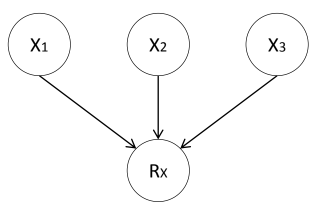
3. Restricted Quasi Bayesian Networks

We simplify restricted Bayesian networks so that only probability value 0 is distinguished from other probability values. This is to concentrate on the qualitative aspects of models of cortical areas. The resulted networks are called *restricted quasi Bayesian networks*. This



X_1	X_2	X_3	$P(R_X = 1 \mid X_1, X_2, X_3)$
0	0	0	1
1	0	0	1
0	1	0	1
0	0	1	1
otherwise			0

Figure 4: The conditional probability table of an exclusive node R_X with its three parent nodes X_1, X_2 and X_3. R_x takes value 1 when at most one parent node is active.



X_1	X_2	X_3	R_X	Interpretation of X
0	0	0	1	v_0
1	0	0	1	v_1
0	1	0	1	v_2
0	0	1	1	v_3

Figure 5: Representing a 4-valued variable X using three binary variables X_1, X_2, X_3, and an exclusive node R_X as their child. Unlisted value combinations of X_1, X_2 and X_3 are excluded because the exclusive node R_X is fixed to 1.

simplification makes it impossible to calculate the most probable explanation (MPE) from partially observed values. Instead, a restricted quasi Bayesian network lists all possible value combinations of observed- and latent variables whose joint probability are greater than 0.

3.1 Properties of Restricted Quasi Bayesian Networks

Restricted quasi Bayesian networks discard quantitative probabilistic calculation, but they still keep some essential properties of Bayesian networks. First, the restrictions to determine the solution are locally defined in the conditional probability tables of the nodes. Secondly, when observed values are assigned to a set of random variables, that information propagates through the whole network and the only value combinations that satisfy the restrictions of all the nodes are determined as solutions.

In other words, restricted quasi Bayesian networks largely ignore an aspect of ordinary Bayesian networks, which is probability distribution manipulation, but still value another aspect, which is optimization of the whole network.

3.2 Nodes in Restricted Quasi Bayesian Networks

We show how OR nodes and gate nodes are simplified in restricted quasi Bayesian networks.

Equation 2 equals to 0 when $w_i = 0$ for all i, but that means none of U_i affects the value of X and thus meaningless. Therefore, $w_i > 0$ for at least one i.

Then Equation 2 is equal to 0 if $U_i = 0$ for all i, and is greater than 0 otherwise. Since restricted quasi Bayesian networks distinguish only 0 from other probabilities, this means noisy-OR is equivalent to Boolean OR within the framework of restricted quasi Bayesian networks.

In a restricted quasi Bayesian network, the controllers of a gate node can be interpreted as "blockers" between the upstream and the gate node. When a controller takes value 1, it blocks the data flow from the upstream to the gate node. If all controllers of a gate node take 0 and the upstream takes 1, the value of the gate node can be 1 with a probability greater than 0. This is equivalent to Boolean AND if we invert the value of the controllers.

4. Application to Syntactic Parser

In this section, we demonstrate that restricted quasi Bayesian networks actually work well as a prototyping tool. For this purpose, we construct a syntactic parser for a context free grammar as a model of the language area.

Table 1 shows a small context free grammar to parse the sentence "Time flies like an arrow", which is syntactically ambiguous. The final part "an arrow" has no syntactical ambiguity, thus can be handled as a single element. Then the number of the elements of this sentence is four.

A restricted quasi Bayesian network that parses syntactically ambiguous sentences with four elements is shown in Figure 6. Note that this network is not specialized for the sentence "Time flies like an arrow." It parses any sentence with four elements, as long as an appropriate grammar is given.

ID	rule			ID	rule		
S0	S	→	NP VP	VP0	VP	→	V PP
NP0	NP	→	time	VP1	VP	→	V NP
NP1	NP	→	flies	VP2	VP	→	VP PP
NP2	NP	→	an arrow	V0	V	→	time
NP3	NP	→	NP NP	V1	V	→	flies
NP4	NP	→	NP PP	V2	V	→	like
PP0	PP	→	P NP	P0	P	→	like

Table 1: A grammar to parse the sentence "Time flies like an arrow." S = sentence; NP = noun phrase; VP = verb phrase; V = verb; PP = prepositional phrase; P = preposition. An imperative sentence is interpreted as an S without the subject, namely VP.

