

Bayesian Columnar Networks for Grounded Cognitive Systems

Florian Röhrbein (florian.roehrbein@honda-ri.de)

Julian Eggert (julian.eggert@honda-ri.de)

Edgar Körner (edgar.koerner@honda-ri.de)

HONDA Research Institute Europe GmbH, Carl-Legien-Str. 30
63071 Offenbach am Main, Germany

Abstract

There are two elementary building blocks whose intimate relationship has been overlooked so far: The cortical column as ubiquitous local processor underlying all cognitive operations and the Bayesian framework with the generic operation of calculating posterior probabilities given likelihoods and prior probabilities. In this contribution we try to bridge this gap by providing a detailed mapping between parts of the cortical column and processing steps involved in Bayesian reasoning. The resulting columnar network is implemented and fully integrated within an intentional vision system, which is also shortly characterized.

Keywords: Cortical column; Bayesian message passing

Introduction

We have developed a relational knowledge representation which combines ideas from classical semantic networks, Bayesian graphical models and the current knowledge on neocortical structures. Within this relational network all semantic entities (properties, objects, concepts ...) are represented by uniform nodes and causal links. The network nodes and their connectivity are strongly motivated by functional models of the cortical column. Our previous work used spreading activation schemas as inference strategy, but here we focus on Bayesian message passing and its possible biological realization. After we have provided detailed arguments for a mapping between Bayesian formulas and columnar layers, a rough overview on the currently implemented system will be given.

Columnar Belief Propagation

Bayesian Networks

Bayesian networks have been successfully applied to numerous application domains and recently also within cognitive science (e.g. Chater, Tenenbaum & Yuille, 2006). These networks are also termed belief networks, probabilistic or causal networks and share simple syntactic elements: A directed acyclic graph (DAG) consisting of a set of nodes, which represent the variables involved, and a set of unidirectional, unlabeled links which are typically interpreted as “directly influences”. Finally a conditional probability table (CPT) is attached to each node which defines the probability for a node given its parents:

$$P(x_i | \text{parents}(x_i)) \quad (1)$$

The basic task Bayesian networks have to solve consists in the computation of posterior probabilities for a set of query variables, given measurements for some evidence variables. For this probabilistic inference many different inference strategies can be used, but most of them are not suitable for an interpretation in biological terms, because they modify the structure of the graph. For example the most popular exact inference algorithm, the clique-tree propagation algorithm, requires even several transformations: First a transformation into an undirected graph, then a triangulation of the resulting moral graph and finally a clustering, which results in a so-called clique-tree, on which the inference is run.

For a biological interpretation any modification of the graphical representation is to be understood as some sort of learning, which clearly is not the case in here. The same argument applies to the many variants of clustering algorithms and also of conditioning algorithms. However, there is a rather old inference strategy, the polytree algorithm, also known as belief propagation or message passing that does not alter the graph structure and has further interesting properties, which will be detailed in the following.

Bayesian Propagation Rules

Belief propagation is an exact inference scheme which requires local computations only. They are carried out within each node and the results are transmitted to all directly connected child and parent nodes. For a recapitulation of the formulas we use the notation introduced by Pearl (1988). These are valid for polytrees, which are DAGs with the restriction that multiple paths between two nodes are not allowed.

An interesting feature of belief propagation is the separation of evidence variables into two disjoint subsets, i.e. the evidence introduced by child nodes and the evidence introduced by parent nodes. This leads to $\lambda(x)$, representing diagnostic support for the proposition $X=x$ and to $\pi(x)$, representing causal support for x . The posterior probability or belief (BEL) of a node x can hence be written as

$$BEL(x) = \alpha \lambda(x) \pi(x). \quad (2)$$

Alpha is a normalization constant and each node's λ and π are calculated according to the following equations:

$$\lambda(x) = \prod_j \lambda_{Y_j}(x). \quad (3)$$

$$\pi(x) = \sum_{u_1, \dots, u_n} P(x | u_1, \dots, u_n) \prod_i \pi_X(u_i) \quad (4)$$

Here two sorts of messages enter the computation: So-called λ -messages (right hand side of equation 3), which are messages received from child nodes (Y_i), and π -messages (right hand side of equation 4), which are received from the node's parents (U_i). The incoming messages are separately combined and, for the computation of π , additionally weighted by the conditional probabilities of node x (provided by the CPT).

