Beyond the Generative Paradigm: Foundations for Computational Abstraction

A Position Paper

Hiroyuki Kido
Cardiff University
Cardiff, United Kingdom
KidoH@cardiff.ac.uk

ABSTRACT

This is a conceptual position paper contrasting abstraction and generation as two opposing AI paradigms. The success of current AI systems can be attributed to their data-first paradigm. However, data is mathematically a product of models, such as mathematical structures, variables, and parameters, underlying these systems. In fact, estimating models from data in AI is formalised as an inverse problem in statistics. In this paper, we argue that the mismatch between the data-first paradigm and this model-first approach is a fundamental cause of various long-standing open problems such as unifying logic and probability, unifying learning and reasoning, unifying symbol grounding and inference grounding, and brain-like AI. We overview abstractive AI as opposed to generative AI and discuss promising future research directions.

KEYWORDS

Abstractive inference, Brain-like AI, Unifying logic and probability, Unifying learning and reasoning, Unifying symbol grounding and inference grounding, Neuroscience, Cognitive science

1 INTRODUCTION

It is widely accepted that AI research today adheres to a data-first paradigm. Table 1 lists Turing Awards for AI research. Over time, the focus of research has shifted from small-scale handcrafted knowledge to large-scale raw data, particularly following the advent of Internet technologies. From a mathematical perspective, however, modern AI research should be seen as a model-first approach rather than data-first. Here, a model represents a mathematical hypothesis, such as structures, variables, and parameters. Indeed, the problem of estimating models from data can be regarded as an inverse problem in statistics. In other words, it is assumed that data are derived from models. For example, the statistical methods most widely used to train machine learning models are maximum likelihood (ML) estimation, maximum a posteriori (MAP) estimation, and Bayesian estimation, defined as follows.

$$h_{ML} = \arg\max_h p(\boldsymbol{d}|h)$$
$$h_{MAP} = \arg\max_h p(h|\boldsymbol{d}) = \arg\max_h p(\boldsymbol{d}|h)p(h)$$
$$p(h|\boldsymbol{d}) = \frac{p(\boldsymbol{d}|h)p(h)}{p(\boldsymbol{d})} = \frac{p(\boldsymbol{d}|h)p(h)}{\sum_h p(\boldsymbol{d}|h)p(h)} \quad (1)$$

Given data, denoted by $\boldsymbol{d}$, they infer a hypothesis, denoted by h, or its probability using the expressions on the right-hand side. The likelihoods on the right-hand side indicate that the data result from the hypothesis. They clearly reflect a model-first approach.

Table 1: Turing awards in AI [17], including the 2024 award

Year	*Rationale*
1969-71	Foundations of the field based on **representation and reasoning**
1994	Developing **expert systems** that encode human knowledge to solve real-world problems
2011	Developing **probabilistic reasoning techniques** that deal with uncertainty in a principled manner
2018	Making **deep learning** (multilayer neural networks), a critical part of modern computing
2024	Developing the conceptual and algorithmic foundations of **reinforcement learning**

We argue that the conflict between the data-first paradigm and the model-first approach needs to be resolved, as it underlies several fundamental challenges in AI, including unifying logic and probability, unifying learning and reasoning, unifying symbol grounding and inference grounding, transparency, hallucinations, and brain-like AI. Although current AI research has achieved remarkable results in areas such as language, speech, and image processing, these fundamental problems remain largely unsolved.

In this paper, we discuss a data-first approach to the data-first paradigm. Specifically, we explore an alternative research direction in which inherently abstract models are derived from inherently concrete data through a process of abstraction, i.e., selective ignorance.[1] We refer to the computational study of abstraction as Computational Abstraction, which focuses on the opposite direction of the current generative paradigm, where data are derived from models. For example, we use Bayes' theorem as follows from the abstraction perspective.

$$p(d|\boldsymbol{h}) = \frac{p(\boldsymbol{h}|d)p(d)}{p(\boldsymbol{h})} = \frac{p(\boldsymbol{h}|d)p(d)}{\sum_d p(\boldsymbol{h}|d)p(d)} \quad (2)$$

Given hypotheses, denoted by $\boldsymbol{h}$, this equation infers the probability of a data point, denoted by d, using the expression on the right-hand side. The likelihoods on the right-hand side indicate that the hypotheses result from the data point. This clearly reflects a data-first approach.

