RESEARCH ARTICLE

Some Afterthoughts and Examples on the Logic of Induction

Uwe Saint-Mont[1]

1 Fachbereich Wirtschafts- und Sozialwissenschaften, Nordhausen University of Applied Sciences, Nordhausen, Germany

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

The aim of Saint-Mont (2020) was to provide a rather general answer to Hume's problem. To this end, induction was treated within a straightforward formal paradigm, i.e., several connected tiers of abstraction.

This note mainly discusses the basic two-tier model and associated examples. It points out that a sound logical solution to the problem of induction boils down to proper accounting, and Bayes' formula in particular. Generalizations are also possible if one evokes suitable assumptions such as some kind of continuity property.

Moreover, the precision added by formal analysis relative to verbal arguments demonstrates that the crucial issue with induction is rather not circularity, but being able to find a suitable general tier that allows for a smooth transition between a less and a more specific situation, avoiding inconsistencies.

Corresponding author: Uwe Saint-Mont, saint-mont@hs-nordhausen.de

1. Introduction

Considering Hume's problem, the basic model introduced in Saint-Mont (2020) consists of two levels of abstraction, with the more general tier on top containing more information than the more specific tier at the bottom:

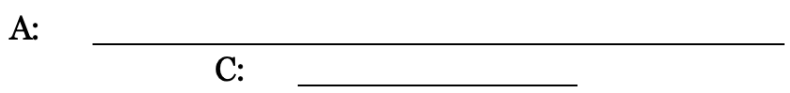


Fig. 1. Basic model with two tiers. *A* (more general), and *C* (less general).

Moving downwards, deduction skips some of the information. Moving upwards, induction leaps from 'less to more'. Given several tiers, the latter article proceeds to figures 6 and 7, i.e., hierarchical and circular structures, in particular

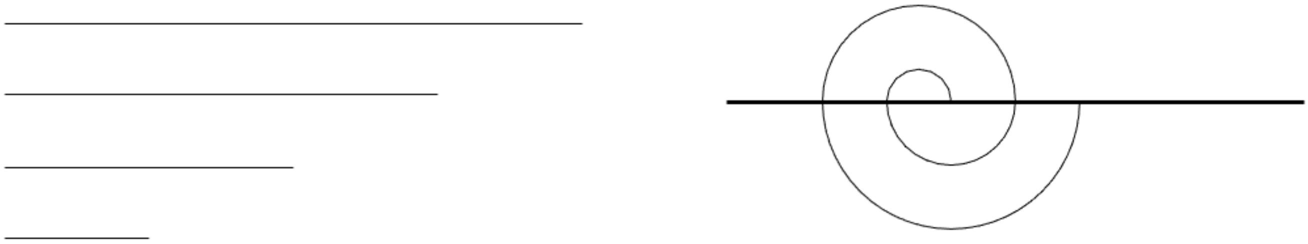
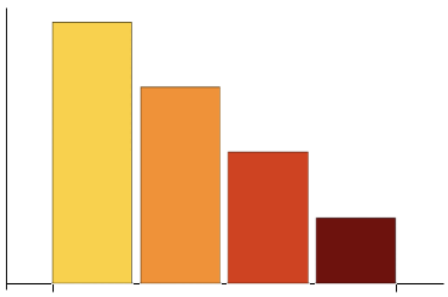
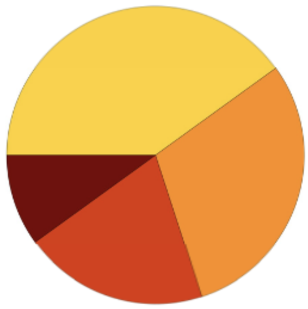


Fig. 2. A hierarchical and a partially recursive structure

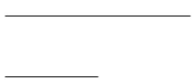
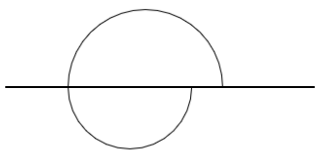
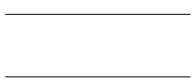
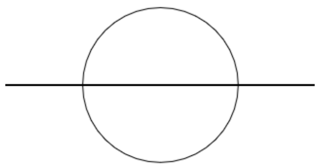
2. A fourfold table

The crucial observation now is that in the above representations curvature is not associated with any kind of meaning. Therefore, the models in Fig. 2 are in fact equivalent – just as a bar chart and a pie chart visualizing the same numbers.





