# How to confirm “all ravens are black”?

**Xi Li**

Center for Logic, Language and Cognition

Department of Philosophy, Peking University

Beijing, 100871, China

xli@pku.edu.cn

**Abstract**

Carnap’s $\lambda$-continuum fails to confirm Raven Paradox. If Solomonoff prior is introduced to Inductive Logic, then it can confirm the hypothesis “all ravens are black” in any computable world as long as all ravens are really black in that world.

**Keywords**

Inductive Logic; Solomonoff Prior; Raven Paradox

## §1 An Introduction to Inductive Logic

In early 20th century, Keynes tried to assign to inductive generalizations, according to available evidence, probabilities that should converge to $1$ as the generalizations are supported by more and more independent events.

Carnap[1] developed the antecedent of the modern inductive logic, in which he tried to use logic to distinguish alternative states of affairs that can be expressed in a given formal language, then define inductive probabilities for sentences by taking advantage of symmetry assumptions concerning such states of affairs. In a deductively valid argument, every possible world in which the premises are true also makes the conclusion true. In a good inductive argument, the set of worlds in which the premises are true and the conclusion false is sufficiently “small”.

Assume the first order language $\mathscr{L}$ contains countable constants $\mathscr{C}$ and $m$ monadic predicates $\mathscr{R} = \{R_1, R_2, \ldots, R_m\}$ with no function symbols nor equality. The constants $\mathscr{C}$ name all the individuals in some Universe though there is no prior assumption that they necessarily name different individuals.

Notation: $f\restriction_X := \{(x,y) \in f : x \in X\}$, and $f_{-1}(X) := \{x : f(x) \in X\}$.

**Definition 1 (Probability on Sentences).**
*A probability on sentences is a non-negative function* $w\colon \mathscr{S} \to [0,1]$ *such that*

$P_1$. $\vDash \psi \implies w(\psi) = 1$

$P_2$. $\psi_1 \vDash \neg\psi_2 \implies w(\psi_1 \vee \psi_2) = w(\psi_1) + w(\psi_2)$

$P_3$. $w(\exists x \psi(x)) = \lim_{n\to\infty} w\left(\bigvee_{i=1}^{n} \psi(a_i)\right)$

**Theorem 1.** *If* $w$ *satisfies* $P_1, P_2$*, then for* $\phi, \psi \in \mathscr{S}$

*(i)* $w(\neg\phi) = 1 - w(\phi)$

*(ii)* $\vDash \neg\phi \implies w(\phi) = 0$

*(iii) The following are equivalent:*

*(a)* $w(\phi) = 1 \implies \vDash \phi$

*(b)* $w(\phi) = 0 \implies \vDash \neg\phi$

*(iv)* $\phi \vDash \psi \implies w(\phi) \le w(\psi)$

*(v)* $\vDash \phi \leftrightarrow \psi \implies w(\phi) = w(\psi)$

*(vi)* $w(\phi) + w(\psi) = w(\phi \wedge \psi) + w(\phi \vee \psi)$

**Theorem 2 (Extension Theorem).**
*Suppose* $w\colon \mathscr{S}_{QF} \to [0,1]$ *over quantifier-free sentences satisfies* $P_1, P_2$*, then* $w$ *has an unique extension to* $w^{+}\colon \mathscr{S} \to [0,1]$ *satisfying* $P_1, P_2, P_3$*.*

Let $Q_i \equiv \bigwedge_{j=1}^{m} \pm R_j$ for $1 \le i \le 2^m =: r$, where $\pm R$ means one of $\{R, \neg R\}$, then $\mathscr{Q} = \{Q_1, \cdots, Q_r\}$ is a $r$-fold classification system of some Universe with domain $\mathscr{C}$, and every individual in the universe has to satisfy one and only one $Q$-predicate which is determined by the state description function $h\colon a_i \mapsto Q_{h_i}$. The set of state descriptions of $\vec{a} = (a_1, \ldots, a_n)$ is

$$\mathcal{H}_{\vec{a}} := \left\{ \bigwedge_{i=1}^{n} Q_{h_i}(a_i) : h\colon \{1, \ldots, n\} \to \{1, \ldots, r\} \right\}$$

Sometimes we write $n_i := |h\restriction_{\{1,\ldots,n\}}{}_{-1}(i)|$ to denote the number of times that event $Q_i$ occurs in $n$ trials $\bigwedge_{j=1}^{n} Q_{h_j}(a_j)$. Carnap takes $\{n_i : 1 \le i \le r\}$ as the structure description.

