Bayesian models and simulations in cognitive science

Giuseppe Boccignone[1], Roberto Cordeschi[2]

[1]Natural Computation Lab
Dipartimento di Ing.dell'Informazione e Ing. Elettrica
Università di Salerno
via Ponte Melillo, 1 -84084 Fisciano (SA), Italy
boccig@unisa.it

[2]Dipartimento di Studi Filosofici ed Epistemologici
Università di Roma "La Sapienza"
via Carlo Fea 2, I-00161 Roma, Italy
cordeschi@caspur.it

Abstract

Bayesian models can be related to cognitive processes in a variety of ways that can be usefully understood in terms of Marr's distinction among three levels of explanation: computational, algorithmic and implementation. In this note, we discuss how an integrated probabilistic account of the different levels of explanation in cognitive science is resulting, at least for the current research practice, in a sort of unpredicted epistemological shift with respect to Marr's original proposal.

1 Introduction

Sophisticated probabilistic models are finding increasingly wide application across the cognitive and brain sciences.

It has been argued [Knill et al., 1996, Chater et al., 2006] that probabilistic models can be related to cognitive processes in a variety of ways. This variety can be usefully understood in terms of Marr's [1982] widely known distinction between three levels at which any agent carrying out a task must be understood, the *what/why* level (computational theory), the *how* level (algorithm), the *physical realization* (implementation):

- *Computational theory.* What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?

- *Representation and algorithm.* How can the computational theory be implemented? What is the representation for the input and output, and what is the algorithm for the transformation?
- *Implementation.* How can the representation and algorithm be realized physically?

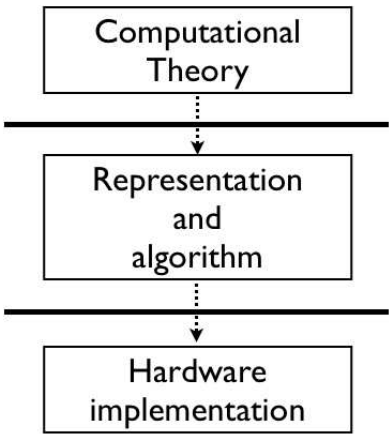

Figure 1: The three levels of explanation suggested by Marr.

Indeed, in research work on theoretical foundations of cognitive science Marr's account has become a sort of paradigm.

In this note, we discuss how an integrated probabilistic account of the different levels of explanation in cognitive science is resulting, at least for the current research practice, in a sort of unpredicted epistemological shift with respect to Marr's original proposal.

From a broader viewpoint such account can be seen as a tenet of philosophy of science that complex systems are to be seen as typically having multiple levels of organization. The behavior of a complex system, e.g., a particular organism, might be explained at various levels of organization, including (but not restricted to) ones which are biochemical, cellular, neurological, psychological.

Although Marr referred to the highest level of analysis as a theory, the term model [Giere, 2004] is more appropriate: the cognitive scientist uses model $\mathcal{M}$ to represent an aspect of the world $\mathcal{W}$ for purposes $\mathcal{P}$ (for details on the development of the concept of model from cybernetics to cognitive science, see Cordeschi [2002]). So, in the following we will refer to this level as the level of the computational model. In general, $\mathcal{M}$ could be many things, but Marr is easily recognized as addressing a kind of *theoretical model* [Giere, 2004, Hartmann and Frigg, 2005].

The intermediate level of analysis can be considered as a computer *simulation*, in the sense of Hartmann [2005], but constrained by model $\mathcal{M}$.

The lowest level is that of the device in which the process is to be physically realized. Clearly, if natural cognition is the focus, then neurophysiology/neuroanatomy is the level that should be addressed.

2 Bayesian explanations

In Chater et al. [2006] and Knill et al. [1996], the Bayesian framework is used to formalize Marr's three-fold hierarchy in two levels: the computational level and the implementation level (embedding both Marr's algorithmic and physical realization levels).