What is relevant, however, is the size $d(A, C)$ of an inductive step, i.e., the distance or difference between two successive tiers. Considering just two tiers, we obtain the following table:

1.	$d > 0$ $\Downarrow$	hierarchy 	spiral 
2.	$d = 0$ $\Updownarrow$	two lines of equal length 	circle 

The first line of the table says that if $d(A, C) > 0$, there are two lines of different length, which are equivalent to a spiral 'in curved terms'. Moreover, the arrow $\Downarrow$ indicates the step from 'more to less', i.e., the implication $A \Rightarrow C$.

The second line of the table considers the (degenerate) case that $d(A, C) = 0$. Thus the two lines have exactly the same length, meaning that A and C are equivalent ($A \Leftrightarrow C$). Moreover, 'in curved terms' the spiral becomes a circle.

Although rather elementary, the last table explains why Hume's idea that any inductive argument in favour of induction is supposed to be viciously circular is wrong. The problem of induction - being the inverse of deduction - is located on the first line. Worthless circular reasoning, however, is located on the second line.

Unfortunately, imprecise verbal reasoning may easily blur this fundamental distinction. For instance, the Stanford Encyclopedia of Philosophy writes on 'The Problem of Induction' (Nov. 22, 2022) that "an argument for a principle may not presuppose the same principle (Non-circularity)". However, it is not clear if the latter sentence refers to the first or the second line: 'non-circularity' points toward an implication, whereas 'the same' rather indicates an equivalence.

If an argument is tautological, we are discussing an equivalence, and thus we find ourselves in the second line of the last table. If, on the other hand, we consider the problem of induction, where necessarily $d(A, C) > 0$, we are talking about the first line. Therefore, for a mathematician, the problem of induction typically boils down to giving some condition B, such that C and B imply A.

For instance, every square S is a rectangle R, thus $S \Rightarrow R$. Given a rectangle, the latter geometrical shape must have at least three sides of equal length in order to qualify as a square. Equivalently, as pointed out in the second footnote of Saint-Mont (2020), the conditions that define a square are stronger (e.g., more numerous) than those describing a rectangle.

Given this perspective, all is well and good: A (closed) perfectly recursive circle is a nice model for an equivalence, and an (open), only partially recursive spiral is an equally fitting model for a hierarchical relation such as an implication. However, mixing up the lines or the four logical possibilities of the last table causes havoc, since identifying an inductive step with a perfectly circular argument means blending the logical functions $\Rightarrow$ and $\Leftrightarrow$. Solving the problem of induction thus becomes a hopeless endeavour, and in doing so, Hume established a conundrum that has confused generations of philosophers.

3. A (de)finite answer

The upshot of the 'new riddle of induction' consists in considering rather arbitrary generalizations. Therefore, in a nutshell, Goodman (1983) refers to the case $d(A, C) = \infty$. Given our point of view, he is correct in mistrusting such 'unbounded' inductive steps – e.g., the paradigmatic (notorious?) emerald changing its colour at any moment in time.

Thus we should consider the finite case $0 < d < \infty$ in more detail. The crucial insight essentially goes back to R.T. Cox (1946) and E.T. Jaynes (2003), who showed that any kind of inductive logic, proceeding from the particular to the general, and which therefore has to consider degrees of certainty, is equivalent to probability theory. More

precisely, any 'reasonable' extension of classical logic leads to the axioms of probability theory (see Greenland (1998), Clayton and Waddington (2017), and Saint-Mont (2011), chap. 4.4, for more details). As a corollary, probability theory should give a general formal solution to Hume's problem.

To this end, suppose there is some current state of knowledge $I(C)$ plus additional information $I(B)$, both finite, at your fingertips. Then the total information you should have is $I(C) + I(B|C) = I(B, C)$. The latter equation determines the unique and thus 'correct distance' between the more informative layer A, consisting of B and C, and the less informative layer consisting of C only, i.e., $d(A, C) = I(B, C) - I(C) = I(B|C)$.

With the usual definition of information $I(p) = \log(1/p)$, exponentiation gives an equivalent expression in terms of probabilities, see equation (2) in Saint-Mont (2020),

$$I(C) + I(B\,|\,C) = I(B, C) \Leftrightarrow p(C) \cdot p(B\,|\,C) = p(B, C).$$

Actually, the last equation is just a formal way of saying that it would be unreasonable if information – or equivalently, (probability) mass – either appeared out of nowhere or vanished into thin air. Rather, honest book-keeping must be sound (coherent) and not allow for such pathologies.