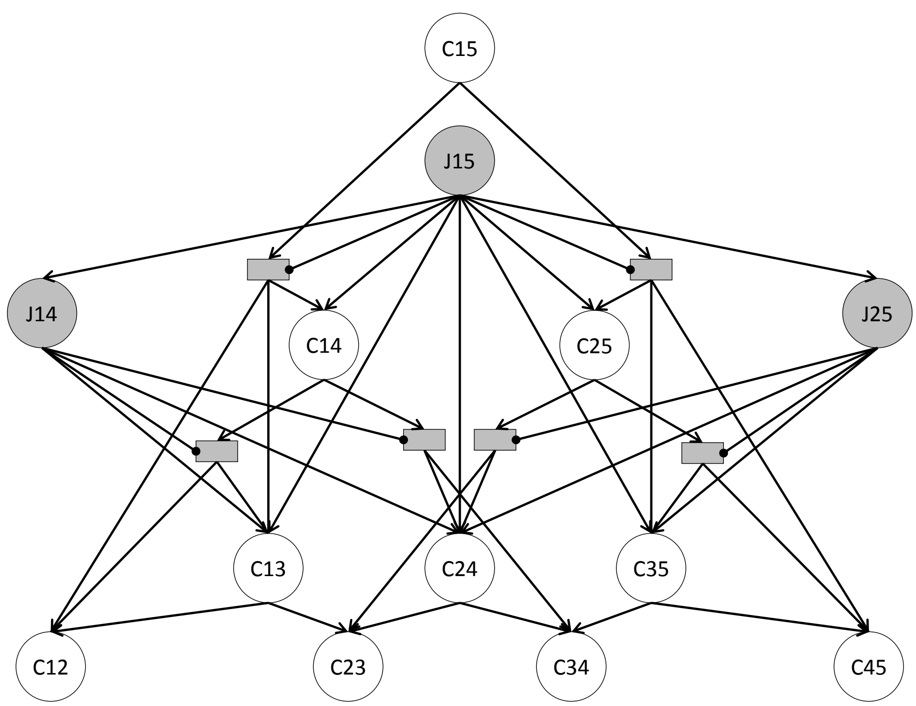
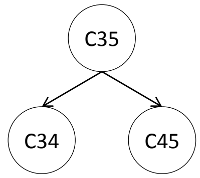


Figure 6: A restricted quasi Bayesian network that parses sentences with four elements.



Node ID	C34	C45	C35
Possible	V → like	NP → an arrow	VP → V NP
combinations	P → like	NP → an arrow	PP → P NP

Figure 7: Parent-children relationship of C nodes (excerpted) and their possible value combinations. When C34 and C45 take unlisted value combinations, C35 takes a special value meaning "unused".

The circles in Figure 6 are OR nodes; white nodes and grey nodes are called C nodes and J nodes, respectively. Each OR node consults the values of its directly connected nodes to determine its own value. The rectangles are shorthand of gate nodes. See Figure 3 for detail.

C nodes correspond to the cells in a CYK parser (Cocke and Schwartz, 1970). Each C node takes a production rule as its value. For example, if the nodes C34 and C45 respectively take "V → like" and "NP → an arrow" as their values, the node C35 takes the value "VP → V NP". See Figure 7.

Each J node works as a controller of gate nodes to determine the structure of the final parse tree. For example, if the word list "C12 C23 C34" in Figure 8 is parsed as [C12 [C23 C34]], the sub-tree under C24 is used but the one under C13 is not. On the other hand, if the same word list is parsed as [[C12 C23] C34], the sub-tree under C13 is used, but the one under C24 is not.

Depending on the values of C13 and C24, J14 selectively blocks the connection between the upstream C14 and the four children C12, C13, C24 and C34 so that one of the following situation is realized.

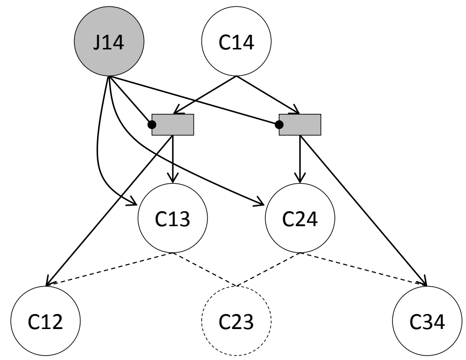
1. Only C12 and C24 are the children of C14.