These incoming λ - and π -messages are used not only for the local computation of the node's belief BEL, but also for the computation of the messages the node sends itself to its parents (equation 5) and children (equation 6):

$$\lambda_X(u_i) = \beta \sum_x \lambda(x) \cdot \sum_{u_k: k \neq i} P(x | u_1, \dots, u_n) \prod_{k \neq i} \pi_X(u_k) \quad (5)$$

$$\pi_{Y_j}(x) = \gamma \left[\prod_{k \neq j} \lambda_{Y_k}(x) \right] \cdot \pi(x) \quad (6)$$

Both equations reveal that the node's output is a normalized combination of both sorts of incoming messages. Note that normalization constants β and γ can be omitted, it suffices to normalize the posterior probabilities (with α in equation 2).

Whereas there is only one π , λ and BEL associated with each node, there are as many π -messages as the node has child nodes and as many λ -messages as the node has parent nodes (since the formulas are not only for trees, but for polytrees). Additionally the CPT, which gives the probability of node x given all parents, has to be stored here.

Representational Constraints

In this subchapter we argue that the cortical column is the biological entity which might best correspond to a belief network's node. Furthermore we show how the calculation of belief propagation might indeed be realized within each cortical column of the brain.

First, any DAG requires localized, interconnected units with a direction defined. For Bayesian networks four further requirements have to be met:

- there has to be a storage place for the CPTs at each entity
- a belief value has to be stored and updated at each entity

- the assumed entities have to represent concepts on various levels like properties, objects and concepts
- for a Bayesian interpretation all links between these entities have to represent dependencies, e.g. "A causes B", if there is a link from A to B

From the work of Fellemann and van Essen (1991) it is well known that a "direction" within the cortex can be defined purely on the basis of the laminar pattern, leading to a complete cortical hierarchy. Therefore it is straightforward to associate a connected pair of network nodes with two neural entities from a connected pair of cortical areas. If we follow this hierarchy from bottom to top, we encounter neurons which are sensitive to increasingly more complex stimuli: In the visual cortex there is a progression from simple features to more complex ones, from object parts to objects and scenes. This path can easily be interpreted in a Bayesian manner and this has been done so already successfully by others (e.g. Yuille & Kersten, 2006). The key idea hereby is that object parts are "caused" by scenes which contain them. Despite these different levels of granularity, the basic neural entity in question has to be uniform and there are even more requirements to be met by this entity in realizing the belief propagation algorithm:

- there should be a continuous, parallel updating mechanism for the computations of λ and π
- two different link types are needed for passing λ - and π -messages
- all connections have to be reciprocal, since between two connected nodes both types of messages need to be send
- the entity has to provide $m+n$ different outputs simultaneously (for each of the m child and n parent nodes)

Clearly, the granular level of a single neuron is too fine to accomplish the requirements of storing the various parameters and calculating a couple of different output messages, but "each minicolumn has all cortical phenotypes, and each has several output channels" (Mountcastle, 2003). It is this entity which is found everywhere in the neocortex and is hence assumed to be the basic building block of the cortical computing machinery. In order to clarify how the inference formulas can be related in detail to certain subsystems of this cortical column, we briefly describe this "local processor" next (for details see e.g. Thomson & Bannister, 2003).

The Cortical Column

The cortical column vertically connects neurons across the six cortical layers with a distinct circuitry: Neurons in layer II/III send their axons to neurons in columns of the next higher cortical area. There the signals mainly target at granular layer IV but also pyramidal cells in layer VI receive afferent input (see Fig. 1). Primary source of

fferent inputs are the deep layers V/VI, which target in the superficial layers (I, II, upper III) in columns of the next lower cortical area (Rockland, 1994).

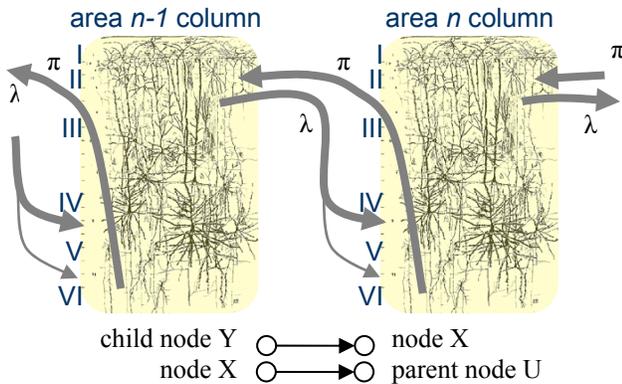


Figure 1: The main connectivity between two columns from successive cortical areas (e.g. visual areas V4 and IT) is shown. Columnar layers are numbered by Roman numerals and the projections between them are labeled according to the type of message transmitted.