Carnap's aim is to find the right $w$.

Carnap believes that the right $w$ should satisfy some symmetry principle. For example, it should be invariant under finite permutations of names.

for any permutation $\sigma$ of $\mathbb{N}^+$,

$$w(\psi(a_1, \dots, a_n)) = w(\psi(a_{\sigma(1)}, \dots, a_{\sigma(n)})) \tag{Ex}$$

for any permutation $\tau$ of $\{1, 2, \dots, r\}$,

$$w\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = w\left(\bigwedge_{i=1}^{n} Q_{\tau(h_i)}(a_i)\right) \tag{Ax}$$

Besides the above symmetry principles there is a stronger postulate—*sufficientness postulate*, which asserts that there exists a series of functions $\{f_i : 1 \le i \le r\}$ such that

$$w\left(Q_j(a_{n+1}) \,\middle|\, \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = f_j(n_j, n) \tag{SP}$$

Principle Ex asserts that $w\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)$ depends only on the vector $\langle n_{h_i} : 1 \le i \le n\rangle$, so that it is independent on the order of observing the individuals, while in the presence of Ex, principle Ax asserts that $w\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)$ depends only on $\{n_i : 1 \le i \le r\}$, and $w(Q_i(a_1)) = 1/r$ for all $1 \le i \le r$.

Considering the indifference principle, there are two intuitive ways to assign prior probability.

(A) All state descriptions have equal weight.

(B) All structure descriptions have equal weight.

Given $n$ individuals, there are $r^n$ possible state descriptions and

$$\binom{n+r-1}{r-1}$$

possible structure descriptions.

According to (A),

$$m^\dagger\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{1}{r^n}$$

and

$$\mathfrak{c}^\dagger\left(Q_j(a_{n+1}) \,\middle|\, \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{m^\dagger\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \wedge Q_j(a_{n+1})\right)}{m^\dagger\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)} = \frac{\frac{1}{r^{n+1}}}{\frac{1}{r^n}} = \frac{1}{r}$$

It is independent of the history $\bigwedge_{i=1}^{n} Q_{h_i}(a_i)$, which means it violates the principle of learning from experience and hence is unacceptable.

According to (B),

$$m^*(n_1, \dots, n_r) = \frac{1}{\binom{n+r-1}{r-1}}$$

Since each structure description can be seen as $\binom{n}{n_1,\ldots,n_r}$ possible state descriptions, and according to principle Ex, every possible state description shares equal portion of its structure description, so we have

$$m^*\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{m^*(n_1,\ldots,n_r)}{\binom{n}{n_1,\ldots,n_r}} = \frac{1}{\binom{n+r-1}{r-1}\binom{n}{n_1,\ldots,n_r}}$$

which depends only on structure description.

Carnap defines his favourite "degree of confirmation" as

$$\mathfrak{c}^*\left(Q_j(a_{n+1}) \,\middle|\, \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{m^*\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \wedge Q_j(a_{n+1})\right)}{m^*\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)} = \frac{n_j+1}{n+r}$$

Carnap's $\lambda$-continuum $\mathfrak{c}_\lambda$ is a generalization of $\mathfrak{c}^*$.

Suppose $(Q_1,\cdots,Q_r)$ are defined so that they have different relative widths $\gamma_i$ such that $\sum_{i=1}^{r}\gamma_i = 1$, Carnap's $\lambda$-continuum is

$$\mathfrak{c}_\lambda\left(Q_j(a_{n+1}) \,\middle|\, \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{n_j+\lambda\gamma_j}{n+\lambda} = \frac{n}{n+\lambda}\frac{n_j}{n} + \frac{\lambda}{n+\lambda}\gamma_j$$

relative to a free parameter $0 < \lambda \leq \infty$ which indicates the weight given to logical or language-dependent factors over and above purely empirical factors (observed frequencies). The parameter $\lambda$ serves as an index of caution in singular inductive inference.

