Inference of Abstraction for Human-like Logical Reasoning

Hiroyuki Kido[0000−0002−7622−4428]

Cardiff University, Cardiff CF10 3AT, UK
KidoH@cardiff.ac.uk

Abstract. Inspired by empirical work in neuroscience for Bayesian approaches to brain function, we give a unified probabilistic account of various types of symbolic reasoning from data. We characterise them in terms of formal logic using the classical consequence relation, an empirical consequence relation, maximal consistent sets, maximal possible sets and maximum likelihood estimation. The theory gives new insights into reasoning towards human-like machine intelligence.

Keywords: Artificial intelligence · Cognition · Symbolic logic · Probability theory · Reasoning and learning · Generative reasoning

1 Introduction

There is growing evidence that the brain is a generative model of environments. The two images shown in Figure 1 would cause the perception that a white triangle and a white square overlay the other objects. A well-accepted explanation of the illusions is that our brains are trained to unconsciously use past experience to see what is likely to happen. Much empirical work argues that Bayesian, or probabilistic generative, models give a clear explanation of how the brain reconciles top-down prediction signals and bottom-up sensory signals, e.g., [8, 13, 7, 12, 14, 9, 5, 23, 22, 3, 10, 1, 20, 29].

An interesting question emerging from this idea is how logical consequence relations, relevant to human higher-order thinking, can be given a Bayesian account. The question is important for the following reasons. First, such an account should result in a simple computational principle that allows logical agents to reason over symbolic knowledge fully from data in an uncertain environment. Second, such a principle is expected to tackle fundamental assumptions of the existing computational models such as statistical relational learning (SRL) [6], Bayesian networks [19], naive Bayes, probabilistic logic programming (PLP) [27], Markov logic networks (MLN) [24], probabilistic logic [18], probabilistic relational models (PRM) [4] and conditional probabilistic logic [25]. For example, they have the implicit assumption that the method used to extract symbolic knowledge from data cannot be applied to the method used to perform logical reasoning over the symbolic knowledge, and vice versa.

In this paper, we simply model how data cause symbolic knowledge in terms of its satisfiability in formal logic. The underlying idea is to see reasoning as

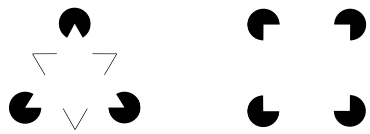

Fig. 1. Kanizsa illusions. It is known that cats select to sit or stand within the illusory square contour just as often as the real square contour [28].

a process of deriving symbolic knowledge from data by abstraction, i.e., selective ignorance. We show that various types of well-grounded symbolic reasoning emerge from direct interaction between data and symbols, not between symbols and symbols. We theoretically characterise them in terms of formal logic.

This paper contributes to new insights into reasoning towards human-like machine intelligence. Symbolic reasoning is essentially a reference to data in our theory. It thus brings up an idea of inference grounding rather than or beyond symbol grounding. Symbolic reasoning can also be seen as interaction between an interpretation in formal logic and its inversion. The inversion, we call inverse interpretation, differentiates our work from the mainstream referred to as inverse entailment [15], inverse resolution [16, 17] and inverse deduction [26], which mainly study dependency between pieces of symbolic knowledge. Our analysis causes reasoning from an impossible source of information, which may be coined as parapossible reasoning, since reasoning from an inconsistent source of information is often referred to as paraconsistent reasoning [21, 2].

In Section 2, we define a generative reasoning model for inference of abstraction. Section 3 gives full logical characterisations of the theory. Section 4 summarises the results.

2 Inference of Abstraction

2.1 Definitions

Let $\{d_1, d_2, ..., d_K\}$ be a multiset of K data. D denotes a random variable of data whose values are all the elements of $\{d_1, d_2, ..., d_K\}$. For all data $d_k (1 \leq k \leq K)$, we define the probability of d_k, denoted by $p(D = d_k)$, as follows.

$$p(D = d_k) = \frac{1}{K}$$

Let L represent a propositional language for simplicity, and $\{m_1, m_2, ..., m_N\}$ be the set of models of L. Each model is a different assignment of a truth value to each atomic formula. Intuitively, each model represents a different state of the world. We assume that each data d_k supports a single model. We use a function m, $\{d_1, d_2, ..., d_K\} \to \{m_1, m_2, ..., m_N\}$, to map each data to the model supported by the data. M denotes a random variable of models whose realisations are all the elements of $\{m_1, m_2, ..., m_N\}$. For all models $m_n (1 \leq n \leq N)$, we define the probability of m_n given d_k, denoted by $p(M = m_n|D = d_k)$, as follows.

$$p(M = m_n|D = d_k) = \begin{cases} 1 & \text{if } m_n = m(d_k) \\ 0 & \text{otherwise} \end{cases}$$

The truth value of a propositional formula and first-order closed formula in classical logic is uniquely determined in a state of the world specified by a model of a language. Let α be a formula in L. We assume that α is a random variable whose realisations are 0 and 1 meaning false and true respectively. We use symbol $[\![\alpha]\!]$ to refer to the models satisfying α. Namely, $[\![\alpha = 1]\!]$ and $[\![\alpha = 0]\!]$ represent the set of models in which α is true and false, respectively. Let $\mu \in [0, 1]$ be a variable, not a random variable. For all formulas $\alpha \in L$, we define the probability of each truth value of α given m_n, denoted by $p(\alpha|M = m_n)$, as follows.

$$p(\alpha = 1|M = m_n) = \begin{cases} \mu & \text{if } m_n \in [\![\alpha = 1]\!] \\ 1 - \mu & \text{otherwise} \end{cases}$$

$$p(\alpha = 0|M = m_n) = \begin{cases} \mu & \text{if } m_n \in [\![\alpha = 0]\!] \\ 1 - \mu & \text{otherwise} \end{cases}$$

Let $[\![\alpha]\!]_{m_n}$ be a function such that $[\![\alpha]\!]_{m_n} = 1$ if $m_n \in [\![\alpha]\!]$ and $[\![\alpha]\!]_{m_n} = 0$ otherwise. The above expressions can be simply written as a Bernoulli distribution with parameter $\mu \in [0, 1]$.

$$p(\alpha|M = m_n) = \mu^{[\![\alpha]\!]_{m_n}} (1 - \mu)^{1-[\![\alpha]\!]_{m_n}}$$

Here, as we will see in the next section, the variable $\mu \in [0, 1]$ plays an important role to relate various types of symbolic reasoning. We will see that $\mu = 1$ relates to the classical consequence relation and its generalisation, and μ approaching 1, denoted by $\mu \to 1$, relates to its generalisation for reasoning from inconsistent sources of information and its further generalisation.

2.2 Properties

In classical logic, the truth value of each formula is determined by a model. In other words, given a model, the truth value of a formula cannot be changed or discarded by the truth value of any other formulas.

Example 1. Let L be a propositional language built with two symbols, $rain$ and wet, meaning 'rain falls' and 'the road gets wet,' respectively. Let $m_n (1 \leq n \leq 4)$ be the models of L. The truth values of the five logical connectives are shown in the following truth table.