Such abstractive inference cannot be categorised within the typical machine learning dichotomies of generative vs. discriminative models or parametric vs. non-parametric models. Data are neither generated nor discriminated, but rather they are what generate

[1] This reflects a materialistic view, consistent with 18th-century mathematicians, who regarded the determination of unobservable forces from observable object trajectories as a direct (or forward) problem [1].

Table 2: Full joint probability distribution

		Sprinkler=0				*Sprinkler*=1			
		Cloud=0		*Cloud*=1		*Cloud*=0		*Cloud*=1	
		Hot=0	*Hot*=1	*Hot*=0	*Hot*=1	*Hot*=0	*Hot*=1	*Hot*=0	*Hot*=1
Rain=1	*Wet*=1	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
	Wet=0	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{16}
Rain=0	*Wet*=1	θ_{17}	θ_{18}	θ_{19}	θ_{20}	θ_{21}	θ_{22}	θ_{23}	θ_{24}
	Wet=0	θ_{25}	θ_{26}	θ_{27}	θ_{28}	θ_{29}	θ_{30}	θ_{31}	θ_{32}

something more abstract. Equation (2) suggests a similarity to a parametric model, as the hypotheses can be viewed as summaries of the data, which are obtained by selectively ignoring certain aspects of the data. At the same time, it also resembles a non-parametric model, since it retains and uses all the data much like instance-based and memory-based learning. Abstractive inference, however, should not be considered a non-parametric model, since non-parametric models do not contradict the generative paradigm in which data are generated from models.

In this paper, we contrast abstraction and generation as two opposing paradigms, and discuss several interesting insights emerging from the former. In Section 2, we critically re-examine Bayesian networks, the probabilistic foundation of generative models, from the data-first perspective. In Section 3, we discuss an alternative approach that begins probabilistic reasoning from the data distribution, which grounds the model distribution discussed during the era of Bayesian networks. In Section 4, we discuss the future prospects of the computational approach to abstraction as a computational model of higher-level cognitive functions of the brain.

2 REVISITING THE CLASSIC

Probability theory underlies modern AI and plays a central role in generative modelling. Bayesian networks (BNs) [14] have had a major impact on the probabilistic approach to generative modelling. To correctly understand BNs, it is important to understand probabilistic inference with the full joint probability distribution that was assumed before their development. Table 2 shows an example of the full joint probability distribution over five discrete random variables: $Cloud$, $Rain$, Hot, $Sprinkler$, and Wet. These variables respectively represent: 'There are clouds', 'It is raining', 'It is hot', 'The sprinkler is on', and 'The grass is wet'.

θ_n is a variable representing a probability value characterising the distribution. Since this is a probability distribution, we have $\theta_{32} = 1 - \sum_{n=1}^{31} \theta_n$. Namely, this distribution is characterised by $31(= 2^5 - 1)$ values. In principle, any probabilistic query can be answered once the full joint probability distribution is available. However, this approach is generally infeasible, as the computational costs, i.e., memory and time, grow exponentially with the number of random variables. For example, given 30 binary random variables, the full joint probability distribution is characterised by $1,073,741,823(= 2^{30} - 1)$ values. The same applies in continuous cases, where exact calculations require multivariate integration.

This issue motivated the development of BNs, which use a directed acyclic graph to encode independence and conditional independence of knowledge.[2] Let $X_i (1 \leq i \leq I)$ be discrete random

[2]The development of BNs was inspired by the top-down and bottom-up processing of human perception in reading [15].

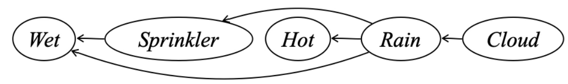

Figure 1: Bayesian network structure

variables. The semantics of BNs defines probabilistic reasoning as follows.

$$p(X_1, ..., X_I) = \prod_{i=1}^{I} p(X_i | X_{i-1}, ..., X_1) \overset{\text{Def.}}{=} \prod_{i=1}^{I} p(X_i | \text{Parents}(X_i))$$

The second expression can be derived by repeatedly applying a valid rule of probability theory, known as the product rule [2, 17]. The third expression applies the semantics of BNs. It is thus important to understand that the semantics of BNs is an approximation of the valid rule of probability theory. For example, consider the BN structure shown in Figure 1. Given this BN structure, the above equations can be expanded as follows, where each random variable is abbreviated by its first letter.

$$p(W, S, H, R, C) = p(W|S, H, R, C)p(S|H, R, C)p(H|R, C)p(R|C)p(C)$$
$$\overset{\text{Def.}}{=} p(W|S, R)p(S|R)p(H|R)p(R|C)p(C)$$

The last equation decomposes the full joint probability distribution into the following conditional probability distributions.

