Inference of Abstraction for Grounded Predicate Logic

Hiroyuki Kido KIDOH@CARDIFF.AC.UK
Cardiff University, Park Place, CF10 3AT, Cardiff, UK

Abstract

An important open question in AI is what simple and natural principle enables a machine to ground logical reasoning in data for meaningful abstraction. This paper explores a conceptually new approach to combining probabilistic reasoning and predicative symbolic reasoning over data. We return to the era of reasoning with a full joint distribution before the advent of Bayesian networks. We then discuss that a full joint distribution over models of exponential size in propositional logic and of infinite size in predicate logic should be simply derived from a full joint distribution over data of linear size. We show that the same process is not only enough to generalise the logical consequence relation of predicate logic but also to provide a new perspective to rethink well-known limitations such as the undecidability of predicate logic, the symbol grounding problem and the principle of explosion. The reproducibility of this theoretical work is fully demonstrated by the included proofs.

1. Introduction

The current artificial intelligence (AI) systems such as large language models (LLMs) (OpenAI, 2025; DeepSeek-AI, 2025) demonstrate a surprising linguistic ability in both what they know and how they articulate it. However, the common view is that they are still not as capable as ordinary people in several areas such as logical reasoning and abstract reasoning. For logical reasoning, it is unclear how the statistical patterns an AI algorithm extracts from finite training data can capture the infinite set of rules of valid inference studied in formal logic. Moreover, it is unlikely that statistical AI modelling how people tend to think can replace formal logic modelling how people ought to think. For abstract reasoning, it is still unclear how a machine should explore and discover abstract concepts and principles from the real world in its own way. Consider the following problems requiring both abstract reasoning and logical reasoning skills.

Example 1. *Carol remembers the following three scenes.*

- *Alice and Bob did not blame each other.*
- *One day Alice blamed Bob, and she blamed herself afterwards.*
- *Alice and Bob blamed each other on another day.*

One day Carol wanted to blame Bob. She hesitated because she learnt that someone will blame those who blame anyone, which can be expressed in a predicate language as follows.

$$\forall x(\exists y(Blames(x,y)) \rightarrow \exists z(Blames(z,x)))$$

This work is licensed under a Creative Commons Attribution International 4.0 License.

Example 2. *Consider the following three data. What number fits in the blank?*

The correct number could be 18 as the following predicate knowledge can be extracted.

$$top \times left + right = bottom$$

Interestingly, the current AI systems such as ChatGPT (OpenAI, 2025) and DeepSeek (DeepSeek-AI, 2025) often answer them incorrectly due to a lack of abstract or logical reasoning skills. Abstraction and Reasoning Corpus (ARC) (Chollet, 2019) is a benchmark test designed to evaluate a machine's ability to extract visual patterns from images. However, the more analytical patterns shown above are beyond its scope.

In this paper, we ask how reasoning in predicate logic should be grounded in data. The underlying idea discussed in this paper is abstraction. Roughly speaking, it is about an inferential process of deriving intrinsically abstract symbols from intrinsically concrete data through selective ignorance. It is not about generalisation where typical inferential procedures, e.g., deductive reasoning, are used backward for general rules from specific examples or facts. This type of reasoning is intensively studied as inverse resolution (Muggleton & Buntine, 1988; Nienhuys-Cheng & Wolf, 1997), inverse deduction (Russell & Norvig, 2020) and inverse entailment (Muggleton, 1995) mainly in inductive logic programming (ILP) (Nienhuys-Cheng & Wolf, 1997). It is also not about parametric learning where intrinsically concrete data are assumed to be generated from probability distributions characterised by their parameters. This idea is prevalent in various applications of machine learning and statistics, e.g., (Bishop, 2006; Tenenbaum et al., 2006; Dasgupta et al., 2020; Lake et al., 2015, 2017). Abstraction is rather relevant to top-down (memory/experience-driven) and bottom-up (sensory-driven) information processing used by neuroscientists and AI researchers as a metaphor for the cognitive process of biological brains, e.g., (Pearl & Russell, 2003; Harnad, 1990; Lee & Mumford, 2003; Hawkins, 2021; Gregory, 1997; Rao & Ballard, 1999; Friston, 2010).

In this paper, we extend the inference of propositional abstraction (Kido, 2025a,b) to the inference of predicative abstraction towards enhanced human-like machine intelligence. The key idea is to use the property of predicate logic and expand the joint probability distribution over data, models of predicate logic and predicate formula, denoted by D, M and α, respectively, as follows.

$$p(D, M, \alpha) = p(\alpha|M, D)p(M|D)p(D) = p(\alpha|M)p(M|D)p(D).$$

Here, the second expression can be derived by applying a valid rule of probability theory. The third expression is an application of the property of formal logic that the truth values of formulas are determined given a model. Probabilistically speaking, the truth values of formulas are conditionally independent of data given a model. The third expression realises the idea that intrinsically abstract knowledge is derived from intrinsically concrete data through abstraction, i.e., selective ignorance.

We show that the inference of predicative abstraction serves as a solution to simple yet important problems such as Examples 1 and 2. The research is not as straightforward as we think because the semantics of predicate logic needs a reformulation in accordance with abstraction. Our theory

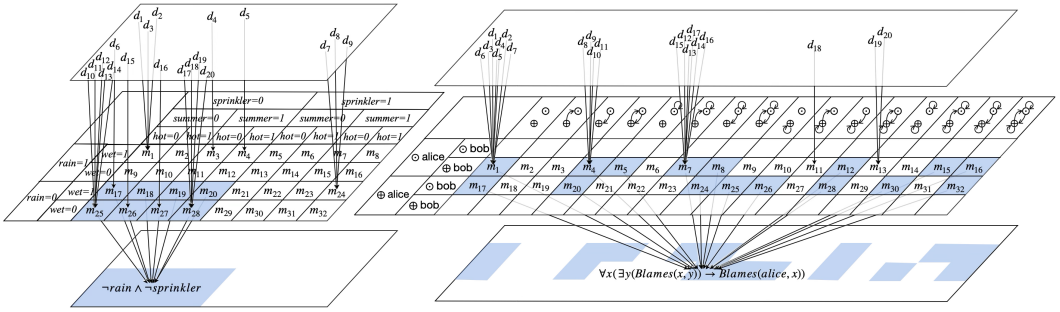


Figure 1. The hierarchy shown on the left is our illustration of the existing work on the inference of propositional abstraction (Kido, 2025a,b). The one shown on the right is an illustration of our work on the inference of predicative abstraction. The top layers are both distributions of data. The middle layer on the left is a distribution of models in propositional logic, i.e., valuations. The one on the right is a distribution of models in predicate logic, i.e., pairs of domains of discourse and valuation functions. The bottom layer on the left is a distribution of the truth values of the propositional formula, whereas the one on the right is the same type of distribution for the predicate formula.

assumes only closed formulas, i.e., predicate formulas without free variables, to balance the expressiveness and simplicity of the theory. The contributions of this paper are summarised as follows.

- We introduce a simple theory of inference that opens up the possibility of combining probability theory and predicate logic in a data-driven manner. Predicate reasoning in our theory always proceeds between data and predicate formulas. This suggests a shift in the traditional view that predicate reasoning proceeds between predicate formulas via rules of inference (see Section 2).

- The theory allows us to see the traditional model-based predicate reasoning as a special case of data-based predicate reasoning studied in this paper. The data-based perspective provides a new opportunity to rethink some existing limitations such as the undecidability of predicate logic, the symbol grounding problem (Harnad, 1990; Russell & Norvig, 2020) and commonsense reasoning (Brewka, 1991; Davis & Marcus, 2015) (see Sections 3.1, 3.2 and 3.3).

- We demonstrate a solution to simple yet essential problems that are often difficult to solve by existing established approaches (see Section 3.4).

2. Proposals

2.1 Data support models

The inference of abstraction for propositional logic (Kido, 2025a,b) is insufficient to handle problems like Examples 1 and 2. We thus propose the inference of abstraction for predicate logic in this section. Let $\{d_1, d_2, ..., d_K\}$ be a multiset of K data and D be a random variable for data taking values from $\{d_1, d_2, ..., d_K\}$.

Definition 1. *Let $d_k \in \{d_1, d_2, ..., d_K\}$. The probability of d_k, denoted by $p(D = d_k)$, is defined as follows.*

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

Namely, $p(D)$ is a uniform distribution. Let $\mathcal{C}$, $\mathcal{V}$, $\mathcal{F}$ and $\mathcal{P}$ be the sets of constants, variables, function symbols and predicate symbols, respectively, and L be the predicate language built with these vocabularies.

