where positive single instances together with projectable predicates count as sufficient criteria of confirmation, or in the context of the deductivist theory of corroboration in which hypotheses are selected to be tested against counterexamples or negative single instances.

(3) **The formal and epistemological completeness of inductive models** according to which a purely inductive agent or intelligence can provide a nonarbitrary description not only of the external world but also of the (inductive) model of mind it inhabits.

(*b*) Concomitant with this analysis, we shall also focus on the import of the problem of induction for philosophy of mind and the project of artificial general intelligence. We shall see that the same predicaments that challenge the epistemic legitimacy of induction also threaten the coherency of those strains of AI in which intelligence is simply equated with predictive induction, and prediction is defined by the information-theoretic concept of compression. It is often assumed that the formal-computational account of Occam's principle of simplicity as put forward by algorithmic information theory—specifically, Ray Solomonoff's account of induction, which is couched in terms of the duality of regularity and compression—circumvents the epistemic quandaries of induction. However, in dispensing with the specificity of the theoretical-semantic *context* in which the principle of simplicity finds its significance as a *pragmatic* tool, the formal generalization of Occam's razor as the cornerstone of all existing computational models of induction not only finds itself faced with the predicaments harboured by the problem of induction, but also results in a number of new complications such as arbitrariness and computational resource problems.

Modelling general intelligence on purely inductive inferences is seen by the current dominant trends in AGI research as an objective realist index of general intelligence. In the same vein, posthumanism built on the assumptions of inductivism and empiricism—i.e., superintelligence can be construed in terms of induction over Big Data—treat inductive models of general intelligence as evidence against an exceptionalism of the conceptualizing human mind.

Such posthumanist accounts of intelligence refuse to see the latter as a sui generis criterion that sets apart general intelligence as a qualitative dimension from quantitative intelligent problem-solving behaviours. Yet, as will be argued, the formal generalization of Occam's razor as a means of granting induction a privileged role capable of replacing all other epistemic activities, along with the equation of general intelligence with induction, turn out to be precisely the fruits of human experiential-cognitive biases.

All in all, my aim is to argue that the force of epistemological scepticism as expressed in the problem of induction can be understood not only in terms of a formidable challenge to entrenched philosophical dogmas and cognitive biases, but also in terms of a razor-sharp critique of purely inductive models of mind and inductivist trends in artificial general intelligence.

Irrespective of their specificity, all models of AGI are built on implicit models of rationality. From the early Carnapian learning machine to Solomonoff prediction as the model of an optimal or universal learning machine to Marcus Hutter's equation of compression with general intelligence and, more recently, the Bayesian program of rational AGI as proposed by Eliezer Yudkowsky, inductivist models are no exception. In this respect, there is certainly a discussion to be had about the sociocultural and political dimension of such trends: What is it exactly in the inductivist models of rationality or approaches to general intelligence that makes them susceptible to appropriation by superintelligence folklores or, worse, by ideologies which champion instrumentalist or even social-Darwinist conceptions of intelligence? Rather than answering this question, I intend to take a different approach: A sociopolitical critique by itself is not, by any means, adequate to challenge such trends in cognitive sciences; nor is a well-constructed rationalist critique, which often devolves into quibbles over whose model of general intelligence or rationality is better. These trends should instead be challenged in terms of their own assumptions and debunked as not only unfounded but also logically erroneous.

A HUMEAN PROVOCATION

To understand the exact nature of Hume's problem of induction, let us first reconstruct it in a more general form and then return to Hume's own exposition of the problem. But before we do so, it would be helpful to provide brief and rudimentary definitions of deduction and induction.

Deduction can be defined as a form of reasoning that links premises to conclusions so that, if the premises are true, following step-by-step logical rules, then the conclusion reached is also *necessarily* true. Deductive inferences are what Hume identifies as demonstrative inferences. In the Humean sense, a demonstrative inference is, strictly speaking, an inference where a pure logical consequence relation is obtained. Such a logical relation has two formal characteristics: (1) there is no increase in the content of the conclusion beyond the content of the premises. Therefore, demonstrative inferences in the Humean sense can be said to be non-ampliative inferences (i.e., they do not augment the content or add anything other than what is already known); (2) the truth of premises is carried over to the conclusion. Accordingly, demonstrative inferences are truth-preserving. It is important to note that non-ampliativity and truth-preservation are—pace some commentators, e.g., Lorenzo Magnani[360]—two separate features and do not by any means entail one another. All that truth-preservation implies is the transferability of the truth of the premises to the conclusion. It does not say anything regarding the augmentation of the content or lack thereof, nor does it exclude the possibility that, if new premises are added, the truth of the conclusion may change. Therefore, Magnani's claim that nonmonotonicity—as for example captured by substructural logics—stands in contrast to the at once non-ampliative and truth-preserving character of demonstrative inferences, is based on a confusion.