S	*R*	*p*(*W*|*S*,*R*)
0	0	θ_1
0	1	θ_2
1	0	θ_3
1	1	θ_4

R	*p*(*S*|*R*)
0	θ_5
1	θ_6

R	*p*(*H*|*R*)
0	θ_7
1	θ_8

C	*p*(*R*|*C*)
0	θ_9
1	θ_{10}

p(*C*)
θ_{11}

The distributions are now characterised by 11 values, reduced from 31 values due to the semantics of BNs.

The impact of this semantics becomes significant as the scale of the problem increases. Consider the 28×28-pixel images in the MNIST database [4]. Let $Digit$ be a random variable taking values in $\{0, 1, ..., 9\}$, representing the possible digits shown in each MNIST image, and let $Pixel_i$ be a random variable taking values in $\{0, 1\}$, representing the colour of pixel i, black and white, respectively. The full joint probability distribution, $p(Digit, Pixel_1,..., Pixel_{28\times28})$, is thus characterised by $10 \times 2^{28\times28} - 1 (\approx 10^{237})$ values. Now, the semantics of BNs defines a naive Bayes as follows.

$$p(Digit, Pixel_1, ..., Pixel_{28\times28}) \overset{\text{Def.}}{=} p(Digit) \prod_{i=1}^{28\times28} p(Pixel_i | Digit)$$

The full joint probability distribution can now be decomposed into the $1 + 28 \times 28$ conditional probability distributions shown below.

p(*Digit*=0)	*p*(*Digit*=1)	…	*p*(*Digit*=8)
θ_1	θ_2	…	θ_9

Digit	*p*(*Pixel*$_1$|*Digit*)
0	θ_{10}
⋮	⋮
9	θ_{19}

Digit	*p*(*Pixel*$_2$|*Digit*)
0	θ_{20}
⋮	⋮
9	θ_{29}

…

Digit	*p*(*Pixel*$_{28\times28}$|*Digit*)
0	θ_{7840}
⋮	⋮
9	θ_{7849}

Note that $p(Digit = 9) = 1 - \sum_{j=1}^{8} p(Digit = j)$. The conditional probability distributions are characterised by $7849(= 9 + 10 \times 28 \times$

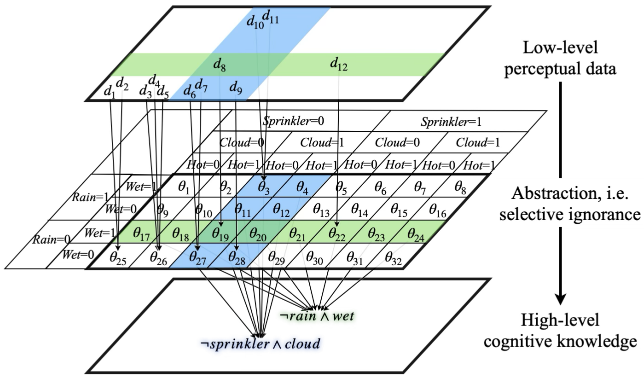

Figure 2: This hierarchy shows that intrinsically abstract propositional formulas (on the bottom layer) should be derived from intrinsically concrete data (on the top layer) by inference of abstraction, i.e., selective ignorance.

$28 (\approx 10^4)$) values, a huge reduction from $10 \times 2^{28\times 28} - 1 (\approx 10^{237})$ due to the semantics of BNs.

However, the fundamental limitation of the semantics of BNs is that independence and conditional independence are assumptions that are useful but unrealistic in practice. Indeed, the BN structures maximising the likelihood almost always result in a fully connected graph (see, e.g., p.745 [17]). Therefore, the independence and conditional independence assumptions are incompatible with the data-first paradigm. It is still important to study what constitutes good BN structures and how to find them efficiently. In the future, however, it will become increasingly important to ask whether a structure is truly necessary and what alternative approaches can realise structure-free yet computationally tractable probabilistic reasoning.