Example 3. *Consider the following vocabularies of a predicate language.*

- *Constants:* $\mathcal{C} = \{alice, bob\}$
- *Variables:* $\mathcal{V} = \{x, y\}$
- *Function symbols:* $\mathcal{F} = \{mentor\}$
- *Predicate symbols:* $\mathcal{P} = \{Blames\}$

The following is a predicate formula meaning that Alice's mentor blames everyone who blames someone.

$$\forall x(\exists y(Blames(x, y)) \rightarrow Blames(mentor(alice), x))$$

In this paper, we assume that the predicate language includes only formulas without free variables, i.e., closed formulas.[1] We exclude open formulas for the following reasons. First, it is inappropriate to view an assignment in predicate logic as an abstraction, or selective ignorance, of data or observations. Its inclusion thus does not fit the underlying idea of the inference of abstraction. Second, a lot of cases such as Examples 1 and 2 do not need open formulas. Its inclusion thus makes our formalism unnecessarily complicated.

As usual, a model in predicate logic is a pair of a domain of discourse and valuation function. The domain of discourse, denoted by u, is a non-empty set of a finite or countably infinite number of entities. The valuation function, denoted by v, is a function that associates constants, function symbols and predicate symbols with u. We use the symbol $ar(x)$ to denote the arity of the function or predicate symbol x. Specifically,

- v maps each constant $c \in \mathcal{C}$ to an entity of u, i.e., $v(c) \in u$.[2]
- v maps each function symbol $f \in \mathcal{F}$ to an $ar(f)$-ary function from u to $u^{ar(f)}$, i.e., $v(f) : u^{ar(f)} \rightarrow u$.
- v maps each predicate symbol $P \in \mathcal{P}$ to a subset of $u^{ar(P)}$, i.e., $v(P) \subseteq u^{ar(P)}$.

1. The truth values of closed formulas depend only on a model, whereas the truth values of open formulas, i.e., formulas with free variables, depend additionally on an assignment, a function mapping each variable to an entity in the domain of discourse.
2. We assume that v is surjective with respect to $\mathcal{C}$, meaning that there is constant $c \in \mathcal{C}$ such that $v(c) = e$, for all entities $e \in u$. This assumption allows us to apply simple semantics of predicate logic.

From the viewpoint of the inference of abstraction, it is important to adopt the perspective that each model represents a different state of the world. For any function symbol $f \in \mathcal{F}$, we write $v(f)(v(t_1), ..., v(t_{ar(f)}))$ as $v(f(t_1, ..., t_{ar(f)}))$, where $t_1, t_2, ...t_{ar(f)}$ are the arguments of f called terms referring to constants, variables or functions.

Let $\{m_1, m_2, ..., m_N\}$ be the set of models of the predicate language L. This set is finite or countably infinite. Let M be a random variable for the models taking values from $\{m_1, m_2, ..., m_N\}$. We assume that each data point supports a single model. We thus assume a function $m : \{d_1, d_2, ..., d_K\} \rightarrow \{m_1, m_2, ..., m_N\}$ such that $m(d_k)$ denotes the model supported by data d_k.

Definition 2. *Let $d_k \in \{d_1, d_2, ..., d_K\}$ and $m_n \in \{m_1, m_2, ..., m_N\}$. The probability of m_n given d_k, denoted by $p(M = m_n|D = d_k)$, is defined as follows.*

$$p(M = m_n|D = d_k) = \begin{cases} 1 & \textit{if } m_n = m(d_k) \\ 0 & \textit{otherwise} \end{cases} \tag{2}$$

We use the symbols $u(m_n)$ and $v(m_n)$ to denote the domain of discourse and the valuation function of the model m_n, i.e., $m_n = \langle u(m_n), v(m_n)\rangle$. Thus, the model supported by data d_k can be written as $m(d_k) = \langle u(m(d_k)), v(m(d_k))\rangle$.

Example 4. *Consider the predicate language L built with the following vocabularies.*

- *Constants: $\mathcal{C} = \{alice, bob\}$*
- *Variables: $\mathcal{V} = \{x, y\}$*
- *Function symbols: $\mathcal{F} = \emptyset$*
- *Predicate symbols: $\mathcal{P} = \{Blames\}$*

The top layer of the hierarchy shown on the right in Figure 1 shows twenty data. Given $u = \{\odot, \oplus\}$, the middle layer shows all the thirty two models $\langle u, v\rangle$ of the language L. The depth of the middle layer shows how the valuation function v associates the constants $alice$ and bob with u. Its width shows how v associates the predicate symbol $Blames$ to u, where an arrow from x to y represents that x blames y. Each blank cell in the middle layer is not a model due to the assumption we made in Footnote 2. The arrow from the top to middle layers represent a function m. The twenty data d_k commonly say that there are two people $\odot$ and $\oplus$ named Alice and Bob, respectively , i.e., $u(m(d_k)) = \{\odot, \oplus\}$, $v(m(d_k))(alice) = \odot$ and $v(m(d_k))(bob) = \oplus$. Each data point, however, supports a different situation. Specifically, the functions m and v are given as follows, for all data d_k.

$$m(d_k) = \begin{cases} m_1 & \textit{if } k \in \{1\text{-}7\} \\ m_4 & \textit{if } k \in \{8\text{-}11\} \\ m_7 & \textit{if } k \in \{12\text{-}17\} \\ m_{11} & \textit{if } k \in \{18\} \\ m_{13} & \textit{if } k \in \{19, 20\} \end{cases} \quad v(m(d_k))(Blames) = \begin{cases} \emptyset & \textit{if } k \in \{1\text{-}7\} \\ \{(\odot, \oplus), (\oplus, \odot)\} & \textit{if } k \in \{8\text{-}11\} \\ \{(\odot, \odot), (\odot, \oplus)\} & \textit{if } k \in \{12\text{-}17\} \\ \{(\odot, \oplus), (\oplus, \oplus)\} & \textit{if } k \in \{18\} \\ \{(\odot, \odot), (\oplus, \oplus)\} & \textit{if } k \in \{19, 20\} \end{cases}$$

2.2 Models support formulas

We are interested in the probability of predicate formula $\alpha \in L$ being true or false. We thus assume that each formula is a random variable taking values from $\{0, 1\}$. For any truth values $v \in \{0, 1\}$, we use the symbol $[\![\alpha = v]\!]$ to denote the set of models where α has the truth value v. We often write $[\![\alpha = v]\!]_{m_n} = 1$ if $m_n \in [\![\alpha = v]\!]$ and $[\![\alpha = v]\!]_{m_n} = 0$ otherwise for the membership of the model. We call formulas with neither logical connectives, such as $\neg$, $\vee$, $\wedge$ and $\rightarrow$, nor quantifiers, such as $\forall$ and $\exists$, atomic formulas. Let $m_n = \langle u, v\rangle$ be a model. As usual, the truth value of an atomic formula without variables is defined as follows.

$$[\![P(t_1, ..., t_{ar(P)})]\!]_{m_n} = \begin{cases} 1 & \text{if } (v(t_1), ..., v(t_{ar(P)})) \in v(P) \\ 0 & \text{otherwise} \end{cases}$$

Let $\alpha, \beta \in L$ be formulas and $m_n = \langle u, v\rangle$ be a model. As usual, the truth values of compound formulas with logical connectives are defined as follows.

$$\begin{aligned}
&[\![\neg\alpha]\!]_{m_n} = 1 \Leftrightarrow [\![\alpha]\!]_{m_n} = 0 \\
&[\![\alpha \wedge \beta]\!]_{m_n} = 1 \Leftrightarrow [\![\alpha]\!]_{m_n} = 1 \text{ and } [\![\beta]\!]_{m_n} = 1 \Leftrightarrow \min\{[\![\alpha]\!]_{m_n}, [\![\beta]\!]_{m_n}\} \\
&[\![\alpha \vee \beta]\!]_{m_n} = 1 \Leftrightarrow [\![\alpha]\!]_{m_n} = 1 \text{ or } [\![\beta]\!]_{m_n} = 1 \Leftrightarrow \max\{[\![\alpha]\!]_{m_n}, [\![\beta]\!]_{m_n}\} \\
&[\![\alpha \rightarrow \beta]\!]_{m_n} = 1 \Leftrightarrow [\![\alpha]\!]_{m_n} = 0 \text{ or } [\![\beta]\!]_{m_n} = 1 \Leftrightarrow \max\{1 - [\![\alpha]\!]_{m_n}, [\![\beta]\!]_{m_n}\}
\end{aligned}$$

Let us use the symbol $\alpha[c/x]$ to denote the formula replacing all the free variables $x \in \mathcal{V}$ in the formula α by the constant $c \in \mathcal{C}$. Here, a variable x is free if there is no quantifier bounding x or x is outside the scope of such quantifiers. As usual, the truth values of compound formulas with quantifiers are defined as follows.[3]

$$\begin{aligned}
&[\![\forall x\ \alpha]\!]_{m_n} = 1 \Leftrightarrow [\![\alpha[c/x]]\!]_{m_n} = 1, \text{ for all } c \in \mathcal{C} \Leftrightarrow \min_{c \in \mathcal{C}}\{[\![\alpha[c/x]]\!]_{m_n}\} \\
&[\![\exists x\ \alpha]\!]_{m_n} = 1 \Leftrightarrow [\![\alpha[c/x]]\!]_{m_n} = 1, \text{ for some } c \in \mathcal{C} \Leftrightarrow \max_{c \in \mathcal{C}}\{[\![\alpha[c/x]]\!]_{m_n}\}
\end{aligned}$$

We say that a formula $\alpha \in L$ is true in a model m_n or m_n satisfies, or supports, α if $[\![\alpha]\!]_{m_n} = 1$. We also write $[\![\alpha]\!]_{m_n} = 1$ as $m_n \in [\![\alpha]\!]$.

