Chapter 5
Induction

Rafal Urbaniak and Diderik Batens

5.1 Introductory Remarks

Inductive reasoning, initially identified with enumerative induction (inferring a universal claim from an incomplete list of particular cases) is nowadays commonly understood more widely as any reasoning based on only partial support that the premises give to the conclusion. This is a tad too sweeping, for this includes any inconclusive reasoning. A more moderate and perhaps more adequate characterization requires that inductive reasoning not only includes generalizations, but also any (ideally, rational) predictions or explanations obtained in absence of suitable deductive premises. Inductive logic is meant to provide guidance in choosing the most supported from a given assembly of conjectures. (Some authors think that this has to be done by capturing the notion of partial support, but this conviction is by no means universally accepted.)

The authors would like to express gratitude to Mathieu Beirlaen and Frederik Van De Putte for reading and commenting on an earlier draft of this paper. Work on this paper was supported by the Special Research Fund of Ghent University through project [BOF07/GOA/019] and by Polish National Science Centre grant 2016/22/E/HS1/00304.

R. Urbaniak (✉)
Centre for Logic and Philosophy of Science, Ghent University, Ghent, Belgium

Institute of Philosophy, Sociology and Journalism, University of Gdańsk, Gdańsk, Poland
e-mail: rafal.urbaniak@ugent.be

D. Batens
Centre for Logic and Philosophy of Science, Ghent University, Ghent, Belgium
e-mail: Diderik.Batens@UGent.be

S. O. Hansson, V. F. Hendricks (eds.), *Introduction to Formal Philosophy*, Springer Undergraduate Texts in Philosophy, https://doi.org/10.1007/978-3-319-77434-3_5

Approaches to inductive reasoning are so varied that it is difficult to find a more specific characterization of all of them. In an attempt to draw at least a partial connection, let us observe that among requirements which such a logic is often expected to satisfy [20] are:

Connection with deduction: Deductive consequence and logical contradiction should fit into an inductive logic as extreme cases of support that a conclusion can obtain from premises.

Objectivity: If premises support the conclusion, this fact depends only on the meaning of the premises and the conclusion.

Connection with probability: Some notion of probability should play an important role in the development of inductive logic.

As we will see later on, the last two requirements are not universally accepted.

To fix the ideas, recall the standard axiomatization of probability theory, as given in 1933 by Kolmogorov [44]:

$$\Pr(p) \geq 0 \qquad \text{for any proposition } p \tag{5.1}$$

$$\Pr(p) = 1 \qquad \text{if } p \text{ is necessary} \tag{5.2}$$

$$\Pr(p \vee q) = \Pr(p) + \Pr(q) \qquad \text{if } p \text{ and } q \text{ exclude each other} \tag{5.3}$$

The first stab at capturing the notion of the support that a piece of evidence E gives to a hypothesis H might be to identify it with the probability of the material conditional $E \rightarrow H$. Alas, this approach does not work. For the probability of $E \rightarrow H$ is the same as the probability of $\neg E \vee H$, which means that even if there is no connection between E and H whatsoever, if the probability of H is high enough or the probability of E is low enough, the probability of $E \rightarrow H$ is still high (at least as high as the probability of H or the probability of $\neg E$). (In fact, *mutatis mutandis*, on this approach you can run any of the paradoxes usually associated with material implication.) Thus, if there is a connection between inductive support and probability, it has to be more sophisticated.

The *received view* is that the degree of confirmation is to be identified with the *conditional probability* of the hypothesis given the evidence, defined by:

$$\Pr(H \mid E) = \frac{\Pr(H \wedge E)}{\Pr(E)} \qquad \text{if } \Pr(E) \neq 0 \tag{5.4}$$

$\Pr(p)$ is usually caled the *absolute probability of* p, as opposed to the *conditional probability* of p given q, noted as $\Pr(p \mid q)$. Probability theory which tells one how probabilities are related is not a full confirmation theory, though. To complete the story we also have to explain and justify the basic assignment of probabilities to propositions involved – their probability measure.

