What can the PGM community contribute to the 'Bayesian Brain' hypothesis?

Johan Kwisthout J.KWISTHOUT@DONDERS.RU.NL
Donders Institute for Brain, Cognition and Behaviour
Radboud University Nijmegen
PO Box 9104, 6500HE Nijmegen, The Netherlands

Abstract

Despite the now common view amongst neuroscientists that the brain effectively approximates Bayesian inferences (known as the 'Bayesian Brain hypothesis'), there are only few researchers in the PGM community currently working in this research area. We believe that this is partially due to a misunderstanding of the theoretical challenges that theoretical neuroscience currently faces and the potential contribution that the PGM community can offer in interdisciplinary research. With this paper we hope to remedy such misunderstandings and invite the community to contribute to the mutual benefit of neuroscience and AI alike.

Keywords: Bayesian Brain hypothesis; neuroscience; interdisciplinary research.

1. Introduction

When discussing recent advances in neuroscience–that postulate that the human brain is at its essence just an approximate Bayesian inferential machine–with scholars in the Probabilistic Graphical Models (PGM) community, our research group occasionally receives lukewarm responses that can best be paraphrased as "I'm just not interested in the brain as an application area of my research". Although there are few things as personal as a research agenda, we still feel that this lack of interest may be at least partially due to a) a misconception of the questions that are currently being addressed in neuroscience and b) lacking some 'insiders insight' in the contribution that the PGM community can offer in interdisciplinary research. With this paper we hope to remedy both.

Our approach here is orthogonal and complementary to the approach put forward by Bielza and Larrañaga (2014) who described the use of Bayesian networks as *tools for* neuroscientific research, such as reconstructing human brain activity from fMRI data (Schoenmakers et al., 2015), spatial component analysis for Alzheimer's disease diagnosis (Illan et al., 2014), or classification of interneurons (Mihaljević et al., 2014). This is an important area of PGM research, but already sufficiently covered in Bielza and Larrañaga's special issue (Bielza and Larrañaga, 2014). In contrast, in our approach we are interested in (computations on) graphical models as *objects of study in* neuroscience, i.e., computational-level explanations of the brain's information processing activity.

We will give a short overview of the increasingly popular 'Bayesian Brain' hypothesis in neuroscience, in particular its 'predictive processing' manifestation. We will then identify three concrete research areas within this topic where contributions from the PGM community can actually have a huge scientific impact. After identifying some potential pitfalls in such interdisciplinary research, including a discussion of the specific (and sometime peculiar) connotations of the neuroscience community with respect to concepts like 'Bayesian,' 'uncertainty,' and 'prior,' we will conclude with an invitation to the community to contribute.

2. The Brain as 'Application Area' for PGM

Herman von Helmholtz (1867) is traditionally seen as the originator of the view of human perception as (statistical) inference to the best explanation of the causes of the perceptual input. The suggestion that the human brain can be seen as performing some approximate Bayesian inference (integrating prior expectations with newly arriving information) was coined as early as 1957 by Edwin T. Jaynes (first published by Jaynes (1988)). Peter Dayan and colleagues further explored these ideas and proposed the notion of the *Bayesian Brain* (Yu and Dayan, 2005), emphasizing on the basis of psychophysical evidence that human perception actually is 'Bayes' optimal' in combining priors and new signals. The *Bayesian coding* hypothesis (Knill and Pouget, 2004) postulates that the brain indeed represents probability distributions in populations of neurons.

In recent years, the *Bayesian Brain* hypothesis has become increasingly popular due to the emergence of Karl Friston's *free energy principle*, providing for a biological and physical foundation; the *predictive processing* view of the brain as a 'prediction machine' that minimizes computational effort by trying to predict its inputs, and the *spiking neural network* research area that shows that probability distributions can be encoded and sampled from using power-efficient networks of spiking neurons. We will elaborate more on these three important recent developments.