3 ABSTRACTIVE INFERENCE

Generative models underpin the current AI paradigm. Its common assumption is that data result from models (see Equation (1) again). In this section, we focus on the direction opposite to the generative paradigm. The perspective we find in the reverse direction is that intrinsically abstract models result from intrinsically concrete data through a process of abstraction, i.e., selective ignorance (see Equation (2) again). We refer to the computational approach in the opposite direction as Computational Abstraction.

In this section, we explain the core idea of abstractive inference [8–11], which is a recent success in Computational Abstraction. Consider Table 2 again. Several entries represent unrealistic situations, such as $(Rain = 1, Wet = 0)$, $(Sprinkler = 1, Wet = 0)$, and $(Rain = 1, Sprinkler = 1)$. Empirically, this would result in $\theta_n = 0$, for all entries highlighted in grey, i.e., $n \in \{5\text{-}16, 29\text{-}32\}$. The idea of abstractive inference is straightforward. Instead of referring to the full joint probability distribution, we directly use the observed data. This approach is promising as the number of data increases only linearly with the number of observations, whereas the size of the full joint probability distribution increases exponentially with the number of random variables.

The term 'abstractive inference' is justified when we consider expressive formal languages, e.g., propositional logic. Figure 2 is an illustration of abstractive inference with a propositional language. The top, middle, and bottom layers represent a data distribution, model distribution, and knowledge distribution, respectively. Note that the middle layer is Table 2. We call the middle layer a model distribution because each entry corresponds to a model in propositional logic, i.e., valuation. A valuation in propositional logic is a function assigning a truth value, 0 or 1, to each of the five propositional variables. Intuitively, each valuation represents a different state of the world. Each arrow indicates that its source supports its target. For example, data d_6 supports a model with probability θ_{27}. We here assume that each data point supports a single model. The assumption is natural because different models are incompatible, and each model is a selective ignorance of certain aspects of more concrete data. The model with probability θ_{27} supports formula $\neg sprinkler \wedge cloud$ as this formula is true in the model. This support relation conforms to the semantics of propositional logic.

Let m be a function that maps each data point to a corresponding model, and m_n be a model with the probability θ_n. Let us evaluate the probability of the formulas according to Figure 2 (see [8–11] for rigorous definitions). We abbreviate each propositional variable by its first letter. Using the sum rule [2] (marginalisation) and the product rule [2] (conditioning), we have

$$p(\neg r \wedge w, \neg s \wedge c) = \sum_{n=1}^{32}\sum_{k=1}^{12} p(\neg r \wedge w, \neg s \wedge c, m_n, d_k) =$$

$$\sum_{n=1}^{32}\sum_{k=1}^{12} p(\neg r \wedge w | \neg s \wedge c, m_n, d_k) p(\neg s \wedge c | m_n, d_k) p(m_n | d_k) p(d_k).$$

Next, the semantics of propositional logic guarantees that formulas are conditionally independent of data given a model.[3] We thus have

$$= \sum_{n=1}^{32}\sum_{k=1}^{12} p(\neg r \wedge w | \neg s \wedge c, m_n) p(\neg s \wedge c | m_n) p(m_n | d_k) p(d_k).$$

For the same reason, formulas are conditionally independent of each other given a model. We thus have

$$= \sum_{n=1}^{32}\sum_{k=1}^{12} p(\neg r \wedge w | m_n) p(\neg s \wedge c | m_n) p(m_n | d_k) p(d_k).$$

Note that this expression corresponds to the arrows in Figure 2. Now, the model summation can be cancelled after swapping the order of the summations. This is because each data point supports a single model, i.e., $p(m_n | d_k) = 0$, for all $m_n \neq m(d_k)$.

$$= \sum_{k=1}^{12} p(d_k) \sum_{n=1}^{32} p(\neg r \wedge w | m_n) p(\neg s \wedge c | m_n) p(m_n | d_k)$$

$$= \sum_{k=1}^{12} p(d_k) p(\neg r \wedge w | m(d_k)) p(\neg s \wedge c | m(d_k))$$

In the simplest case, we have the uniform probability $p(d_k) = 1/12$ and $p(x | m(d_k)) = 1$ iff propositional formula x is true in the model

[3] In ordinary logic, once a model is given, the truth value of any formula is fully determined and independent of any additional formulas or data (see, e.g., [8] for formal proof).