	$rain$	wet	$\neg rain$	$rain \land wet$	$rain \lor wet$	$rain \to wet$	$rain \leftrightarrow wet$
m_1	0	0	1	0	0	1	1
m_2	0	1	1	0	1	1	0
m_3	1	0	0	0	1	0	0
m_4	1	1	0	1	1	1	1

Each row represents a different state of the world characterised by the two symbols *rain* and *wet*. It is clear from the truth table that the truth value of each formula is determined given a model.

In probability theory, the truth value of a formula α_1 is thus conditionally independent of the truth value of another formula α_2 given a model M, i.e., $p(\alpha_1|\alpha_2, M, D) = p(\alpha_1|M, D)$ or equivalently $p(\alpha_1, \alpha_2|M, D) = p(\alpha_1|M, D)p(\alpha_2|M, D)$. Let $\Gamma \subseteq L$ be a finite theory. We therefore have

$$p(\Gamma|M, D) = \prod_{\alpha \in \Gamma} p(\alpha|M, D). \tag{1}$$

Moreover, in classical logic, the truth value of a formula depends on models but not data. Thus, in probability theory, the truth value of a formula α is conditionally independent of data D given a model M, i.e., $p(\alpha|M, D) = p(\alpha|M)$. We thus have

$$\prod_{\alpha \in \Gamma} p(\alpha|M, D) = \prod_{\alpha \in \Gamma} p(\alpha|M). \tag{2}$$

Therefore, the full joint distribution $p(\Gamma, M, D)$ can be written as follows.

$$p(\Gamma, M, D) = p(\Gamma|M, D)p(M|D)p(D) = \prod_{\alpha \in \Gamma} p(\alpha|M)p(M|D)p(D) \tag{3}$$

Here, the second equation is derived by the product rule (or chain rule) of probability theory, and the third equation by Equations (1) and (2). The full joint distribution $p(\Gamma, M, D)$ serves as a probabilistic model of logical reasoning from data. We refer to the full joint distribution and reasoning from the distribution as a generative reasoning model and generative reasoning, respectively. We often represent $p(\Gamma, M, D)$ as $p(\Gamma, M, D; \mu)$ if our discussion is relevant to μ. We use symbol ';' to represent that μ is a variable, but not a random variable. In this paper, we assume a finite number of realisations of each random variable.

The full joint distribution implies that we can no longer discuss only the probabilities of individual formulas, but they are derived from data. For example, the probability of $\alpha \in L$ is calculated as follows.

$$p(\alpha) = \sum_m \sum_d p(\alpha, m, d) = \sum_m p(\alpha|m) \sum_d p(m|d)p(d) \tag{4}$$

Here, the second equation is derived by the sum rule of probability theory, and the third equation by Equation (3).

Proposition 1 (Negation). *Let $p(\Gamma, M, D; \mu)$ be a generative reasoning model. For all $\alpha \in \Gamma$, $p(\alpha = 0) = p(\neg\alpha = 1)$ holds.*

Proof. For all models m, α is false in m if and only if $\neg\alpha$ is true in m. Thus, $[\![\alpha = 0]\!] = [\![\neg\alpha = 1]\!]$ is the case. Therefore,

$$p(\alpha = 0) = \sum_m p(\alpha = 0|m)p(m) = \sum_m \mu^{[\![\alpha=0]\!]_m}(1-\mu)^{1-[\![\alpha=0]\!]_m} p(m)$$
$$= \sum_m \mu^{[\![\neg\alpha=1]\!]_m}(1-\mu)^{1-[\![\neg\alpha=1]\!]_m} p(m) = \sum_m p(\neg\alpha = 1|m)p(m) = p(\neg\alpha = 1).$$

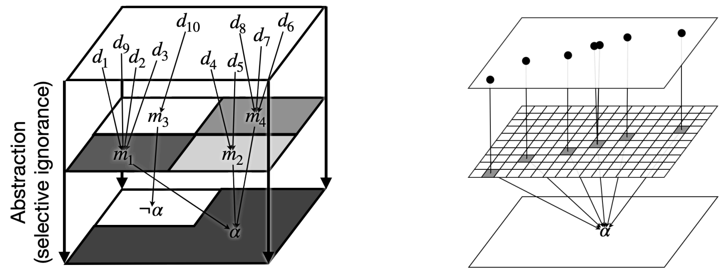

Fig. 2. Right: A schematic of how the probability distribution over data determines the probability distribution over logical formulas. For simplicity, an arrow is omitted if the formula at the end of the arrow is false in the model at the start of the arrow or if the model at the end of the arrow is not supported by the data at the start of the arrow. Left: The summation over models can be eliminated, since each model without data support has a zero probability.

This holds regardless of the value of μ. □

Hence, we replace $\alpha = 0$ by $\neg\alpha = 1$ and abbreviate $\neg\alpha = 1$ to $\neg\alpha$. We also abbreviate $M = m_n$ to m_n and $D = d_k$ to d_k.

The hierarchy shown on the left in Figure 2 illustrates Equation (4). The top layer of the hierarchy is a probability distribution over data, the middle layer is a probability distribution over states of the world, often referred to as models in formal logic, and the bottom layer is a probability distribution over a logical formula α. A darker colour indicates a higher probability. Each element of a lower layer is an abstraction, selective ignorance, of the linked element of the upper.

Example 2 (Continued). Let $d_k (1 \leq k \leq 10)$ be ten data about rain and road conditions. Figure 2 shows which data support which models characterised by the two symbols *rain* and *wet*. The probability of $rain \to wet$ can be calculated using Equation (4) as follows.

$$p(rain \to wet) = \sum_{n=1}^{4} p(rain \to wet|m_n) \sum_{k=1}^{10} p(m_n|d_k)p(d_k)$$

$$= \mu \sum_{k=1}^{10} p(m_1|d_k)\frac{1}{10} + \mu \sum_{k=1}^{10} p(m_2|d_k)\frac{1}{10} + (1-\mu) \sum_{k=1}^{10} p(m_3|d_k)\frac{1}{10}$$

$$+ \mu \sum_{k=1}^{10} p(m_4|d_k)\frac{1}{10} = \frac{4}{10}\mu + \frac{2}{10}\mu + \frac{1}{10}(1-\mu) + \frac{3}{10}\mu = \frac{1}{10} + \frac{8}{10}\mu$$

Therefore, $p(rain \to wet) = 9/10$ when $\mu = 1$ or $\mu \to 1$, i.e., μ approaching 1. The calculation is fully visualised by Figure 2 where α represents $rain \to wet$.