Example 5 (Continued)**.** *Let us find the models where everyone blames someone, i.e.,* $\forall x \exists y\ Blames(x, y)$*, is true. Each atomic formula has the following truth value (see Figure 1).*

$$\begin{aligned}
&[\![Blames(alice, alice)]\!] = \{m_n | n \in \{5\text{-}8, 13\text{-}16, 25\text{-}32\}\} \\
&[\![Blames(alice, bob)]\!] = \{m_n | n \in \{3, 4, 7, 8, 11, 12, 15, 16, 18, 20, 22, 24, 26, 28, 30, 32\}\} \\
&[\![Blames(bob, alice)]\!] = \{m_n | n \in \{2, 4, 6, 8, 10, 12, 14, 16, 19, 20, 23, 24, 27, 28, 31, 32\}\} \\
&[\![Blames(bob, bob)]\!] = \{m_n | n \in \{9\text{-}16, 21\text{-}24, 29\text{-}32\}\}
\end{aligned}$$

3. This definition is based on the assumption we made in Footnote 2. With this assumption, we can define the truth solely in terms of constants, without referring to the domain of discourse.

The compound formula, $\forall x \exists y\ Blames(x, y)$, thus has the following truth value.

$$
\begin{aligned}
&[\![\forall x \exists y\ Blames(x, y)]\!]_{m_n} = \min_{c_1 \in \mathcal{C}} \left\{ [\![\exists y\ Blames(c_1, y)]\!]_{m_n} \right\} \\
&= \min_{c_1 \in \mathcal{C}} \left\{ \max_{c_2 \in \mathcal{C}} \left\{ [\![Blames(c_1, c_2)]\!]_{m_n} \right\} \right\} \\
&= \min \left\{ \begin{array}{l} \max \left\{ [\![Blames(alice, alice)]\!]_{m_n}, [\![Blames(alice, bob)]\!]_{m_n} \right\}, \\ \max \left\{ [\![Blames(bob, alice)]\!]_{m_n}, [\![Blames(bob, bob)]\!]_{m_n} \right\} \end{array} \right\} \\
&= \begin{cases} 1 & \textit{if } n \in \{3\text{-}8, 11\text{-}16, 18, 20, 22, 24\text{-}32\} \cap \{2, 4, 6, 8\text{-}16, 19\text{-}24, 27\text{-}32\}, \textit{ i.e.,} \\ & \textit{if } n \in \{4, 6, 8, 11\text{-}16, 20, 22, 24, 27\text{-}32\} \\ 0 & \textit{otherwise} \end{cases}
\end{aligned}
$$

The probability of the truth of a formula is defined using the semantics of predicate logic.

Definition 3. *Let $\mu \in [0.5, 1]$, $d_k \in \{d_1, d_2, ..., d_K\}$, $m_n \in \{m_1, m_2, ..., m_N\}$, $\alpha_1, \alpha_2, ..., \alpha_I \in L$ and $v_1, v_2, ..., v_I \in \{0, 1\}$. The probability of $\alpha_1 = v_1$ given $\alpha_2 = v_2, ..., \alpha_I = v_I$, $M = m_n$ and $D = d_k$, denoted by $p(\alpha_1 = v_1 | \alpha_2 = v_2, ..., \alpha_I = v_I, M = m_n, D = d_k)$, is defined as follows.*

$$
p(\alpha_1 = v_1 | \alpha_2 = v_2, ..., \alpha_I = v_I, M = m_n, D = d_k) = \begin{cases} \mu & \textit{if } m_n \in [\![\alpha_1 = v_1]\!] \\ 1 - \mu & \textit{otherwise} \end{cases}
$$

If we adopt the convention that $0^0 = 1$ then Definition 3 can be expressed as a Bernoulli distribution with parameter μ.

$$
p(\alpha_1 = v_1 | \alpha_2 = v_2, ..., \alpha_I = v_I, M = m_n, D = d_k) = \mu^{[\![\alpha_1 = v_1]\!]_{m_n}} (1 - \mu)^{1 - [\![\alpha_1 = v_1]\!]_{m_n}}
$$

In Figure 1, the arrows from the middle to the bottom layer of the hierarchy shown on the right indicate that the predicate formula is true in these models. The following probabilistic property of conditional independence comes directly from the property of predicate logic.

Proposition 1. *Let $\alpha_1, \alpha_2 \in L$. α_1 is conditionally independent of α_2 and D given M, i.e., $p(\alpha_1 | \alpha_2, M, D) = p(\alpha_1 | M)$.*

Proof. See Appendix. □

From the equation in Proposition 1, we can simplify Definition 3 as follows.

$$
p(\alpha_1 = v_1 | M = m_n) = \mu^{[\![\alpha_1 = v_1]\!]} (1 - \mu)^{1 - [\![\alpha_1 = v_1]\!]} \tag{3}
$$

The following example shows why we need to assume $\mu \in [0.5, 1]$, rather than simply $\mu = 1$.

Example 6 (Continued.)**.** *Let $\mu = 1$ and $\alpha = Blames(alice, bob) \wedge \neg Blames(alice, bob)$. Using the definition of conditional probability, the sum rule, the product rule and Proposition 1, we have*

$$
\begin{aligned}
p(M = m_4 | \alpha) &= \frac{p(M = m_4, \alpha)}{p(\alpha)} = \frac{p(M = m_4, \alpha)}{\sum_{n=1}^{32} p(M = m_n, \alpha)} \\
&= \frac{p(\alpha | M = m_4) p(M = m_4)}{\sum_{n=1}^{32} p(\alpha | M = m_n) p(M = m_n)} = \frac{(1 - \mu) p(M = m_4)}{\sum_{n=1}^{32} (1 - \mu) p(M = m_n)} = \frac{0}{0}.
\end{aligned}
$$

Namely, the value is undefined due to division by zero. However, given $\mu \neq 1$ such as μ approaches one, denoted by $\mu \to 1$, the undefined value can be replaced by a reasonable one.

$$p(M = m_4|\alpha) = \lim_{\mu \to 1} \frac{(1-\mu)p(M = m_4)}{\sum_{n=1}^{32}(1-\mu)p(M = m_n)} = \frac{p(M = m_4)}{\sum_{n=1}^{32} p(M = m_n)} = p(M = m_4)$$

Here, $\sum_{n=1}^{32} p(M = m_n) = 1$ as $p(M)$ is the probability distribution over all the models.

In Section 3, we will discuss that $\mu = 1$ corresponds to the logical consequence relation and $\mu \to 1$ corresponds to its natural generalisations. In Figure 1, each arrow between the middle and bottom layers of the hierarchy shown on the right shows that the model satisfies or supports the formula, $\forall x(\exists y(Blames(x, y)) \to Blames(alice, x))$.

2.3 Predicate reasoning

We can now discuss probabilistic reasoning with predicate language. The following property is useful to simply our notation.

Proposition 2. *Let $\alpha \in L$. $p(\alpha = 1) = p(\neg\alpha = 0)$.*

Proof. See Appendix. □

Therefore, we can write $\neg\alpha = 0$ as $\alpha = 1$ and abbreviate $\alpha = 1$ as α, for all $\alpha \in L$. We also abbreviate $M = m_n$ and $D = d_k$ as m_n and d_k, respectively.

Now, using the sum rule, the product rule and the conditional independence, i.e., Proposition 1, the probability of $\alpha, \beta \in L$ can be expressed as follows.

$$\begin{aligned} p(\alpha, \beta) &= \sum_k^K \sum_n^N p(\alpha, \beta, m_n, d_k) = \sum_k^K \sum_n^N p(\alpha|\beta, m_n, d_k)p(\beta|m_n, d_k)p(m_n|d_k)p(d_k) \\ &= \sum_k^K \sum_n^N p(\alpha|m_n)p(\beta|m_n)p(m_n|d_k)p(d_k) \end{aligned}$$

Since $p(d_k) = 1/K$, i.e., Definition 1, and our assumption that each data point supports a single model, we finally have

$$p(\alpha, \beta) = \frac{1}{K} \sum_k^K \sum_n^N p(\alpha|m_n)p(\beta|m_n)p(m_n|d_k) = \frac{1}{K} \sum_k^K p(\alpha|m(d_k))p(\beta|m(d_k)). \tag{4}$$

We here used $\sum_n^N p(\alpha|m_n)p(\beta|m_n)p(m_n|d_k) = p(\alpha|m(d_k))p(\beta|m(d_k))$. This fact is crucially important in terms of decidability and computational complexity since N can be countably infinite.