2.1 The free energy principle

Friston's *free energy principle* (Friston, 2009, 2010) postulates that any biological system that 'resists a tendency to disorder' – be it a single cell or a social network – effectively aims to minimize free energy. In thermodynamics, free energy is the amount of energy that is potentially available, but not put to effective use. In information theory, it is a measure on the discrepancy between our observation of the world and our model of the world, which becomes manifest as the *prediction error* between predicted and observed world state. A biological system that aims to defy disorder seeks to lower expected entropy (the average of surprise of future outcomes). It can do so by minimizing prediction error, that is, aiming to make the predicted world state match the observed world state (adapting one's models of the world), or vice versa (changing one's sensory input by acting upon the world). Because biological systems must remain within certain boundaries to exist, their models of what the world should look like (e.g., have access to a sufficient, but not excess, amount of oxygen to maintain homeostasis) and how they currently perceive the world (e.g., shortage of oxygen) should match, and if not, actions are taken to minimize this prediction error (e.g., breathe faster and deeper). Friston (2009, p.295) summarizes this by postulating that *(i) agents resist a natural tendency to disorder by minimizing a free-energy bound on surprise; (ii) this entails acting on the environment to avoid surprises, which (iii) rests on making Bayesian inferences about the world*.

2.2 Predictive processing

The Predictive Processing account proposes that the brain continuously predicts its inputs in a hierarchical cascade of (increasingly more concrete) probabilistic predictions (Clark, 2013, 2015; Hohwy, 2013). For example, when observing a bowler on a bowling lane, contextual information ("this bowler already hit three strikes in this game") will generate predictions for the result of the throw ("many pins will fall down"). Based on that expectation, more specific predictions will be made for the throwing kinematics, the ball trajectory, where the ball will hit the pins, etc. Ultimately this will generate predictions for sensory inputs to, e.g., the retina. Violations of predictions (a miss) will yield prediction errors that need to be 'explained away' by updating ones hypotheses ("even good bowlers will sometimes fail to throw a strike"), taking new contextual information into consideration ("the bowler seems to have injured his wrist whilst throwing") etc. Predictions are made with a specific precision, reflecting uncertainty about outcomes due to limited exposure (i.e, reducible uncertainty) or due to inherent stochasticity of the data-generating process (i.e., irreducible uncertainty). Prediction errors are used to update the generative models to minimize the reducible uncertainty.

The computations 'under the hood' of this conceptual description can be described and analyzed as various computations on causal Bayesian networks, such as the computation of posterior probability distributions, updating hyperparameters of distributions, and tuning of selected parameters of the network (Kwisthout et al., 2017). Despite its popularity as a unifying theory, it is far from clear what the brain's approximation algorithms actually look like; in Clark's words: *What do the local approximations to Bayesian reasoning look like as we depart further and further from the safe shores of basic perception and motor control? What new forms of representation are then required, and how do they behave in the context of the hierarchical predictive coding regime (Clark, 2013, p.201)?*

2.3 Networks of spiking neurons

One of the most promising computational models of neuronal computation in general is the recurrent *network of spiking neurons* model (Maass, 2014). These biologically inspired networks mimic Boltzmann machines (neural networks that represent a probability distribution that can be sampled from), with a key difference that the neurons are not outputting a zero or one state, but a *spike*; a brief burst of energy. These networks are energy-efficient and stochastic in nature and they can represent, and reason with, arbitrary probability distributions by means of stochastic sampling in winner-take-all microcircuits (Buesing et al., 2011; Pecevski et al., 2011; Habenschuss et al., 2013). It has been proposed that such sampling methods (like MCMC sampling) are the most promising techniques to describe actual stochastic inferences in the brain (Tenenbaum et al., 2011). Because of their efficiency – the brain uses a mere 25W of energy – these networks are potentially crucial for future generations of computer hardware by utilizing (rather than trying to filter) the noise that is inherent at the nano-scale (Hamilton et al., 2014). No free lunch is offered, though: As approximate Bayesian inference is an intractable problem (Dagum and Luby, 1993; Kwisthout, 2018), there will be problem instances where the convergence time of the network will grow exponentially with the input size, in particular in networks with extreme probability distributions (Maass, 2014).

In terms of Marr's levels of explanation (Marr, 1982), one can see the free energy principle as aiming to answer the 'why' of the Bayesian Brain hypothesis, the predictive processing account describes 'what' is actually being computed, whereas the 'spiking neurons' community studies the algorithmic 'how' aspect of approximate Bayesian computations in the brain. Where the free energy/predictive processing and the networks of spiking neurons communities were traditionally relatively isolated – as a proxy, one could see them as exponents of the *UK*, respectively *Continental* approach towards theoretical neuroscience – there have been recent mutual research events (for example at the European Institute for Theoretical Neuroscience in Paris) that try to bridge the gap between both communities.