In contrast to deductive-demonstrative inferences, inductive inferences cannot be as neatly formulated. However, roughly speaking, induction is a form of inference in which premises provide strong support (whether in

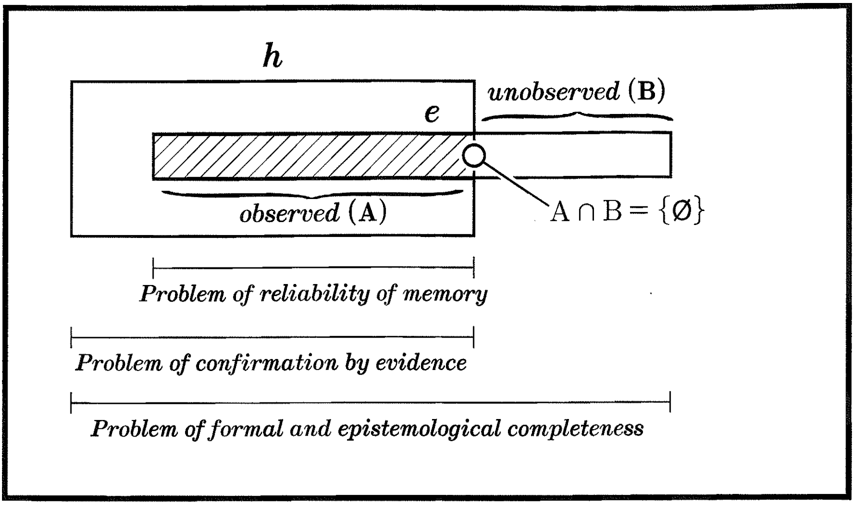
360 L. Magnani, *Abductive Cognition: The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning* (Dordrecht: Springer, 2009).

causal, statistical, or computational terms) for the *outcome*—as distinguished from the deductive *conclusion*—of the inference. Whereas the truth of the conclusion in deductive reasoning is logically certain, the truth of the outcome in inductive reasoning is only *probable* in proportion to the supporting evidence. Hence, as evidence piles up, the degree of supporting valid statements for a hypothesis indicate that false hypotheses are—as a matter of generalization—probably false and, in the same vein, that true hypotheses are—as a matter of generalization—probably true. But this dependency on evidence also means that inductive inference is contingent and non-monotonic. Nonmonotonicity means that the addition of new premises can fundamentally change the truth of the conclusion, either drastically raising or lowering the degree of support already established for the inductive outcome. The significance of induction is that it permits the differentiation of laws from non-laws. This is precisely where the problem of induction surfaces.

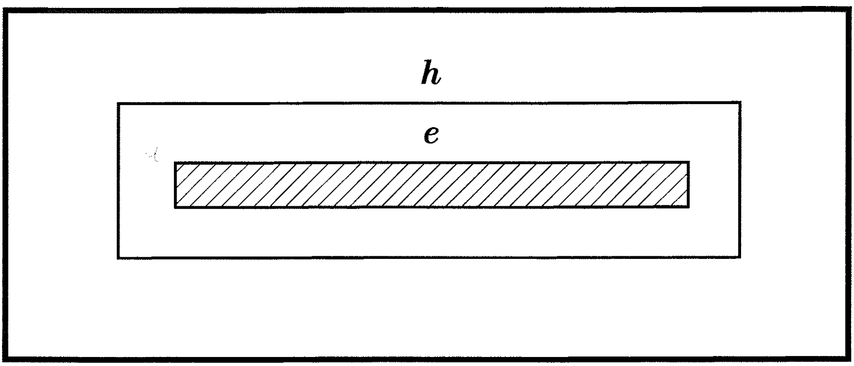
Now, with these clarifications, Hume's problem of induction can be formulated quite generally without being narrowed down to a special class of nondemonstrative inferences (e.g., induction by enumeration) as follows:

(A) Our knowledge of the world must be, at least at some level, based upon what we observe or perceive, insofar as purely logical reasoning by itself cannot arrive at knowledge. We shall call this the problem of synthetic content of knowledge about the world.