Suppose $\vec{a} = (a_1,\ldots,a_n)$, and the state description of $\vec{a}$ is $\Theta(\vec{a}) \equiv \bigwedge_{i=1}^{n} Q_{h_i}(a_i)$, which can be assigned a degree of confirmation.

$$\begin{aligned}
\mathfrak{c}_\lambda(\Theta(\vec{a})) &= \mathfrak{c}_\lambda\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)\\
&= \prod_{i=0}^{n-1}\mathfrak{c}_\lambda\left(Q_{h_{i+1}}(a_{i+1}) \,\middle|\, \bigwedge_{j=0}^{i} Q_{h_j}(a_j)\right)\\
&= \prod_{i=0}^{n-1}\frac{i_{h_{i+1}}+\lambda\gamma_{h_{i+1}}}{i+\lambda}
\end{aligned}$$

When $\lambda = 0$, let

$$\mathfrak{c}_0(\Theta(\vec{a})) := \begin{cases}\frac{1}{r} & \text{if } h_1 = \ldots = h_n\\ 0 & \text{otherwise}\end{cases} \tag{1.1}$$

If we let $\gamma_i = 1/r$ for all $1 \leq i \leq r$, then

$$\begin{aligned}
\mathfrak{c}_\lambda(\Theta(\vec{a})) &= \mathfrak{c}_\lambda\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)\\
&= \prod_{j=1}^{n}\mathfrak{c}_\lambda\left(Q_{h_j}(a_j) \,\middle|\, \bigwedge_{i=j+1}^{n} Q_{h_i}(a_i)\right)
\end{aligned}$$

$$= \prod_{j=1}^{n} \left( \frac{|h\restriction_{\{j+1,\dots,n\}}{}_{-1}(h_j)| + \dfrac{\lambda}{r}}{n - j + \lambda} \right)$$

$$= \frac{\prod_{i=1}^{r} \prod_{j=0}^{n_i - 1} \left( j + \dfrac{\lambda}{r} \right)}{\prod_{j=0}^{n-1} (j + \lambda)} \tag{1.2}$$

from 1.1 and 1.2 we can see that Carnap's $\lambda$-continuum is invariant under the two symmetry principles Ex and Ax. By adding $\lambda = r$, then $\mathfrak{c}_r = \mathfrak{c}^*$.

Actually, Carnap proved the following theorem.

**Theorem 3.** *Suppose language $\mathscr{L}$ has at least two predicates i.e. $m \geq 2$, then the probability function $w$ on $\mathscr{L}$ satisfies Ex, SP if and only if $w = \mathfrak{c}_\lambda$ for some $0 \leq \lambda \leq \infty$.*
*Namely,*

$$f_i(n_i, n) = \frac{n_i + \lambda\gamma_i}{n + \lambda}$$

*where $\gamma_i = f_i(0,0)$ and $\lambda = \dfrac{f_i(0,1)}{f_i(0,0) - f_i(0,1)}$.*
*By adding Ax, $\forall i : \gamma_i = \dfrac{1}{r}$.*

The symmetry principles Ex and Ax says that the temporal order of events is irrelevant, but in reality, the temporal order is of great significance. So the right 'degree of confirmation' should go against them. If the temporal order is taken into consideration, the subscript of $a$ should represent the time stamp, and the conjunction $\bigwedge_{i=1}^{n} Q_{h_i}(a_i)$ indicates the time series of observations $\langle Q_{h_1}(a_1), \dots, Q_{h_n}(a_n) \rangle$.

In 1963, Putnam[5, 6] criticized Carnap's program. He took Carnap's inductive logic as a design for a 'learning machine', and the task of inductive logic is to construct a 'universal learning machine'. If there is a correct definition of 'degree of confirmation', then a machine which predicted in accordance with the degree of confirmation would be the cleverest possible learning machine. Then he suggested that sort of 'degree of confirmation' can be defined according to some 'simplicity order', and either there are better and better 'degree of confirmation' functions, but no 'best possible', or else there is a 'best possible' but it is not computable by a machine.

But, what is the correct language? What is the correct 'degree of confirmation'? How many evidences are strong enough to hold our belief?

By choosing the smallest model class that contains the true environment and the universal (mixture) prior beliefs of the environments that reflect the simplicity criterion, Solomonoff [8] solved the problem.

# §2 Solomonoff Prior in Inductive Logic

We assume the reader is familiar with the basics of Kolmogorov Complexity. Preliminaries can be found in Li and Vitányi[3] or Hutter[2].

### Solonomoff Prior

Every specific state description function $h$ determines an unique state description—an unique universe, so any program $p$ generates an universe. We identify $h_{1:n}$ with history $\bigwedge_{i=1}^{n} Q_{h_i}(a_i)$ and identify $p$ with the universe $\bigwedge_{i=1}^{\infty} Q_{h_i}(a_i)$ if $U(p) = h_{1:\infty}$ without confusion.