Equation (1) remains true if one takes expectations, which yields the chain rule of entropy, $H(B, C) = H(C) + H(B|C)$, see Cover and Thomas (2006), pp. 17f, 22. Moreover, notice that the intersection $A = B \cap C$ is smaller than the sets it consists of. However, its corresponding amount of information $I(A) = -\log(p(A))$ is larger, and the above figures and tables should be interpreted in that way.

Extending (1) to the standard Bayesian framework of statistics, C corresponds to the prior distribution, and A to the posterior distribution of some parameter θ. Here, sound book-keeping boils down to Bayes' theorem, i.e., the posterior can be computed with the help of the prior and the likelihood function of the data B. The information about θ may be measured with the reciprocal of the variance of the former distributions and is called 'precision'. (Thus, high precision is tantamount to small variance, which means that much is known about the parameter in question.) The success of Bayesian statistics is due to the fact that the precision typically improves if the number of observations increases. For details, see almost every book on Bayesian statistics, for instance Johnson et al. (2022), in particular the chapter on 'conjugate families'.

Coherent book-keeping is also the essence of the third axiom of probability theory, $p(B \cup C) = p(B) + p(C)$, where B and C are mutually exclusive events. Therefore, it is the foundation of equation (1), and it is crucial in avoiding so-called 'Dutch book arguments' (Greenland 1998).

In sum, R.A. Fisher (1966), p. 4, was absolutely right when he concluded:

> *We may at once admit that any inference from the particular to the general must be attended with some degree of uncertainty, but this is not the same as to admit that such inference cannot be absolutely rigorous, for the nature and degree of the uncertainty may itself be capable of rigorous expression.*
>
> *In the theory of probability.. we have the classic example proving this possibility.. The mere fact that inductive inferences are uncertain cannot, therefore, be accepted as precluding perfectly rigorous and unequivocal inference.*

4. Complexity, Prediction, and Bayes

Probability theory, combined with the crucial concept of information, gives a logically sound answer to Hume's problem. Complexity is a closely related concept that allows for a certain extension of the classical train of thought:

If a sequence $xy... = (x_1,..., x_m, y_1,..., y_n,...)$ that starts with the string xy has some kind of regularity to it, a non-trivial educated guess about y given x should be(come) possible. In other words, the problem of inductive inference translates into predicting y upon knowing x.

In general, given a stochastic process $X_1, X_2,...$, let $\pi(X_1 = x_1,..., X_m = x_m) = \pi(x)$ be the (prior) probability that some sequence starting with x occurs. Then the theory of probability advises us to calculate the conditional probability $\pi(y|x)$ that y occurs given x, which is given by Bayes' formula $\pi(y|x) = \pi(xy)/\pi(x)$, or, in other words, to apply equation (1).

Although correct, this solution has a major flaw: in most cases, $\pi(\cdot)$ is unknown or even unknowable (Li and Vitányi (2008), pp. 349ff). Therefore, the question arises whether it is possible – in general – to approximate $\pi(\cdot)$ in a satisfactory way.

Given certain conditions, the answer is 'yes': First, suppose that there is just a finite alphabet $x_i \in \{a, b,..., z\}$, which w.l.o.g. may be reduced to just two symbols, 0 and 1, say. Second, it is very helpful to evoke the idea of algorithmic complexity $Km(x)$, which is small if there is much structure in the data, and large if there is not. That is, a short computer program is able to produce data with much regularities, whereas a rather long program is needed to create a complicated string of data - given that both are executed on an appropriate 'Turing machine' (see Li and Vitányi (2008), chapters 4, 5, in particular section 5.2). The argument then runs as follows:[1]

i. A reference monotone Turing machine U is supplied with fair coin tosses. Then $M(x)$ is the so-called 'universal probability' that U's output starts with x (ibid., pp. 298, 302, 304).
ii. 'The conditional probability $M(y|x)$ suffices to approximate $\pi(y|x)$' (p. 360). More precisely (p. 358, emphasis in the original):

> *We can view M as a mixture of hypotheses... with greater weights for the simpler ones... Solomonoff's inductive formula M (y|x), to estimate the actual probabilities π(y|x) to predict outcomes y given a sequence of observed outcomes x, can be viewed as a mathematical form of Occam's razor: find all rules fitting the data and then predict y according to the universal distribution on them.*

iii. The upshot is that 'the problem with Bayes's rule has always been the determination of the prior. Using M universally gets rid of that problem and is provably perfect' (pp. 360f). That is, evoking appropriate versions of equation (1), Li and Vitányi's theorem 5.2.3., pp. 361f, demonstrates that prediction works on a set of π-measure one (and so, in this sense, perfectly).