Proposition 2 (Linear-time reasoning). *Let $p(\Gamma, M, D; \mu)$ be a generative reasoning model. For all $\alpha \in \Gamma$, $p(\alpha) = \sum_d p(d)p(\alpha|m(d))$ holds.*

Proof. Equation (4) can be expanded as follows.

$$\sum_m p(\alpha|m) \sum_d p(m|d)p(d) = \sum_d p(d) \sum_m p(\alpha|m)p(m|d) = \sum_d p(d)p(\alpha|m(d))$$

The third expression is derived by the fact that, for all data d, the probability of each model m except one supported by d is zero, i.e., $p(m|d) = 0$. □

Proposition 2 is crucially important because, in contrast to Example 2, the number of models is generally much larger than the number of data. Indeed, the number of models exponentially increases with respect to the number of propositional symbols. For example, 30 propositional symbols cause 2^{30} models. The hierarchy shown on the right in Figure 2 illustrates this effect. The following fact justifies the model distribution (see [11] for the proof).

Proposition 3 (Maximum likelihood estimation). *Let $p(\Gamma, M, D)$ be a generative reasoning model. $p(M)$ represents maximum likelihood estimates.*

3 Correctness

3.1 Reasoning from consistent sources of information

In the previous section, we defined the generative reasoning model $p(\Gamma, M, D)$ and looked at its basic probabilistic properties. In this section, we answer how it generalises classical and non-classical reasoning in terms of the data-based perspective rather than the traditional model-based perspective. We define models without support from data as being impossible.

Definition 1 (Possible models). *Let m be a model of a language L. m is possible if $p(m) \neq 0$ and impossible otherwise.*

For $\Delta \subseteq L$, we use symbol $[\![\Delta]\!]$ to denote the set of all the possible models of Δ, i.e., $[\![\Delta]\!] = \{m \in [\![\Delta]\!] | p(m) \neq 0\}$. We also use symbol $[\![\Delta]\!]_m$ such that $[\![\Delta]\!]_m = 1$ if $m \in [\![\Delta]\!]$ and $[\![\Delta]\!]_m = 0$ otherwise. Obviously, $[\![\Delta]\!] \subseteq [\![\Delta]\!]$, for all $\Delta \subseteq L$, and $[\![\Delta]\!] = [\![\Delta]\!]$ if all models are possible. If Δ is inconsistent, $[\![\Delta]\!] = [\![\Delta]\!] = \emptyset$. If Δ is an empty set or if it only includes tautologies then every model satisfies all the formulas in the possibly empty Δ, and thus $[\![\Delta]\!]$ includes all the models.

This section looks at generative reasoning models with $\mu = 1$, $p(\Gamma, M, D; \mu = 1)$, for reasoning from a consistent source of information. The following theorem relates the probability of a formula to the probability of its models.

Theorem 1. *Let $p(\Gamma, M, D; \mu = 1)$ be a generative reasoning model, and $\alpha \in \Gamma$ and $\Delta \subseteq \Gamma$ such that $[\![\Delta]\!] = [\![\Delta]\!]$.*

$$p(\alpha|\Delta) = \begin{cases} \dfrac{\sum_{m \in [\![\Delta]\!] \cap [\![\alpha]\!]} p(m)}{\sum_{m \in [\![\Delta]\!]} p(m)} & \text{if } [\![\Delta]\!] \neq \emptyset \\ \text{undefined} & \text{otherwise} \end{cases}$$

Table 1. Some inconsistencies between generative reasoning and classical reasoning.

Generative reasoning	Classical reasoning	Rationale
$p(wet\|rain, \neg rain) \neq 1$	$rain, \neg rain \models wet$	$[\![rain, \neg rain]\!] = \emptyset$
$p(wet\|rain) = 1$	$rain \not\models wet$	$[\![rain]\!] \neq [\![\![rain]\!]\!]$
$p(\neg rain \vee wet) = 1$	$\not\models \neg rain \vee wet$	$[\![\emptyset]\!] \neq [\![\![\emptyset]\!]\!]$

Proof. Let $|\Delta|$ denote the cardinality of Δ. Dividing models into the ones satisfying all the formulas in Δ and the others, we have

$$p(\alpha|\Delta) = \frac{\sum_m p(\alpha|m)p(\Delta|m)p(m)}{\sum_m p(\Delta|m)p(m)}$$
$$= \frac{\sum_{m\in[\![\Delta]\!]} p(m)p(\alpha|m)\mu^{|\Delta|} + \sum_{m\notin[\![\Delta]\!]} p(m)p(\alpha|m)p(\Delta|m)}{\sum_{m\in[\![\Delta]\!]} p(m)\mu^{|\Delta|} + \sum_{m\notin[\![\Delta]\!]} p(m)p(\Delta|m)}.$$

By definition, $p(\Delta|m) = \prod_{\beta\in\Delta} p(\beta|m) = \prod_{\beta\in\Delta} \mu^{[\![\beta]\!]_m}(1-\mu)^{1-[\![\beta]\!]_m}$. For all $m \notin [\![\Delta]\!]$, there is $\beta \in \Delta$ such that $[\![\beta]\!]_m = 0$. Therefore, $p(\Delta|m) = 0$ when $\mu = 1$, for all $m \notin [\![\Delta]\!]$. We thus have

$$p(\alpha|\Delta) = \frac{\sum_{m\in[\![\Delta]\!]} p(m)p(\alpha|m)1^{|\Delta|}}{\sum_{m\in[\![\Delta]\!]} p(m)1^{|\Delta|}} = \frac{\sum_{m\in[\![\Delta]\!]} p(m)1^{[\![\alpha]\!]_m}0^{1-[\![\alpha]\!]_m}}{\sum_{m\in[\![\Delta]\!]} p(m)}.$$

We obtain the theorem, since $1^{[\![\alpha]\!]_m}0^{1-[\![\alpha]\!]_m} = 1^1 0^0 = 1$ if $m \in [\![\alpha]\!]$ and $1^{[\![\alpha]\!]_m}0^{1-[\![\alpha]\!]_m} = 1^0 0^1 = 0$ if $m \notin [\![\alpha]\!]$. In addition, if $[\![\Delta]\!] = \emptyset$ then $p(\alpha|\Delta)$ is undefined due to division by zero. □

Recall that a formula α is a logical consequence of a set Δ of formulas, denoted by $\Delta \models \alpha$, in classical logic iff (if and only if) α is true in every model in which Δ is true, i.e., $[\![\Delta]\!] \subseteq [\![\alpha]\!]$. The following Corollary shows the relationship between the generative reasoning model and the classical consequence relation $\models$.

Corollary 1. *Let $p(\Gamma, M, D; \mu = 1)$ be a generative reasoning model, and $\alpha \in \Gamma$ and $\Delta \subseteq \Gamma$ such that $[\![\Delta]\!] = [\![\![\Delta]\!]\!]$ and $[\![\Delta]\!] \neq \emptyset$. $p(\alpha|\Delta) = 1$ iff $\Delta \models \alpha$.*

Proof. By the assumptions $[\![\Delta]\!] = [\![\![\Delta]\!]\!]$ and $[\![\Delta]\!] \neq \emptyset$, $p(m)$ is non zero, for all m in the non-empty set $[\![\Delta]\!]$. The assumptions thus prohibit a division by zero in Theorem 1. Therefore, $\frac{\sum_{m\in[\![\Delta]\!]\cap[\![\alpha]\!]} p(m)}{\sum_{m\in[\![\Delta]\!]} p(m)} = 1$ iff $[\![\alpha]\!] \supseteq [\![\Delta]\!]$, i.e., $\Delta \models \alpha$. □

The following example shows the importance of the assumptions of $[\![\Delta]\!] = [\![\![\Delta]\!]\!]$ and $[\![\Delta]\!] \neq \emptyset$ in Corollary 1.