Our beliefs and hence probabilities are a result of our personal history. To be able to update beliefs consistently we must first decide on the set of all explanations that may be possible. In order to find the true governing process behind our entire reality we consider all possible universes in a certain sense. The actual universe is just one of a large number of possible universes. Each universe is in one of possible states; the probability assigned to each state is then the proportion of the possible universes in which that state is attained. Each new measurement eliminates some fraction of the universes in a given state, depending on how likely or unlikely that state was to actually produce that measurement; the surviving universes then have a new posterior probability distribution, which is related to the prior distribution by Bayes' formula.

**Definition 2 (Universal Probability).**

$$\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \sum_{p:U(p)=h_{1:n}*} 2^{-|p|}$$

*Where $U$ is a universal monotone Turing machine.*

It can be regarded as the limit of the relative frequency of the consistent possible worlds over all possible worlds:

$$\begin{aligned}
\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) &= \sum_{p} 2^{-|p|} [\![U(p) = h_{1:n}*]\!] \\
&= \lim_{n\to\infty} \frac{\sum_{p:|p|\le n} 2^{n-|p|} \left[\!\!\left[\bigwedge_{i=1}^{n} Q_{\langle U(p)\rangle_i}(a_i) \equiv \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right]\!\!\right]}{2^n} \\
&\approx \lim_{n\to\infty} \frac{\left|\left\{p : |p| = n \;\&\; \bigwedge_{i=1}^{n} Q_{\langle U(p)\rangle_i}(a_i) \equiv \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right\}\right|}{2^n}
\end{aligned}$$

It means that $\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)$ is the frequentist probability that the program of a universal monotone Turing machine $U$ generates $\bigwedge_{i=1}^{n} Q_{h_i}(a_i)$ when provided with uniform random noise (fair coin flips) on the input tape.

$$\textbf{probability} = \frac{|\text{consistent universes}|}{|\text{all possible universes}|}$$

The benevolence of God is represented in the way he plays dice. God does not play dice directly with us, but plays dice indirectly through some Universal Turing machine to offer us the freedom to realize any possible regular world. And God's dice is absolutely fair, which means God never play tricks. God offers us the fullest freedom to chose the most perfect world, that is to say, the one which is at the same time the simplest in hypothesis and the richest in phenomena.

**Lemma 1.** *For every $\nu \in \mathcal{M}_U$ there exists some monotone Turing machine $T$ such that*

$$\nu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \sum_{p:T(p)=h_{1:n}*} 2^{-|p|} \quad \textit{and} \quad K(\nu) \overset{+}{=} |\langle T\rangle|$$

*where $T(p) = U(\langle T\rangle p)$.*

**Lemma 2.** *For $\nu \in \mathcal{M}_U$,*

$$\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \overset{+}{\geq} 2^{-K(\nu)} \nu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)$$

*Proof.*

$$\begin{aligned}
\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) &= \sum_{p:U(p)=h_{1:n}*} 2^{-|p|} \\
&\geq \sum_{q:U(\langle T\rangle q)=h_{1:n}*} 2^{-|\langle T\rangle q|} \\
&= 2^{-|\langle T\rangle|} \sum_{q:T(q)=h_{1:n}*} 2^{-|q|} \\
&\overset{\times}{=} 2^{-K(\nu)} \nu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) && \text{[Lemma 1]}
\end{aligned}$$

□

Define

$$\mathfrak{c}'_M(\top) := 1$$

$$\begin{aligned}
\mathfrak{c}'_M\left(\bigwedge_{i=1}^{t} Q_{h_i}(a_i)\right) &:= \mathfrak{c}'_M\left(\bigwedge_{i=1}^{t-1} Q_{h_i}(a_i)\right) \frac{\mathfrak{c}_M\left(\bigwedge_{i=1}^{t} Q_{h_i}(a_i)\right)}{\sum_{1\leq k\leq r} \mathfrak{c}_M\left(\bigwedge_{i=1}^{t-1} Q_{h_i}(a_i) \wedge Q_k(a_t)\right)} \\
&= \frac{\mathfrak{c}_M\left(\bigwedge_{i=1}^{t} Q_{h_i}(a_i)\right)}{\mathfrak{c}_M(\top)} \prod_{i=1}^{t} \frac{\mathfrak{c}_M\left(\bigwedge_{j=1}^{i-1} Q_{h_j}(a_j)\right)}{\sum_{1\leq k\leq r} \mathfrak{c}_M\left(\bigwedge_{j=1}^{i-1} Q_{h_j}(a_j) \wedge Q_k(a_i)\right)}
\end{aligned}$$