The first mathematically developed proposal following this path was put forward by Carnap [11], and we start with presenting his approach (meant to satisfy all three above-mentioned requirements) in Sect. 5.2. In Sect. 5.3 we will briefly survey

Reichenbach's attempt to satisfy all three requirements. In Sect. 5.4 we discuss one of the main theories on today's market, Bayesianism, which drops the second requirement. Next, in Sect. 5.5, we discuss Popper's approach (which is a serious attempt to drop the third requirement). Finally, in Sect. 5.6 we discuss the adaptive approach to inductive generalization, which proceeds qualitatively and drops the third requirement, not taking any degrees of confirmation as necessary for inductive inferences.

A very important issue which we will not discuss in detail is the philosophical problem of finding a general justification of inductive methods. The problem, raised by Hume [33], has received enough attention in the literature (see for instance the survey by Vickers [75]) and we could not do it justice in this short essay meant to focus on formal methods (one exception is Sect. 5.3, where we look at an attempt of justifying induction by means of certain results about a formal method). Another thing which we won't mention are causal and abductive inferences. They do fall under our general notion of induction, but we decided to focus on more crucial phenomena in the development of formal methods of induction instead.

5.2 Carnap and Induction

5.2.1 Preliminaries

The main notion which Carnap's approach to induction [11] is meant to explicate is the logical notion of *the degree of confirmation of a hypothesis H by a given body of evidence E* : $c(H, E)$. If E is the conjunction of the available observational data, $c(H, E)$ expresses the degree of confidence or belief that one should assign to H.[1]

Consider a first-order language containing a finite number of logically independent monadic predicates, a finite number of individual constants and standard Boolean connectives. A *literal* in such a language is either an atomic formula or its negation.

A *state description* in such a language is a conjunction which for any predicate and any constant contains exactly one literal composed of them (e.g. either Ga or $\neg Ga$ but not both). Thus, a state description for any property and any object says whether this object has this property. Every sentence is logically equivalent to the disjunction of the state descriptions which entail it. If every object in the domain is named by a different individual constant, then the set of all state descriptions exhausts the possible states of the domain as describable in the language.

[1] Early defenders of the logical approach include Keynes [43] and Johnson [36, 37].

A *structure description* generated from a given state description ϕ is the set of all state descriptions that result from ϕ by a permutation of individual constants (sometimes, it's identified with the disjunction of its elements). Structure descriptions can be interpreted as encoding information about the numerical distribution of properties among objects.

For instance, take a language with only one predicate G and only two constants a and b. There are four state descriptions:

$$\text{(a) } Ga \wedge Gb, \text{ (b) } Ga \wedge \neg Gb, \text{ (c) } \neg Ga \wedge Gb, \text{ (d) } \neg Ga \wedge \neg Gb.$$

and three structure descriptions:

(A)	$\{Ga \wedge Gb\}$	('all objects have property G'),
(BC)	$\{Ga \wedge \neg Gb, \neg Ga \wedge Gb\}$	('exactly one object has property G'),
(D)	$\{\neg Ga \wedge \neg Gb\}$	('no object has property G').

A *probability measure* assigns probabilities to state descriptions, so that the sum of the probability measures of all state descriptions is 1. As state descriptions are mutually exclusive, the probability measure of a disjunction of state descriptions is the sum of the probability measures of all disjuncts. Each sentence is equivalent to a disjunction of state descriptions, so the probability measure covers all sentences. Given a probability measure m, $c(H, E)$ (the degree of confirmation of H by E) can be defined by:

$$c(H, E) = \frac{m(H \wedge E)}{m(E)} \tag{5.5}$$

That is, the degree to which evidence E confirms hypothesis H is the proportion of the probability of the hypothesis and the evidence to the probability of the evidence.[2] Thus, various confirmation functions arise from various probability measures.

5.2.2 Probability Measures $m^\dagger$ and m^*

One way to define a probability measure, introduced by Carnap, is to divide the probabilities equally among the state descriptions. If there are k available (up to logical equivalence) state descriptions, and exactly n of those state descriptions logically imply sentence H, the probability of H is defined by $m^\dagger(H) = n/k$.