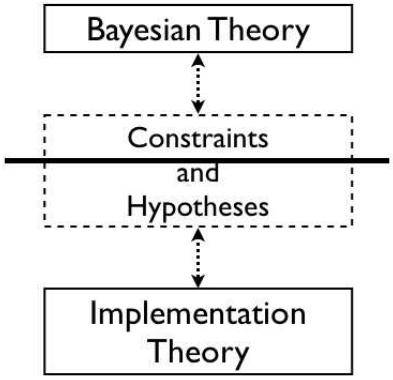

Figure 2: The levels of explanation according to Yuille and Kersten (adapted from [Knill et al., 1996]).

Interestingly enough, both levels are denoted "theories" [Chater et al., 2006] and, differently from Marr, a tight interaction between computational (here Bayesian) theory and implementation theory is assumed.

Beyond such claims and independently of methodological issues, current research practice in Bayesian cognition seems to reflect and widen such epistemological shift.

Much cognitive processing is naturally interpreted as uncertain inference which can be shaped by probabilistic methods at the computational level. At such level the Bayesian framework is exploited to specify the so called generative model, via the joint probability density function (pdf).

The joint pdf $P(\{X_k\}_{k=1}^K)$ of the random variables (RVs) $X_1, \cdots, X_K$ is defined and constraints accounted for through a graphical model $\mathcal{GM}$ (e.g., a Bayesian network, or a Markov Random Field), an iconic representation of RVs dependencies where each node of the graph represents a RV, and arrows denote conditional dependencies between variables.

To make this point clear, consider the following example (a similar one is discussed in Kersten and Yuille [2003]). We want to model the ability of the visual system in recognizing that the object represented in the two images shown in Fig. 3 is one and the same despite considerable image variation caused by a viewpoint change (object invariance).

Figure 3: The same object seen under different views.

More precisely, the issue here is to make a decision about the object O, when the image I being observed takes the value I^* (this is also known as the evidence), while discounting the "noisy" change of view V; in probabilistic terms, the problem amounts to infer the marginal probability $P(O|I = I^*)$.

Thus, knowledge is represented by the joint probability $P(O, V, I)$. Note that such pdf is symmetrical with respect to the three variables O, V, I. In fact any three variable distribution $P(X_1, X_2, X_3)$ may be written, by application of product rule of probability, in any of the 6 ways

$$P(X_{i_1}|X_{i_2}, X_{i_3})P(X_{i_2}|X_{i_3})P(X_{i_3})$$

where (i_1, i_2, i_3) is any of the 6 permutations of $(1, 2, 3)$. Hence, whilst all graphically different, they all represent the same distribution which does not make any conditional independence statements.

In the example we are considering, we can assume that the data I are determined (generated) by the object O plus a possible view change V; further, it is plausible to assume that the target object O does not directly influence the set of possible views V, neither views affect object model O (conditional independence $P(O, V) = P(O)P(V)$). In other terms, the problem at hand suggest a joint probability decomposition of the type

$$P(O, V, I) = P(I|O, V)P(O, V) = P(I|O, V)P(O)P(V) \tag{1}$$

whose equivalent iconic representation is the graph $\mathcal{GM} = \{(O, V) \rightarrow I\}$. Thus, the generative model of object appearance under a change of view can be depicted as in Fig. 4.

(A common practice is to work the way round: design the graph structure that best represent the addressed problem, and then straightforwardly turn such representation in the corresponding joint pdf decomposition).

Recall that we want to infer the marginal probability $P(O|I = I^*)$ given the evidence (observation) $I = I^*$. For simplicity, assume that probabilities $P(I|O, V), P(O), P(V)$ specifying the generative model in Eq. 1 are actually known, either in parametric form or in terms of conditional probability tables.

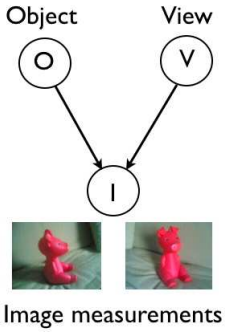

Figure 4: A directed graphical model constraining the $P(O, V, I)$ distribution. The image pair represent actual observations $I = I^1, I = I^2$.