$m(d_k)$.[4] As shown in Figure 2, $p(\neg r \wedge w|m(d_k)) = 1$ if $k \in \{8, 12\}$, and $p(\neg s \wedge c|m(d_k)) = 1$ if $k \in$ {6-11}. Therefore, we have

$$= \frac{1}{12} \sum_{k \in \{8\}} 1 = \frac{1}{12}.$$

This is the essence of abstractive inference with a propositional language. For conditional probabilities, the same procedure should be applied to both the numerator and the denominator. For example,

$$p(\neg r \wedge w|\neg s \wedge c) = \frac{p(\neg r \wedge w, \neg s \wedge c)}{p(\neg s \wedge c)} = \frac{\frac{1}{12}\sum_{k \in \{8\}} 1}{\frac{1}{12}\sum_{k \in \{6\text{-}11\}} 1} = \frac{1}{6}$$

In sum, the natural combination of logical semantics and probability theory provides a simple formalisation of the idea of abstraction, i.e., selective ignorance.

Abstractive inference can be justified in terms of computation, statistics, logic, and machine learning. First, given the function m, its computation scales essentially linearly with respect to the number of data. As discussed above, this is because we can cancel the model summation that increases exponentially with the number of propositional variables.[5] Without the function m, its computation still scales essentially linearly with respect to the product of the number of data and the number of propositional variables. This is because we only need to determine the truth value of each propositional variable under each data. Second, it is known that abstractive inference over propositional and first-order formulas generalises the logical consequence relations (see [9, 10]). Third, the probability distribution over models derived by abstractive inference is equivalent to maximum likelihood estimation (see [11]). Fourth, abstractive inference over formulas can be seen as a sort of Bayesian learning [17] with the following form (see [8]).

$$p(\neg r \wedge w|\neg s \wedge c) = \sum_{k=1}^{12} p(\neg r \wedge w|d_k) p(d_k|\neg s \wedge c)$$

Here, the target of marginalisation is data. This is in contrast to standard Bayesian learning, where the target of marginalisation is parameters. The above equation shows that symbolic reasoning is grounded in data in the sense that reasoning from symbols to symbols is essentially reference to data. Finally, it is known that the above equation serves as a general framework for various problems across machine learning (e.g., MNIST [11]), probability (e.g., Markov chains and hidden Markov models [8]) and logic (e.g., propositional logic [10] and first-order logic [9]).

4 HUMAN-LIKE ABSTRACTIVE MACHINES

The past two sections motivate us to conceive Computational Abstraction as a potential computational theory of the human brain, including higher-level cognitive functions. This at least offers a unique perspective as current prominent brain theories in neuroscience and cognitive science fall within the generative paradigm. For example, Bayesian brain hypothesis [12], predictive coding [16], free-energy principle [7], hierarchical Bayesian inference [13], and theory-based induction [18] commonly assume that biological brains have a model of the world that explains concrete sensory input. Moreover, most of them primarily target lower-level perceptual functions of the brain, and there is very little research on higher-level cognitive functions, except theory-based induction [18] in cognitive science. While they focus on inference within a given generative model, Computational Abstraction has the potential to help the community understand how generative models and their associated inference emerge through abstraction from data. It could also provide a way to approach the problem from the opposite perspective, in which low-level perceptual data serve as the primary source of high-level cognitive functions (see Figure 2).

This research direction is highly relevant to the AI community because biological brains hold the key to addressing difficult AI problems, such as unifying logic and probability[6], the separation of learning and reasoning, the symbol grounding and inference grounding problems, knowledge acquisition bottleneck, hallucinations, blackboxness, and massive computational resources. Surprisingly, none of them appear to be a problem for the brain. Moreover, agents are expected to play a more active and creative role in science and the arts in the near future. For creativity to be valuable, agents should acquire and use concepts grounded in their own experiences. They also need the ability to reason logically in order to critically evaluate the value of creativity. At present, modern language models are good at explaining how people tend to think, but not very good at explaining how people ought to think in a normative sense. Filling this gap is a fundamental challenge not only for creative agents but also for moral agents, and abstractive agents have the potential to contribute to this area.