2.4 Organization of this paper

All these developments support the 'Bayesian' view of the brain as it is currently dominant in contemporary neuroscience. We believe that this opens up a significant area of research for the PGM community. In the remainder of this paper we will further elaborate on this. We will show how a formal and computational background can help to bring conceptual clarity and formal rigidity to the field; how neuroscience is in urgent need for new algorithms, implementations, and complexity analyses that computer scientists and AI practitioners can provide, and where new questions in the 'meta'-theory of learning and modifying Bayesian networks emerge.

3. Conceptual Clarity and Rigidity

An important area where researchers with a strong background in computational and formal modeling can make vital contributions is in offering conceptual clarity and formal rigidity, translating verbal theories into complete and consistent computational models, thus exposing ambiguities and gaps in the theory and explicating 'design choices' and their computational consequences (Otworowska et al., 2015). Examples are in the formal explication of the role and nature of the underlying principles of predictive processing (Phillips, 2017; Kay and Phillips, 2011; Thornton, 2017), critically assessing the validity of simplifying assumptions (Otworowska et al., 2014, see also Figure 1), and in exposing the consequences of alternative readings of vague or conflicting verbal models (Kwisthout and van Rooij, under review). On top of this, the specific background of researchers in the PGM community can contribute significantly to the theory itself, generating new theoretical and empirical questions. The following case study will further exemplify this.

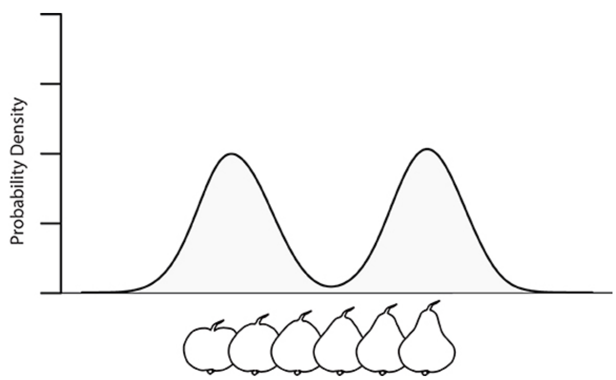

Figure 1: Recognition density for apples and pears based on the shape of the bulbous cone. Observe that, based on the frequency of apple-shapes, pear-shapes, and 'intermediate shapes' in the world, this recognition density cannot be assumed to be a simple Gaussian density; a violation of the Laplace assumption in Friston (2010). Picture reprinted (with permission) from Otworowska et al. (2014).

In the predictive processing theory, precision-weighted stochastic predictions are compared with actual observations and only the residual (non-predicted) signal is processed by the brain. Here, 'processed' means that by belief revision or by intervention the model and the reality are adapted to converge; a process denoted by prediction error minimization. For example, when we are tossing a coin to decide which team will start a match, initially we have uniform probability distributions predicting who wins the toss (the home or away team), what the outcome of the coin toss is (heads or tails), and what visual stimuli we observe (either side of the coin). Note that these predictions are uncertain due to the inherent stochasticity of tossing coins, and will inevitably induce a prediction error when the coin lands as this will generate one bit of information that could not yet be predicted. This information is propagated 'upwards' by the prediction error minimization mechanism: the outcome is updated to 'heads,' which induces a prediction error with the original uniform prediction; in turn, the winner of the toss is updated to the away team to minimize this prediction error. Prediction error minimization is thus the mechanism by which information is processed in the brain.

Prediction errors, however, are dependent on the state space of the prediction and its granularity (the number of categories distinguished). In the absence of a coin we might have used a regular die and predict 'odd' or 'even' instead. We thus *lower* the typical state space of a die throw. Similarly, we might think of different sets of predicted inputs made by a couple strolling through the forest on a Sunday afternoon and an arborist looking for potentially hazardous situations in the same forest. From a modeling perspective: When we move from Gaussian densities to describe predictions in early vision or simple motor control to discrete probability distributions to describe higher cognitive capacities, we need to define what our categories are, and the granularity of our categories determines the prediction error. If we interpret the outcome of a die throw as odd or even, the prediction error decreases from 2.58 bits to 1 bit. This observation—made from an information-theoretic point of view—led to a further refinement of the predictive processing account with the notion of *levels of detail* of models and predictions (Figure 2), and spawned various research projects (Kiverstein et al., 2017; Kwisthout et al., 2017; Pink-Hashkes et al., 2017).