Given the causal interpretation motivated above, the connections between columns along the ascending pathway correspond to links from nodes to their parent nodes. The resulting cortical bottom-up (BU) information flow can hence be understood as a realization of the passing of λ -messages. Accordingly, the inter-columnar projections along the descending pathway correspond to links from nodes to child nodes and the cortical top-down (TD) flow realizes the passing of π -messages. However, not only these BU- and TD-outputs, but also substantial parts of the formulas and intermediate computation stages can be matched to columnar layers.

Resulting Mapping

Before we detail the parallels between message passing and columnar circuitry, it should be emphasized that we do not claim the calculation of posterior probabilities being the only role of the cortical column. However we do believe that Bayesian message passing might very well be realized within the columnar machinery. Also, we are totally aware of the complexity of the cortical column and hereby refer to a quite abstract functional view on columnar layers. Fig. 2 provides an overview of the proposed mapping between Bayesian nodes and biological columns.

BU inputs enter twice

Information from lower cortical areas enters the column (at least) twice, at layer IV as the primary target of the ascending pathway and also at layer VI (see Fig. 1). Analogously, in the Bayesian model each λ -message λ_{y_j} is needed in two equations: For the computation of λ (Equ. 3)

and for the computation of the outgoing π -messages π_{y_j} . The latter is not apparent from Equ. 6., but becomes obvious if we insert (2) in (6):

$$\pi_{y_j}(x) = BEL(x) / \lambda_{y_j}(x). \quad (7)$$

This BU stream can also be inspected in Fig. 2 by following the outputs of child nodes Y_1, Y_2, Y_3 to node X.

BU / TD asymmetry

There is a well documented asymmetry between BU- and TD-inputs to the cortex: Whereas forward connections from lower areas are quite focused, the feedback signals from higher areas are generally much more widespread (e.g. Suzuki, Saleem & Tanaka, 2000). Interestingly, a similar pattern can be found within the message passing formulas:

For an analysis we have to check how often a particular λ -message is needed during the node's computation compared to the number of accesses to a particular π -message. Following the equations, each ingoing π -message is needed once for the computation of π (Equ. 4) and $n-1$ times for the computation of the outgoing λ -messages (Equ. 5). Contrary, each ingoing λ -message is needed once for the computation of λ (Equ. 3) and once for the computation of the outgoing π -messages (Equ. 7). It is crucial, of course, that Equ. 5 cannot be further simplified (which is due to the additional weighting with CPT entries).

The impact of this asymmetry becomes evident for a realistic number of child and parent nodes: With a growing number of child nodes the number of a particular BU-input remains constant (and equals 2), whereas the number of a particular TD-input grows linearly with the number n of parent nodes (and equals n).

Local computations

Within the cortical hierarchy columns from successive areas like V2 and V4 are highly interconnected. Also the formulas show a local connectivity pattern: Here only information from direct neighbors is used for the computations at each node and the result is sent only to nodes in the next upper and lower layer. It should be noted, however, that in biology there are also "shortcuts" between areas more far apart, especially between those which have a common border (like e.g. V2 and PO).

Storage of CPT in dendritic trees

It is well known that there is a very dense wiring in the upper layers II/III, which is also a prerequisite for storing the local CPT. As can be seen in Fig. 2, these layers are just beneath those processes, which need values from the CPT: λ_x and π . But there are two more interesting facts from biology: First, horizontal inputs to layer II/III are mostly modulatory (Ichinose & Murakoshi, 1996), which would fit to the "weighting behaviour" of the CPT entries. Secondly, the synapses of these connections are those that change first

during training (Rioult-Pedotti, Friedman & Donoghue, 2000), a property highly desirable, because all CPTs have to be learned sometime.

Intertwined processing streams

There could be separate processing streams for BU and TD information and the role of each column/node would then consist in fusing all locally available information of one sort. This is not the case in the neocortex, where complex interactions of BU and TD information can be observed, and also not in the Bayesian model: Equ. 5 and 6 show that there is a non-trivial mixing of λ - and π -messages instead.