Example 3 (Continued). Suppose that the probability distribution is given by $p(M) = (m_1, m_2, m_3, m_4) = (0.5, 0.2, 0, 0.3)$. Table 1 exemplifies differences between the generative reasoning and classical consequence relation. The last column explains why the generative reasoning is inconsistent with the classical

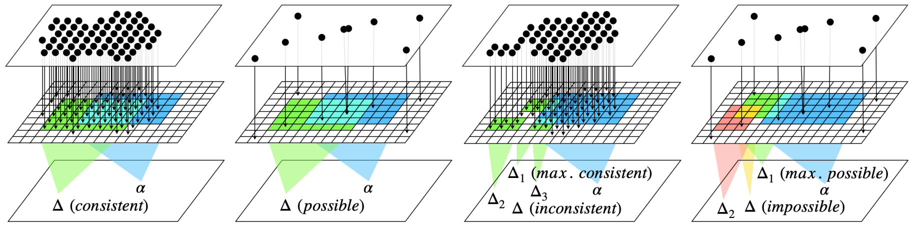

Fig. 3. The left two graphs illustrate reasoning of $\alpha \in \varGamma$ from $\varDelta \subseteq \varGamma$ using $p(\varGamma, M, D; \mu = 1)$. The leftmost shows the assumptions of $\llbracket \varDelta \rrbracket = \lBrack \varDelta \rBrack$ and $\llbracket \varDelta \rrbracket \neq \emptyset$. Each arrow from a datum to model, denoted respectively by a black circle on the top layer and a cell on the middle layer, represents that the datum supports the model. Each model with an incoming arrow thus has a non-zero probability. A model is coloured in green (resp. blue) if all the formulas in $\varDelta$ are (resp. α) true in the model. The second shows the assumption of $\lBrack \varDelta \rBrack \neq \emptyset$. The right two graphs illustrate reasoning of $\alpha \in \varGamma$ from $\varDelta \subseteq \varGamma$ using $p(\varGamma, M, D; \mu \to 1)$. The third shows the assumption of $(\!(\varDelta)\!) = (\!(\!(\varDelta)\!)\!)$. $\varDelta_1$, $\varDelta_2$ and $\varDelta_3$ are the cardinality-maximal consistent subsets of $\varDelta$. The rightmost shows no assumption. $\varDelta_1$ and $\varDelta_2$ are the cardinality-maximal possible subsets of $\varDelta$.

consequence. In particular, the rationale of the last example comes from the fact that Theorem 1 explains $p(\neg rain \vee wet)$ as $p(\neg rain \vee wet|\emptyset)$.

$$p(\neg rain \vee wet) = p(\neg rain \vee wet|\emptyset) = \frac{\sum_{m \in \lBrack \emptyset \rBrack \cap \llbracket \neg rain \vee wet \rrbracket} p(m)}{\sum_{m \in \lBrack \emptyset \rBrack} p(m)}$$

$$= \frac{\sum_{m \in \llbracket \neg rain \vee wet \rrbracket} p(m)}{\sum_{m} p(m)} = \sum_{m \in \llbracket \neg rain \vee wet \rrbracket} p(m) = 1.$$

Here, $\llbracket \emptyset \rrbracket = \{m_1, m_2, m_3, m_4\}$ but $\lBrack \emptyset \rBrack = \{m_1, m_2, m_4\}$.

Figure 3 illustrates the assumptions of $\llbracket \varDelta \rrbracket = \lBrack \varDelta \rBrack$ and $\llbracket \varDelta \rrbracket \neq \emptyset$ for reasoning of $\alpha \in L$ from $\varDelta \subseteq L$ using the generative reasoning model $p(L, M, D; \mu = 1)$. Both α and $\varDelta$ are consistent, since there is at least one model satisfying α and all the formulas in $\varDelta$, i.e., $\llbracket \alpha \rrbracket \neq \emptyset$ and $\llbracket \varDelta \rrbracket \neq \emptyset$. Such models are highlighted on the middle layer in blue and green, respectively. Figure 3 also shows that every model satisfying all the formulas in $\varDelta$ is possible, since there is at least one datum that supports each model of $\varDelta$, i.e., $\llbracket \varDelta \rrbracket = \lBrack \varDelta \rBrack$.

3.2 Reasoning from possible sources of information

Theorem 1 and Corollary 1 depend on the assumption of $\llbracket \varDelta \rrbracket = \lBrack \varDelta \rBrack$. In this section, we cancel the assumption to fully generalise our discussions in Section 3.1. The following theorem relates the probability of a formula to the probability of its possible models.

Theorem 2. *Let $p(\Gamma, M, D; \mu = 1)$ be a generative reasoning model, and $\alpha \in \Gamma$ and $\Delta \subseteq \Gamma$.*

$$p(\alpha|\Delta) = \begin{cases} \dfrac{\sum_{m \in [\![\![\Delta]\!]\!] \cap [\![\![\alpha]\!]\!]} p(m)}{\sum_{m \in [\![\![\Delta]\!]\!]} p(m)} & \textit{if } [\![\![\Delta]\!]\!] \neq \emptyset \\ \textit{undefined} & \textit{otherwise} \end{cases}$$

Proof. $p(m) = 0$, for all $m \in [\![\Delta]\!] \setminus [\![\![\Delta]\!]\!]$ and $m \in [\![\alpha]\!] \setminus [\![\![\alpha]\!]\!]$. From Theorem 1,

$$\frac{\sum_{m \in [\![\Delta]\!] \cap [\![\alpha]\!]} p(m)}{\sum_{m \in [\![\Delta]\!]} p(m)} = \frac{\sum_{m \in [\![\![\Delta]\!]\!] \cap [\![\![\alpha]\!]\!]} p(m)}{\sum_{m \in [\![\![\Delta]\!]\!]} p(m)}.$$

The condition of $[\![\Delta]\!] \neq \emptyset$ should be replaced by $[\![\![\Delta]\!]\!] \neq \emptyset$, since there is a possibility of $[\![\Delta]\!] \neq \emptyset$ and $[\![\![\Delta]\!]\!] = \emptyset$. Given the condition of $[\![\Delta]\!] \neq \emptyset$, this causes a probability undefined due to a division by zero. □

In Section 3.1, we used the classical consequence relation in Corollary 1 for a logical characterisation of Theorem 1. In this section, we define an alternative consequence relation for a logical characterisation of Theorem 2.