4. Theory, Algorithms, and Analysis

Most, if not all, computational problems in Bayesian networks are intractable. For example, inference is PP-complete (Littman et al., 1998), which implies that there cannot exist efficient approximation algorithms in general, unless BPP equals PP; casting a possible shadow over the biological plausibility of the Bayesian brain hypothesis. It has been suggested (e.g., (Clark, 2013, p.25, p.31)) that processing only the prediction error is less computationally demanding as processing the entire input and that predictive processing thus allows for a tractable implementation of the Bayesian Brain hypothesis. This assumption, however, does not (by and of its own) render inferences tractable. It was shown that processing even a single bit of prediction error is an NP-hard problem (Kwisthout, 2014). Recent developments in the area of *fixed-parameter tractability* allow for the analysis of stochastic computations where the probability of answering incorrectly is parameterized, rather than the computation time (Kwisthout, 2015, 2018). This allows for the study of so-called *fixed error randomized tractable* approximations, relative to 'ecologically valid' parameters, viz. parameters that can plausibly be assumed to be small in the computations as performed by the brain. In a separate paper submitted to this conference we show that the (relative) size of the prediction error plays virtually no role at all in tractability considerations: approximations are intractable or tractable, relative to a set of parameters, irrespective of the size of the prediction error (Donselaar, 2018). This effectively defies Clark's assertion; the biological validity of constraining parameters that *do* render