(B) Despite A, we take our factual knowledge of the world to exceed what we have acquired through mere observation and sensory experience. However, here a problem arises. Let us call it *problem 1*: How can we justify that our knowledge of the unobserved is really knowledge? At this point, the central issue is the problem of justification rather than the problem of discovery, in the sense that, for now, it does not matter how we have attained this supposed knowledge. Thus, the Kantian claim that Hume confuses *quid facti* and *quid juris*—the origination of knowledge claims and the justification of knowledge claims—and that the problem of induction applies only to the former, simply misses the point.



Induction



Deduction

Induction and deduction as defined over hypotheses (h) and evidence (e)

(C) Since the justification of the alleged knowledge in B cannot be a matter of logical demonstration, it must then be obtained by way of arguments whose premises are observations of the kind cited in A, and the conclusion must be a kind of knowledge which goes beyond observation (i.e., knowledge of the kind mentioned in B).

(D) Two tentative solutions can be provided for the problem of justification characterized in C:

> (D-1) Justification by demonstrative arguments: But as argued earlier, demonstrative inferences do not augment the content (i.e., they are not ampliative). That is to say, if the premises are the observed, nothing beyond the content of the observed is yielded in the conclusion. The knowledge of the unobserved (B) is included in the observed premises (A). But this contradicts B, since the kind of knowledge it assumes must go beyond mere observation. In other words, the content of the statements regarding the alleged knowledge of the unobserved is not contained in the content of the knowledge of the observed. The first solution, however, does not permit this because in demonstrative arguments, the content of the conclusion is necessarily included or contained in the content of the premises.
>
> (D-2) Justification by nondemonstrative arguments: If the non-ampliative nature of the demonstrative inference does not allow the transition from the observed to the unobserved, then the solution to the problem of justification would have to involve ampliative inferences. But is there such a thing as an ampliative inference? The answer is yes: take a logically invalid inference in which the conclusion has an augmented or stronger content in comparison to the conjunction of premises. However, this is not what the inductivist qua proponent of the alleged knowledge of the unobserved is looking for. In logic, invalid logical inferences may be of some interest, but when it comes to knowledge justification, they are merely absurd. Therefore, the ampliative inference *by itself* is not a sufficient condition for the kind of argument that

can answer the problem of justification. The solution to the problem of justification would require an inference in which not only is the content of the conclusion stronger than that of the premises, but also the tentative truth in the content of the premises—knowledge of the observed—can be transferred to the conclusion (i.e., the knowledge of the unobserved) (cf. the earlier note on the distinction between ampliative and truth-preserving features). In short, the solution to the problem of justification would require a truth-preserving (of demonstrative arguments) and ampliative (of nondemonstrative arguments) inference. This is the generalized form of Hume's problem of induction: Is there such an inference—generally speaking and with no reference to specific inductive rules—one that is both truth-preserving and ampliative, demonstrative and nondemonstrative?

(E) There is no such nondemonstrative inference as characterized in D-2. An inference is either ampliative such that the content is extended—but then there is no guarantee that the conclusion is true even if all premises are true; or it is truth-preserving—in which case the conclusion is true and the inference is valid but the content is not augmented. Said differently, the problem of justification concerning the inferential transition of the knowledge of the observed to the knowledge of the unobserved will end up with either content-augmenting but non-truth-preserving inferences, or truth-preserving but non-content-augmenting inferences. Neither case can justify the transition of the observed premises to the unobserved conclusion.