Definition 2 (Empirical consequence). *Let $\Delta \subseteq \Gamma$ and $\alpha \in \Gamma$. α is an empirical consequence of Δ, denoted by $\Delta \mathrel{|}\!\equiv \alpha$, if $[\![\![\Delta]\!]\!] \subseteq [\![\![\alpha]\!]\!]$.*

If $\Delta \models \alpha$ then $\Delta \mathrel{|}\!\equiv \alpha$, but not vice versa, for all $\Delta \subseteq \Gamma$ and $\alpha \in \Gamma$. The following Corollary shows the relationship between the generative reasoning model and the empirical consequence relation $\mathrel{|}\!\equiv$.

Corollary 2. *Let $p(\Gamma, M, D; \mu = 1)$ be a generative reasoning model, and $\alpha \in \Gamma$ and $\Delta \subseteq \Gamma$ such that $[\![\![\Delta]\!]\!] \neq \emptyset$. $p(\alpha|\Delta) = 1$ iff $\Delta \mathrel{|}\!\equiv \alpha$.*

Proof. $\Delta \mathrel{|}\!\equiv \alpha$ iff $[\![\![\Delta]\!]\!] \subseteq [\![\![\alpha]\!]\!]$. $p(m) \neq 0$, for all $m \in [\![\![\Delta]\!]\!]$. Thus, from Theorem 2, $p(\alpha|\Delta) = 1$ iff $[\![\![\Delta]\!]\!] \subseteq [\![\![\alpha]\!]\!]$. □

Note that Theorem 2 and Corollary 2 no longer depend on the assumption of $[\![\Delta]\!] = [\![\![\Delta]\!]\!]$ required in Theorem 1 and Corollary 1. Figure 3 illustrates the assumption of $[\![\![\Delta]\!]\!] \neq \emptyset$ for reasoning of $\alpha \in L$ from $\Delta \subseteq L$ using the generative reasoning model $p(L, M, D; \mu = 1)$. It shows that both α and Δ are consistent, i.e., $[\![\alpha]\!] \neq \emptyset$ and $[\![\Delta]\!] \neq \emptyset$, since there is at least one model for both α and Δ satisfying the formulas. It also shows that Δ and α are possible, i.e., $[\![\![\Delta]\!]\!] \neq \emptyset$ and $[\![\![\alpha]\!]\!] \neq \emptyset$, since there is at least one model for both Δ and α supported by data.

3.3 Reasoning from inconsistent sources of information

Theorem 1 and Corollary 1 assume $[\![\Delta]\!] \neq \emptyset$ in practice. The conditional probability $p(\alpha|\Delta)$ is undefined otherwise. This section aims to cancel the assumption to fully generalise our discussions in Section 3.1 so that we can reason also

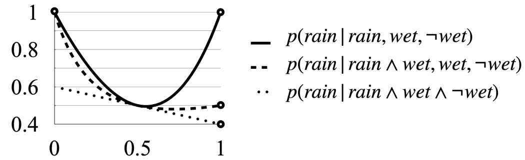

Fig. 4. Three examples of reasoning from inconsistency. The probability versus μ.

from an inconsistent source of information. To this end, we look at the generative reasoning model $p(\Gamma, M, D; \mu \to 1)$, rather than $p(\Gamma, M, D; \mu = 1)$, where $\mu \to 1$ represents μ approaching one, i.e., $\lim_{\mu \to 1}$. The following example shows the intuition of how the limit works and how it naturally generalises reasoning regardless of the consistency of its premises.

Example 4 (Continued). Consider the three conditional probabilities given different inconsistent premises shown in Figure 4. Suppose that the probability distribution is given by $p(M) = (m_1, m_2, m_3, m_4) = (0.4, 0.2, 0.1, 0.3)$. The conditional probability shown on the top right is expanded as follows.

$$p(rain|rain, wet, \neg wet) = \frac{\sum_m p(rain|m)^2 p(wet|m) p(\neg wet|m) p(m)}{\sum_m p(rain|m) p(wet|m) p(\neg wet|m) p(m)}$$
$$= \frac{(p(m_1) + p(m_2))\mu(1-\mu)^3 + (p(m_3) + p(m_4))\mu^3(1-\mu)}{(p(m_1) + p(m_2))\mu(1-\mu)^2 + (p(m_3) + p(m_4))\mu^2(1-\mu)}$$
$$= \frac{0.6\mu(1-\mu)^3 + 0.4\mu^3(1-\mu)}{0.6\mu(1-\mu)^2 + 0.4\mu^2(1-\mu)}$$

The graph with the solid line in Figure 4 shows $p(rain|rain, wet, \neg wet)$ given different μ values. The graph also includes the other two conditional probabilities calculated in the same manner. Each of the open circles represents an undefined value. This means that no substitution gives a probability, even though the curve approaches a certain probability. The certain probability can only be obtained by the use of limit. Indeed, given $\mu \to 1$, the three conditional probabilities turn out to be 1, 0.5 and 0.4, respectively.

Everything is entailed from an inconsistent set of formulas in formal logic. The use of limit is a reasonable alternative, since it allows us to consider what if there is a very tiny chance of the formula being true. The mathematical correctness of the solution can be shown using maximal consistent sets and maximal possible sets.

Definition 3 (Maximal consistent sets). *Let $S, \Delta \subseteq L$. $S \subseteq \Delta$ is a maximal consistent subset of Δ if $[\![S]\!] \neq \emptyset$ and $[\![S \cup \{\alpha\}]\!] = \emptyset$, for all $\alpha \in \Delta \setminus S$.*

We refer to a maximal consistent subset as a cardinality-maximal consistent subset when the set has the maximum cardinality. We use symbol $MCS(\Delta)$

to denote the set of the cardinality-maximal consistent subsets of $\Delta \subseteq L$. We use symbol $(\!(\Delta)\!)$ to denote the set of the models of the cardinality-maximal consistent subsets of Δ, i.e., $(\!(\Delta)\!) = \bigcup_{S \in MCS(\Delta)} [\![S]\!]$.

Example 5 (Continued). Consider the model distribution shown in Example 4 and $\Delta = \{\, rain,\, wet,\, rain \rightarrow wet,\, \neg wet\}$. It gives the following three maximal consistent subsets of Δ: $S_1 = \{rain, wet, rain \rightarrow wet\}$, $S_2 = \{rain, \neg wet\}$ and $S_3 = \{rain \rightarrow wet, \neg wet\}$. Only S_1 is the cardinality-maximal consistent subset of Δ, i.e., $MCS(\Delta) = \{S_1\}$. Therefore, $(\!(\Delta)\!) = \bigcup_{S \in MCS(\Delta)} [\![S]\!] = [\![S_1]\!] = \{m_4\}$.

Definition 4 (Maximal possible sets). *Let $S, \Delta \subseteq L$. $S \subseteq \Delta$ is a maximal possible subset of Δ if $[\![\![S]\!]\!] \neq \emptyset$ and $[\![\![S \cup \{\alpha\}]\!]\!] = \emptyset$, for all $\alpha \in \Delta \setminus S$.*

Similarly, we refer to a maximal possible subset as a cardinality-maximal possible subset when the set has the maximum cardinality. We use symbol $MPS(\Delta)$ to denote the set of the cardinality-maximal possible subsets of $\Delta \subseteq L$. We use symbol $(\!(\!(\Delta)\!)\!)$ to denote the set of possible models of the cardinality-maximal possible subsets of Δ, i.e., $(\!(\!(\Delta)\!)\!) = \bigcup_{S \in MPS(\Delta)} [\![\![S]\!]\!]$.