Inference on variable O can be performed via the Belief Propagation (BP) algorithm described in the Appendix.

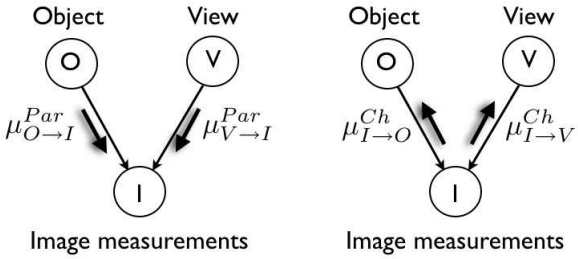

Figure 5: Message passing in BP: from parents O, V to child I (left) and from child to parents (right).

The BP algorithm is an efficient way to solve inference problems based on passing local messages. By applying the procedure described in the Appendix, in this simple case we can straightforwardly pass messages from parents O, V to the evidential node I, as

$$\mu^{Par}_{O\to I}(O) = P(O),$$
$$\mu^{Par}_{V\to I}(V) = P(V).$$

These messages forward prior probabilities of objects and views

To pass a message from node I to parent node O we need to gather information from node V, and similarly to pass a message from node I to parent node V we need to obtain information from parent node O, namely:

$$\mu^{Ch}_{I\to O}(O) = \sum_V P(I = I^*|O, V)\mu^{Par}_{V\to I}(V) = \sum_V P(I = I^*|O, V)P(V),$$

$$\mu^{Ch}_{I\to V}(V) = \sum_O P(I = I^*|O, V)\mu^{Par}_{O\to I}(O) = \sum_O P(I = I^*|O, V)P(O).$$

Note that $\sum_X P(X)$ denotes a sum with respect to values taken by RV X, e.g., $P(X = x_1) + P(X = x_2) + P(X = x_3)$.

Eventually the probability $P(O|evidence)$ can be approximated as:

$$P(O|I = I^*) \propto P(O) \cdot \sum_V P(I = I^*|O, V)P(V). \tag{2}$$

Summing up, the (Bayesian) computational model can be precisely defined as the pair $\mathcal{M} = \langle P(\{X_k\}_{k=1}^K), \mathcal{GM}\rangle$ (and, obviously, by exactly specifying the form/parameters of conditional distributions)

Model $\mathcal{M}$ together with probabilistic calculus, could in principle be used for Bayesian inference and estimating any variable X_k. Note that by inference we simply mean the computation of these marginal probabilities (beliefs). It is worth noting that for a small Bayesian network as that used before, we could have easily performed direct marginalization. However, when scaled-up to real-world problems, exact Bayesian computations are intractable and approximate algorithms have been developed for both learning and inference (e.g., the Expectation-Maximization algorithm, the Belief Propagation algorithm [MacKay, 2004, Lee and Mumford, 2003], and also Appendix A). The virtue of the BP algorithm is that we can use it to compute approximate marginal probabilities in a time that grows only linearly with the number of nodes in the system. In this perspective BP can be properly considered as a *simulation* of the inference process in the sense of Winsberg [2004]. In other terms, Marr's algorithmic level now performs a computer simulation by using the $\mathcal{GM}$ as a representation (data structure).

Eventually, for what concerns Marr's implementation level, current debate is focusing on whether the brain itself should be viewed in probabilistic terms and the nervous system as implementing probabilistic calculations [Lee and Mumford, 2003, Kersten and Yuille, 2003, Paulin, 2005]. This issue will become evident in the case study presented in the following Section.

3 A case study

One clear and elegant example of Bayesian methodology is provided by Rao [2005] who addresses the issue that neurons in cortical areas V2 and V4 can be modulated by attention to particular locations of an image I (the retinal input).