There are several neuroscience findings that can be naturally explained by viewing the brain as an abstractive machine, not as a generative machine. For example, it is known that rats in a maze replay and reverse-replay individual past experiences in their hippocampus [3], seemingly previewing and reviewing their experiences. The hippocampus might retain experiences as accurately as possible. For prediction, a highly flexible and efficient mechanism should decide what to replay and reverse-replay in the brain, because the exact same experience never occurs twice. Here, abstraction may play an important role in ignoring irrelevant details. Notably, the replay and reverse-replay occur even for a single event [6]. Their cognitive skill appears distinct from the inductive generalisation approaches commonly studied in machine learning. It also cannot be easily explained by standard generative models, which typically require massive computational resources, including large datasets, numerous parameters, and substantial energy. Moreover, reasoning relies on the trained models themselves rather than on the data used for training. By contrast, learning and reasoning in brains may rely fundamentally on memory of past experiences and abstraction-guided referencing of that memory. This perspective may suggest a new direction in AI that does not rely on backpropagation, an AI learning method often considered biologically implausible.

[4] $p(x|m(d_k)) \in [0.5, 1]$ is assumed in general. This gives a simple yet rigorous way to avoid division by zero in conditional probability (i.e., the inability to reason from impossible information) (see [8, 10]), to prevent the principle of explosion in ordinary logic (i.e., the inability to reason from inconsistent information) (see [9, 10]), and to mitigate overfitting in machine learning (see [11]).

[5] Independence and conditional independence needed to be introduced in Bayesian networks because probabilistic reasoning was assumed to begin from the middle layer in Figure 2. Researchers at that time did not overlook the top layer. It simply did not exist in the 1980s, before the data-first paradigm. For the same reason, we think logicians at the time did not find the abstractive relation between the middle and bottom layers.

[6] This basic problem has remained mathematically unsolved for over 300 years [5].

REFERENCES

[1] E. J. Aiton. 1964. The inverse problem of central forces. *Annals of Science* 20(1) (1964), 81–99.

[2] Christopher M. Bishop. 2006. *Pattern Recognition and Machine Learning*. Springer New York, NY, 1 New York Plaza, Suite 4600, New York, NY 10004-1562.

[3] Margaret F Carr, Shantanu P Jadhav, and Loren M Frank. 2011. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neuroscience* 14 (2011), 147–153.

[4] Li Deng. 2012. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine* 29, 6 (2012), 141–142.

[5] Pedro Domingos. 2015. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World.* Basic Books, New York City.

[6] David J. Foster and Matthew A. Wilson. 2006. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* 440 (2006), 680–683.

[7] Karl Friston. 2012. The history of the future of the Bayesian brain. *Neuroimage* 62-248(2) (2012), 1230–1233.

[8] Hiroyuki Kido. 2025. A data-transparent probabilistic model of temporal propositional abstraction. In *Twelfth Annual Conference on Advances in Cognitive Systems*. 193–212.

[9] Hiroyuki Kido. 2025. Inference of abstraction for grounded predicate logic. In *Twelfth Annual Conference on Advances in Cognitive Systems*. 213–232.

[10] Hiroyuki Kido. 2025. Inference of Abstraction for Human-Like Logical Reasoning. In *Machine Learning, Optimization, and Data Science*. Springer Nature Switzerland, Cham, 191–206.

[11] Hiroyuki Kido. 2025. Inference of Abstraction for Human-like Probabilistic Reasoning. In *Machine Learning, Optimization, and Data Science*. Springer Nature Switzerland, Cham, 116–131.

[12] David C. Knill and Whitman Richards. 1996. *Perception as Bayesian Inference*. Cambridge University Press, Cambridge University.

[13] Tai Sing Lee and David Mumford. 2003. Hierarchical Bayesian inference in the visual cortex. *Journal of Optical Society of America* 20 (2003), 1434–1448.

[14] Judea Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann; 1st edition, Burlington, Massachusetts.

[15] Judea Pearl. 2018. *A Personal Journey into Bayesian Networks*. Technical Report R-476, <http://ftp.cs.ucla.edu/pub/stat_ser/r476.pdf>. Department of Computer Science, University of California, Los Angeles, CA.

[16] Rajesh P. N. Rao and Dana H. Ballard. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* 2 (1999), 79–87.

[17] Stuart Russell and Peter Norvig. 2020. *Artificial Intelligence : A Modern Approach, Fourth Edition.* Pearson Education, Inc., London, England.

[18] Joshua B. Tenenbaum, Thomas L. Griffiths, and Charles Kemp. 2006. Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences* 10(7) (2006), 309–318.