Couching the solution in terms of complexity, $Km(x)$ is the length of the shortest program that if fed to U, produces an output that starts with x. Since $Km(x)$ is monotone, it turns out that it is a good idea to 'predict by data compression' (p. 362), which means to choose an extrapolation y^* that minimizes $D(y) = Km(xy) - Km(x)$.

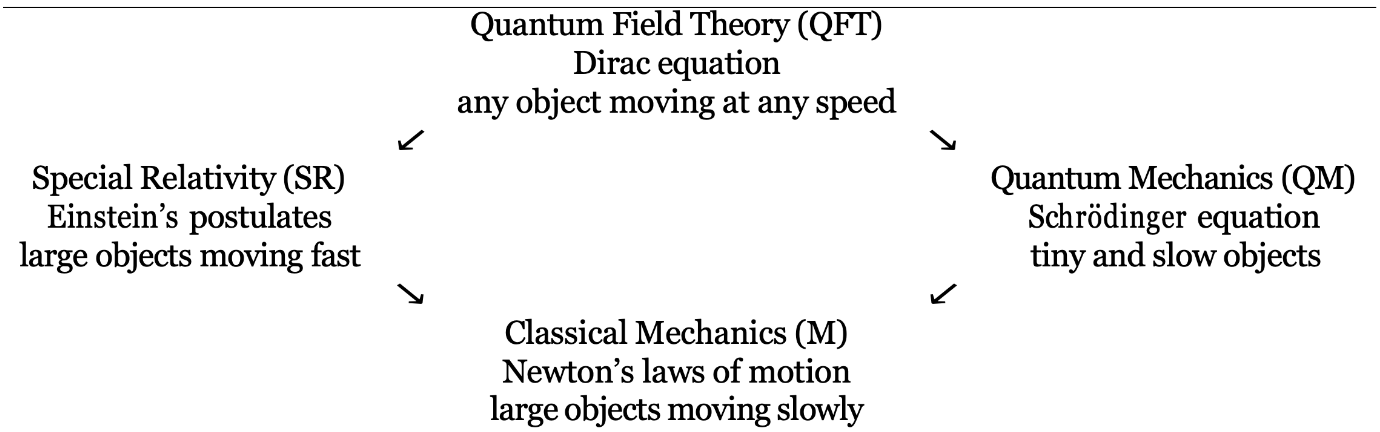
In plain English, $Km(x)$ represents the structure in x, and xy is at least as complex. If we think of y as a hypothesis or a prediction (what is going to happen in the future, given x), Occam's razor chooses a y^* that best fits the given structure. In other words, y^* should minimize the 'deviation' from x in the sense that the increase in complexity $D(y)$ is as small as possible.

In sum, there is a 'formal general theory for prediction' due to Solomonoff (1964), which essentially is an extended version of Bayes' rule. Hutter (2007), his emphasis, explains:

> *A Bayesian considers a set of environments = hypotheses = models M which includes the true data-generating probability distribution π. Solomonoff's rigorous, essentially unique, formal, and universal solution to this problem is to consider a single large universal class M_U suitable for all induction problems. The corresponding universal prior w^U is biased towards simple environments in such a way that it dominates.. all other priors. This leads to an a priori probability M (x) which is equivalent to the probability that a universal Turing machine with random input tape outputs x, and the shortest program computing x produces the most likely continuation (prediction) of x.*

5. Extending Physical Laws

The evolution of our most basic physical theories may also be understood in the above framework. To this end, physical laws may be interpreted as mathematical formulas that apply to a rather restricted area or in a more general setting. Following Zee (2023), p. 4, and placing - as usual - scenarios that are more general further up, we obtain



Zee (ibid., p. 142, my emphasis) writes:

> *Established theories in physics... are not overthrown so much as **extended and generalized**. The Dirac equation did not overthrow Schrödinger's equation any more than Schrödinger or Einstein overthrew Newton. It is almost the opposite: that the Dirac equation has to **reduce** to Schrödinger's equation **under the appropriate circumstances** imposes a powerful constraint on what it could possibly be.*

In other words, QFT contains more information than any other theory mentioned in the above diagram; it is the most general and powerful. Thus, given particular circumstances, more specific laws follow. The historical development, however, was just the other way around. Einstein's SR extended M to fast-moving objects, and QM generalized M to tiny objects. In a process that Post (1971) named 'conservative induction' and exemplified for SR, the concrete received theory was cautiously modified to work in a more general setting. This is in stark contrast to Popper's idea that hypotheses are free inventions of the human mind. (Since he rejects any kind of induction there is no other place they could have come from.)