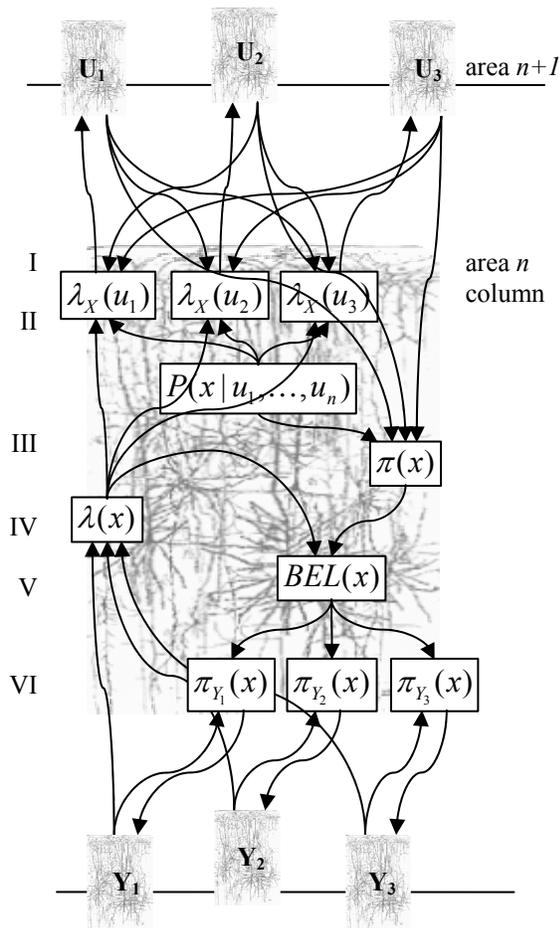


Figure 2: Message passing parameters as they might be localized within a cortical column. Depicted is node x with connections to child nodes Y_1, Y_2, Y_3 and parent nodes U_1, U_2, U_3 . Cortical layers are numbered on the left. Arrows indicate computational dependencies, not shown are normalization factors.

The canonical microcircuit

In all mammals a basic intra-columnar wiring pattern has been found: A canonical functional microcircuit, which projects from layer IV to layers II/III and from there to lower layers V and VI (Callaway, 1998). The arrangement

of processing steps in Fig. 2 shows a similar circuit. It starts at layer IV with the computation of λ , which is needed for the calculation of λ_x in layer II. The result is passed to layer III for the calculation of π and then to BEL at layer V and λ_y at layer VI.

Behavioral Output

The most important parameter, though, might be the behavioural output of the cortical column. Neurons highly associated with this role are pyramidal cells in layer V, since they project to action and attention related subcortical structures (e.g. Lomber & Payne, 2000). In Bayesian models the “output” of a node is just the posterior probability BEL and should hence be strongly associated with cortical layer V. This layer is near to the proposed locations of those calculations which are needed for or depend on BEL: λ , π and π_{Y_j} . Therefore localizing BEL at layer V might also be inferred from a wiring length constraint.

Discussion

The question of relating Bayesian formulas to neurobiological findings has been addressed by several researches in the recent years. But, whereas we identify the cortical column as the basic unit for belief computations, others target at single neurons (e.g. Rao, 2004; Deneve, 2005; Yu & Dayan, 2005) or at whole cortical areas (e.g. Lee & Mumford, 2003; George & Hawkins, 2005). The latter at least mentions the cortical column but the exact relationship remains obscure. Others are unsure at all on which computational level the cortical column should be located (Minsky, 2007, p.293).

The drawback of message passing schemes described herein is that it only works for polytrees (i.e. singly connected networks), but there are approximate inference strategies, which are extensions of standard message passing and are applicable for arbitrary DAGs, e.g. loopy belief propagation (Murphy, Weiss & Jordan, 1999). The extension proposed by Pearl, loop cutset conditioning, is an exact inference procedure for DAGs, but it changes the connectivity of the network and therefore is not of interest in here.

Moreover, exact inference in Bayesian Networks is NP-hard, which makes it necessary to use approximations. Since the cognitive system sketched in here will ultimately be integrated into our humanoid robot ASIMO, real-time algorithms are of special importance. An interesting modification of the standard message-passing therefore is the anytime algorithm “localized partial evaluation” (Draper, 1995). Instead of point-valued messages it uses interval-valued messages by computing upper and lower bounds of nodes within the “neighborhood” of only the query nodes. As long as we restrict our representation to polytrees, we profit from the polynomial complexity (in number of nodes) of standard belief propagation. Here the number of propagation steps depends on the length of the longest path in the network.

System Integration

Our graphical representation is embedded twofold: It makes use of public available databases on common sense knowledge to allow for high-level reasoning and it is embedded in a top-down controlled intentional vision system which connects concepts with real-world measurements (Eggert, Rebhan & Körner, 2007).