Example 6 (Continued). Suppose that the probability distribution is given by $p(M) = (m_1, m_2, m_3, m_4) = (0.9, 0.1, 0, 0)$. Consider $\Delta = \{\, rain,\, wet,\, rain \rightarrow wet, \neg wet\}$. It gives the following two maximal possible subsets of Δ: $S_1 = \{wet, rain \rightarrow wet\}$ and $S_2 = \{rain \rightarrow wet, \neg wet\}$. Both S_1 and S_2 are the cardinality-maximal possible subsets of Δ, i.e., $MPS(\Delta) = \{S_1, S_2\}$. Only m_2 is the possible model of S_1 and m_1 is the possible model of S_2. Namely, $[\![\![S_1]\!]\!] = \{m_2\}$ and $[\![\![S_2]\!]\!] = \{m_1\}$. Therefore, $(\!(\!(\Delta)\!)\!) = \bigcup_{S \in MPS(\Delta)} [\![\![S]\!]\!] = \{m_1, m_2\}$.

Obviously, $(\!(\Delta)\!) = [\![\Delta]\!]$ if there is a model of Δ, i.e., $[\![\Delta]\!] \neq \emptyset$. Similarly, $(\!(\!(\Delta)\!)\!) = [\![\![\Delta]\!]\!]$ if there is a possible model of Δ, i.e., $[\![\![\Delta]\!]\!] \neq \emptyset$. Note that if Δ is an empty set or Δ only includes tautologies then every model satisfies all the formulas in the possibly empty Δ. $[\![\Delta]\!]$ is thus the set of all models, and therefore $[\![\Delta]\!] \neq \emptyset$. Moreover, $[\![\![\Delta]\!]\!] \neq \emptyset$, since $p(M)$ is a probability distribution, and thus, there is at least one model m such that $p(m) \neq 0$.

Example 7. Let $\Delta_1 = \{\alpha, \neg\alpha\}$ and $\Delta_2 = \{\alpha \wedge \neg\alpha\}$, for $\alpha \in L$. $(\!(\Delta_1)\!) = \bigcup_{S \in MCS(\Delta_1)} [\![S]\!] = \bigcup_{S \in \{\{\alpha\},\{\neg\alpha\}\}} [\![S]\!] = [\![\alpha]\!] \cup [\![\neg\alpha]\!] = \mathcal{M}$. $(\!(\Delta_2)\!) = \bigcup_{S \in MCS(\Delta_2)} [\![S]\!] = \bigcup_{S \in \{\emptyset\}} [\![S]\!] = [\![\emptyset]\!] = \mathcal{M}$. Here, $\mathcal{M}$ denotes all the models associated with L.

Example 8 (Continued). Suppose that the probability distribution is given by $p(M) = (m_1, m_2, m_3, m_4) = (0.5, 0.2, 0, 0.3)$. Let $\Delta_1 = \{rain, \neg rain\}$ and $\Delta_2 = \{rain \wedge \neg rain\}$. $(\!(\!(\Delta_1)\!)\!) = \bigcup_{S \in MPS(\Delta_1)} [\![\![S]\!]\!] = [\![\![rain]\!]\!] \cup [\![\![\neg rain]\!]\!] = \{m_1, m_2, m_4\}$. $(\!(\!(\Delta_2)\!)\!) = \bigcup_{S \in MPS(\Delta_2)} [\![\![S]\!]\!] = [\![\![\emptyset]\!]\!] = \{m_1, m_2, m_4\}$.

The generative reasoning model $p(\Gamma, M, D; \mu \rightarrow 1)$ has the following property.

Theorem 3. *Let $p(\Gamma, M, D; \mu \rightarrow 1)$ be a generative reasoning model, and $\alpha \in \Gamma$ and $\Delta \subseteq \Gamma$ such that $(\!(\Delta)\!) = (\!(\!(\Delta)\!)\!)$.*

$$p(\alpha|\Delta) = \frac{\sum_{m \in (\!(\Delta)\!) \cap [\![\alpha]\!]} p(m)}{\sum_{m \in (\!(\Delta)\!)} p(m)}$$

Proof. We use symbol $|\Delta|$ to denote the number of formulas in Δ and symbol $|\Delta|_m$ to denote the number of formulas in Δ that are true in m, i.e., $|\Delta|_m = \sum_{\beta\in\Delta}[\![\beta]\!]_m$. Dividing models into $(\!(\Delta)\!)$ and the others, we have

$$\begin{aligned} p(\alpha|\Delta) &= \lim_{\mu\to 1}\frac{\sum_m p(\alpha|m)p(m)p(\Delta|m)}{\sum_m p(m)p(\Delta|m)} \\ &= \lim_{\mu\to 1}\frac{\sum_{\hat{m}\in(\!(\Delta)\!)} p(\alpha|\hat{m})p(\hat{m})p(\Delta|\hat{m}) + \sum_{m\notin(\!(\Delta)\!)} p(\alpha|m)p(m)p(\Delta|m)}{\sum_{\hat{m}\in(\!(\Delta)\!)} p(\hat{m})p(\Delta|\hat{m}) + \sum_{m\notin(\!(\Delta)\!)} p(m)p(\Delta|m)}. \end{aligned}$$

Now, $p(\Delta|m)$ can be developed as follows, for all m.

$$\begin{aligned} p(\Delta|m) &= \prod_{\beta\in\Delta} p(\beta|m) = \prod_{\beta\in\Delta}\mu^{[\![\beta]\!]_m}(1-\mu)^{1-[\![\beta]\!]_m} \\ &= \mu^{\sum_{\beta\in\Delta}[\![\beta]\!]_m}(1-\mu)^{\sum_{\beta\in\Delta}(1-[\![\beta]\!]_m)} = \mu^{|\Delta|_m}(1-\mu)^{|\Delta|-|\Delta|_m} \end{aligned}$$

Therefore, $p(\alpha|\Delta) = \lim_{\mu\to 1}\frac{W+X}{Y+Z}$ where

$$\begin{aligned} W &= \textstyle\sum_{\hat{m}\in(\!(\Delta)\!)} p(\alpha|\hat{m})p(\hat{m})\mu^{|\Delta|_{\hat{m}}}(1-\mu)^{|\Delta|-|\Delta|_{\hat{m}}} \\ X &= \textstyle\sum_{m\notin(\!(\Delta)\!)} p(\alpha|m)p(m)\mu^{|\Delta|_m}(1-\mu)^{|\Delta|-|\Delta|_m} \\ Y &= \textstyle\sum_{\hat{m}\in(\!(\Delta)\!)} p(\hat{m})\mu^{|\Delta|_{\hat{m}}}(1-\mu)^{|\Delta|-|\Delta|_{\hat{m}}} \\ Z &= \textstyle\sum_{m\notin(\!(\Delta)\!)} p(m)\mu^{|\Delta|_m}(1-\mu)^{|\Delta|-|\Delta|_m}. \end{aligned}$$