Moreover, since it took several decades, many geniuses, and even more Nobel prizes to get from classical physics before 1900 to QFT in the 1970es, it can be safely assumed that the 'construction problem',

$$\begin{array}{c} ? \\ \uparrow \\ C \end{array}$$

see Fig. 12 in Saint-Mont (2020), is far from trivial. Just the opposite: there are countless ways in which *C* (for instance, classical mechanics) might be generalized, but it is extremely difficult to find a route that leads to the next level. More specifically, Zee (2023), pp. 84f, remarks that there are three mathematically equivalent ways to describe classical mechanics (Hamiltonian, Lagrangian, Newtonian), and also three ways to describe QM (Dirac-Fermi, Heisenberg, Schrödinger). Yet only the underlined formalisms can be generalized in an appropriate manner.

Arbitrary hypotheses or some slight extension of an established theory most certainly will not work. What is actually needed is fundamental insight that may be translated into a new abstract framework - building on *C* but also vastly exceeding it. Consistently, it took decades to understand that 'field' is a core concept, for instance, since fields are able to generate waves, particles, and forces. On the other hand, 'strings' and their theory are a less successful example (Woit 2006-2024).

It may be mentioned in passing that the most appropriate generalizations of important mathematical concepts, structures or operations – such as the natural numbers, Zermelo-Fraenkel set theory, continuity, or integration – are also not straightforward.

6. Educated Guesses with the Axiom of Choice

Alexander Paseau made me aware of the following related problem: Suppose f is a function defined on A. If somebody knows all values of f on $C \subseteq A$, what can he say about $f(y)$ if $y \in A \setminus C$? For example, if C and A are the intervals $C = [a, b)$ and $A = [a, b]$, what do we know about $f(b)$?

By the very definition of a function, a function's values may be chosen in an arbitrary way. Thus $f(x)$ does not imply anything about $f(y)$ if $x \neq y$. Vice versa, in order to have some idea about $f(y)$, given $f(x)$, there has to be some kind of 'connection' between the values of f.

Perhaps the most spectacular example of this kind if given by the principle of analytic continuation of holomorphic functions, which are very smooth. If one knows the values of such a function locally, i.e., in a small open set U_x about x, it is typically possible to extend this 'germ' to a much larger domain,[2] which also shows that analyticity is a very strong assumption.

More elementary, a function f on $C = [a, b)$ is continuous at b, if

$$\lim_{x \to b} f(x) = f(b).$$

In terms of induction this means that there is a unique extension of f from the domain C to A, that is, the values $f(C)$ determine $f(b)$. Mathematically, the crucial point is that the limit on the left-hand side exists, since one may then set $f(b) := \lim_{x \to b} f(x)$. Since, if f is rather irregular - which is quite natural for an arbitrary function - the limit cannot be defined. (A standard example is the function $f(x) = 1$ if x is a rational number, and $f(x) = 0$ if x is irrational.) Nonetheless, discontinuity at b may also just mean that $\lim_{x \to b} f(x) \neq f(b)$.

Against this background, it is rather surprising that Hardin and Taylor (2008a) could prove the remarkable fact that, given an arbitrary function f on $[a, b)$, there is a guessing strategy that is able to predict the correct value $f(b)$ almost surely. In their own words, they state more precisely and more generally:

> *... given the values of the function on an interval $(-\infty, t)$, the strategy produces a guess for the values of the function on $[t, \infty)$, and at all but countably many t, there is an $\varepsilon > 0$ such that the prediction is valid on $[t, t + \varepsilon)$.*

For the non-trivial technical details see the original article, and also Paseau (2011), Pawlowski (2017). Although the latter authors raise some reasonable concerns, it is nevertheless remarkable that an inductive step that is almost always correct seems to be possible.

Hardin and Taylor (2008a) point out that their result rests on the Axiom of Choice (AoC), since they have to choose an order $\preceq$ on the set $[f]_t$ of all guessing strategies that are compatible with the past $(-\infty, t)$, and select the $\preceq$-least element of the latter set (which they interpret as a formalization of Occam's razor).

Consistently, the 'μ-strategy' thus defined does not imply any constructive algorithm that would almost certainly give a correct guess. This is reassuring, since if that were the case, Hardin and Taylor's result would collide with the very definition of an arbitrary function, i.e. that $f(b)$ can be any value. In other words, their result shows the enormous (pathological?) strength of the AoC, which almost leads to a contradiction.