Each state description in our example is assigned the $m^\dagger$-value of $1/4$. So, $m^\dagger(\neg Ga) = m^\dagger(c) + m^\dagger(d) = 1/4 + 1/4 = 1/2$. As it turns out, the degree of confirmation of $\neg Ga$ by Gb is also $1/2$:

[2] In case no evidence is available, hypothesis H is evaluated against any logical theorem $\top$, so that $c(H, \emptyset) = c(H, \top) = m(H)$.

$$c^{\dagger}(\neg Ga, Gb) = \frac{m^{\dagger}(\neg Ga \wedge Gb)}{m^{\dagger}(Gb)} = \frac{m^{\dagger}(c)}{m^{\dagger}(a) + m^{\dagger}(c)} = \frac{1/4}{1/2} = 1/2.$$

Analogous calculations show that $m^{\dagger}(Ga) = c^{\dagger}(Ga, Gb) = 1/2$. But this shows that observing Gb has no impact on the degree of belief one should assign to Ga.

The problem with this independence generalizes. Even for a thousand objects $a_1, a_2, \ldots, a_{1000}$ the following will hold:

$$c(P(a_1), P(a_2) \wedge \cdots \wedge P(a_{1000})) = c(\neg P(a_1), P(a_2) \wedge \cdots \wedge P(a_{1000})) = m(P(a_1)).$$

But this means that no amount of evidence will have any impact on the level of confirmation of $P(a_1)$.

This led Carnap to consider a different probability measure, m^*. The method of assigning m^* is quite simple: first divide probability 1 equally among the available (up to logical equivalence) structure descriptions, thus building in the assumption that each structure description is equally probable. Then, divide the probability of each structure description equally among its members.

In our example, each of three structure descriptions is assigned probability measure $1/3$. Since (A) and (D) contain exactly one state description, each of those state descriptions is assigned probability measure $1/3$. On the other hand, each element of (BC) obtains the value $1/6$.

To see how this probability measure favors homogenous descriptions and deals with the independence issue, compare the probability measure of $\neg Ga$ with the confirmation of the hypothesis that $\neg Ga$ on the evidence that Gb (intuitively, the latter should be lower). $\neg Ga$ holds in (c) and (d) and hence $m^*(\neg Ga) = m^*(c) + m^*(d) = 1/3 + 1/6 = 1/2$. On the other hand:

$$c^*(\neg Ga, Gb) = \frac{m^*(\neg Ga \wedge Gb)}{m^*(Gb)} = \frac{m^*(c)}{m^*(a) + m^*(c)} = \frac{1/6}{3/6} = 1/3.$$

As expected, $c^*(\neg Ga, Gb) < m^*(\neg Ga)$. Similarly, $c^*(Ga, Gb) = 2/3 > m^*(Ga) = 1/2$, so Gb (partially) confirms Ga and (partially) disconfirms $\neg Ga$.

5.2.3 The λ-Continuum of Confirmation Functions

As it turns out, there is a wide variety of confirmation functions [12]. To see how such a variety arises, consider the following. If $F_1, F_2, \ldots, F_k$ are all the monadic predicates of a given language, we say that a Q-formula predicated of a constant a is of the form:

$$\pm F_1 a \wedge \pm F_2 a \wedge \cdots \wedge \pm F_k a$$

where each $\pm$ stands either for a negation or for nothing. Q-formulas of such a language can be enumerated, let's pick the i-th one and call it Q_i. One of the key confirmation assignments that we would like to calculate is that of $c(H_{Q_i}, E_Q)$ where H_{Q_i} is the Q_i-formula predicated of a certain constant a and E_Q is a conjunction of certain Q-formulas predicated of some constants different from a. (That is, we would like to be able to measure how complete information about certain objects observed so far confirms a given complete description of a new object.)