According to the classical *what/where* model [Mishkin et al., 1983], Rao assumes that neurons in V4 area encode feature F (preferred stimulus) while spatial locations L are encoded within the parietal cortex area; V1 and V2 areas play the role of encoding an intermediate representation, say C.

The hierarchical organization of visual attention processing is summarized in Fig. 6. The same figure shows the corresponding graphical model, where F, L, C, I play the role of random variables.

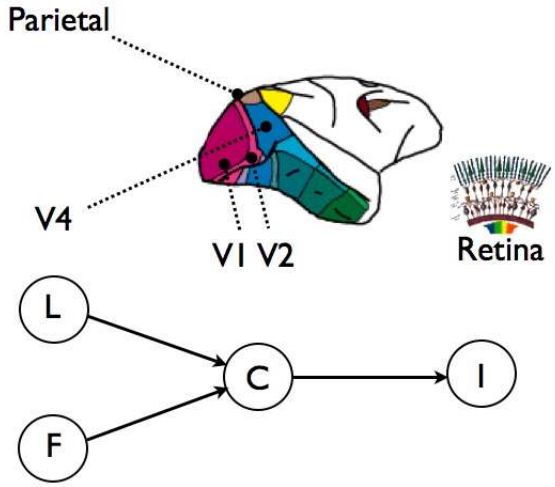

Figure 6: The underlying hierarchical neural architecture (top) supporting visual attention and the corresponding graphical model (bottom).

Marr's levels of explanations for such a problem can be restated as follows.

Computational level. The computational problem is now from our point of view that of estimating the probabilities of gazing at object features F and/or spatial locations L, given an observed input image $I = I^*$, namely $P(L|I = I^*), P(F|I = I^*)$.

Thus, model $\mathcal{M}$ is expressed by the joint pdf $P(F, L, C, I)$, where C is an intermediate RV, and by RV dependencies $\mathcal{GM} = \{(F, L) \rightarrow C \rightarrow I\}$ (cfr. Fig. 6).

Algorithm. At the algorithmic level, Belief Propagation is put into run on the graph $\mathcal{GM}$ by computing the steps described in Appendix A until convergence.

At convergence, it is easy to show that the marginals of interest related to non-evidential variables F, L, C, under the evidence $I = I^*$, can be estimated as:

$$P(C|I = I^*) \propto P(I = I^*|C) \sum_{F,C} P(C|L, F)P(F)P(L) \tag{3}$$

$$P(L|I = I^*) \propto P(L) \sum_{F,C} P(C|L,F)P(F)P(I = I^*|C) \tag{4}$$

$$P(F|I = I^*) \propto P(F) \sum_{F,C} P(C|L,F)P(C)P(I = I^*|C) \tag{5}$$

Note that $P(L), P(F)$ represent prior probabilities of focussing attention on certain locations or specific features.

Implementation. At the neural level, a recurrently connected network of spiking neurons is assumed; such network is just another kind of graphical model, say $\widetilde{\mathcal{GM}}$.

More precisely, $\widetilde{\mathcal{GM}}$ is a recurrent network of standard leaky integrator neurons, where for each neuron k the dynamics of the instantaneous membrane potential $v_k(t)$ is given by:

$$\tau \frac{dv_k(t)}{dt} = \overbrace{-v_k(t)}^{\texttt{neuron membrane potential}} + \overbrace{f(\sum_j w_{kj} i_j(t))}^{\texttt{external injected current}} + \overbrace{g(\sum_j u_{kj} v_j(t))}^{\texttt{recurrent connections}} ; \tag{6}$$

here w_{kj}, u_{kj} represent input neuron $i_j(t)$ and recurrent connection synaptic strengths, respectively; f, g are (non-linear) dendritic filtering functions.

The connection between BP estimates given in Eqs. 3, 4, 5, and neuron dynamics (Eq. 6) is derived as follows.