Obviously, for any state description $\Theta, \Theta'$,

(i) $\mathfrak{c}'_M\left(\Theta(a_1,\ldots,a_n)\right) \geq 0$

(ii) $\mathfrak{c}'_M(\top) = 1$

(iii) $\mathfrak{c}'_M\left(\Theta(a_1,\ldots,a_n)\right) = \sum_{\Theta'(a_1,\ldots,a_{n+1}) \vDash \Theta(a_1,\ldots,a_n)} \mathfrak{c}'_M\left(\Theta'(a_1,\ldots,a_{n+1})\right)$

For any quantifier-free sentence $\psi(\vec{a})$, let

$$\mathfrak{c}'_M\left(\psi(\vec{a})\right) := \sum_{\Theta(\vec{b}) \vDash \psi(\vec{a})} \mathfrak{c}'_M\left(\Theta(\vec{b})\right)$$

where $\vec{b}$ is sufficiently large that all of the $\vec{a}$ are amongst $\vec{b}$, and $\bigvee_{\Theta(\vec{b}) \vDash \psi(\vec{a})} \Theta(\vec{b})$ is the *full disjunctive normal form* of $\psi(\vec{a})$.

$$\psi(\vec{a}) \equiv \bigvee_{\Theta(\vec{b}) \vDash \psi(\vec{a})} \Theta(\vec{b}) \tag{DNF}$$

It is easy to see, $\mathfrak{c}'_M$ satisfies $P_1, P_2$, and according to Theorem2, $\mathfrak{c}'_M$ has an unique extension over all of the sentences $\mathscr{S}$ of $\mathscr{L}$. Then $\mathfrak{c}'_M$ induces a confirmation function

$$c'_M(\psi_H|\psi_E) = \frac{\mathfrak{c}'_M(\psi_E \wedge \psi_H)}{\mathfrak{c}'_M(\psi_E)}$$

In fact, any $w$ satisfying (i),(ii),(iii) can extend to a probability function on $\mathscr{L}$.

**Theorem 4** (**Completeness Theorem**)**.**
*If the universe is generated by a computable stochastic process $\mu$, $\mathfrak{c}'_M$ predicts well with few errors.*

$$\begin{aligned}
&\sum_{l=1}^{\infty} \sum_{h_{1:l} \in \{1,\ldots,r\}^l} \mu\left(\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right) \cdot \\
&\left(\sum_{h_{l+1:t} \in \{1,\ldots,r\}^{t-l}} \left|\mathfrak{c}'_M\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i) \,\middle|\, \bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right) - \mu\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i) \,\middle|\, \bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)\right|\right)^2 \\
&\leq 2(t-l) D_{1:\infty}(\mu \| \mathfrak{c}_M) \\
&\overset{+}{\leq} 2(t-l) K(\mu) \ln 2 \\
&< \infty
\end{aligned}$$

*where*

$$D_{1:n}(\mu\|\rho) := \sum_{h_{1:n} \in \{1,\ldots,r\}^n} \mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \ln \frac{\mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)}{\rho\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)}$$