Example 7. *Let α be $\forall x(\exists y(Blames(x,y)) \rightarrow Blames(alice,x))$ shown in Figure 1. The probability of the formula being true can be evaluated using Equation (4).*

$$\begin{aligned} p(\alpha) &= \frac{1}{20}\sum_{k=1}^{20} p(\alpha|m(d_k)) = \frac{1}{20}\sum_{k=1}^{20} \mu^{[\![\alpha]\!]_{m(d_k)}}(1-\mu)^{1-[\![\alpha]\!]_{m(d_k)}} \\ &= \frac{1}{20}\{\sum_{k\in\{1\text{-}17\}} \mu^1(1-\mu)^0 + \sum_{k\in\{18\text{-}20\}} \mu^0(1-\mu)^1\} \\ &= \frac{1}{20}\{\sum_{k\in\{1\text{-}17\}} \mu + \sum_{k\in\{18\text{-}20\}} (1-\mu)\} = \frac{17\mu + 3(1-\mu)}{20} \end{aligned}$$

Therefore, $p(\alpha) = 17/20$ when $\mu = 1$. This result is intuitive as it is the number of data supporting models where α is true, out of all the twenty data.

3. Evaluations

3.1 Reasoning as learning

The common view in statistics is that observed data are generated from probability distributions characterised by their parameters. Maximum likelihood estimation (MLE) is the most commonly used statistical method to estimate the values of unobserved parameters only from observed data. MLE is defined as $\hat{\Theta} = \arg\max_{\Theta} p(d_1, d_2, ..., d_K|\Theta)$, where each d_k is an observed data point and Θ is the set of parameters of a probability distribution.

Proposition 3. *Let $\{d_1, d_2, ..., d_K\}$ be a multiset of K data and Θ be the parameters of a categorical distribution. $p(M) = \hat{\Theta}$ if and only if $\hat{\Theta}$ maximises the likelihood of data, i.e., $\hat{\Theta} = \arg\max_{\Theta} p(d_1, d_2, ..., d_K|\Theta)$.*

Proof. See Appendix. □

Example 8. *Consider the twenty data and thirty two models shown on the top layers of the both hierarchies in Figure 1. Let K be the total number of data, and K_n be the number of data in the nth model. We then have $p(m_1) = \frac{7}{20}$, $p(m_4) = \frac{4}{20}$, $p(m_7) = \frac{6}{20}$, $p(m_{11}) = \frac{1}{20}$, $p(m_{13}) = \frac{2}{20}$ and $p(m_n) = 0$, for all the remaining models m_n.*

3.2 Reasoning from possible information

This section aims to logically characterise the inference of predicative abstraction with $\mu = 1$. We focus on the relation between models and formulas by marginalising out data, i.e., $p(\alpha, M) = \sum_k p(\alpha, M, D = d_k)$. As usual, we use the symbol $[\![\Delta]\!]$ to denote the set of models where all the formulas in $\Delta \subseteq L$ are true, i.e., $[\![\Delta]\!] = \bigcap_{\alpha\in\Delta}[\![\alpha]\!]$. A model with a non-zero probability is called possible. We use the symbol $[\![\![\Delta]\!]\!]$ to denote the set of possible models where all the formulas in $\Delta \subseteq L$ are true, i.e., $[\![\![\Delta]\!]\!] = \{m_n \in [\![\Delta]\!] | p(m_n) \neq 0\}$. We write $m_n \in [\![\Delta]\!]$ and $m_n \in [\![\![\Delta]\!]\!]$ as $[\![\Delta]\!]_{m_n} = 1$ and $[\![\![\Delta]\!]\!]_{m_n} = 1$, respectively. Note that $[\![\emptyset]\!]$ and $[\![\![\emptyset]\!]\!]$ are the sets of all models and

all possible models, respectively. We use the empirical consequence relation originally defined for propositional logic.

Definition 4 ((Kido, 2025a))**.** *Let $\alpha \in L$ and $\Delta \subseteq L$. α is an empirical consequence of Δ, denoted by $\Delta \equiv \alpha$, if $[\![\Delta]\!] \subseteq [\![\alpha]\!]$.*

As usual, α is a logical consequence of Δ, denoted by $\Delta \models \alpha$, if $[\![\Delta]\!] \subseteq [\![\alpha]\!]$. The empirical consequence relation $\equiv$ is thus a probabilistic generalisation of the logical consequence relation $\models$.

We write $p_{\mu=1}(\alpha)$ and $p_{\mu\to 1}(\alpha)$ when we want to specify the value of μ used in the evaluation. We often omit it when the value of μ is obvious from the context. We can now logically characterise the inference of predicative abstraction with $\mu = 1$.

Theorem 1. *Let $\alpha \in L$ and $\Delta \subseteq L$ such that $[\![\Delta]\!] \neq \emptyset$.*

$$p_{\mu=1}(\alpha|\Delta) = \frac{\sum_n [\![\alpha]\!]_{m_n} [\![\Delta]\!]_{m_n} p(m_n)}{\sum_n [\![\Delta]\!]_{m_n} p(m_n)}$$

Proof. See Appendix. □

The assumption of $[\![\Delta]\!] \neq \emptyset$ guarantees that $p_{\mu=1}(\alpha|\Delta)$ involves no division by zero, which causes an undefined value.

Example 9 (Continued.)**.** *Given that everyone blames someone and that everyone is blamed by someone, what is the probability that Alice blames Bob? Each model determines the truth value of each atomic formula as follows (see Figure 1).*

$$\begin{aligned}
&[\![Blames(alice, alice)]\!] = \{m_7, m_{13}\} \\
&[\![Blames(alice, bob)]\!] = \{m_4, m_7, m_{11}\} \\
&[\![Blames(bob, alice)]\!] = \{m_4\} \\
&[\![Blames(bob, bob)]\!] = \{m_{11}, m_{13}\}
\end{aligned}$$

Thus, each compound formula has the following truth value.

$$\begin{aligned}
[\![\forall x \exists y\ Blames(x, y)]\!]_{m_n} &= \min_{c_1 \in \mathcal{C}} \left\{ \max_{c_2 \in \mathcal{C}} \left\{ [\![Blames(c_1, c_2)]\!]_{m_n} \right\} \right\} \\
&= \min \left\{ \begin{array}{l} \max \left\{ [\![Blames(alice, alice)]\!]_{m_n}, [\![Blames(alice, bob)]\!]_{m_n} \right\}, \\ \max \left\{ [\![Blames(bob, alice)]\!]_{m_n}, [\![Blames(bob, bob)]\!]_{m_n} \right\} \end{array} \right\} \\
&= \begin{cases} 1 & \textit{if } n \in \{4, 11, 13\} \\ 0 & \textit{otherwise} \end{cases} \\
[\![\forall y \exists x\ Blames(x, y)]\!]_{m_n} &= \min_{c_1 \in \mathcal{C}} \left\{ \max_{c_2 \in \mathcal{C}} \left\{ [\![Blames(c_2, c_1)]\!]_{m_n} \right\} \right\} \\
&= \min \left\{ \begin{array}{l} \max \left\{ [\![Blames(alice, alice)]\!]_{m_n}, [\![Blames(bob, alice)]\!]_{m_n} \right\}, \\ \max \left\{ [\![Blames(alice, bob)]\!]_{m_n}, [\![Blames(bob, bob)]\!]_{m_n} \right\} \end{array} \right\} \\
&= \begin{cases} 1 & \textit{if } n \in \{4, 7, 13\} \\ 0 & \textit{otherwise} \end{cases}
\end{aligned}$$

Now, we can apply Theorem 1 to find the probability.

$$
\begin{aligned}
&p(Blames(alice, bob) | \forall x \exists y\ Blames(x, y), \forall y \exists x\ Blames(x, y)) \\
&= \frac{\sum_n [\![Blames(alice, bob)]\!]_{m_n} [\![\forall x \exists y\ Blames(x, y), \forall y \exists x\ Blames(x, y)]\!]_{m_n} p(m_n)}{\sum_n [\![\forall x \exists y\ Blames(x, y), \forall y \exists x\ Blames(x, y)]\!]_{m_n} p(m_n)} \\
&= \frac{\sum_{n \in \{4\}} p(m_n)}{\sum_{n \in \{4,13\}} p(m_n)} = \frac{\sum_k p(m_4 | d_k)}{\sum_k \{p(m_4 | d_k) + p(m_{13} | d_k)\}} = \frac{4}{4+2} = \frac{2}{3}
\end{aligned}
$$

We added the marginalised data in the last line. $p(d_k)$ is canceled out due to its uniformity.