First, neuron's dynamics is discretized via a finite difference scheme obtaining

$$v_k(t+1) = f(\sum_j w_{kj} i_j(t)) + g(\sum_j \widetilde{u}_{kj} v_j(t)) \tag{7}$$

Second, BP is assumed to be performed in the log domain; considering for instance, Eq. 3, it can be written as:

$$\log P(C|I = I^*) \propto \log P(I = I^*|C) + \log \sum_{F,C} P(C|L,F)P(F)P(L) \tag{8}$$

By comparing the discrete membrane potential equation (Eq. 7) with the log BP equation (Eq. 8) the following correspondences hold:

$$v_k(t+1) \propto \log P(C|I = I^*) \tag{9}$$

$$f(\sum_j w_{kj} i_j(t)) = \log P(I = I^*|C) \tag{10}$$

$$g(\sum_j \widetilde{u}_{kj} v_j(t)) = \log \sum_{F,C} P(C|L,F)P(F)P(L) \tag{11}$$

In particular Eq. 9, which can be rewritten exactly by introducing appropriate constants, $v_k(t+1) \propto c \log P(C|I = I^*) + T$ straightforwardly provides

the following relation:

$$P(C|I = I^*) = \rho \exp\{\frac{v_k(t+1) - T}{c}\} \tag{12}$$

which defines the spiking probability of a neuron of type C in terms of the distance between the membrane potential v and the membrane threshold T

Similar relations can be obtained for the other BP equations and F, L variables.

In other terms, Rao's approach provides a new interpretation of the spiking probability of cortical neurons in terms of posterior probabilities and is used to simulate and predict behaviors of neurons within the V4 cortical area under different conditions [Rao, 2005].

From an architectural standpoint, if for instance feature F can take one among m values/states $\{f_1, f_2, \cdots, f_m\}$, then the F node in $\mathcal{GM}$ is implemented by m leaky integrate-and-fire neurons in $\widetilde{\mathcal{GM}}$. The same holds for nodes L, C. This shows how the design of $\mathcal{GM}$ top-down constrains the design of $\widetilde{\mathcal{GM}}$.

4 Discussion

The case study presented above suggests a remarkable conceptual shift with respect to Marr's three levels proposal.

In the framework of a Bayesian approach, the computational model $\mathcal{M} = \langle P(\{X_k\}_{k=1}^K), \mathcal{GM}\rangle$ is fully specified when the graphical model $\mathcal{GM}$ has been defined. Such graphical model indeed is an *iconic model* of the world $\mathcal{W}$ [Giere, 2004, Hartmann and Frigg, 2005]; in the case at hand (cfr. Fig. 6) conditional dependencies are derived so as to reflect functional dependencies occurring among brain areas. More precisely, $\mathcal{GM}$ is iconic in the sense that the hierarchy $(F, L) \rightarrow C \rightarrow I$ reflects the (Parietal,V4) $\rightarrow$ V2/V1 $\rightarrow$ Retina hierarchy (cfr. Fig. 6). Thus, computational model $\mathcal{M}$, related to cognitive level activity, is specified in terms of the underlying neural architecture as illustrated in Fig. 6, where $\mathcal{GM}$ is conceived as a blueprint for the *biological* neural architecture. In turn, once $\mathcal{GM}$ has been obtained it can be used to top-down constrain the design of the *artificial* neural architecture.

In this respect, the notion of architecture becomes a central tenet in the Bayesian approach in defiance of Marr's methodological effort to provide, in the vein of traditional AI, a careful separation between "hardware" and "software" issues.

As a consequence, the algorithmic level rather than encompassing, as for Marr, specific procedures/routines to solve problems stated at the computational level, becomes a general purpose simulation step: it does not depend on *what* is computed (except for structural constraints imposed by $\mathcal{GM}$), and it is usually used for inference and learning (much like learning algorithms in artificial neural nets). Indeed, when BP equations are put into run they provide a simulation (approximate inference), say $\mathcal{S}$ of model $\mathcal{M}$.

Two further aspects deserve some comments.