*Proof.*

$$
\begin{aligned}
&\sum_{l=1}^{\infty} \sum_{h_{1:l}\in\{1,\dots,r\}^l} \mu\left(\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)\cdot \\
&\left(\sum_{h_{l+1:t}\in\{1,\dots,r\}^{t-l}} \left| \mathfrak{c}'_M\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i) \,\middle|\, \bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right) - \mu\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i) \,\middle|\, \bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)\right|\right)^2 \\
&\overset{(a)}{\le} 2\sum_{l=1}^{\infty} \sum_{h_{1:l}\in\{1,\dots,r\}^l} \mu\left(\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)\cdot \\
&\qquad \sum_{h_{l+1:t}\in\{1,\dots,r\}^{t-l}} \mu\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i) \,\middle|\, \bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right) \ln \frac{\mu\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i) \,\middle|\, \bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i) \,\middle|\, \bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)} \\
&= 2\sum_{l=1}^{\infty} \sum_{h_{1:t}\in\{1,\dots,r\}^l} \mu\left(\bigwedge_{i=1}^{t} Q_{h_i}(a_i)\right) \sum_{m=l}^{t-1} \ln \frac{\mu\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)} \\
&= 2\sum_{l=1}^{\infty} \sum_{m=l}^{t-1} \sum_{h_{1:m+1}\in\{1,\dots,r\}^{m+1}} \mu\left(\bigwedge_{i=1}^{m+1} Q_{h_i}(a_i)\right)\cdot \\
&\qquad \left(\sum_{h_{m+2:t}\in\{1,\dots,r\}^{t-m-1}} \mu\left(\bigwedge_{i=m+2}^{t} Q_{h_i}(a_i) \,\middle|\, \bigwedge_{i=1}^{m+1} Q_{h_i}(a_i)\right)\right) \ln \frac{\mu\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)} \\
&\le 2(t-l) \sum_{m=1}^{\infty} \sum_{h_{1:m+1}\in\{1,\dots,r\}^{m+1}} \mu\left(\bigwedge_{i=1}^{m+1} Q_{h_i}(a_i)\right) \ln \frac{\mu\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)} \\
&= 2(t-l) \lim_{n\to\infty} \sum_{m=1}^{n} \sum_{h_{1:m+1}\in\{1,\dots,r\}^{m+1}} \left(\sum_{h_{m+2:n}\in\{1,\dots,r\}^{n-m-1}} \mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)\right) \ln \frac{\mu\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)} \\
&= 2(t-l) \lim_{n\to\infty} \sum_{m=1}^{n} \sum_{h_{1:n}\in\{1,\dots,r\}^{n}} \mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \ln \frac{\mu\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)} \\
&= 2(t-l) \lim_{n\to\infty} \sum_{h_{1:n}\in\{1,\dots,r\}^{n}} \mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \sum_{m=1}^{n} \ln \frac{\mu\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}
\end{aligned}
$$

$$= 2(t-l)\lim_{n\to\infty}\sum_{h_{1:n}\in\{1,\dots,r\}^n}\mu\left(\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right)\ln\prod_{m=1}^{n}\frac{\mu\left(Q_{h_{m+1}}(a_{m+1})\,\middle|\,\bigwedge_{i=1}^{m}Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1})\,\middle|\,\bigwedge_{i=1}^{m}Q_{h_i}(a_i)\right)}$$

$$= 2(t-l)\lim_{n\to\infty}\sum_{h_{1:n}\in\{1,\dots,r\}^n}\mu\left(\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right)\ln\frac{\mu\left(\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right)}$$

$$\leq 2(t-l)\lim_{n\to\infty}\sum_{h_{1:n}\in\{1,\dots,r\}^n}\mu\left(\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right)\ln\frac{\mu\left(\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right)}{\mathfrak{c}_M\left(\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right)}$$

$$\overset{+}{\leq} 2(t-l)K(\mu)\ln 2$$

$$< \infty$$

where $\overset{(a)}{\leq}$ follows from Entropy Inequality and the last inequality $\overset{+}{\leq}$ follows from Lemma2. □

Compare Carnap's $\lambda$-continuum $\mathfrak{c}_\lambda$ with $\mathfrak{c}'_M$. With zero-knowledge ($n = 0$), Carnap would use

$$\mathfrak{c}_\lambda\left(Q_j(a_1)\right) = \frac{0+\lambda\gamma_j}{0+\lambda} = \gamma_j$$

to estimate the future, while Solomonoff would prefer

$$\mathfrak{c}'_M\left(Q_{h_1}(a_1)\right) = \frac{\mathfrak{c}_M\left(Q_{h_1}(a_1)\right)}{\sum_{1\leq j\leq r}\mathfrak{c}_M\left(Q_j(a_1)\right)}$$

with sufficient experiences ($n$ large enough), Carnap would use

$$\mathfrak{c}_\lambda\left(Q_j(a_{n+1})\,\middle|\,\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right) = \frac{n_j+\lambda\gamma_j}{n+\lambda} \approx \frac{n_j}{n}$$

—the frequency of the phenomena, while Solomonoff would still use

$$\mathfrak{c}'_M\left(Q_j(a_{n+1})\,\middle|\,\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right)$$

—the normalization of the frequency of the consistent universes/causes. In other words, Carnap would like to *know how* while Solomonoff would like to *know why*. Besides, Carnap's $\lambda$-continuum $\mathfrak{c}_\lambda$ fails to confirm "all ravens are black" while $\mathfrak{c}'_M$ is qualified as a solution.

## §3 How to Confirm "All Ravens are Black"?

The Raven Paradox induces two problems: (i) Can a non-black non-raven justify "all ravens are black"? (ii) How many witnesses can confirm the fact that "all ravens are black"? We only take care of problem (ii) in this paper.