Relational Network

The basic representational dimensions used in our model of short- and long-term memory (STM, LTM) are associated with two bodies of knowledge which are of outstanding interest for most cognitive tasks: knowledge about hierarchical relationships and ontological knowledge about properties and subclass relations. As we argued in Röhrbein, Eggert and Körner (2007a), these semantic links can be associated with different functional subsystems of the cortical column. For the representation of arbitrary relations (like “gives”, “made of” etc.) we use so-called relational prototypes, which function as a template and which are connected with all instances of that type (Röhrbein, Eggert & Körner, 2007b).

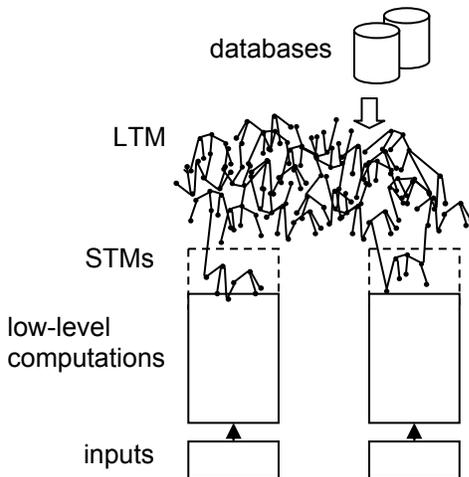


Figure 3: Sketch of the overall system. Inputs are processed in many different channels and interact with stored LTM representations via STM entities. Several databases provide the knowledge for the relational LTM network

All these network links differ in two important aspects to common semantic network links: First, we only use a very restricted set of basic link types, which are biologically justified, since they can be associated with specific cortical source and target populations. Second, these links do not vary from node to node, but are common to all nodes. This is in contrast to the conception of STM and LTM by Navalpakkam and Itti (1995), who commit that their symbolic knowledge architecture is not biological.

Coupling with ontological knowledge

One important aspect of our approach is that we use only one uniform node type in the network, which thus is the representational entity of all concepts in both STM and LTM. The motivation for this homogenous layout is that the basic structure of the biological column is widely independent of the cortical site (although there are variations, e.g. Barone & Kennedy, 2000). For the knowledge instantiation of the relational LTM we made use of several databases (see Röhrbein et al., 2007a), for instance about 200.000 assertions which were collected by the Open Mind Common Sense project at M.I.T.

Here we make use only of knowledge about hierarchical relationships which is usually expressed in meronymies and holonymies, but also in relations like “is located in” or in the temporal domain (“happens during” etc). These are quite essential as it is well known that hierarchies are used all over the neocortex as the core organization principle to deal with the nested structure of the surrounding world.

Grounding with Saliency Data

For the grounding of concepts with sensory data, STM nodes for “proto-objects” are provided, which are linked to sensory measurements and properties like size and shape (Fig. 4). These entities are structured in a way that resembles the biological separation in a feature extraction stream and a pathway for saliency computation.

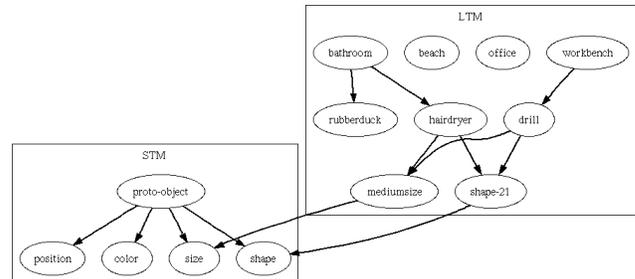


Figure 4: The ontological knowledge in LTM includes properties like color and size which can be linked to appropriate objects in STM. The figure, which is generated from our system, is motivated by an example from Barr (2004) on context influences.

In any probabilistic inference system the basic task amounts to computing the posterior probability distribution for a set of query variables, given new information about some evidence variables. In our semantic Bayesian network several types of reasoning can be performed, depending on these two sets of variables: *Diagnostic reasoning*, which asks for an object or scene given one or several measurements. *Predictive reasoning*, which starts with evidence for a particular scene and makes predictions about objects to be recognized and features to be measured. *Intercausal reasoning*, where we have evidence for a feature and a corresponding object and are interested in the

probability of alternative object hypotheses. Here the evidence for the particular object already “explains away” the evidence for the alternatives. *Combined reasoning*, which asks for an object given evidence for both the current scene and one or several measurements.