As Carnap suggests, there are at least two important factors in our assessment of $c(H_{Q_i}, E_Q)$. One is the empirical factor of the relative frequency of Q_is in E_Q: s_i/s (where s_i is the number of occurrences (modulo logical equivalence) of Q_i in E and s is the number of non-equivalent Q-formulas in E). The other factor is the logical one: the logical factor of Q_i equals $1/K$, where K is the number of all Q-predicates of the language. Following Carnap, $c(H_{Q_i}, E_Q)$ should be somewhere between these two values. A convenient way of representing this is to take it to be their weighted mean defined by:

$$c(H_{Q_i}, E_Q) = \frac{\frac{w_1 s_i}{s} + \frac{w_2}{K}}{w_1 + w_2} \tag{5.6}$$

where w_1 and w_2 are weights. Actually, since what matters is the ratio of the weights, one of them can be parametrized. Carnap suggested parametrizing w_1 and taking it to be s, thus making sure that the empirical factor gains weight as more observations are being made. The other weight is usually represented as λ:

$$c(H_{Q_i}, E_Q) = \frac{s_i + \lambda/K}{s + \lambda} \tag{5.7}$$

Any choice of λ in (5.7) gives a new confirmation function in the sense of (5.5). Consider what happens when we take $\lambda = 0$. In this case

$$c(H_{Q_i}, E_Q) = \frac{s_i + 0/K}{s + 0} = s_i/s$$

For instance, suppose there are only three constants a, b, c and only one predicate F and that we so far observed only two of them, which turned out to be F. What are the confirmation values of the hypothesis that the last object will also be F and of the opposite hypothesis, if $\lambda = 0$?

$$c(Fc, Fa \wedge Fb) = s_F/s = 2/2 = 1$$
$$c(\neg Fc, Fa \wedge Fb) = S_{\neg F}/s = 0/2 = 0$$

If however, one observed object is F and another one isn't, we get:

$$c(Fc, Fa \wedge \neg Fb) = s_F/s = 1/2$$

In this sense, (5.7) for $\lambda = 0$ assigns maximal role to the evidence and no role whatsoever to the logical possibilities (it corresponds to Reichenbach's straight rule — see Sect. 5.3).

For comparison, consider what happens as λ approaches ∞: in the limit (5.7) yields $1/K$. Thus, in our example, no matter whether we observed any other objects which are F, the confirmation of the hypothesis that the next object will be F is just the prior logical probability of that hypothesis[3]:

$$c(Fc, Fa \wedge Fb) = c(Fc, \neg Fa \wedge \neg Fb) = m(Fc) = 1/2.$$

So taking $\lambda = \infty$ assigns maximal importance to the logical factor and no role to the evidence and does not allow for learning from experience. In fact, the confirmation function thus defined is $c^{\dagger}$, which we already discussed.

The above choices of λ are two extremes of a continuum of confirmation functions (the lower λ, the more important the impact of the evidence on the confirmation value of the hypothesis). Where is the c^* in this continuum? It is obtained by equating λ to K, in which case (5.7) yields the following:

$$c(H_{Q_i}, E_Q) = \frac{s_i + K/K}{s + K} = \frac{s_1 + 1}{s + K}. \tag{5.8}$$

5.2.4 Challenges and Tweaks

One difficulty is that the above framework provides a variety of probability measures without indicating why we should prefer any of them over the others. Hájek [28] and Glaister ([22]: 569) see this as a serious challenge. Vickers [75] is more moderate: given certain basic restrictions,[4] even if the confirmation function is not unique, quite a few useful claims hold no matter which non-extreme function we pick. Initially, Carnap felt quite strongly about m^*, but eventually this embarrassment of riches motivated Carnap to accept a somewhat subjectivist attitude consisting in saying that there is a wide variety of options which remain open, even after all methodological considerations have been brought it.[5] Some others, like Fitelson [20], see nothing wrong in relativizing confirmation to probability measures and using the logically objective 'given such-and-such probability measure, the confirmation degree in this case is…' (Fitelson compares this to special relativity theory in which it is not velocity but rather velocity with respect to a frame of reference that is objective.)

[3] This holds as long as the evidence does not contain any constant occurring in the hypothesis.

[4] Most notably, regularity (every state description has non-zero probability) and symmetry (complete permutations of individual constants and predicates of the same type do not change the value of the function).

[5] See [79] for historical remarks.