$(\!(\Delta)\!) = \bigcup_{S\in MCS(\Delta)}[\![S]\!] \neq \emptyset$, for all $\Delta \subseteq L$. Since $\hat{m} \in (\!(\Delta)\!)$ is a model of a cardinality-maximal consistent subset of Δ, $|\Delta|_{\hat{m}}$ has the same value, for all $\hat{m} \in (\!(\Delta)\!)$. Therefore, the fraction can be simplified by dividing the denominator and numerator by $(1-\mu)^{|\Delta|-|\Delta|_{\hat{m}}}$. We thus have $p(\alpha|\Delta) = \lim_{\mu\to 1}\frac{W'+X'}{Y'+Z'}$ where

$$\begin{aligned} W' &= \textstyle\sum_{\hat{m}\in(\!(\Delta)\!)} p(\alpha|\hat{m})p(\hat{m})\mu^{|\Delta|_{\hat{m}}} \\ X' &= \textstyle\sum_{m\notin(\!(\Delta)\!)} p(\alpha|m)p(m)\mu^{|\Delta|_m}(1-\mu)^{|\Delta|_{\hat{m}}-|\Delta|_m} \\ Y' &= \textstyle\sum_{\hat{m}\in(\!(\Delta)\!)} p(\hat{m})\mu^{|\Delta|_{\hat{m}}} \\ Z' &= \textstyle\sum_{m\notin(\!(\Delta)\!)} p(m)\mu^{|\Delta|_m}(1-\mu)^{|\Delta|_{\hat{m}}-|\Delta|_m}. \end{aligned}$$

Applying the limit operation, we can cancel out X' and Z'.

$$p(\alpha|\Delta) = \frac{\sum_{\hat{m}\in(\!(\Delta)\!)} p(\alpha|\hat{m})p(\hat{m})}{\sum_{\hat{m}\in(\!(\Delta)\!)} p(\hat{m})} = \frac{\sum_{\hat{m}\in(\!(\Delta)\!)} 1^{[\![\alpha]\!]_{\hat{m}}}0^{1-[\![\alpha]\!]_{\hat{m}}}p(\hat{m})}{\sum_{\hat{m}\in(\!(\Delta)\!)} p(\hat{m})}.$$

We have the theorem, since $1^{[\![\alpha]\!]_{\hat{m}}}0^{1-[\![\alpha]\!]_{\hat{m}}} = 1^1 0^0 = 1$ if $\hat{m} \in [\![\alpha]\!]$ and $1^{[\![\alpha]\!]_{\hat{m}}}0^{1-[\![\alpha]\!]_{\hat{m}}} = 1^0 0^1 = 0$ if $\hat{m} \notin [\![\alpha]\!]$. □

The following Corollary shows the relationship between the generative reasoning model $p(\Gamma, M, D; \mu \to 1)$ and the classical consequence relation with maximal consistent sets.

Corollary 3. *Let $p(\Gamma, M, D; \mu \to 1)$ be a generative reasoning model, and $\alpha \in \Gamma$ and $\Delta \subseteq \Gamma$ such that $(\!(\Delta)\!) = (\!(\!(\Delta)\!)\!)$. $p(\alpha|\Delta) = 1$ iff $S \models \alpha$, for all cardinality-maximal consistent subsets S of Δ.*

Proof. By the assumption of $(\!(\Delta)\!) = (\!(\!(\Delta)\!)\!)$, $p(m)$ is non zero, for all $m \in (\!(\Delta)\!)$. From Theorem 3, thus $p(\alpha|\Delta) = 1$ iff $[\![\alpha]\!] \supseteq (\!(\Delta)\!)$. Since $(\!(\Delta)\!) = \bigcup_{S \in MCS(\Delta)} [\![S]\!]$, $p(\alpha|\Delta) = 1$ iff $[\![\alpha]\!] \supseteq \bigcup_{S \in MCS(\Delta)} [\![S]\!]$. Namely, $p(\alpha|\Delta) = 1$ iff $[\![\alpha]\!] \supseteq [\![S]\!]$, for all cardinality-maximal consistent subsets S of Δ. □

Note that Theorem 3 and Corollary 3 no longer depend on the assumption of $[\![\Delta]\!] \neq \emptyset$ required in Section 3.1 for Theorem 1 and Corollary 1. Figure 3 illustrates the assumption of $(\!(\Delta)\!) = (\!(\!(\Delta)\!)\!)$ for reasoning of $\alpha \in \Gamma$ from inconsistent $\Delta \subseteq \Gamma$ using the generative reasoning model $p(\Gamma, M, D; \mu \to 1)$. It shows that Δ has no model satisfying all its formulas. It also shows that every model satisfying all the formulas in a cardinality-maximal consistent subset of Δ is possible.

3.4 Reasoning from impossible sources of information

Theorem 3 and Corollary 3 depend on the assumption of $(\!(\Delta)\!) = (\!(\!(\Delta)\!)\!)$. In this section, we cancel the assumptions to fully generalise our discussions in Section 3.3. The generative reasoning model $p(\Gamma, M, D; \mu \to 1)$ has the following property.

Theorem 4. *Let $p(\Gamma, M, D; \mu \to 1)$ be a generative reasoning model, and $\alpha \in \Gamma$ and $\Delta \subseteq \Gamma$.*

$$p(\alpha|\Delta) = \frac{\sum_{m \in (\!(\!(\Delta)\!)\!) \cap [\![\alpha]\!]} p(m)}{\sum_{m \in (\!(\!(\Delta)\!)\!)} p(m)}$$

Proof. The proof is almost same as Theorem 3. Only difference is to divide models into $(\!(\!(\Delta)\!)\!)$ rather than $(\!(\Delta)\!)$. □

The following Corollary shows the relationship between the generative reasoning model $p(\Gamma, M, D; \mu \to 1)$ and the empirical consequence relation with maximal possible sets.

Corollary 4. *Let $p(\Gamma, M, D; \mu \to 1)$ be a generative reasoning model, and $\alpha \in \Gamma$ and $\Delta \subseteq \Gamma$. $p(\alpha|\Delta) = 1$ iff $S \mathrel{|\!\!\!\equiv} \alpha$, for all cardinality-maximal possible subsets S of Δ.*

Proof. From Theorem 4, $p(\alpha|\Delta) = 1$ iff $(\!(\!(\Delta)\!)\!) \subseteq [\![\alpha]\!]$. Since $(\!(\!(\Delta)\!)\!) = \bigcup_{S \in MPS(\Delta)} [\![S]\!]$, $p(\alpha|\Delta) = 1$ iff $\bigcup_{S \in MPS(\Delta)} [\![S]\!] \subseteq [\![\alpha]\!]$. Therefore, $p(\alpha|\Delta) = 1$ iff $[\![S]\!] \subseteq [\![\alpha]\!]$, for all $S \in MPS(\Delta)$. □

Note that Theorem 4 and Corollary 4 no longer depend on the assumption of $(\!(\Delta)\!) = (\!(\!(\Delta)\!)\!)$ required in Section 3.3 for Theorem 3 and Corollary 3. Figure 3 illustrates reasoning of $\alpha \in \Gamma$ from impossible $\Delta \subseteq \Gamma$ using the generative reasoning model $p(\Gamma, M, D; \mu \to 1)$. It illustrates the most general situation without the assumptions discussed in the previous sections.