The following fact shows that Theorem 1 is a generalisation of the empirical consequence relation. The proof can be found in Appendix A.

Corollary 1. *Let $\alpha \in L$ and $\Delta \subseteq L$ such that $[\![\Delta]\!] \neq \emptyset$. $p_{\mu=1}(\alpha|\Delta) = 1$ if and only if $\Delta \models \alpha$.*

3.3 Reasoning from impossible information

This section aims to logically characterise the inference of predicative abstraction with $\mu \to 1$. We use the maximal possible set originally defined for propositional logic.

Definition 5 ((Kido, 2025a))**.** *Let $\Delta \subseteq L$. $S \subseteq \Delta$ is a maximal possible subset of Δ if $[\![S]\!] \neq \emptyset$ and $[\![S \cup \{\alpha\}]\!] = \emptyset$, for all $\alpha \in \Delta \setminus S$.*

As usual, $S \subseteq \Delta$ is a maximal consistent subset of Δ if $[\![S]\!] \neq \emptyset$ and $[\![S \cup \{\alpha\}]\!] = \emptyset$, for all $\alpha \in \Delta \setminus S$. A maximal possible set is thus a probabilistic generalisation of a maximal consistent set. We use the symbol $MPS(\Delta)$ and $MCS(\Delta)$ to denote the set of the cardinality-maximal possible subsets of Δ and the set of cardinality-maximal consistent subsets of Δ, respectively.

Example 10. *Let us discuss examples of maximal consistent sets and maximal possible sets in propositional logic. Consider the hierarchy shown on the left in Figure 1. Let $\Delta = \{rain, sprinkler, sprinkler \to wet, hot, wet, \neg wet\}$. The set of the maximal consistent subsets of Δ is $\{S_1, S_2, S_3\}$ given as follows.*

- $S_1 = \{rain, sprinkler, sprinkler \to wet, hot, wet\}$ *where* $[\![S_1]\!] = \{m_6, m_8\}$
- $S_2 = \{rain, sprinkler, hot, \neg wet\}$ *where* $[\![S_2]\!] = \{m_{14}, m_{16}\}$
- $S_3 = \{rain, sprinkler \to wet, hot, \neg wet\}$ *where* $[\![S_3]\!] = \{m_{10}, m_{12}\}$

Therefore, the set of the cardinality-maximal consistent subsets of Δ is $\{S_1\}$, i.e., $MCS(\Delta) = \{S_1\}$. Meanwhile, the set of the maximal possible subsets of Δ is $\{S_4, S_5, S_6\}$ given as follows.

- $S_4 = \{rain, sprinkler \to wet, hot, wet\}$ *where* $[\![S_4]\!] = \{m_4\}$
- $S_5 = \{sprinkler, sprinkler \to wet, hot, wet\}$ *where* $[\![S_5]\!] = \{m_{24}\}$
- $S_6 = \{sprinkler \to wet, hot, \neg wet\}$ *where* $[\![S_6]\!] = \{m_{26}, m_{28}\}$

Therefore, the set of the cardinality-maximal possible subsets of Δ *is* $\{S_4, S_5\}$*, i.e.,* $MPS(\Delta) = \{S_4, S_5\}$.

We often use the symbol $(\!(\!(\Delta)\!)\!)$ to denote the set of possible models where all the formulas in a cardinality-maximal possible subset of Δ are true, i.e., $(\!(\!(\Delta)\!)\!) = \bigcup_{S \in MPS(\Delta)} [\![S]\!]$. We also use the symbol $(\!(\Delta)\!)$ to denote the set of models where all the formulas in a cardinality-maximal consistent subset of Δ are true, i.e., $(\!(\Delta)\!) = \bigcup_{S \in MCS(\Delta)} [\![S]\!]$.

Theorem 2. *Let* $\alpha \in L$ *and* $\Delta \subseteq L$.

$$p_{\mu \to 1}(\alpha|\Delta) = \frac{\sum_n [\![\alpha]\!]_{m_n} (\bigcup_{S \in MPS(\Delta)} [\![S]\!])_{m_n} p(m_n)}{\sum_n (\bigcup_{S \in MPS(\Delta)} [\![S]\!])_{m_n} p(m_n)}$$

Proof. See Appendix. □

The denominator of Theorem 2 cannot be zero. For example, $(\!(\!(\emptyset)\!)\!) = \bigcup_{S \in MPS(\emptyset)} [\![S]\!] = [\![\emptyset]\!]$, where $[\![\emptyset]\!]$ is the set of all possible models. All models cannot be impossible by definition.

Example 11 (Continued.)**.** *Given that everyone blames someone, everyone is blamed by someone, and someone blames everyone, what is the probability that Alice blames Bob? Let us use* α *to denote* $Blames(x, y)$*. Each compound formula has the following truth value.*

$$[\![\forall x \exists y\ \alpha]\!] = \{m_4, m_{11}, m_{13}\} \textit{ (See Example 9.)}$$
$$[\![\forall y \exists x\ \alpha]\!] = \{m_4, m_7, m_{13}\} \textit{ (See Example 9.)}$$
$$[\![\exists x \forall y\ \alpha]\!]_{m_n} = \max_{c_1 \in \mathcal{C}} \left\{ [\![\forall y\ Blames(c_1, y)]\!]_{m_n} \right\} = \max_{c_1 \in \mathcal{C}} \left\{ \min_{c_2 \in \mathcal{C}} \left\{ [\![Blames(c_1, c_2)]\!]_{m_n} \right\} \right\}$$
$$= \max \left\{ \begin{array}{l} \min \left\{ [\![Blames(alice, alice)]\!]_{m_n}, [\![Blames(alice, bob)]\!]_{m_n} \right\}, \\ \min \left\{ [\![Blames(bob, alice)]\!]_{m_n}, [\![Blames(bob, bob)]\!]_{m_n} \right\} \end{array} \right\} = \begin{cases} 1 & \textit{if } n \in \{7\} \\ 0 & \textit{otherwise} \end{cases}$$

Now, we can apply Theorem 2 to find the probability.

$$\begin{aligned} & p(Blames(alice, bob) | \forall x \exists y\ \alpha, \forall y \exists x\ \alpha, \exists x \forall y\ \alpha) \\ = & \frac{\sum_n [\![Blames(alice, bob)]\!]_{m_n} (\!(\!(\forall x \exists y\ \alpha, \forall y \exists x\ \alpha, \exists x \forall y\ \alpha)\!)\!)_{m_n} p(m_n)}{\sum_n (\!(\!(\forall x \exists y\ \alpha, \forall y \exists x\ \alpha, \exists x \forall y\ \alpha)\!)\!)_{m_n} p(m_n)} \\ = & \frac{\sum_n [\![Blames(alice, bob)]\!]_{m_n} ([\![\forall x \exists y\ \alpha, \forall y \exists x\ \alpha]\!] \cup [\![\forall y \exists x\ \alpha, \exists x \forall y\ \alpha]\!])_{m_n} p(m_n)}{\sum_n ([\![\forall x \exists y\ \alpha, \forall y \exists x\ \alpha]\!] \cup [\![\forall y \exists x\ \alpha, \exists x \forall y\ \alpha]\!])_{m_n} p(m_n)} \\ = & \frac{\sum_{n \in \{4,7\}} p(m_n)}{\sum_{n \in \{4,7,13\}} p(m_n)} = \frac{\sum_k \{p(m_4|d_k) + p(m_7|d_k)\}}{\sum_k \{p(m_4|d_k) + p(m_7|d_k) + p(m_{13}|d_k)\}} = \frac{4+6}{4+6+2} = \frac{5}{6} \end{aligned}$$

We added the marginalised data in the last line. $p(d_k)$ *is canceled out due to its uniformity.*

Theorem 2 implies the following properties, and their proofs can be found in Appendix A. The certain reasoning, i.e., reasoning with a probability of one, can be characterised in terms of the empirical consequence relation and its application to cardinality-maximal possible sets.

Corollary 2. *Let $\alpha \in L$ and $\Delta \subseteq L$. $p_{\mu\to 1}(\alpha|\Delta) = 1$ if and only if $S \models\!\!\!\equiv \alpha$, for all cardinality-maximal possible subsets S of Δ.*

Let us assume $(\!(\!(\Delta)\!)\!) = (\!(\Delta)\!)$. This is the case when all the models satisfying S are possible, for all cardinality-maximal consistent subsets S of Δ, i.e., $[\![S]\!] = [\![S]\!]$, for all $S \in MCS(\Delta)$. Then, the certain reasoning can be characterised in terms of the logical consequence relation and its application to cardinality-maximal consistent sets.