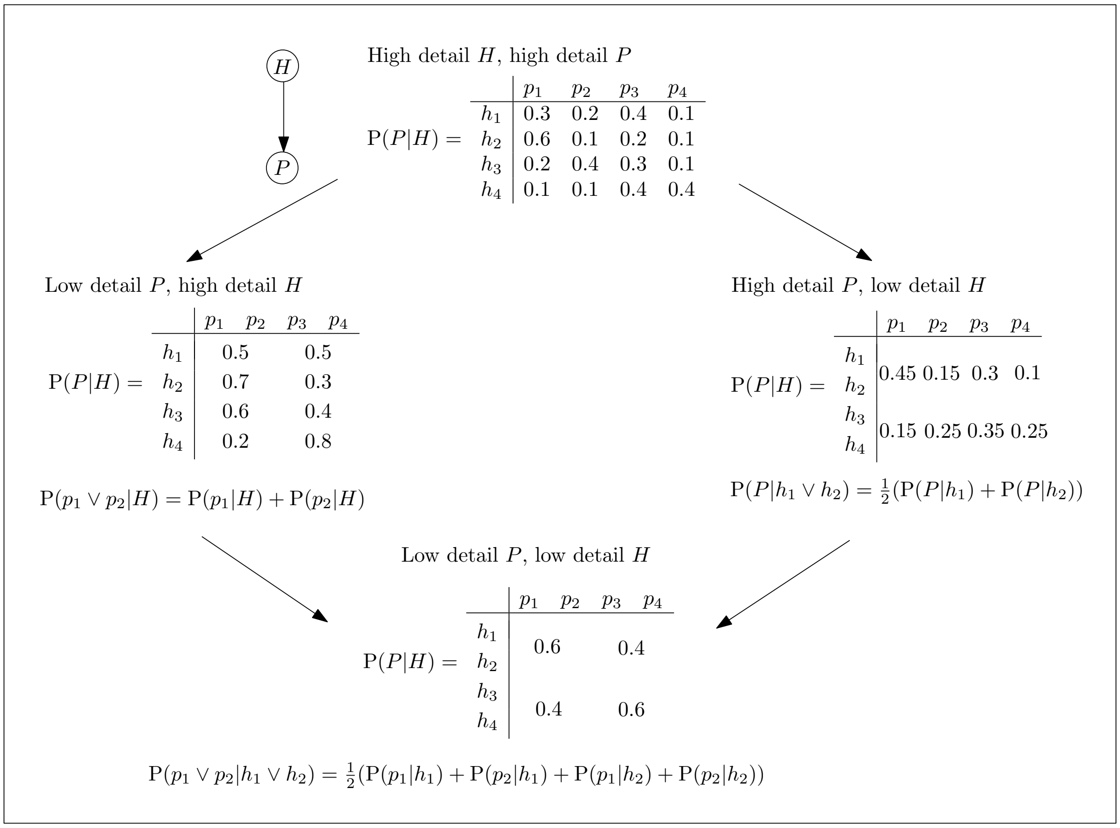

Figure 2: A formalization of the relationship between different levels of detail of hypotheses and predictions. Observe that actual hypotheses, as well as predictions, can be *clustered*, re-defining the conditional probability distributions in a straightforward way.

---

approximation tractable, such as the local variance bound (Dagum and Luby, 1997), is currently under investigation.

Apart from process-level considerations (under what constraints can the approximations postulated by predictive processing be tractable), one can study the properties and plausibility of neuronal implementations of such approximations using networks of spiking neurons. Crucial properties here are the power efficiency of such networks (Maass, 2014), the nature of the *noise* in the brain and its consequences for efficient sampling (Habenschuss et al., 2013), and the general question how many resources are needed for effective computations (Maass, 2000). Computational complexity theory offers an indication of the resources needed for a particular computational problem to be solved, as a function of the input size of a problem. These resources most notably, time and memory are typically fairly coarse and built on a theoretical abstract model of computation: Turing machines. Here, the ‘time’ resource refers to the number of state transitions in the machine, and the ‘memory’ resource refers to the number of memory cells on the tape that are used. It has been proposed by a working group at the Dagstuhl seminar on Resource-Bounded Problem Solving (seminar 14341) to have a more refined, brain-focused model of computation in the brain, based on networks of spiking neurons, and have complexity measures based on brain resources, such as spiking rates, network size, and connectivity (Haxhimusa et al., 2014). The development of such a model of computation would allow for seminal contributions to the Bayesian Brain hypothesis by analyzing the fundamental limits of brain computations.

5. Meta-theory of Bayesian Networks

When a prediction error is to be accounted for, one can either update ones current beliefs about the actual hypotheses, act upon the world in order to bring the reality closer to the desired state, or try to reduce uncertainty by observing hidden variables. These predictive processing sub-processes (belief revision, intervention, and adding observations) correspond to aspects of parameter tuning and sensitivity analysis (Coupé et al., 2000), counterfactual and prospective reasoning (Pearl, 2000), and selecting evidence (van der Gaag and Bodlaender, 2011). Several conceptual issues are still not resolved; for example, how counterfactual models can be built up and how we can use structure equation models to reason about *what* action we should undertake. Algorithmic and analytical aspects of these problems are of direct relevance to the Bayesian Brain hypothesis.

When learning a Bayesian network from data one might reconstruct the structure of the network, the probability distributions, and even the distributions over hidden variables. Crucially, though, one needs to settle beforehand on the variables and their state space. This is to be contrasted with how generative models in the Bayesian brain hypothesis are actually constructed: Here, one somehow needs to 'learn' new variables and the values they can take, both for potential causes and their observable manifestations. The question then arises *when* a Bayesian learner realizes that the current model is insufficient and new hypotheses should be formed, as well as *what* these hypotheses should look like (Carroll and Kemp, 2013). This process can be coined as *model revision* (Figure 3, right panel) where a model is structurally adapted because of an unexpected high prediction error. This problem comes on top of *model updating* by Jeffrey updating (Jeffrey, 1965, see also Figure 3, left panel), where just the probability distributions are updated in the light of new evidence, *and* on top of *model refinement* strategies (Figure 3, middle panel) where the level of detail of models increases to overcome plateauing of the prediction error. The problem of adding new variables and values of variables to a network in the light of unresolvable prediction error is a major open problem in the theory.

Another vital open problem in the predictive processing account relates to the trade-off between making predictions that are very *detailed* and predictions that are likely to be *correct*. For example, when predicting the outcome of a throw at a bowling lane, a prediction over a distribution containing values like 'pin four will be hit by the ball from the left side and will topple over pins seven and eight' is very detailed, but probably always gives a huge prediction error. On the other hand, a prediction like 'the ball will hit the pins and some will fall' is likely to be correct, but as a prediction not very informative. There are reasons to believe that particular neurotransmitters control this *level of detail* (Pink-Hashkes et al., 2017), but from a more meta-perspective it is completely open how causal Bayesian models can be 'flexible' in their granularity and how algorithms on such models may trade-off information gain and prediction error.