2. Only C13 and C34 are the children of C14.

3. C14 has no children and is not used by other nodes.

So far, we have explained as if the information flows from bottom to top (words to sentence). However, as restricted quasi Bayesian networks inherit the information propagation property of ordinary Bayesian networks, the actual information flow is bidirectional. We can, for example, fix the topmost syntactical category and let the network select appropriate word sequences.

Thanks to the restriction to the node types, the number of the parameters of the network grows only polynomially against the number of words. However, if we straightfor-



J14	C13	C24	C12	C13	C24	C34
2	unused	any	Y	N	Y	N
3	any	unused	N	Y	N	Y
unused	any	any	N	N	N	N

Figure 8: Possible value combinations of J14, C13 and C24, and the resulting connection between C14 and other C nodes. Y means that the C node of that column is a child of C14, N means the other case.

Node	Possible value	Original rule
C12	NP0	NP → time
	V0	V → time
C23	NP1	NP → flies
	V1	V → flies
C34	V2	V → like
	P0	P → like
C45	NP2	NP → an arrow
C15	S0	S → NP VP
	VP0	VP → V PP
	VP1	VP → V NP
	VP2	VP → VP PP

Table 2: Possible values for the highest and lowest C nodes in Figure 6.

wardly implement a CYK parser using a Bayesian network, the number of parameters grow exponentially against the number of words (Pynadath and Wellman, 1996).

5. Experiment

We implemented a domain specific language (DSL) for building restricted quasi Bayesian networks. The structure and the conditional probability tables of the target network are specified in the DSL.

Parent	Child 1	Child 2
S0	NPx	VPx
NP3	NPx	NPx
NP4	NPx	PP0
VP0	Vx	PP0
VP1	Vx	NPx
VP2	VPx	PP0
PP0	P0	NPx

Table 3: The possible value combinations for the intermediate C nodes. The parent takes the indicated value only when the value combination of the two children matches a row. The suffix x means a wild card.

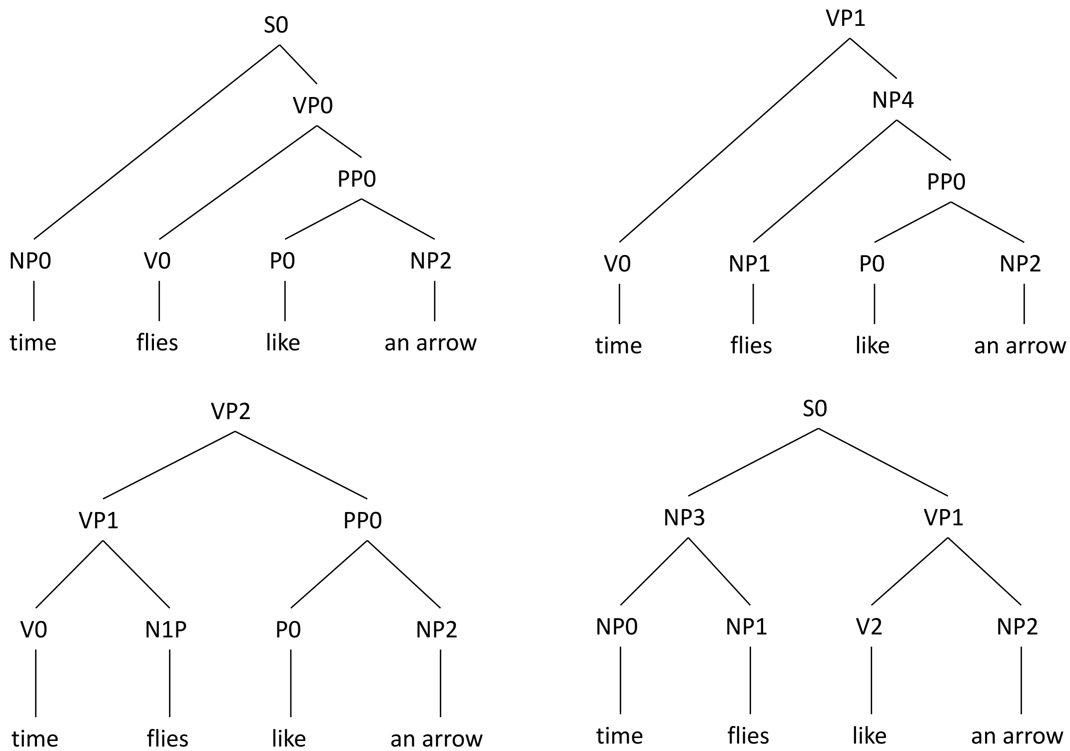


Figure 9: Four parse trees for "Time flies like an arrow." (Upper left) Time proceeds as quickly as an arrow proceeds. (Upper right) Measure the speed of flies that resemble an arrow. (Lower left) Measure the speed of flies in the same way that you measure the speed of an arrow. / Measure the speed of flies in the same way that an arrow measures the speed of flies. (Lower Right) Flies of a particular kind, time-flies, are fond of an arrow. The paraphrases are taken from Pinker (1994).