(F) At this point, the inductivist might argue that, in order to resolve the problem of justifying the transition from the observed to the unobserved, we must abandon the absurd idea of a truth-preserving ampliative inference and instead replace it with a probability inference. That is to say, we should understand inductive inferences strictly in terms of probability inferences. But this supposed solution does not work either. For either what is *probable* is construed in terms of frequency, so that the more probable is understood as what has so far occurred more often—in which case we are again confronted with the same problem: How do we know that a past frequency distribution will hold in the future? To arrive at and justify such knowledge, we must again look for a truth-preserving ampliative inference, of the very sort that the inductivist has claimed to be a futile enterprise. Or else the probable is construed in a different sense, which then raises the question of why we should anticipate that the more probable—in whatever sense it has been formulated—will be realized rather than the improbable. In attempting to answer this question, we face the same problem we initially sought to resolve. Thus, the nature of the problem of induction turns not to be limited to the truth-preserving schema of justification, which some commentators have claimed to be 'the outcome of the strictures imposed by the deductive paradigm underlying the classical view of scientific demonstration'.[361]

(G) Considering E and F, the nature of Hume's problem of induction is revealed to be—contrary to common interpretations—not about the origination of knowledge claims (*quid facti*) or even the justification of induction (problem 1), but about a far more serious predicament: we cannot extrapolate knowledge of the unobserved from knowledge of the observed.

If we extend the conclusion reached in G from what has been observed so far and what has not been observed yet to knowledge of the past and the future, Hume's problem of induction then boils down to the claim that *we cannot possibly gain knowledge of the future*. It should be clarified that this claim does not mean that our knowledge of the future can never be certain, either in the deductive sense or in the probabilistic sense—something that any inductivist would accept—but rather that our contention regarding the possibility of having such knowledge is irrational. To put it more bluntly, there simply cannot be any (inductive) knowledge of the future to be deemed certain or uncertain, determinate or wholly indeterminate in any sense.

361 N.B. Goethe, 'Two Ways of Thinking About Induction', in *Induction, Algorithmic Learning Theory, and Philosophy* (Dordrecht: Springer, 2007), 238.

In a contemporary formulation, we can express such knowledge within a system of inductive logic where we interpret probability in the sense of degree of confirmation—the most fundamental concept of inductive logic which characterizes 'the status of any scientific hypothesis (e.g., a prediction or a law), with respect to given evidence'.[362] Such a purely logical concept of probability is distinguished from probability in the sense of 'the relative frequency of a kind of event in a long sequence of events',[363] which can be entirely couched in statistical terms. We can respectively call these two different but related senses of probability, the logical or confirmationist and the frequentist accounts of probability, or, using Carnap's classification, probability_1 and probability_2. In the sense of the logical concept of probability_1, knowledge of the unobserved or the future can be expressed by way of Carnap's formula $c(h, e)=r$, where

- c is a confirmation function (or alternatively, a belief function in the context of formal learning theory) whose arguments are the effective hypothesis h and the empirical data e. Theorems of inductive logic which hold for all regular c-functions can be such theorems as those of Bayes and classical probability theory.

- h is an effective or computable hypothesis expressing supposed universal laws or a singular statement or assertion about the future such that

— h is expressible in a language L rich enough to support a description of space-time order and elementary number theory;

— If it is a consequence of h that $M(x_i)$ is true (where M is a molecular predicate of L and x_i an individual constant running through the names of all the individuals), then it can be said that $h\ M(x_i)$ is provable or alternatively, computable in L;

362 R. Carnap, *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950), viii.

363 Ibid.

– h is equivalent to a set of (computable) sentences of the forms $M(x_i)$ and $\neg M(x_i)$. For example, $M(x_1)$ can be read as Is-Green(*this emerald*). In this sense, 'x_1 is green' or 'x_1 is not green' means that the position x_1 is occupied (or is not occupied) by something green, or that green occurs (or does not occur) at x_1. More accurately, the predicate is attached to the description of the individual's arrangement (e.g., this emerald) in the space-time continuum.

– The outcome of the inductive reasoning is concerned not so much with the acceptance, rejection or temporary suspension of h as with finding the numerical value of the probability of h on the basis of e. This means that even though a thought or a decision—or more generally judgements about h—are not explicitly framed as a probability statement, they can nevertheless be reconstructed as a probability statement.

Now, if a hypothesis implies that each individual satisfies the molecular predicate $M_1(x)\, M_2(x)$, then for each i, $M_1(x_i)\, M_2(x_i)$ should be deducible from h in L, in order for h to qualify as an effective hypothesis.

- e is a statement about the past or a report about the observed, evidence or empirical data.