It is well known that the AoC can be formulated in many equivalent ways, in particular, Zorn's lemma and the well-ordering theorem. The former lemma states that every totally ordered subset of a partially ordered set contains a maximal element; the latter theorem says that every set can be well-ordered, such that every non-empty subset has a least element. Note that these reformulations are similar to continuity, which guarantees that the set $f(C)$ can be extrapolated to $f(b)$.

7. Educated Guesses may lead astray

Hardin and Taylor (ibid.) present a discrete version of their puzzle, which they attribute to Gabay and O'Connor:

> *A prison warden lines up infinitely many prisoners, numbered 0, 1, 2,..., in a large prison yard, and places a black or white hat on each of the prisoner's head. Each prisoner can see only the hats of the higher-numbered prisoners. Each prisoner must venture a guess about his own hat color, without hearing the guess of any other prisoner (though they may collaborate on a strategy before the hats are placed).*

Their general result works for ordinals in ω_1: "Our solution is that, if the ordering of the prisoners is any ordinal, then the prisoners can guarantee that all but finitely many guess correctly, by applying the μ-strategy."

More details can be found in Hardin and Taylor (2008b), where the authors explain that their warden selects the colours of the hats at random, that is, owing to the result of independent tosses of a fair coin.[3] Moreover, the ordering of the prisoners is not crucial, since one only needs to assume that every prisoner is able to see all but a finite number of the hats (ibid., Theorem 4, attributed to Gabay and O'Connor). In the most elementary case, $n = 3$ and a prisoner can see the hats of the other two prisoners, but not the hat on his own head. Moreover, he may venture a guess or pass.

In this special case, the optimal strategy is to guess if the prisoner sees two hats with identical colours, and to claim that the colour of his hat is just the opposite of the colour he sees. This strategy wins with probability $p_3 = 6/8 = 3/4$, see the next table.

A	B	C	Guess	Decision	Probability
0	0	0	all prisoners guess 1	all are wrong	1/8
0	0	1	C guesses 1	correct	
0	1	0	B guesses 1	correct	3/8
1	0	0	A guesses 1	correct	
0	1	1	A guesses 0	correct	
1	0	1	B guesses 0	correct	3/8
1	1	0	C guesses 0	correct	
1	1	1	all prisoners guess 0	all are wrong	1/8

In general, since the binomial $B(n, 1/2)$ is symmetric and unimodal it is obviously a good idea to suppose that the hat on one's own head has the less frequent colour observed. The difficult part consists in figuring out who should answer. For general n the answer is unknown, but at least if $n = 2^k - 1$, a binary error-correcting Hufman code gives the optimum answer and yields $p_n = n/(n + 1)$, see Buhler (2002). Passing to the limit, $\lim_{n\to\infty} p_n = 1$, and we arrive at the result stated at the beginning of this section.

Although formally correct, the problem with this solution is that the continuity at the point infinity need not make sense: Owing to the (strong) law of large numbers, a typical prisoner i sees a random binary sequence $b_{i+1}, b_{i+2},\ldots$ (or even $\mathbf{b}_{-i} = b_1, b_2,\ldots, b_{i-1}, b_{i+1}, b_{i+2},\ldots$ since a finite number of realizations does not matter), and both possible values are equally probable in the sense that the relative frequency of the symbol 1 goes to $1/2$ if $n \to \infty$. With the notation of section 4 we have $\pi(B_i = b_i|\mathbf{b}_{-i}) = \pi(B_i = b_i) = 1/2$, and $\pi(y|x) = \pi(y) = 2^{-n}$ for every binary vector $y = (y_1,\ldots, y_n)$ of length n.

In the limit, i.e., given an infinite sequence of random coin tosses, there is no such thing as a less frequent symbol on which to place a bet. Rather, the prisoners face the problem of having to decide in favour of a symbol, though all they see is a perfectly balanced situation - zeros and ones is equal proportion, no matter where or how they look.

Thus, although the amount of information grows with n, and thus 'the larger n the better' – in the limit, given a random binary sequence, there is no preference or pattern that could be exploited, and therefore $p_\infty = 1/2 \neq \lim_{n\to\infty} p_n = 1$. The right-hand side of Fig. 13 in Saint-Mont (2020) that is reproduced below depicts this situation: On the one hand, it is reasonable to use equation (2) and to pass to the limit. On the other hand, considering the limit situation in its own right, the solution is different.