First note that Eq. 7 is the discrete form of the differential equation (Eq. 6) that models a leaky integrate-and-fire neuron. Recall that discretization turns differential equations, which relate continuous rates of change over infinitesimal intervals, into difference equations, which relate rates of change over finite, or discrete, intervals. The values that these difference equations give can then be calculated by a digital computer, in discrete time steps. In other words, as pointed out by Winsberg [2004], "finite differencing", namely the transformation of the differential equations into difference equations constitutes a simulation, say $\widetilde{\mathcal{S}}$ of $\widetilde{\mathcal{M}}$, which in the case is the system of differential equations defined on a recurrent network $\widetilde{\mathcal{GM}}$ of leaky integrate-and-fire neurons.

Second, BP equations also provide the starting point for reduction to what Marr would call the implementation level. In fact, reduction is achieved by setting the "correspondence rules" as in Eq. 9, 10, 11. Interestingly enough, the derivation of model $\widetilde{\mathcal{M}}$ from $\mathcal{M}$ (represented by arrow (1) in Fig. 7 below), is achieved as a sort of inter-theoretic reduction [Nagel, 1961] but where the bridge principle (see Eq. 12) involves equations used in simulations $\widetilde{\mathcal{S}}$ and $\mathcal{S}$ (cfr. arrow (3) in Fig. 7) rather than a straightforward logical derivation *á la* Nagel [1961] of the laws or principles of the reduced theory (in this case $\mathcal{M}$) from the laws or principles of the reducing theory $\widetilde{\mathcal{M}}$.

Eventually, it is clear that different from Marr, the implementation level itself is a kind of theoretical model $\widetilde{\mathcal{M}} = \langle P(\{i,t\}_{i=1}^{N}), \widetilde{\mathcal{GM}} \rangle$, related in this case to neural level activity, where $\widetilde{\mathcal{GM}}$ represents the neural architecture whose design is constrained by $\mathcal{GM}$.

The state of affairs that has been so far achieved can be generalized, beyond the specific example, and summarized at a glance as in Fig. 7

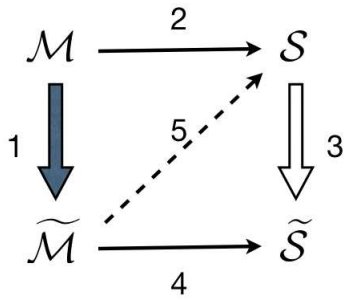

Figure 7: The Bayesian computational model $\mathcal{M}$ and the neural model $\widetilde{\mathcal{M}}$; $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ represent software simulations of models $\mathcal{M}$ and $\widetilde{\mathcal{M}}$, respectively. Arrows (1) and (3) specify ways of reduction from $\mathcal{M}$ to $\widetilde{\mathcal{M}}$.

The same figure also suggests that in the context we have described, simulation $\mathcal{S}$ may also be conceived as a coarse grained simulation of the underlying neural model (arrow (5) in Fig. 7).

5 Final remarks

The conceptual shift with respect to Marr's original proposal is readily apparent.

The Bayesian practice reshapes Marr's original account in a quite different story, where levels of explanation become a hierarchy of models (in Rao's case $\mathcal{M} \rightarrow \widetilde{\mathcal{M}}$ interlaced via simulation) and where the notion of architecture becomes a central one for all levels.

Also, by recalling again Fig. 7, note that process behavior at the neural level (differential equations) can in principle be obtained either straightforwardly (at the same level) by running deterministic finite difference equations ($\widetilde{S}$) or through a coarse grained simulation S performed on probabilities: in this perspective much of the controversial debate on dynamical system hypothesis as opposed to higher level computations becomes an ill-posed question.

Further levels of reduction could be achieved by noting that the firing rate model of the neuron described by the dynamical system given by Eq. 6, is a simplified model of neuron ($\widetilde{\mathcal{M}}$) [Koch, 1999]. It can be easily related to the standard cable equation, which in turn can be obtained through mathematical linearization of reaction-diffusion system of partial differential equations controlling the spatiotemporal dynamics of calcium ions and bound buffer concentrations at the *chemical level* [Koch, 1999].