- r is a real number denoting the quantitative degree of confirmation or the degree of belief such that $0 \leq r \leq 1$. It is represented by a measure function P (i.e., an a priori probability distribution). In this sense, $c(h, e) \leq 1$, $c(h, e) \geq 0.9$ and $c(h, e) \geq 0.5$ mean that, depending on the particular system of inductive logic and the inductive theory of confirmation, the degree by which the statement expressed by h is *confirmed* by the statement expressed by e approaches 1, converges on or remains greater than 0.9 or, in the weakest condition of confirmation, becomes and remains greater than 0.5. Expressing the value of r in terms of a limit function as approaching or remaining greater means that either of the above scenarios admits of exceptions.

- For an initial n-membered segment of a series with the molecular property M that has been observed m times, we can anticipate that the relative frequency of the members observed with the characteristic M—or the m-membered sub-class of M—in the entire series should correspond to m/n. For example, in the case of 0.5, for every n we can expect to find an m such that, if the next m-membered individuals $(x_{n+1}, x_{n+2}, \ldots, x_{n+m})$ are all M, then the degree of confirmation of the effective hypothesis $M(x_{n+m+1})$ is greater than $1/2$ irrespective of the characteristic of the first n-individuals. Suppose n=8, then it must be possible to find an m—let us say m=10^7—such that we can state that x_9, $x_{10}, \ldots, x_{10000000}$ being all green, then it is probable more than one-half that $x_{100000001}$ will also be green *even if* $x_1, x_2, \ldots, x_8$ are not green.

- The criterion for the adequacy of the measure function is that for every true computable hypothesis h, the instance confirmation $P(M(x_{n+1})) \mid M(x_0), \ldots, M(x_n)$ should at the very least converge on and exceed 0.5 after sufficiently many confirming or positive instances x_0, ..., x_n. Let us call this condition CP1.

- In order for any measure function to satisfy the weak condition of effective computability so as to qualify as an explicit inductive method, it must satisfy the following condition CP2: For any true computable hypothesis $M(x_{n+m+1})$ and for every n, it *must be possible to find, i.e., compute* an m such that if $M(x_{n+1}), \ldots, M(x_{n+m})$ hold, then $P(M(x_{n+m}))$ $\mid M(x_0), \ldots, M(x_{n+m})$ is greater than 0.5.

- $c(h, e)$ is a statement that can be *analytically* proved in L.

It is important to note that even though e is based on the observed relative (statistical) frequency (i.e., probability$_2$) and indeed contains empirical content, $c(h, e)$=r or the logical probability$_1$ statement does not contain

e nor is it derived from e. What the probability$_1$ statement contains is a *reference* to the evidence e and its empirical content.[364]

Hume's problem of induction challenges the claim that there can ever be $c(h, e)$ as knowledge of the future. Why? Because, irrespective of the specificities of the system of inductive logic and the theory of confirmation, the c-function is either analytic or ampliative. If it is analytic, the possibility that it can provide us with knowledge about the future is precluded. This is because e (i.e., the nonanalytic source of knowledge about the past or the observed) together with an analytic statement—that is, $c(h, e)=r$—cannot provide us with knowledge or information of any kind about future or unobserved (h). If on the other hand such information is indeed possible, then, in so far as e-statements are about the past and h about the future, $c(h, e)=r$ cannot then be an analytic statement and cannot be analytically proved in L. In other words, the h-statement that refers to the future cannot be considered reasonable or rational if the only factual information it is based on is the past. The observed past and the unobserved future concern with two *disjoint* classes that cannot be bridged by any probability logic. Once again, this raises the question: Why should we expect information about the future to continue the trends of the past in any sense?

Furthermore, as Wolfgang Stegmüller has pointed out through a fictional conversation between Hume and Rudolf Carnap as a champion of inductivism, the choice of the particular c-function is quite arbitrary.[365] An inductivist like Carnap might say, 'It quite suffices for a rational procedure that there is in the long run a higher probability of success'.[366] That is, the inductive model of rationality—which for an inductivist is the *only* viable model of rationality—is based on the claim that, given a sufficiently long

364 'Thus our empirical knowledge does not constitute a part of the content of the probability$_1$ statement (which would make this statement empirical) but rather of the sentence e which is dealt with in the probability$_1$ statement. Thus the latter, although referring to empirical knowledge, remains itself purely logical.' Ibid., 32.