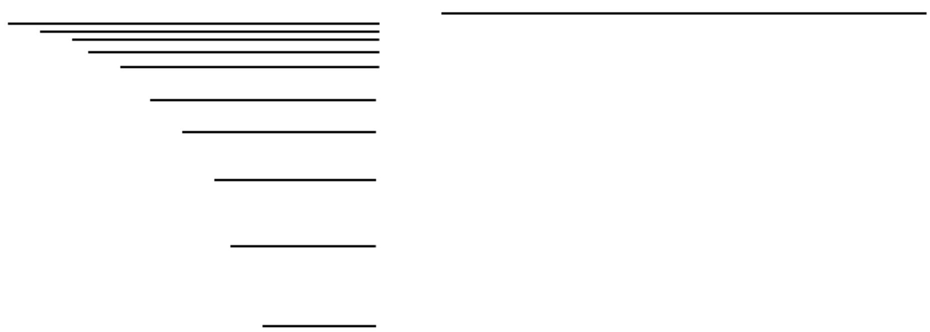


Fig. 3. An inductive failure: inconsistency.

The same phenomenon occurs in a related situation of prophet theory (Hill and Kertz 1982): Suppose $\mathbf{X}_n = (X_1, \ldots, X_n)$ is a vector of iid, [0, 1]-valued random variables. Then a so-called prophet, who knows all realizations of the random variables, typically obtains $M(\mathbf{X}_n) = E(\sup_{1 \le i \le n} X_i)$. She is better off than a statistician, who observes the random variables sequentially and gets $V(\mathbf{X}_n) = \sup_T EX_T$, where T is a stopping rule. Just as before the maximum attainable difference $d_n = \sup_{\mathbf{X}n} (M(\mathbf{X}_n) - V(\mathbf{X}_n))$ grows with n and converges to a positive number if $n \to \infty$.

However, in the limit, i.e., given a sequence of iid random variables $\mathbf{X} = X_1, X_2, \ldots$, the difference vanishes, since the statistician may wait just as long as the prophet waits, and thus obtain the prophet's value. In other words, for the prophet to have an advantage, it is crucial that n be finite. Moreover, $d_\infty = M(\mathbf{X}) - V(\mathbf{X}) = 0 \neq \lim_{n\to\infty} d_n > 0$.

8. Summary

The elementary model of two levels of abstraction, though simple, provides valuable insight:

i. It is important not to confuse the tiers, in particular if arguing verbally.
ii. If the inductive step is in some sense finite, multiplicative probabilities, additive information, and complexity (or traditionally, the dual concept of simplicity) arise naturally.
iii. Hume's problem can be solved with the help of those concepts, and in many frameworks probability updates take centre stage (cf. Kelly (1996), Solomonoff (1964), and most textbooks on formal learning theory or Bayesian statistics). Li and Vitányi (2008), p. 344, are thus right when they conclude that Bayes' formula is 'superficially trivial but philosophically deep.'
iv. Essentially, Bayes' solution is a consequence of sound book-keeping (equation (1)), although the latter elementary principle may be hidden under a formidable superstructure (see, for instance, Li and Vitányi (2008), pp. 264ff).

Finally, some words of caution: There is no such thing as a free lunch (Stavros et al. 2019). Since C contains fewer information than A, the step from C to A can only be logically justified if certain assumptions are met. In particular, the Bayesian mechanism is an algorithm that may fail. It works fine in an essentially finite setting such as coding theory. However, it has taken decades to develop the formal details of (general) nonparametric Bayesian statistics, since there are many pitfalls that need to be avoided (see for instance Ghosal and van der Vaart (2017), and the references given there).

Moreover, a positive answer such as that given by Hardin and Taylor may not be as far-reaching as one might like it to be. It is indeed possible to extrapolate with large (maximum) probability of success - but only in principle, since the objects involved are typically not accessible. Rather, their amazing result demonstrates the surprising consequences of their crucial assumption – the axiom of choice.

All things considered, general, simple, and formally sound answers to the conundrum of induction are available. Fundamental epistemic pessimism with respect to inductive steps no longer seems to be justified.

On the one hand, there are various formalisms – be they probabilistic or couched in terms of simplicity / complexity and information – where the problem of induction essentially boils down to sound book-keeping (Bayes, Occam, Kolmogorov, Solomonoff).

On the other hand and contrary to what Hume claims, the crucial problem of inductive reasoning does not seem to be some kind of vicious circularity (cf. section 2). Rather, the fundamental difficulty (cf. section 5) consists in finding an appropriate general framework (law, pattern, or comprehensive theory), given a more specific situation (data, facts, or a restricted received theory).