Thus, in principle, computations can be carried out by using probabilities, or at a lower level using membrane potential as the crucial variable, controlled by the cable equation, or further at the lower level by taking into account concentration of calcium or or other substances governed by reaction-diffusion equations. As pointed out by Koch (see Koch [1999], p. 279) *"the principal difference are the relevant spatial and temporal scales dictated by the different physical parameters, as well as the dynamical range of the [...] sets of parameters"*.

Eventually, it is worth noting that at any level l, a model $\mathcal{M}^l$ can certainly be algorithmically simulated, via $\mathcal{S}^l$, but equivalently could undergo a simulation $\mathcal{A}^l$, via a physical analog (actually Marr's original physical realization level). For instance, at the neuron level one could implement the circuit corresponding to the leaky integrate and fire neuron [Koch, 1999]; at a higher level, inference on the graphical model could be realized through a special purpose hardware message-passing architecture.

To conclude, the complete picture could be depicted as in Fig 8.

In some sense, by taking a Bayesian stance, Marr's proposal seems more related to different kinds of model simulation (namely, $\mathcal{A}^l$ and $\mathcal{S}^l$) at a given level l of explanation, rather than actually involving different levels of explanation. For instance, considering Fig. 8, Marr's proposal can be accounted for by a single horizontal level, whilst the $\mathcal{M}^{l+1} \rightarrow \mathcal{M}^l \rightarrow \mathcal{M}^{l-1}$ hierarchy straightforwardly denotes the hierarchy of explanations.

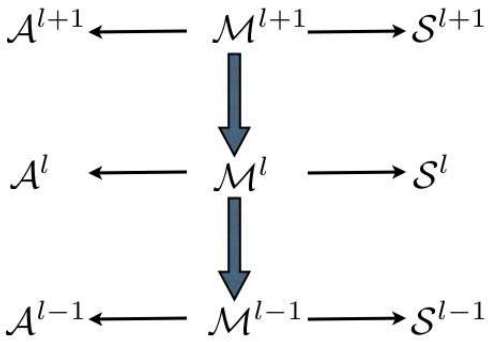

Figure 8: Hierarchy of models corresponding to hierarchy of explanations.

A The Belief Propagation algorithm for directed graphs

A general node X_n has messages coming in from parents $Par(X_n)$ and from children $Ch(X_n)$. We can collect all the messages from parents that will be sent through X_n to any subsequent children as

$$\mu^{Par}_{X_n}(X_n) = \sum_{Par(X_n)} P(X_n|Par(X_n)) \prod_{X_i \in Par(X_n)} \mu^{Par}_{X_i \to X_n}(X_i) \tag{13}$$

Similarly, we can collect all the information coming from the children of node X_n that can subsequently be passed to any parents of X_n

$$\mu^{Ch}_{X_n}(X_n) = \prod_{X_i \in Ch(X_n)} \mu^{Ch}_{X_i \to X_n}(X_n) \tag{14}$$

The bottom-up messages from children are defined as

$$\mu^{Ch}_{X_i \to X_n}(X_n) = \sum_{X_i} \mu^{Ch}_{X_i}(X_i) \sum_{X_k \in Par(X_i) \setminus X_n} P(X_i|Par(X_i)) \prod_{X_k \in Par(X_i) \setminus X_n} \mu^{Par}_{X_k \to X_i}(X_i), \tag{15}$$

and the top-down messages from parents

$$\mu^{Par}_{X_k \to X_n}(X_k) = \mu^{Par}_{X_k}(X_k) \prod_{X_i \in Ch(X_k) \setminus X_n} \mu^{Ch}_{X_i \to X_k}(X_k) \tag{16}$$

The structure of the above equations is that to pass a message from a node X_1 to a child node X_2 , we need to take into account information from all the parents of X_1 and all the children of X_1, except X_2 . Similarly, to pass a message from node X_2 to a parent node X_1, we need to gather information from all the children of node X_2 and all the parents of X_2 , except X_1.