If a search task now is provided by triggering a concept like “hairdryer”, an instance of the prototype entity is generated and several feature nodes have values attached, which are extracted from the triggered LTM representations. This target object can then be matched against stored STM entities, which have all the same structure. If the match is insufficient, e.g. because features are missing, this STM entity can now be enriched by a highly selective modulation of the saliency computation and a direct feature extraction. This is possible due to the information attached to the target object. If no object at all is found in STM a new default entity is used to trigger this top-down processing. For the instantiation of root and leaf nodes and other details we refer to Rebhan et al. (2008).

References

- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5, 617-629.
- Barone, P. & Kennedy, H. (2000). Non-uniformity of Neocortex: Areal Heterogeneity of NADPH-diaphorase Reactive Neurons in Adult Macaque Monkeys. *Cerebral Cortex*, 10 (2), 160-174.
- Callaway, E.M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience* 2: 47-74.
- Chater, N., Tenenbaum, J.B. & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences* Vo. 10(7): 287-291.
- Deneve, S. (2005). Bayesian inference in spiking neurons. *Advances in Neural Information Processing Systems* 17: 353-360.
- Draper, D. (1995). *Localized partial evaluation of Bayesian belief networks*. PhD thesis, Computer Science Department, University of Washington, Seattle.
- Eggert, J., Rebhan, S. & Körner, E. (2007). First steps towards an intentional vision system. *Proceedings of the 5th International Conference on Computer Vision Systems (ICVS)*, Bielefeld University, Germany.
- Felleman D.J. & Van Essen D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*. 1(1):1-47.
- George, D. & Hawkins, J. (2005) A hierarchical Bayesian model of invariant pattern recognition in the visual cortex. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*.
- Ichinose, T. & Murakoshi, T. (1996). Electrophysiological elucidation of pathways of intrinsic horizontal connections in rat visual cortex. *Neuroscience*, 73: 25-37.
- Lee, T.S. & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America*, 20(7):1434-1448.
- Lomber, S.G. & Payne, B.R. (2000) Translaminar differentiation of visually guided behaviors revealed by restricted cerebral cooling deactivation. *Cerebral Cortex*, 10, 1066-1077.
- Minsky, M. (2006). *The emotion machine*. Simon and Schuster Publisher, New York.
- Mountcastle, V.B. (2003). Introduction to special issue about computation in cortical columns. *Cerebral Cortex*, 13, 2-4.
- Murphy, K.P., Weiss, Y. & Jordan, M. (1999). Loopy belief propagation for approximate inference: an empirical study. In *Proceedings of Uncertainty in AI*, pages 467-475.
- Navalpakkam, V. & Itti, L. (1995). Modeling the influence of task on attention. *Vision Research*, 45, 205-231.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.
- Rao, R.P.N. (2004). Bayesian Computation in Recurrent Neural Circuits. *Neural Computation*, 16(1): 1-38.
- Rebhan, S., Röhrbein, F., Eggert, J. & Körner, E. (2008). Attention modulation using short- and long-term knowledge, in *6th Int. Conf. Computer Vision Systems*, accepted.
- Rioult-Pedotti, M.S., Friedman, D. & Donoghue, J.P. (2000). Learning-induced LTP in neocortex. *Science*, 290: 533-536.
- Rockland, K.S. (1994). The organization of feedback connections from area V2 (18) to V1 (17). In *Cerebral Cortex*, vol.10, A. Peters and K. S. Rockland, eds., New York: Plenum Press.
- Röhrbein, F., Eggert, J. & Körner, E. (2007a). A Cortex-Inspired Neural-Symbolic Network for Knowledge Representation. In: *Proceedings of the IJCAI-07 Workshop on Neural-Symbolic Learning and Reasoning (NeSy)*. CEUR Workshop Proceedings.
- Röhrbein, F., Eggert, J. & Körner, E. (2007b). Prototypical Relations for Cortex-Inspired Semantic Representations. *Proceedings of the 8th International Conference on Cognitive Modeling (ICCM)*, Ann Arbor, U.S.A. 307-312.
- Suzuki, W., Saleem, K.S. & Tanaka, K. (2000). Divergent backward projections from the anterior part of the inferotemporal cortex (area TE) in the macaque. *Journal of Comparative Neurology* 422, 206-228.
- Thomson, A.M. & Bannister, A.P. (2003). Interlaminar connections in the neocortex. *Cerebral Cortex*, 13, 5-14.
- Yu, A.J. & Dayan, P. (2005). Inference, attention, and decision in a Bayesian neural architecture. In *Advances in Neural Information Processing Systems* 17. MIT Press, Cambridge, MA.
- Yuille, A. & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10(7): 301-308.