365 Stegmüller, *Collected Papers on Epistemology, Philosophy of Science and History of Philosophy,* vol. 2, 117–19.

366 Ibid., 119.

but finite time, it is reasonable or rational to believe that the inductive model (whether of the mind, intelligence, or a scientific theory) will be vindicated by a higher probability of success. Hume, however, would then challenge Carnap's claim by saying:

> Why should one be rational in your sense of the word 'rational'? The possible answer: 'Because it is just rational to be rational', would, of course, amount to a mere sophism; for in the latter part of the sentence you would be referring to your concept of rationality. The whole sentence then would be taken to mean that it is rational to accept your concept of rationality. But that is exactly the very thing that is in question. Finally, our common acquaintance will not be unaware of the fact that there are infinitely many different possibilities for defining the concept of confirmation and, hence, the concept of rationality.[367]

Now, if we take a system of inductive logic as a design for a learning machine (i.e., a computational agent that can extrapolate empirical regularities from supplied *e*-statements), we can treat $c(h, e)=r$ as a principle of inductive logic upon which a Carnapian computational learning machine or an AGI can be designed. However, such an inductivist machine cannot in any sense be called rational or based on reasonable principles. The generalization of Hume's problem of induction would count as a clear refutation of such a purely inductive model of general intelligence or purported model of rationality or mind. Of course, a more astute inductivist might object that the *c*-functions 'correspond to "learning machines" of very low power'.[368] In other words, that we should give up on the idea of modelling the mind or general intelligence on induction as the degree of confirmation, and instead model it on the recursion-theoretical account of induction. As we shall see, not only is this solution plagued with the aforementioned quandaries of prediction, it is also, in so far as the problem of induction is not limited

367 Ibid.

368 Putnam, *Philosophical Papers* (2 vols. Cambridge: Cambridge University Press, 1979), vol. 1, 297.

to the predictive induction or *c*-functions, plagued with the problems of retrodiction and formalization, which are in fact more serious to the extent that, in being less pronounced, they encroach upon more fundamental assumptions than the possibility of inductively inferring knowledge of the future from knowledge of the past.

However, the sophistication and subtlety of Carnap's thesis on the possibility of constructing an inductive learning machine deserves to be fully recognized and also defended against Putnam's and other critics' occasional manhandling of Carnap's view. Carnap quite explicitly rejects the idea of a mechanical device—a computing machine—that, upon being fed observational reports, can produce suitable hypotheses.[369] Put differently, what Carnap takes to be an inductive learning machine has far more modest ambitions.[370] Carnap does not believe that an inductive learning machine can be said to be an artificial general intelligence since, by itself, an inductively modelled learning machine, just like a purely deductivist learning machine, can neither find the suitable hypothesis nor examine whether or not a given hypothesis is suitable. The Carnapian learning machine is restricted to the determination of $c(h, e)$ under the condition that a limited range of suitable hypotheses has been provided externally—presumably by an ordinary rational judge—which is to say, it must not contain variables with an infinite range of values. As we shall see, even this modest view of an inductive learning machine modelled on a Turing-computable notion of the quantitative degree of confirmation faces serious problems. Nevertheless, it is necessary to acknowledge that the Carnapian learning machine is, properly speaking, a machine—an AI rather than AGI—whose scope is explicitly limited. This is of course in contrast to the claims made in the name of Solomonoff's Carnap-inspired induction and in the context

369 See Carnap, *Logical Foundations of Probability*, 193.

370 'I am completely in agreement that an inductive machine of this *kind* is not possible. However, I think we must be careful not to draw too far-reaching negative consequences from this fact. I do not believe that this fact excludes the possibility of a system of inductive logic with exact rules or the possibility of an inductive machine with a different, more limited, aim.' Ibid.

of algorithmic information theory by the likes of Marcus Hutter and Paul Vitányi in support of the idea of a constructible universal learning machine that can be genuinely said to be a general intelligence.