Such schedule is formalized in the Belief Propagation algorithm:

for all evidential nodes X_i **do**
  $\mu_{X_i}^{Par}(X_i) = 1$ for node X_i in the evidential state, 0 otherwise.
  $\mu_{X_i}^{Ch}(X_i) = 1$ for node X_i in the evidential state, 0 otherwise.
for all non-evidential nodes X_i with no parents **do**
  $\mu_{X_i}^{Par}(X_i) = P(X_i)$
for all non-evidential nodes X_i with no children **do**
  $\mu_{X_i}^{Ch}(X_i) = 1$
for every non-evidential node X_i **do**
  repeat
    if X_i has received the μ^{Par}messages from all its parents **then**
      calculate $\mu_{X_i}^{Par}(X_i)$.
    if X_i has received μ^{Ch} messages from all its children **then**
      calculate $\mu_{X_i}^{Ch}(X_i)$.
    if $\mu_{X_i}^{Par}(X_i)$ has been calculated **then**
      for every child X_j of X_i such that X_i has received the μ^{Ch} messages from all of its other children **do**
        calculate and send the message $\mu_{X_i \to X_j}^{Par}(X_i)$
    if $\mu_{X_i}^{Ch}(X_i)$ has been calculated **then**
      for every parent X_j of X_i such that X_i has received the μ^{Par} messages from all of its other parents **do**
        calculate and send message $\mu_{X_i \to X_j}^{Ch}(X_j)$
  until all the μ^{Ch} and μ^{Par} messages between any two adjacent nodes have been calculated
for all non-evidential nodes X_i **do**
  compute the marginal $P(X_i|evidence) \simeq \mu_{X_i}^{Par}(X_i) \cdot \mu_{X_i}^{Ch}(X_i)$

References

N. Chater, J.B. Tenenbaum, and A. Yuille. Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, 10(7):287–291, July 2006.

R. Cordeschi. *The Discovery of the Artificial: Behavior, Mind and Machines Before and Beyond Cybernetics*. Kluwer Academic Publishers, 2002.

R.N. Giere. How models are used to represent reality. *Philosophy of Science*, 71:742 – 752, 2004.

S. Hartmann and R. Frigg. *Scientific Models*, volume 2, pages 740–749. Routledge, New York, 2005.

D. Kersten and A. Yuille. Bayesian models of object perception. *Current Opinion in Neurobiology*, 13:150–158, 2003.

D.C. Knill, D. Kersten, and A. Yuille. *Introduction: A Bayesian formulation of visual perception*, pages 1–21. Cambridge University Press, 1996.

C. Koch. *Biophysics of Computation - Information Processing in Single Neurons*. Oxford University Press, New York, 1999.

T. S. Lee and D. Mumford. Hierarchical bayesian inference in the visual cortex. *J. Opt. Soc. Am. A*, 20(7):1434–1448, 2003.

D.J.C. MacKay. *Information Theory, inference and Learning Algorithms*. Cambridge University Press, Cambridge, UK, 2004.

D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, New York, 1982.

M. Mishkin, LG Ungerleider, and KA Macko. Object vision and spatial vision: Two cortical pathways. *Trends Neurosci*, 6(10):414–417, 1983.

E. Nagel. *The Structure of Science*. Harcourt, Brace, and World, New York, 1961.

M.G. Paulin. Evolution of the cerebellum as a neuronal machine for bayesian state estimation. *J. Neural Eng.*, 2:219–234, 2005.

R.P.N. Rao. Bayesian inference and attentional modulation in the visual cortex. *Neuroreport*, 16:1843–1848, 2005.

E. Winsberg. Sanctioning models: The epistemology of simulation. *Science in Context*, 12(2):275–292, 2004.