Corollary 3. *Let $\alpha \in L$ and $\Delta \subseteq L$ such that $(\!(\!(\Delta)\!)\!) = (\!(\Delta)\!)$. $p_{\mu\to 1}(\alpha|\Delta) = 1$ if and only if $S \models \alpha$, for all cardinality-maximal consistent subsets S of Δ.*

Let us assume that there is a possible model satisfying Δ, i.e., $[\![\Delta]\!] \neq \emptyset$. Then, the certain reasoning can be characterised in terms of the empirical consequence relation.

Corollary 4. *Let $\alpha \in L$ and $\Delta \subseteq L$ such that $[\![\Delta]\!] \neq \emptyset$. $p_{\mu\to 1}(\alpha|\Delta) = 1$ if and only if $\Delta \models\!\!\!\equiv \alpha$.*

Let us assume $(\!(\!(\Delta)\!)\!) = (\!(\Delta)\!)$ and $[\![\Delta]\!] \neq \emptyset$. The certain reasoning can then be characterised in terms of the logical consequence relation.

Corollary 5. *Let $\alpha \in L$ and $\Delta \subseteq L$ such that $(\!(\!(\Delta)\!)\!) = (\!(\Delta)\!)$ and $[\![\Delta]\!] \neq \emptyset$. $p_{\mu\to 1}(\alpha|\Delta) = 1$ if and only if $\Delta \models \alpha$.*

The following fact shows that the assumption of $[\![\Delta]\!] \neq \emptyset$ makes $\mu = 1$ and $\mu \to 1$ equivalent.

Corollary 6. *Let $\alpha \in L$ and $\Delta \subseteq L$ such that $[\![\Delta]\!] \neq \emptyset$. $p_{\mu\to 1}(\alpha|\Delta) = p_{\mu=1}(\alpha|\Delta)$.*

The following is the summary of Sections 3.2 and 3.3. There is an arrow from property x to property y when x is more general than y. Each arrow specifies the assumption of y that makes it less general than x.

$$\text{Theorem 1} \xrightarrow{p(\alpha|\Delta)=1} \text{Corollary 1}$$

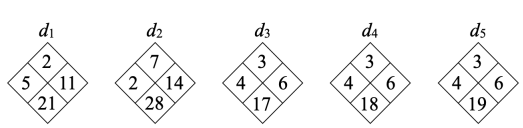
$$\text{Theorem 2} \xrightarrow{p(\alpha|\Delta)=1} \text{Corollary 2} \xrightarrow{(\!(\!(\Delta)\!)\!)=(\!(\Delta)\!)} \text{Corollary 3} \xrightarrow{[\![\Delta]\!]\neq\emptyset} \text{Corollary 5}$$

$$\text{Corollary 2} \xrightarrow{[\![\Delta]\!]\neq\emptyset} \text{Corollary 4} \xrightarrow{(\!(\!(\Delta)\!)\!)=(\!(\Delta)\!)} \text{Corollary 5}$$

3.4 Applicability

This section aims to show the applicability of the inference of predicative abstraction. We discuss a solution to a simple example, similar to Examples 1 and 2, that is often difficult to solve by existing AI approaches. Consider the following simple arithmetic quiz for testing one's abstract and logical thinking skills.

Example 12. *Which one of the following data d_3, d_4 and d_5 is a companion of d_1 and d_2?*



The correct answer could be d_4. Only d_1, d_2 and d_4 satisfy the following arithmetic rule.

$$\textit{top} \times \textit{left} + \textit{right} = \textit{bottom}$$

We assume the following vocabularies for the usual arithmetic operations.

$$\begin{aligned}
&\mathcal{C} = \{\text{`}i\text{'}, \textit{left, right, top, bottom} | i \in \mathbb{R}\} \\
&\mathcal{F} = \{\text{`}+\text{'}, \text{`}-\text{'}, \text{`}\times\text{'}, \text{`}\div\text{'}\}, \text{where } ar(\cdot) = 2, \text{for all `}\cdot\text{'} \in \mathcal{F} \\
&\mathcal{P} = \{\text{`}=\text{'}\}, \text{where } ar(\text{`}=\text{'}) = 2
\end{aligned}$$

$\mathcal{C}$ comprises the constants for the real numbers $i \in \mathbb{R}$ and for the four numbers located to the left, right, top and bottom. We assume no variables, i.e., $\mathcal{V} = \emptyset$. Here, *left, right, top, bottom* are not variables since each of them has a single value for each data point. We omit the symbol ‘’ when it is obvious that its inside is a vocabulary of a predicate language. We use the infix notation for readability. Let w, x, y and z be the numbers on the left, right, top and bottom of a data point, respectively. Since we are interested in the usual arithmetic operations, we assume the function m that maps each d_k to the model $m(d_k) = \langle u(m(d_k)), v(m(d_k))\rangle$ given as follows.

$$\begin{aligned}
&u(m(d_k)) = \mathbb{R} \\
&v(m(d_k))(\text{`}i\text{'}) = i, \text{for all } i \in u(m(d_k)) \\
&v(m(d_k))(\textit{left}) = w \\
&v(m(d_k))(\textit{right}) = x \\
&v(m(d_k))(\textit{top}) = y \\
&v(m(d_k))(\textit{bottom}) = z \\
&v(m(d_k))(\text{`}\cdot\text{'})(i, j) = i \cdot j, \text{for all `}\cdot\text{'} \in \mathcal{F} \\
&v(m(d_k))(\text{`}=\text{'}) = \{(i, i) | i \in u(m(d_k))\}
\end{aligned}$$

Note that only the assignments to *left*, *right*, *top*, *bottom* depend on individual data, and the other assignments depend only on the arithmetic rule.

Our solution is based on two modes of the inference of abstraction. The first mode is the inference over d_1 and d_2, which aims to extract an abstract rule from the concrete data using the following equation.

$$\hat{\alpha} = \arg\max_{\alpha \in L} p(\alpha) \tag{5}$$

The second mode is the inference over d_3, d_4 and d_5, which aims to apply the extracted abstract rule to the concrete data using the following equation.

$$\hat{\beta} = \arg\max_{\beta \in L} p(\beta | \hat{\alpha}) \tag{6}$$

Equations (5) and (6) are computationally intractable, since the predicate language generally has the infinite number of formulas. It is beyond the scope of this paper to fully discuss how to explore

the infinite language space. We here simply search for a formula with a small number of logical connectives for a simple explanation, which conforms to Occam's razor. Now, let $\mu = 1$. $\hat{\alpha} = (\textit{top} \times \textit{left} + \textit{right} = \textit{bottom})$ satisfies Equation (5).

$$\begin{aligned} p(\hat{\alpha}) &= \sum_{n=1}^{\infty}\sum_{k=1}^{2} p(\hat{\alpha}, m_n, d_k) = \sum_{n=1}^{\infty}\sum_{k=1}^{2} p(\hat{\alpha}|m_n)p(m_n|d_k)p(d_k) \\ &= \sum_{k=1}^{2} p(\hat{\alpha}|m(d_k))p(d_k) = \frac{1}{2}\sum_{k=1}^{2} [\![\hat{\alpha}]\!]_{m(d_k)} = 1 \end{aligned}$$

Now, $\hat{\beta} = (\textit{bottom} = 18)$ satisfies Equation (6).

$$\begin{aligned} p(\hat{\beta}|\hat{\alpha}) &= \frac{\sum_{n=1}^{\infty}\sum_{k=3}^{5} p(\hat{\beta}, \hat{\alpha}, m_n, d_k)}{\sum_{n=1}^{\infty}\sum_{k=3}^{5} p(\hat{\alpha}, m_n, d_k)} = \frac{\sum_{n=1}^{\infty}\sum_{k=3}^{5} p(\hat{\beta}|m_n)p(\hat{\alpha}|m_n)p(m_n|d_k)p(d_k)}{\sum_{n=1}^{\infty}\sum_{k=3}^{5} p(\hat{\alpha}|m_n)p(m_n|d_k)p(d_k)} \\ &= \frac{\sum_{k=3}^{5} p(\hat{\beta}|m(d_k))p(\hat{\alpha}|m(d_k))}{\sum_{k=3}^{5} p(\hat{\alpha}|m(d_k))} = \frac{\sum_{k=3}^{5} [\![\hat{\beta}]\!]_{m(d_k)} [\![\hat{\alpha}]\!]_{m(d_k)}}{\sum_{k=3}^{5} [\![\hat{\alpha}]\!]_{m(d_k)}} = \frac{1}{1} \end{aligned}$$

4. Conclusions and Future Work

In this paper, we asked how predicate reasoning should be grounded in data for meaningful abstraction. We proposed the inference of abstraction by simply modelling the idea that an intrinsically abstract predicate formula is a selective ignorance of the models of a predicate language and each of the models is a selective ignorance of intrinsically concrete observed data. We showed that the idea is not only enough to characterise the logical consequence relation of predicate logic but also to generalise it for the empirical consequence relation and its application to cardinality-maximal possible sets. The simple yet unconventional idea suggests a fresh perspective to rethink various important issues such as symbol grounding, computationally tractable predicate reasoning, commonsense reasoning, and reasoning from an inconsistent and impossible source of information.