UNIFORMITY, REGULARITY, AND MEMORY

Having examined Hume's problem of induction in its general form, we can now proceed to briefly look at Hume's own argument. According to Hume, inductive reasoning is grounded on the principle of the uniformity of nature as its premise—that is, unobserved events are similar to observed events, so that generalizations obtained from past observed occurrences can be applied to future unobserved occurrences. In Hume's words, 'that instances of which we have had no experience, must resemble those of which we have had experience, and that the course of nature continues always uniformly the same'.[371] But this principle itself is a conclusion reached by induction, and cannot be proved by the understanding or by deductive reasoning. It cannot be proved by deduction because anything that can be proved deductively is necessarily true. But the principle of uniformity is not necessarily true since the deductive framework admits, without logical contradiction, counterexamples for events which have not yet been experienced, in which a true antecedent (past patterns of events) is consistent with the denial of a consequent (future patterns of events not similar to the past).

Thus, if the principle of uniformity cannot be proved through deduction, and if therefore the validity of induction cannot be established deductively, then it must be proved by causal-probabilistic or inductive reasoning. Yet the validity of such reasoning is precisely what we sought to prove. To justify the principle of uniformity and induction by inductive reasoning is simply question-begging (i.e., a fallacy in which the conclusion is granted for the premises). Therefore, it follows that induction cannot be proved inductively either, because this would count as vicious circularity.

---

371 Hume, *A Treatise of Human Nature*, 390.

For this reason, Hume's problem of induction comes down to the idea that experience cannot provide the idea of an effect by way of understanding or reason (i.e., deductive and causal inferences), but only by way of the impression of its cause, which requires 'a certain association and relation of perceptions'.[372] Understanding cannot produce cause-effect relations or matters-of-fact since such relations are obtained via the inductive generalization of observations. Matters of fact rely on causal relations and causal relations can only be obtained inductively. But the validity of inductions themselves cannot be corroborated deductively, nor can they be explained inductively. Therefore, what is problematic is not only the derivation of uncertain conclusions from premises by way of induction, but also, and more gravely, the very inductive principle by which such uncertain conclusions are reached.

Hume's problem of induction, accordingly, challenges the validity of our predictions—the validity of the connection between that which has been observed in the past and that which has not yet been observed. We cannot employ deductive reasoning to justify such a connection since there are no valid rules of deductive inference for predictive inferences. Hume's own resolution of this predicament was that our observations of patterns of events—one kind of event following another kind of event—create a habit of regularity in the mind. Our predictions are then reliable to the extent that they draw on reliable causal habits or regularities formed in the mind through the function of memory that allows us to correlate an impression with its reproduction and anticipation. For example, if we have the impression (or remember) that A resulted in B, and if we also witness at a later time and in another situation that 'an A of the same kind resulted in a B of the same kind', then we anticipate a nomological relation between A and B: B is the effect of A, as the cause of which we have an *impression*.

However, rather than settling the problem of induction, Hume's resolution (i.e., the reliability of habits of regularity accessible through memory) inadvertently reveals a more disquieting aspect of the problem: that it challenges not only our predictive inductions about the future, but also

372 Ibid.

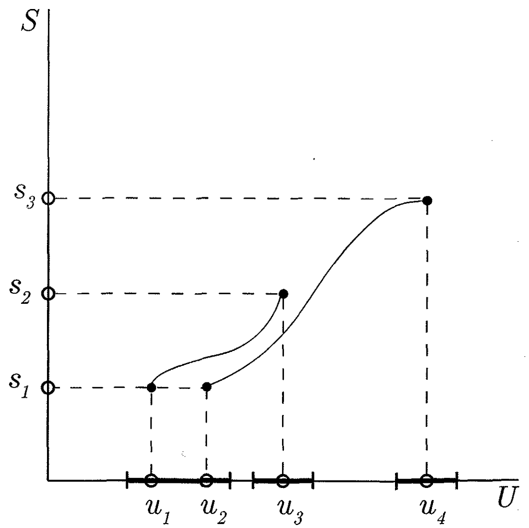
our retrodictive or memory-driven knowledge of the past. In this sense, Hume's problem is as much about the empirical reliability or factuality of the *e*-statements or information about the past as it is about the derivation of the future-oriented *h*-statement (unobserved future) from the *e*-statements (observed past) whose firm status is as much questionable as the former. A proponent of the inductivist model of the mind, general intelligence, or scientific theories thinks that all he must do is to make sure that the evil Humean demon does not get through the door, not realising that the demon is already in the basement. It is a dogmatic assumption to conclude that, so long as one can manage to take care of the reliability of predictive claims—either through a better theory of confirmation, a better Bayesian inference, or a more adequate formal-computational reformulation of inductive reasoning—one does not need to worry about the reliability of empirical reports referring to the past. In other words, a puritan inductivist who believes that general intelligence or the construction of theories can be sufficiently modelled on inductive inferences alone takes for granted the reliability of the information about the past, namely, *e*-statements. Yet for a so-called ideal inductive judge modelled on the alleged sufficiency of induction alone, the predicaments of induction hold as much for empirical data referring to the past as they do for assertions about the future.