Another problem with Carnap's inductive logic is that it is not very successful at handling reasoning by analogy. Intuitively speaking, the more primitive properties two objects share, the more likely it should be that they would agree on other properties. Yet, c^* fails to capture this intuition.[6]

For instance, suppose we have a language with two predicates F and G and two constants a and b. If reasoning by analogy worked, then the fact that $Fa \wedge Fb \wedge Ga$ should give more support to the hypothesis that Gb than just the evidence Ga:

$$c^*(Gb, Fa \wedge Fb \wedge Ga) > c^*(Gb, Ga) \tag{5.9}$$

And yet, (5.9) fails, because in this case both degrees of confirmation are equal:

$$c^*(Gb, Ga) = \frac{m^*(Ga \wedge Gb)}{m^*(Ga)} = \frac{1/3}{1/2} = 2/3$$

$$c^*(Gb, Fa \wedge Fb \wedge Ga) = \frac{m^*(Fa \wedge Fb \wedge Ga \wedge Gb)}{m^*(Fa \wedge Fb \wedge Ga)} = \frac{1/9}{3/18} = 2/3$$

Carnap attempted to deal with such issues [13] (he introduced yet another parameter apart from λ, usually called η), but the success is quite limited. Some attempts to deal with analogical reasoning within a (widely) Carnapian framework are [15, 44, 53, 70] and [50].

Once we generalize the notions to infinite domains, Carnap's inductive methods a priori assign zero probabilities to universal generalizations. This is considered a problem [2] because usually laws of nature are taken to be universal, and if it were true that no finite evidence can provide support for any universal statement, this would go against our intuitions that certain scientific hypotheses are better confirmed than others. The requirements put on confirmation functions can be modified to allow for non-zero probabilities of universal generalizations [77], and some attempts to give a systematic account of non-zero probabilities of universal claims have been put forward. Most notable are those by Hintikka [30], who introduced yet another parameter α dependent on the number of constants available in the language to contribute to the non-zero confirmation of universal claims (the theory has been extended in Hintikka and Niiniluoto [31]) and Kemeny [39], who even with almost-zero confirmation degrees of universal hypotheses allowed to compare their support in model-theoretic terms.[7] Hintikka's approach only enables one to assign non-zero probabilities to really general hypotheses, such as 'all G are F', but not to objective probabilistic sentences like 'the ratio of F within the set of Gs is r'.

[6] The problem was noticed already by Kemeny [40]. See however ([3]: 92–96) and [51] for more details.

[7] See also [3, 5, 78] and [57] for more detailed accounts.

Fitelson [20] worries that on a Carnapian account nothing warrants the relevance of the evidence to the hypothesis, and irrelevant evidence may highly confirm a hypothesis just because the hypothesis is highly likely or the evidence highly unlikely. He suggests [19] that the only historically proposed definition of confirmation that obeys certain basic relevance requirements is that of Kemeny and Oppenheim [42], which identifies it with

$$\frac{\Pr(E|H) - \Pr(E|\neg H)}{\Pr(E|H) + \Pr(E|\neg H)}.$$

Thus, he suggests, relevance requirements help to deal with the initial embarrassment of riches.[8]

A challenge to a purely syntactic approach to confirmation has been posed by Goodman [26]. Say we have drawn a marble from a certain bowl on each of the past ninety days and they all have been red. Thus, it seems, the evidence that the first ninety marbles were red increases the confirmation of the hypothesis that the next one will be red as well. But take another predicate, S, defined as 'drawn up to today and red, or drawn after today and blue.' Our evidence tells us that the ninety marbles observed so far were S, and so, if Carnapian theory was straightforwardly adequate, that the next one will be S too. But this is clearly not the case: our evidence does not confirm the hypothesis that the next marble will be blue.[9] The main lesson to be drawn is that which predicates can be sensibly used in inductive reasonings is an extralogical issue.

Carnap set out to solve the problem of confirmation in terms of logical probability, apparently expecting that there would be a single adequate probability measure. After 1952 it turned out that he had to justify the choice of a probability measure. The only sensible way of achieving this which he saw was in terms of empirically motivated methodological considerations. In a sense this turned his program upside down. For instance, choosing different w_1 in (5.6) leads to a new variety of measures and parameterizing on s in (5.7) already presupposes that induction is justified (for example, weighing it with 1/s leads to an anti-inductive measure).