An interesting direction in future work is to study how to define the function m so that it accurately maps data to models, as depicted by the arrows from the top to middle layers in Figure 1. A more fundamental question is how an agent should decide which vocabularies to use in its predicate language for creative abstraction of data.

References

Bishop, C. M. (2006). *Pattern recognition and machine learning*. 1 New York Plaza, Suite 4600, New York, NY 10004-1562: Springer New York, NY.

Brewka, G. (1991). *Nonmonotonic reasoning: Logical foundations of commonsense*. Cambridge, England: Cambridge University Press.

Chollet, F. (2019). On the measure of intelligence. *arXiv:1911.01547v2 [cs.AI]* (pp. 1–64).

Dasgupta, I., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2020). A theory of learning to infer. *Psychol Rev.*, *127(3)*, 412–441.

Davis, E., & Marcus, G. (2015). Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, *58(9)*, 92–103.

DeepSeek-AI (2025). DeepSeek-R1: Incentivizing reasoning capability in llms via reinforcement learning. https://arxiv.org/pdf/2501.12948, Retrieved Feb 2025.

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

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philos Trans R Soc Lond B Biol Sci*, *352*, 1121–1127.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42(1-3)*, 335–346.

Hawkins, J. (2021). *A thousand brains: A new theory of intelligence*. New York City: Basic Books.

Kido, H. (2025a). Inference of abstraction for human-like logical reasoning. *Machine Learning, Optimization, and Data Science* (pp. 191–206). Cham: Springer Nature Switzerland.

Kido, H. (2025b). Inference of abstraction for human-like probabilistic reasoning. *Machine Learning, Optimization, and Data Science* (pp. 116–131). Cham: Springer Nature Switzerland.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350(6266)*, 1332–1338.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40(e253)*, 1–72.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of Optical Society of America*, *20*, 1434–1448.

Muggleton, S. (1995). Inverse entailment and progol. *New Generation Computing*, *13*, 245–286.

Muggleton, S., & Buntine, W. (1988). Machine invention of first-order predicates by inverting resolution. *Proc. 5th International Conference on Machine Learning* (pp. 339–352).

Nienhuys-Cheng, S. H., & Wolf, R. D. (1997). *Foundation of inductive logic programming*. Heidelberg: Springer Berlin, Heidelberg.

OpenAI (2025). ChatGPT. https://chat.openai.com/chat, Retrieved Feb 2025.

Pearl, J., & Russell, S. (2003). *Handbook of brain theory and neural networks*, chapter Bayesian Networks, (pp. 157–160). Cambridge, Massachusetts: MIT Press.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.

Russell, S., & Norvig, P. (2020). *Artificial intelligence : A modern approach, fourth edition*. London, England: Pearson Education, Inc.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10(7)*, 309–318.

Appendix A. Proofs

Proposition 1. Using the definition of conditional probability, the right-hand side can be written as

$$p(\alpha_1|M) = \frac{p(\alpha_1, M)}{p(M)}.$$

Using the sum rule (Bishop, 2006) and the product rule (Bishop, 2006) of probability theory, its numerator can be expanded as

$$\begin{aligned} p(\alpha_1, M) &= \sum_{v_2} \sum_{d_k} p(\alpha_1, \alpha_2 = v_2, M, D = d_k) \\ &= \sum_{v_2} \sum_{d_k} p(\alpha_1|\alpha_2 = v_2, M, D = d_k) p(\alpha_2 = v_2, M, D = d_k). \end{aligned}$$

Now, it is obvious from Definition 3 that neither α_2 nor D affects the value of $p(\alpha_1|\alpha_2 = v_2, M, D = d_k)$. We can thus move it outward. Using the sum rule, we have

$$p(\alpha_1, M) = p(\alpha_1|\alpha_2, M, D) \sum_{v_2} \sum_{d_k} p(\alpha_2 = v_2, M, D = d_k) = p(\alpha_1|\alpha_2, M, D) p(M).$$

Taking into accout the denominator, the original expression can be written as

$$p(\alpha_1|M) = \frac{p(\alpha_1|\alpha_2, M, D) p(M)}{p(M)} = p(\alpha_1|\alpha_2, M, D).$$

□

Proposition 2. α is true in a model iff $\neg\alpha$ is false in the model. Thus, $[\![\alpha = 1]\!] = [\![\neg\alpha = 0]\!]$. Using the sum rule, the product rule and Proposition 1, we have

$$\begin{aligned} p(\alpha = 1) &= \sum_{m_n} p(\alpha = 1|M = m_n) p(M = m_n) = \sum_{m_n} \mu^{[\![\alpha=1]\!]_{m_n}} (1-\mu)^{1-[\![\alpha=1]\!]_{m_n}} p(M = m_n) \\ &= \sum_{m_n} \mu^{[\![\neg\alpha=0]\!]_{m_n}} (1-\mu)^{1-[\![\neg\alpha=0]\!]_{m_n}} p(M = m_n) \\ &= \sum_{m_n} p(\neg\alpha = 0|M = m_n) p(M = m_n) = p(\neg\alpha = 0). \end{aligned}$$

This holds regardless of the value of μ. □

Proposition 3. Let K be the total number of data, and K_n be the number of data in the nth category or model. The maximum likelihood estimate of the parameter θ_n for a categorical distribution is simply known as the relative frequency of data, i.e.,

$$\theta_n = \frac{\text{The number of data in the } n\text{th category}}{\text{The total number of data}} = \frac{K_n}{K}.$$

Therefore, the maximum likelihood estimate is given by

$$\hat{\Theta} = (\frac{K_1}{K}, \frac{K_2}{K}, ..., \frac{K_N}{K}).$$

Let m_n be a model of predicate logic. Using the sum and product rules, we have

$$p(m_n) = \sum_k p(m_n, d_k) = \sum_k p(m_n|d_k)p(d_k) = \frac{1}{K}\sum_k p(m_n|d_k) = \frac{K_n}{K}.$$

Therefore, we have $p(M) = \hat{\Theta}$. □

Theorem 1. Using the definition of conditional probability and the conditional independence we showed in the previous section, we have

$$p(\alpha|\Delta) = \frac{p(\alpha, \Delta)}{p(\Delta)} = \frac{\sum_n p(\alpha, \Delta, m_n)}{\sum_n p(\Delta, m_n)} = \frac{\sum_n p(\alpha|m_n)p(\Delta|m_n)p(m_n)}{\sum_n p(\Delta|m_n)p(m_n)}.$$

Dividing models into possible ones, i.e., $[\![\Delta]\!]$, and the others, we have

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

Since $\mu = 1$, $p(\Delta|m_n)$ can be expanded as follows.

$$p(\Delta|m_n) = \prod_{\beta \in \Delta} p(\beta|m_n) = \prod_{\beta \in \Delta} 1^{[\![\beta]\!]_{m_n}} 0^{1-[\![\beta]\!]_{m_n}} = \begin{cases} 1 & \text{if } m_n \in [\![\Delta]\!] \\ 0 & \text{otherwise} \end{cases}$$

Thus, $p(\Delta|m_n) = 0$, for all $m_n \notin [\![\Delta]\!]$. Moreover, $p(m_n) = 0$, for all $m_n \in [\![\Delta]\!] \setminus [\![\Delta]\!]$. These two facts imply that $\sum_{m_n \notin [\![\Delta]\!]} p(\Delta|m_n)p(m_n) = 0$. Since $p(\Delta|m_n) = 1$, for all $m_n \in [\![\Delta]\!]$, we have

$$p(\alpha|\Delta) = \frac{\sum_{m_n \in [\![\Delta]\!]} p(\alpha|m_n)p(m_n)}{\sum_{m_n \in [\![\Delta]\!]} p(m_n)} = \frac{\sum_n [\![\Delta]\!]_{m_n} p(\alpha|m_n)p(m_n)}{\sum_n [\![\Delta]\!]_{m_n} p(m_n)}.$$

Since $\mu = 1$, $p(\alpha|m_n)$ can be developed as follows.