"All ravens are black" can be expressed by $\forall x(R(x) \to B(x))$.

Since

$$\begin{aligned} w(\forall x\psi(x)) &= 1 - w(\exists x\neg\psi(x)) \\ &= 1 - \lim_{n\to\infty} w\left(\bigvee_{i=1}^{n} \neg\psi(a_i)\right) \\ &= \lim_{n\to\infty}\left(1 - w\left(\bigvee_{i=1}^{n} \neg\psi(a_i)\right)\right) \\ &= \lim_{n\to\infty} w\left(\bigwedge_{i=1}^{n} \psi(a_i)\right) \end{aligned}$$

To solve the Raven Paradox, we only have to make sure that $\lim_{n\to\infty} w\left(\bigwedge_{i=1}^{n}(R(a_i) \to B(a_i))\right) > 0$.

$$\begin{aligned} &\mathfrak{c}'_M(\forall x(R(x) \to B(x)) \\ &= \lim_{n\to\infty} \mathfrak{c}'_M\left(\bigwedge_{i=1}^{n}(R(a_i) \to B(a_i))\right) \\ &= \lim_{n\to\infty} \mathfrak{c}'_M\left(\bigwedge_{i=1}^{n}(\neg R(a_i) \vee B(a_i))\right) \\ &= \lim_{n\to\infty} \mathfrak{c}'_M\left(\bigvee_{j=1}^{2^n}\left(\bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j\right) \\ &= \lim_{n\to\infty} \mathfrak{c}'_M\left(\bigvee_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \models \bigvee_{j=1}^{2^n}\left(\bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j} \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \\ &= \lim_{n\to\infty} \sum_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \models \bigvee_{j=1}^{2^n}\left(\bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j} \mathfrak{c}'_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \\ &= \lim_{n\to\infty} \sum_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \models \bigvee_{j=1}^{2^n}\left(\bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j} \frac{\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)}{\mathfrak{c}_M(\top)} \prod_{i=1}^{n} \frac{\mathfrak{c}_M\left(\bigwedge_{j=1}^{i-1} Q_{h_j}(a_j)\right)}{\sum_{1\le k\le r} \mathfrak{c}_M\left(\bigwedge_{j=1}^{i-1} Q_{h_j}(a_j) \wedge Q_k(a_i)\right)} \\ &= \lim_{n\to\infty} \sum_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \models \bigvee_{j=1}^{2^n}\left(\bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j} \frac{\sum_{p:U(p)=h_{1:n}*} 2^{-|p|}}{\sum_{p\in dom(U)} 2^{-|p|}} \prod_{i=1}^{n} \frac{\sum_{p:U(p)=h_{<i}*} 2^{-|p|}}{\sum_{1\le k\le r} \sum_{p:U(p)=h_{<i}k*} 2^{-|p|}} \end{aligned}$$

where $\psi_1/\psi_2$ means that we mutual exclusively choose either $\psi_1$ or $\psi_2$.

Since

$$\forall i : \frac{\sum_{p:U(p)=h_{<i}*} 2^{-|p|}}{\sum_{1\le k\le r} \sum_{p:U(p)=h_{<i}k*} 2^{-|p|}} \ge 1$$

Hence $\mathfrak{c}'_M\left(\forall x(R(x) \to B(x))\right) > 0$ if there exists some computable universe $h_{1:\infty}$ such that

$$\forall n \left( \bigwedge_{i=1}^{n} Q_{h_i}(a_i) \vDash \bigwedge_{i=1}^{n} (\neg R(a_i)/B(a_i)) \right)$$

In other words, if all of the ravens in a computable universe are black, $\mathfrak{c}'_M$ can confirm that "all ravens are black".

$$\begin{aligned}
&\lim_{n\to\infty} \mathfrak{c}'_M\left(\forall x(R(x) \to B(x)) \,\middle|\, \bigwedge_{i=1}^{n} (\neg R(a_i) \vee B(a_i))\right) \\
&= \lim_{n\to\infty} \frac{\mathfrak{c}'_M\left(\forall x(R(x) \to B(x))\right)}{\mathfrak{c}'_M\left(\bigwedge_{i=1}^{n} (R(a_i) \to B(a_i))\right)} \\
&= \frac{\mathfrak{c}'_M\left(\forall x(R(x) \to B(x))\right)}{\lim_{n\to\infty} \mathfrak{c}'_M\left(\bigwedge_{i=1}^{n} (R(a_i) \to B(a_i))\right)} \\
&= \frac{\mathfrak{c}'_M\left(\forall x(R(x) \to B(x))\right)}{\mathfrak{c}'_M\left(\forall x(R(x) \to B(x))\right)} \\
&= 1
\end{aligned}$$