Such an ideal inductive judge would be particularly vulnerable to the problem of the unreliability of knowledge of the past derived through memory or, in Humean terms, the regularity-forming memory of impressions of causes. The strongest version of the problematic nature of the reliability hypothesis of memory was given by Russell's 'five minutes ago' paradox, which we had the occasion to examine briefly in chapter 4.

As Meir Hemmo and Orly Shenker have elaborated, the 'five minutes ago' paradox can be conceptually reframed using the Boltzmannian notions of microstate (complexion) and macrostate (distribution of state):[373] Suppose at time t_1 an observer S remembers an event that took place at an earlier time t_0. Let us say, S observes at t_0 a partially deflated ball, and at the same time remembers that at an earlier time t_{-1} the ball was fully inflated. This

373 Hemmo and Shenker, *The Road to Maxwell's Demon*.

memory is occasioned by the microstate s_1 of the nervous system of S at time t_0. The microstate memory s_1 is compatible with at least two microstates of the rest of the universe U, u_1 and u_2, at t_0. These two microstates are in the same *macrostate* (each macrostate is represented on the U-axis by a bracket). If we were to retrodict from the microstate s_1u_1 of $S{\times}U$ the microstates of S and U at t_1, then we would have been able to find them in the microstate s_2u_3 wherein the observer experiences a fully inflated ball and U is in a macrostate compatible with this experience. However, in the case of the microstate s_1u_2 at t_0 the scenario changes. For if we were to retrodict from it the microstates of S and U at t_1, we would have found them in the microstate s_3u_4—that is, where the observer experiences a fully deflated ball. The false retrodiction from s_1u_2 at t_0—as the consequence of many-to-one or possibly many-to-many correlations between the observer's memory states and the rest of the universe—is what is captured by the 'five minutes ago' paradox.



Hemmo and Shenker's macrostate-microstate view of the five minutes ago paradox

The gist of the 'five minutes ago' paradox consists of two parts: (1) Memory-beliefs are constituted by what is happening now, not by the past time to which the said memory-beliefs appear to refer. In so far as everything that forms memory-beliefs is happening now, there is no *logical* or *a priori* necessity that what is being remembered (the reference of the memory-belief) should have actually occurred, or even that the past should have existed at all. (2) There is no logical reason to expect that memory states are in one-to-one correspondence with the rest of the universe. There can be both many-to-one and many-to-many correlations between memory states and external states of affairs. Therefore, what we remember as the impression of a cause, a past event, or an observation, may very well be a false memory—either a different memory or a memory of another impression of a cause. Accordingly, our knowledge of the past or of the impressions of causes can also be problematic at the level of logical plausibility and statistical improbability, which does not imply impossibility. Consequently, it is not only the justification of our predictions regarding events not yet experienced or observed that faces difficulty, but also our memories of past impressions that have shaped our regularities and habits of mind.

THE DISSOLUTION OF HUME'S PROBLEM AND ITS REBIRTH

The Humean problem of induction undergoes a radical change, first at the hands of Nelson Goodman in the context of the new riddle of induction, and subsequently those of Hilary Putnam in the context of Gödel's incompleteness theorems.[374]

Goodman observes that Hume's version of the problem of induction is, at its core, not about the justification of induction but rather about how evidence can inductively confirm or deductively corroborate law-like generalizations. Before moving forward, let us first formulate the Hempelian confirmation problem that motivates Goodman's problem of induction: A positive instance which describes the same set of observations can always

374 See Goodman, 'The New Riddle of Induction', in *Fact, Fiction, and Forecast*, 59–83; and Putnam, *Representation and Reality* (Cambridge, MA: MIT Press, 1988).