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

Therefore,

$$p(\alpha|\Delta) = \frac{\sum_n [\![\alpha]\!]_{m_n} [\![\Delta]\!]_{m_n} p(m_n)}{\sum_n [\![\Delta]\!]_{m_n} p(m_n)}.$$

□

Corollary 1. $p(m_n) = 0$, for all $m_n \in [\![\alpha]\!] \setminus [\![\alpha]\!]$. From Theorem 1, we thus have

$$p_{\mu=1}(\alpha|\Delta) = \frac{\sum_n [\![\alpha]\!]_{m_n} [\![\Delta]\!]_{m_n} p(m_n)}{\sum_n [\![\Delta]\!]_{m_n} p(m_n)} = \frac{\sum_n [\![\alpha]\!]_{m_n} [\![\Delta]\!]_{m_n} p(m_n)}{\sum_n [\![\Delta]\!]_{m_n} p(m_n)}.$$

The above equation turns out to be one if and only if $[\![\Delta]\!] \subseteq [\![\alpha]\!]$, i.e., $\Delta \models \alpha$. □

Theorem 2. Let the symbols $|\Delta|$ and $|\Delta|_{m_n}$ denote the number of formulas in Δ and the number of formulas in Δ that are true in the model m_n, i.e., $|\Delta|_{m_n} = \sum_{\beta \in \Delta} [\![\beta]\!]_{m_n}$, respectively. Using the definition of conditional probability and the conditional independence we showed in the previous section, we have

$$p(\alpha|\Delta) = \frac{p(\alpha, \Delta)}{p(\Delta)} = \frac{\sum_n p(\alpha, \Delta, m_n)}{\sum_n p(\Delta, m_n)} = \frac{\sum_n p(\alpha|m_n) p(\Delta|m_n) p(m_n)}{\sum_n p(\Delta|m_n) p(m_n)}.$$

Dividing models into ones in $(\!(\!(\Delta)\!)\!)$ and the others, we have

$$p(\alpha|\Delta) = \frac{\sum_{\hat{m}_n \in (\!(\!(\Delta)\!)\!)} p(\alpha|\hat{m}_n) p(\Delta|\hat{m}_n) p(\hat{m}_n) + \sum_{m_n \notin (\!(\!(\Delta)\!)\!)} p(\alpha|m_n) p(\Delta|m_n) p(m_n)}{\sum_{\hat{m}_n \in (\!(\!(\Delta)\!)\!)} p(\Delta|\hat{m}_n) p(\hat{m}_n) + \sum_{m_n \notin (\!(\!(\Delta)\!)\!)} p(\Delta|m_n) p(m_n)}.$$

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

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

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

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

If $m_n \notin (\!(\!(\Delta)\!)\!)$ then m_n is impossible or m_n is a possible model of a subset of Δ that is not a cardinality-maximal possible subset of Δ. Therefore, $p(m_n) = 0$ or there is $\hat{m}_n \in (\!(\!(\Delta)\!)\!)$ such that $|\Delta|_{m_n} < |\Delta|_{\hat{m}_n}$. $|\Delta|_{\hat{m}_1} = |\Delta|_{\hat{m}_2}$ by definition, for all $\hat{m}_1, \hat{m}_2 \in (\!(\!(\Delta)\!)\!)$. The fraction thus can be simplified by dividing the denominator and numerator by $(1-\mu)^{|\Delta| - |\Delta|_{\hat{m}_n}}$. We thus have $p(\alpha|\Delta) = \lim_{\mu \to 1} \frac{W'+X'}{Y'+Z'}$ where

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

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

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

By convention, $1^{[\![\alpha]\!]_{\hat{m}_n}} 0^{1-[\![\alpha]\!]_{\hat{m}_n}} = 1^1 0^0 = 1$ if $\hat{m}_n \in [\![\alpha]\!]$ and $1^{[\![\alpha]\!]_{\hat{m}_n}} 0^{1-[\![\alpha]\!]_{\hat{m}_n}} = 1^0 0^1 = 0$ otherwise. Therefore,

$$p(\alpha|\Delta) = \frac{\sum_{\hat{m}_n \in (\!(\!(\Delta)\!)\!)} [\![\alpha]\!]_{\hat{m}_n} p(\hat{m}_n)}{\sum_{\hat{m}_n \in (\!(\!(\Delta)\!)\!)} p(\hat{m}_n)} = \frac{\sum_n [\![\alpha]\!]_{m_n} (\!(\!(\Delta)\!)\!)_{m_n} p(m_n)}{\sum_n (\!(\!(\Delta)\!)\!)_{m_n} p(m_n)}. \tag{7}$$

□

Corollary 2. Recall that $S \models \alpha$ is defined as $[\![S]\!] \subseteq [\![\alpha]\!]$. $[\![S]\!] \subseteq [\![\alpha]\!]$, for all $S \in MPS(\Delta)$ iff $\bigcup_{S \in MPS(\Delta)} [\![S]\!] \subseteq [\![\alpha]\!]$, i.e., $(\!(\!(\Delta)\!)\!) \subseteq [\![\alpha]\!]$. Since $\sum_{m_n \in (\!(\!(\Delta)\!)\!)} [\![\alpha]\!]_{m_n} p(m_n) = \sum_{m_n \in (\!(\!(\Delta)\!)\!)} [\![\alpha]\!]_{m_n} p(m_n)$, Equation (7) can be further expanded as follows, where the resulting value is set to one.

$$p(\alpha|\Delta) = \frac{\sum_n [\![\alpha]\!]_{m_n} (\!(\!(\Delta)\!)\!)_{m_n} p(m_n)}{\sum_n (\!(\!(\Delta)\!)\!)_{m_n} p(m_n)} = 1 \tag{8}$$

There is thus no model m_n such that $m_n \in [\![\alpha]\!] \setminus (\!(\!(\Delta)\!)\!)$. Therefore, $(\!(\!(\Delta)\!)\!) \subseteq [\![\alpha]\!]$. □

Corollary 3. Since $(\!(\!(\Delta)\!)\!) = (\!(\Delta)\!)$, Equation (7) can be expanded as follows, where the resulting value is set to one.

$$p(\alpha|\Delta) = \frac{\sum_n [\![\alpha]\!]_{m_n} (\!(\Delta)\!)_{m_n} p(m_n)}{\sum_n (\!(\Delta)\!)_{m_n} p(m_n)} = 1 \tag{9}$$

Since $(\!(\!(\Delta)\!)\!) = (\!(\Delta)\!)$, $p(m_n) \neq 0$, for all $m_n \in (\!(\Delta)\!)$. Thus, Equation (9) holds iff $(\!(\Delta)\!) \subseteq [\![\alpha]\!]$, i.e., $\bigcup_{S \in MCS(\Delta)} [\![S]\!] \subseteq [\![\alpha]\!]$. This holds iff $[\![S]\!] \subseteq [\![\alpha]\!]$, for all $S \in MCS(\Delta)$. □

Corollary 4. Since $[\![\Delta]\!] \neq \emptyset$, $MPS(\Delta) = \Delta$. Thus, $(\!(\!(\Delta)\!)\!) = \bigcup_{S \in MPS(\Delta)} [\![S]\!] = [\![\Delta]\!]$. Therefore, Equation (8) can be written as follows.

$$p(\alpha|\Delta) = \frac{\sum_n [\![\alpha]\!]_{m_n} [\![\Delta]\!]_{m_n} p(m_n)}{\sum_n [\![\Delta]\!]_{m_n} p(m_n)} = 1 \tag{10}$$

The denominator cannot be zero because of the assumption of $[\![\Delta]\!] \neq \emptyset$. From Equation (10), $p(\alpha|\Delta) = 1$ iff $[\![\Delta]\!] \subseteq [\![\alpha]\!]$. □

Corollary 5. Since $[\![\Delta]\!] \neq \emptyset$, $MPS(\Delta) = \Delta$. Thus, $(\!(\!(\Delta)\!)\!) = \bigcup_{S \in MPS(\Delta)} [\![S]\!] = [\![\Delta]\!]$. Namely, there is a possible model where all the elements of Δ are true. Therefore, there is a model where all the elements of Δ are true, i.e., $(\!(\Delta)\!) = [\![\Delta]\!]$. Since $(\!(\!(\Delta)\!)\!) = (\!(\Delta)\!)$, we therefore have $(\!(\!(\Delta)\!)\!) = [\![\Delta]\!] = (\!(\Delta)\!) = [\![\Delta]\!]$. From Equation (7), we have

$$p(\alpha|\Delta) = \frac{\sum_n [\![\alpha]\!]_{m_n} [\![\Delta]\!]_{m_n} p(m_n)}{\sum_n [\![\Delta]\!]_{m_n} p(m_n)} = 1. \tag{11}$$

The denominator cannot be zero as $[\![\Delta]\!] = [\![\Delta]\!] \neq \emptyset$. We thus have $p(\alpha|\Delta) = 1$ iff $[\![\Delta]\!] \subseteq [\![\alpha]\!]$. □

Corollary 6. This is obvious from Corollary 1 and Corollary 4. □