Table 2. The summary of the logical grounds. Due to the assumption strictness, uncertain parapossible reasoning is the most generalised type of reasoning whereas certain consistent reasoning is the most specialised type.

Reasoning type	Logical ground	Grounding assumptions
Consistent	$p(\alpha \mid \Delta) = 1$ iff $\Delta \models \alpha$	$[\![\Delta]\!] \neq \emptyset$, $[\![\Delta]\!] = [\![\Delta]\!]$
Possible	$p(\alpha \mid \Delta) = 1$ iff $\Delta \equiv \alpha$	$[\![\Delta]\!] \neq \emptyset$
Paraconsistent	$p(\alpha \mid \Delta) = 1$ iff $\forall S \in MCS(\Delta).S \models \alpha$	$(\!(\Delta)\!) = (\!(\!(\Delta)\!)\!)$
Parapossible	$p(\alpha \mid \Delta) = 1$ iff $\forall S \in MPS(\Delta).S \equiv \alpha$	No assumption
Uncertain	$p(\alpha \mid \Delta) \in [0, 1]$ generalises the above	Same as above

4 Conclusions

Symbolic knowledge is an abstraction of data. This simple idea caused a probabilistic model of how data cause symbolic knowledge in terms of its satisfiability in formal logic. Table 2 summarises the logical grounds and their assumptions of all the types of logical reasoning studied in this paper. They are all based on the simple principle that intrinsically abstract symbolic knowledge is caused from intrinsically concrete data. The principle opposes the prevailing idea in formal logic that knowledge is caused from preceding knowledge by rules of inference. The principle focuses on what machine learning and statistics usually do not deal with. The central question studied in this paper is how symbolic knowledge is caused from data. This differs from machine learning and statistics primarily asking how data are caused from the parameters of probability distributions. Finally, the principle can be seen as a solution to inference grounding in the sense that reasoning from symbols to symbols always occurs via data as its references.

References

1. Adams, R.A., Huys, Q.J.M., Roiser, J.P.: Computational psychiatry: towards a mathematically informed understanding of mental illness. Journal of Neurology, Neurosurgery & Psychiatry **87(1)**, 53–63 (2016)
2. Carnielli, W., Coniglio, M.E., Marcos, J.: Logics of Formal Inconsistency, vol. 14, pp. 1–93. Springer Dordrecht, Dordrecht, Netherlands, handbook of philosophical logic, 2nd edn. (2007)
3. Caucheteux, C., Gramfort, A., King, J.R.: Evidence of a predictive coding hierarchy in the human brain listening to speech. Nature Human Behaviour **7**, 430–441 (2023)
4. Friedman, N., Getoor, L., Koller, D., Pfeffer, A.: Learning probabilistic relational models. In: Proc. 16th Int. Joint Conf. on Artif. Intell. pp. 1297–1304 (1996)
5. Friston, K.: The free-energy principle: a unified brain theory? Nature Reviews Neuroscience **11**, 127–138 (2010)
6. Getoor, L., Taskar, B.: Introduction to Statistical Relational Learning. MIT Press, Cambridge, MA (2007)
7. Gregory, R.L.: Knowledge in perception and illusion. Philos Trans R Soc Lond B Biol Sci **352**(1358), 1121–1127 (1997)
8. von Helmholtz, H.: Helmholtz's treatise on physiological optics. Optical Society of America **3** (1925)

9. Hohwy, J.: The Predictive Mind. Oxford University Press, University of Oxford (2014)
10. Itti, L., Baldi, P.: Bayesian surprise attracts human attention. Vision Research **49(10)**, 1295–1306 (2009)
11. Kido, H.: Inference of abstraction for human-like probabilistic reasoning. In: The 10th Int Conf on machine Learning, Optimization & Data science - LOD and Symposium on Artificial Intelligence & Neuroscience (ACAIN 2024) (2024)
12. Knill, D.C., Pouget, A.: The Bayesian brain: the role of uncertainty in neural coding and computation. Trends in Neurosciences **27**, 712–719 (2004)
13. Knill, D.C., Richards, W.: Perception as Bayesian Inference. Cambridge University Press, Cambridge University (1996)
14. Lee, T.S., Mumford, D.: Hierarchical Bayesian inference in the visual cortex. Journal of Optical Society of America **20**, 1434–1448 (2003)
15. Muggleton, S.: Inverse entailment and progol. New Generation Computing **13**, 245–286 (1995)
16. Muggleton, S., Buntine, W.: Machine invention of first-order predicates by inverting resolution. In: Proc. 5th International Conference on Machine Learning. pp. 339–352 (1988)
17. Nienhuys-Cheng, S.H., Wolf, R.D.: Foundation of Inductive Logic Programming. Springer Berlin, Heidelberg, Heidelberg (1997)
18. Nilsson, N.J.: Probabilistic logic. Artificial Intelligence **28**, 71–87 (1986)
19. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann; 1st edition, Burlington, Massachusetts (1988)
20. Pellicano, E., Burr, D.: When the world becomes 'too real': a Bayesian explanation of autistic perception. Trends in Cognitive Sciences **16(10)**, 504–510 (2012)
21. Priest, G.: Paraconsistent Logic, vol. 6, pp. 287–393. Springer Dordrecht, Dordrecht, Netherlands, handbook of philosophical logic, 2nd edn. (2002)
22. Rao, R.P.N.: Bayesian inference and attentional modulation in the visual cortex. Neuroreport **16(16)**, 1843–1848 (2005)
23. Rao, R.P.N., Ballard, D.H.: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience **2**, 79–87 (1999)
24. Richardson, M., Domingos, P.: Markov logic networks. Machine Learning **62**, 107–136 (2006)
25. Rödder, W.: Conditional logic and the principle of entropy. Artificial Intelligence **117**, 83–106 (2000)
26. Russell, S., Norvig, P.: Artificial Intelligence : A Modern Approach, Fourth Edition. Pearson Education, Inc., London, England (2020)
27. Sato, T.: A statistical learning method for logic programs with distribution semantics. In: Proc. 12th int. conf. on logic programming. pp. 715–729 (1995)
28. Smith, G.E., Chouinard, P.A., Byosiere, S.E.: If i fits i sits: A citizen science investigation into illusory contour susceptibility in domestic cats (*Felis silvestris catus*). Applied Animal Behaviour Science **240**, 105338 (2021)
29. Tenenbaum, J.B., Griffiths, T.L., Kemp, C.: Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences **10(7)**, 309–318 (2006)