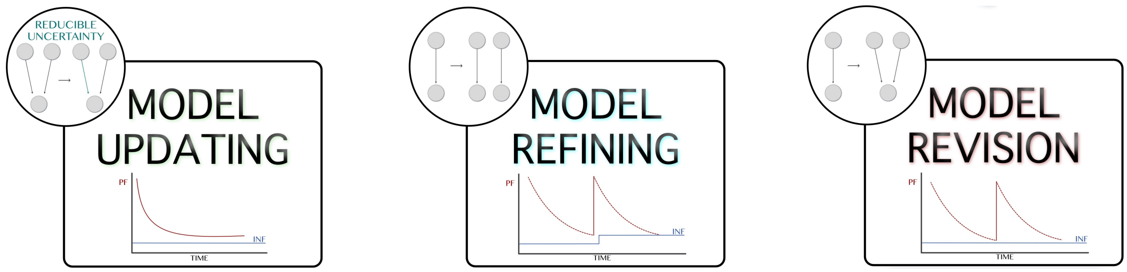

Figure 3: Model updating, model refinement, and model revision processing relative to the prediction error and the amount of irreducible uncertainty. See the main text for explanation of these strategies.

6. Potential Pitfalls

In the previous sections we highlighted several research areas and tentative research questions where the PGM community can substantially contribute to the 'Bayesian Brain' with a potential for considerable impact. Notwithstanding this potential, there are also pitfalls to avoid that are inherent risks of interdisciplinary work, in particular when the research fields have different cultures and tradition and use specific terminology that may be misunderstood. Here we enumerate a few potential pitfalls.

- '**Terminology**' — An informal quiz at the interdisciplinary Lorentz Center workshop 'Perspectives on Human Probabilistic Inference'[1] on the association that participants had with the word 'Bayesian' was illuminative to us. For some participants *Bayesian* was a synonym of *probabilistic*, for others it concerned the semantics of probability distributions (*subjective*, as contrasted with *frequentist*), yet others associated *Bayesian* with *Bayes' rule* for updating distributions. In cognitive science communities, *Bayesian* is often synonymous with *optimal* models and contrasted with *heuristic* explanations. Despite the traditional interpretation of 'Bayesian' as 'subjective degrees of belief' (Jaynes, 2003), it is not uncommon for proponents of the Bayesian Brain hypothesis to have a strong frequentist view on probabilities as describing the objective state of the world (Fiorillo, 2012). Similarly diverse (and sometimes counterintuitive) associations could be elicited for terms like 'prior,' 'uncertainty,' 'information,' and 'structure.' The bottom line is to be aware of potential misunderstandings and to be explicit of one's intended meaning of such terms in communication with neuroscientists.
- '**Culture and tradition**' — In computer science and artificial intelligence, acceptance of a paper to a prestigious conference such as AAAI, UAI, NIPS or STOC is distinctive. Many scholars focus their publication strategy on such conferences, rather than journal papers. In neuroscience, a conference publication is close to irrelevant when it comes to evaluating research output; much more emphasis is put on the impact factor of the journals one is publishing in. Culture and tradition put emphasis on different 'golden standards' of excellence in

1. `http://www.lorentzcenter.nl/lc/web/2014/627/info.php3?wsid=627&venue=Oort`

research, validity of research methodology, and importance of research topics. Awareness of such issues and an open mind may help avoid or solve misunderstandings.

- '**Interdisciplinary**' — Members of interdisciplinary teams have different backgrounds and distinct areas of expertise; that is exactly the main benefit of having interdisciplinary collaborations at all. There is a fine line between 'nitpicking on details' versus 'allowing crucial misconceptions to exist' in interdisciplinary collaborations, and it requires some expertise to see what is important and what not. For example, it is rarely important to insist on the distinction between NP-hardness and NP-completeness of a problem, but the difference between an observation and an intervention in (causal) Bayesian networks may well be important to clarify. Don't assume your neuroscience collaborators share your background, and don't be afraid to ask for clarification about what seems obvious to them. But do understand that a major intellectual effort will be spent on thoroughly understanding each other where this is important for scientific progress.