Table 2 shows the possible values (values with probabilities greater than 0) for the highest and the lowest C nodes in Figure 6. The possible value combinations for other C nodes are given in Table 3, which follows the grammar specified in Table 1.

Figure 9 shows graphical representation of the possible value combinations of C nodes. We can confirm that all four possible parse trees are obtained.

6. Summary and Future Work

We proposed restricted quasi Bayesian networks as an agile prototyping tool for creating computational models of individual cortical areas. We also demonstrated their usefulness through constructing a context free grammar parser.

Since restricted quasi Bayesian networks ignore the quantitative aspect of probability distribution manipulation, it would be difficult to use them to solve practical problems. Learning ability and probability distribution manipulation are absolutely necessary for practical problems, but combinatorial explosion must be avoided. Therefore we shall need *real* restricted Bayesian networks (without "quasi").

We are planning to implement the architectures of prototypes using real restricted Bayesian networks, then make them learn from real, massive data, and finally evaluate their performance.

Acknowledgments

This paper is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

Scharat Chikkerur, Thomas Serre, Cheston Tan, and Tomaso Poggio. What and where: A bayesian inference theory of attention. *Vision Research*, 50(22):2233–2247, 2010.

John Cocke and Jacob T. Schwartz. Programming languages and their compilers: Preliminary notes. Technical report, Courant Institute of Mathematical Sciences, New York University, 1970.

Salvador Dura-Bernal, Thomas Wennekers, and Susan L. Denham. Top-down feedback in an hmax-like cortical model of object perception based on hierarchical bayesian networks and belief propagation. *PLoS ONE*, 7(11), 2012.

Dileep George and Jeff Hawkins. A hierarchical bayesian model of invariant pattern recognition in the visual cortex. In *2005 International Joint Conference on Neural Networks (IJCNN)*, 2005.

David Heckerman. Causal independence for knowledge acquisition and inference. In *Proceedings of the Ninth International Conference on Uncertainty in Artificial Intelligence*, pages 122–127, 1993.

Haruo Hosoya. Multinominal bayesian learning for modeling classical and nonclassical receptive field properties. *Neural Computation*, 24(8), 2012.

Yuuji Ichisugi. The cerebral cortex model that self-organizes conditional probability tables and executes belief propagation. In *2007 International Joint Conference on Neural Networks (IJCNN)*, 2007.

Yuuji Ichisugi. Recognition model of cerebral cortex based on approximate belief revision algorithm. In *2011 International Joint Conference on Neural Networks (IJCNN)*, 2011.

Tai Sing Lee and David Mumford. Hierarchical bayesian inference in the visual cortex. *J. Opt. Soc. Am. A*, 20(7):1434–1448, Jul 2003.

Shai Litvak and Shimon Ullman. Cortical circuitry implementing graphical models. *Neural Computation*, 21(11):3010–3056, 2009.

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988. ISBN 0-934613-73-7.

Steven Pinker. *The Language Instinct*. William Morrow and Campany, 1994.

Xaq Pitkow and Dora E. Angelaki. Inference in the brain: Statistics flowing in redundant population codes. *Neuron*, 94(5):943–953, 2017.

David V. Pynadath and Michael P. Wellman. Generalized queries on probabilistic context-free grammars. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence and Eighth Innovative Applications of Artificial Intelligence Conference, AAAI 96, IAAI 96, Portland, Oregon, August 4-8, 1996, Volume 2.*, pages 1285–1290, 1996.

Rajkumar Vasudeva Raju and Xaq Pitkow. Inference by reparameterization in neural population codes. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 2029–2037, 2016.

Rajesh P.N. Rao. Bayesian inference and attention modulation in the visual cortex. *Neuroreport*, 16(16):1843–1848, 2005.

Florian Röhrbein, Julian Eggert, and Edgar Körner. Bayesian columnar networks for grounded cognitive system. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, pages 1423–1428, 2008.