For Carnap's $\lambda$-continuum $\mathfrak{c}_\lambda$,

$$\begin{aligned}
&\mathfrak{c}_\lambda\left(\forall x(R(x) \to B(x))\right) \\
&= \lim_{n\to\infty} \sum_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \vDash \bigvee_{j=1}^{2^n} \left(\bigwedge_{i=1}^{n} (\neg R(a_i)/B(a_i))\right)_j} \mathfrak{c}_\lambda\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \\
&= \lim_{n\to\infty} \sum_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \vDash \bigvee_{j=1}^{2^n} \left(\bigwedge_{i=1}^{n} (\neg R(a_i)/B(a_i))\right)_j} \prod_{i=0}^{n-1} \mathfrak{c}_\lambda\left(Q_{h_{i+1}}(a_{i+1}) \,\middle|\, \bigwedge_{j=1}^{i} Q_{h_j}(a_j)\right) \\
&= \lim_{n\to\infty} \sum_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \vDash \bigvee_{j=1}^{2^n} \left(\bigwedge_{i=1}^{n} (\neg R(a_i)/B(a_i))\right)_j} \prod_{i=0}^{n-1} \frac{i_{h_{i+1}} + \lambda\gamma_{h_{i+1}}}{i + \lambda} \\
&\le \lim_{n\to\infty} \prod_{i=0}^{n-1} \frac{i + \lambda(1 - \min_{1\le t\le r} \gamma_t)}{i + \lambda} \\
&= 0
\end{aligned}$$

The last step follows from

$$\prod_{n\ge1} a_n = 0 \iff \sum_{n\ge1} (1 - a_n) = \infty \quad \text{for} \quad \forall n: \ 0 < a_n \le 1$$

The reason that $\mathfrak{c}_\lambda$ fails to confirm universal generalization is that the speed of convergence is too slow.

$$\sum_{i=0}^{\infty} \frac{\lambda \cdot \min_{1\le t\le r} \gamma_t}{i + \lambda} = \infty$$

If we use

$$\mathfrak{c}'_\lambda\left(Q_j(a_{n+1})\,\middle|\,\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) := \frac{\dfrac{n_j^2+\lambda\gamma_j}{n^2+\lambda}}{\sum\limits_{1\le k\le r}\dfrac{n_k^2+\lambda\gamma_k}{n^2+\lambda}} = \frac{n_j^2+\lambda\gamma_j}{\sum\limits_{1\le k\le r} n_k^2+\lambda}$$

rather than $\mathfrak{c}_\lambda$, then the convergence is guaranteed. And $\mathfrak{c}'_\lambda$ agrees with the principle Ex but violates the postulate SP and the frequency interpretation.

In summary, $\mathfrak{c}'_M$ can confirm "all ravens are black" while $\mathfrak{c}_\lambda$ can not. Although $\mathfrak{c}'_M$ can confirm it yet the temporal order should not be neglected.

# References


[1] R. Carnap. Notes on Probability and Induction. *Synthese*, vol.25, no.3/4, pp. 269-298, 1973.

[2] M. Hutter. Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability. *Springer*, 2005.

[3] M. Li and P. M. B. Vitányi. An Introduction to Kolmogorov Complexity and its Applications. *Springer, Berlin, 3rd edition*, 2008.

[4] J. Paris and A. Vencovska. Pure inductive logic. *Cambridge University Press*. 2014.

[5] H. Putnam. 'Degree of confirmation' and inductive logic. *in P. A. Schilpp (ed.), The Philosophy of Rudolf Carnap (La Salle, 111., The Open Court Publishing Co., 1963).*

[6] H. Putnam. Probability and confirmation. *The Voice of America, Forum Philosophy of Science,10*, U.S. Information Agency, 1963.

[7] S. Rathmanner and M. Hutter. A philosophical treatise of universal induction. *Entropy*, 13(6):1076–1136, 2011.

[8] R. J. Solomonoff. Complexity-based induction systems: Comparisons and convergence theorems. *IEEE Transactions on Information Theory*, IT-24:422–432, 1978.