- '**Selling your work**' — An elegant intractability proof or a new formalization of a verbal theory is typically not sufficient for publication in neuroscience outlets. In order to get published one should aim to understand the problems that neuroscientists care about, make clear why your contribution is instrumental in solving these problems, and write in a way that connects to their background and expectations. It might be difficult to convince one's departmental chair or (grant) reviewers of the relevance of this work. Our approach is to seek for niches that both allow for a significant PGM contribution *and* solve crucial problems with respect to the Bayesian Brain.

7. Conclusion

Despite the potential pitfalls we identified in the previous section, we strongly believe computer scientists and AI practitioners working in the PGM area can make a vital interdisciplinary contribution to contemporary theoretical neuroscience. With this paper we hope to have given an overview of crucial open problems in the Bayesian Brain hypothesis and a sketch of the contributions that the PGM community can offer. We conclude this paper with this quote from Karl Friston that (probably inadvertently) illustrates the importance of research on probabilistic graphical models for theoretical neuroscience: *Life (...) is an inevitable and emergent property of any (ergodic) random dynamical system that possesses a Markov blanket* (Friston, 2013). We would like to invite the community to bring their toolbox of computational and formal modeling and help to advance this fascinating research area — who knows what else may emerge!

Acknowledgments

First ideas for this paper emerged from several discussions at the workshop "Perspectives on Human Probabilistic Inference" organized in May 2014 at the Lorentz Center in Leiden, The Netherlands. A preliminary version of this paper was presented at the 2016 Benelux Conference in Artificial Intelligence.

References

C. Bielza and P. Larrañaga. Bayesian networks in neuroscience: A survey. *Frontiers in Computational Neuroscience*, 8:Article 131, 2014.

L. Buesing, J. Bill, B. Nessler, and W. Maass. Neural dynamics as sampling: A model for stochastic computation in recurrent networks of spiking neurons. *PLoS Computational Biology*, 7(11): e1002211, 2011.

C. D. Carroll and C. Kemp. Hypothesis space checking in intuitive reasoning. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 2013.

A. Clark. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013.

A. Clark. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press, 2015.

V. M. H. Coupé, F. V. Jensen, U. B. Kjærulff, and L. C. van der Gaag. A computational architecture for n-way sensitivity analysis of Bayesian networks. Technical report, Aalborg University, 2000.

P. Dagum and M. Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60(1):141–153, 1993.

P. Dagum and M. Luby. An optimal approximation algorithm for Bayesian inference. *Artificial Intelligence*, 93:1–27, 1997.

N. Donselaar. Parameterized hardness of active inference. In *Proceedings of PGM'18*, 2018.

C. Fiorillo. Beyond Bayes: On the need for a unified and Jaynesian definition of probability and information within neuroscience. *Information 2012, 3(2)*, 3(2):175–203, 2012.

K. Friston. The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, 13(7):293–301, 2009.

K. Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11 (2):127–138, 2010.

K. Friston. Life as we know it. *Journal of the Royal Society Interface*, 10(86):20130475, 2013.

S. Habenschuss, Z. Jonke, and W. Maass. Stochastic computations in cortical microcircuit models. *PLoS Computational Biology*, 9(11):e1003037, 2013.

T. Hamilton, S. Afshar, A. van Schaik, and J. Tapson. Stochastic electronics: A neuro-inspired design paradigm for integrated circuits. *Proceedings of the IEEE*, 5:843–859, 2014.

Y. Haxhimusa, I. van Rooij, S. Varma, and H. T. Wareham. Resource-bounded problem solving (dagstuhl seminar 14341). *Dagstuhl Reports*, 4(8), 2014.

J. Hohwy. *The Predictive Mind*. Oxford University Press, 2013.

I. Illan, J. Górriz, J. Ramírez, and A. Meyer-Base. Spatial component analysis of MRI data for Alzheimer's disease diagnosis: a Bayesian network approach. *Frontiers in Computational Neuroscience*, 8:156, 2014.

E. Jaynes. How does the brain do plausible reasoning? In G. J. Erickson and C. R. Smith, editors, *Maximum-Entropy and Bayesian Methods in Science and Engineering*, 1988.

E. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

R. Jeffrey. *The Logic of Decision*. University of Chicago Press, 1965.

J. W. Kay and W. A. Phillips. Coherent infomax as a computational goal for neural systems. *Bulletin of Mathematical Biology*, 73(2):344–372, 2011.

J. Kiverstein, M. Miller, and E. Rietveld. The feeling of grip: Novelty, error dynamics, and the predictive brain. *Synthese*, 2017. doi: https://doi.org/10.1007/s11229-017-1583-9.

D. Knill and A. Pouget. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–719, 2004.

J. Kwisthout. Minimizing relative entropy in hierarchical predictive coding. In L. van der Gaag and A. Feelders, editors, *Proceedings of PGM'14*, LNCS 8754, pages 254–270, 2014.

J. Kwisthout. Tree-width and the computational complexity of MAP approximations in Bayesian networks. *Journal of Artificial Intelligence Research*, 53:699–720, 2015.

J. Kwisthout. Approximate inference in Bayesian networks: Parameterized complexity results. *International Journal of Approximate Reasoning*, 93:119–131, 2018.

J. Kwisthout and I. van Rooij. Predictive processing and the Bayesian brain: Intractability hurdles that are yet to overcome. *Computational Brain and Behavior*, under review.

J. Kwisthout, H. Bekkering, and I. van Rooij. To be precise, the details don't matter: On predictive processing, precision, and level of detail of predictions. *Brain and Cognition*, 112:84–91, 2017.

M. L. Littman, J. Goldsmith, and M. Mundhenk. The computational complexity of probabilistic planning. *Journal of Artificial Intelligence Research*, 9:1–36, 1998.

W. Maass. Neural computation: a research topic for theoretical computer science? Some thoughts and pointers. In *Bulletin of the European Association for Theoretical Computer Science (EATCS)*, volume 72. 2000.

W. Maass. Noise as a resource for computation and learning in networks of spiking neurons. *Proceedings of the IEEE*, 102(5):860–880, 2014.

D. Marr. *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman, 1982.

B. Mihaljević, C. Bielza, R. Benavides-Piccione, J. DeFelipe, and P. Larrañaga. Multi-dimensional classification of GABAergic interneurons with Bayesian network-modeled label uncertainty. *Frontiers in Computational Neuroscience*, 8:150, 2014.

M. Otworowska, J. Kwisthout, and I. van Rooij. Counter-factual mathematics of counterfactual predictive models. *Frontiers in Consciousness Research*, 5:801, 2014.

M. Otworowska, J. Riemens, C. Kamphuis, P. Wolfert, L. Vuurpijl, and J. Kwisthout. The robo-havioral methodology: Developing neuroscience theories with FOES. In *Proceedings of the 27th Benelux Conference on AI (BNAIC'15)*, 2015.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge: MIT Press, 2000.

D. Pecevski, L. Bueling, and W. Maass. Probabilistic inference in general graphical models through sampling in stochastic networks of spiking neurons. *PLoS Computational Biology*, 7(12):1–25, 2011.

W. Phillips. Cognitive functions of intracellular mechanisms for contextual amplification. *Brain and Cognition*, 12:39–53, 2017.

S. Pink-Hashkes, I. van Rooij, and J. Kwisthout. Perception is in the details: A predictive coding account of the psychedelic phenomenon. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, pages 2907–2912, 2017.

S. Schoenmakers, U. Güçlü, M. van Gerven, and T. Heskes. Gaussian mixture models and semantic gating improve reconstructions from human brain activity. *Frontiers in Computational Neuroscience*, 8:173, 2015.

J. B. Tenenbaum, C. Kemp, T. Griffiths, and N. Goodman. How to grow a mind: Statistics, structure, and abstraction. *Science*, 331:1279–1285, 2011.

C. Thornton. Predictive processing simplified: The infotropic machine. *Brain and Cognition*, 112: 13–24, 2017.

L. C. van der Gaag and H. L. Bodlaender. On stopping evidence gathering for diagnostic Bayesian networks. In *Proceedings of the Eleventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 6717 of *LNCS*, pages 170–181, 2011.

H. von Helmholtz. *Handbuch der Physiologischen Optik*. Leipzig: Leopold Voss, 1867.

A. J. Yu and P. Dayan. Uncertainty, neuromodulation, and attention. *Neuron*, 46:681–692, 2005.