Footnotes

[1] In what follows, the notation of Li and Vitányi has been slightly adapted. In order to avoid confusion with the 'μ-strategy' considered later, we write $\pi(\cdot)$ instead of $\mu(\cdot)$.

[2] e.g. https://encyclopediaofmath.org/index.php?title=Analytic_continuation

[3] More precisely, $B_0, B_1, B_2, \ldots$ are independent and identically distributed Bernoulli random variables with success probability $1/2$, the realization of B_j is b_j, and $b_j \in \{0, 1\}$.

Statements and Declarations

Conflict of Interest: The author states that there is no conflict of interest.

Ethical Approval: not applicable

Funding: none

Availability of data and materials: not applicable

References

- Buhler, J.P. (2002). Hat tricks. The Mathematical Intelligencer 24, 44–49. https://doi.org/10.1007/BF03025323
- Clayton, A.; and T. Waddington (2017). Bridging the intuition gap in Cox's theorem: A Jaynesian argument for universality. *Int. J. of Approximate Reasoning***80**, 36-51.
- Cover, T.M.; and Thomas, J.A. (2006). Elements of Information Theory. (2nd ed.) New York: Wiley.
- Cox, R.T. (1946). Probability, Frequency, and Reasonable Expectation. *American J. of Physics* **14**, 1-13.
- Fisher, R.A. (1966). The Design of Experiments. (8th ed.). New York: Hafner Publishing Company.
- Ghosal, S.; and A. van der Vaart (2017). Fundamentals of Nonparametric Bayesian Inference. Cambridge: Cambridge Univ. Press.
- Goodman, N. (1983). Fact, fiction, and forecast. Harvard University Press. Greenland, S. (1998). Probability Logic and Probabilistic Induction. *Epidemiology* **9(3)**, 322-332.
- Hardin, C.; and A. Taylor (2008a). A Peculiar Connection Between the Axiom of Choice and Predicting the Future. *American Mathematical Monthly* **115**, 91–96.
- Hardin, C.; and A. Taylor (2008b). An Introduction to Infinite Hat Problems *The Mathematical Intelligencer* **30(4)**, 20–25.
- Hill, T.P.; and R.P. Kertz (1982). Comparison of stop rule and supremum expectations of i.i.d. random variables *Ann. Prob.* **10 (2)**, 336-345.
- Hutter, M. (2007). On Universal Prediction and Bayesian Confirmation. *Theoretical Computer Science* **384**, 33-48.
- Jaynes, E.T. (2003). Probability Theory. The Logic of Science. Cambridge: Cambridge University Press.
- Johnson, A.A.; Ott, M.Q.; and M. Dogucu (2022). Bayes Rules! An Introduction to Applied Bayesian Modeling. Boca Raton, FL: CRC Press.
- Kelly, K.T. (1996). The logic or reliable inquiry. New York, Oxford: Oxford Univ. Press.
- Li, M.; and P. Vitányi (2008). An Introduction to Kolmogorov Complexity and its Applications. (3rd ed.) New York: Springer.
- Paseau, A. (2011). Proving Induction. Australasian Journal of Logic**9**, 1–17.
- Pawlowski, P. (2017). Philosophical Aspects of an Alleged Connection Between the Axiom of Choice and Predicting the Future. Chapter 10 in Urbaniak, R.; and G. Payetter (eds). Applications of Formal Philosophy. The Road Less Travelled. Springer, 213–220.
- Post, H.R. (1971). Correspondence, Invariance and Heuristics: In Praise of Conservative Induction.*Studies in history and philosophy of science* **2(3)**, 213-255.
- Saint-Mont, U. (2011). Statistik im Forschungsprozess. Heidelberg: Physika.
- Saint-Mont, U. (2020). Induction: a logical analysis. Foundations of Science https://doi.org/10.1007/s10699-020-09683-z
- Solomonoff, R. (1964). A Formal Theory of Inductive Inference, Parts I and II.*Information and Control* **7**, 1-22, 224-254.
- Stavros, P.A., Stamatios-Aggelos N.A., Panos M.P.; and M.N. Vrahatis (2019). No free lunch theorem: a review. *Approximation and Optimization*, 57–82.
- Woit, P. (2006-2024). Not Even Wrong. The Failure of String Theory and the Continuing Challenge to Unify the Laws of

Physics. Basic Books and www.math.columbia.edu/ woit/wordpress/.

- Zee, A. (2023). Quantum Field Theory, As Simple As Possible. Princeton and Oxford: Princeton Univ. Press.