Despite the difficulties, Carnap's contributions were among the first technically elaborate attempts to explicate the notions involved. The Carnapian program encountered its difficulties, but their very appearance motivated researchers to follow many different paths and led to a variety of ongoing research projects.

[8] In fact, many other attempts of redefining c have been observed. See [18, 32] and [1] for a variety of options.

[9] The paradox is slightly better known in the version from 1953, where Goodman speaks of 'blue' and 'grue' [27].

5.3 Reichenbach's Straight Rule and Pragmatic Justification of Induction

Reichenbach identifies the probability of an event with the limit of the relative frequency of events of the same kind.[10] Given that we normally observe only finite sequences of events, the question arises as to how we are to assess the relative frequency at the limit and how our general strategy of achieving this is to be justified. Reichenbach's response to the first question is that we should apply what he calls the *straight rule* (SR), which roughly speaking, says that one should take observed relative frequencies to be the limiting relative frequencies (and adjust as new observations are made) [4].

Reichenbach [4, 65] attempted to motivate the acceptance of SR, and hence induction by the following considerations. Either there is an inductive method which succeeds, or there is none. If there is none, we do not lose anything by using SR. If there is one, then SR will succeed as well. This justification turns out to be problematic ([3]: 152–153), for there are many inductive methods which agree with SR on past success ratio, vary from it in the predictions about the future which they legitimize at any finite point [67], and converge to the same value. Reichenbach provides no way of picking SR from among all its rivals. Even if it is SR which in fact makes the right predictions, when assessed in terms of past successes it does not stand out from a crowd of so far equally successful methods (although, for a defense of SR against this qualm see [38]).

A related difficulty is that the type of convergence involved in SR is somewhat weak because if one wants to obtain knowledge about infinitely many probabilistic relations there might be no single upper limit on the number of observations that have to be made even if for each such relation an upper limit exists ([17]: 375).[11]

5.4 Bayesian Approaches to Induction

5.4.1 Bayesianism and Subjective Probability

The embarrassment of riches which haunts the Carnapian objectivist program is embraced by Bayesians. While the logical approach faces the difficulty of finding a justification for a specific choice of initial probabilities, the personalistic Bayesians take the choice of initial probabilities to be an extralogical (and personal) issue. For them, an important task of a formal theory of inductive reasoning is to explain how,

[10] Reichenbach developed a slightly unorthodox probability calculus, see [17] for details.

[11] Reichenbach's approach also has to face all the challenges which haunt any frequentist approaches to probability (like the need for a sensible account of the probabilities of singular events). For a discussion, see [3, 28].

given certain initial degrees of beliefs, one has to revise their commitment when faced with new evidence.

Bayesians take personal probabilities (degrees of beliefs, also called subjective probabilities or credences) to be strongly connected with bets [16, 64]. Suppose you bet an amount n on a certain outcome S and I bet $3n$ against S. If S takes place, you win $4n$ (gaining $3n$) and I lose $3n$. If S does not take place, I win $4n$ (gaining n) and you lose n. In such a case we say that the stake is $4n$ (the sum of all bets), your betting rate is $1/4$ and my betting rate against S is $3/4$. In general, a betting rate is just the bet divided by the stake. (A conditional bet is just like that, with the difference that if the condition is not satisfied, the bet is off.) A bet on S at rate k is called fair if there is no advantage in betting on S at rate k rather than against S at rate $1 - k$. The degree of your belief in S is within the Bayesian framework identified with what you consider the fair betting rate on S.[12]

An important role in updating beliefs in face of new evidence is played by a theorem of probability theory called Bayes' Theorem. Before we describe how Bayesian updating works, let us introduce the theorem.

5.4.2 Understanding and Applying Bayes' Theorem

Bayes' Theorem in its simple formulation states:

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E)} \tag{5.10}$$

The denominator can be rewritten in terms of conditional probabilities. By the law of total probability, if $A_1, \ldots A_n$ are mutually disjoint hypotheses such that the sum of their probabilities is 1,

$$\Pr(E) = \Pr(E|A_1)\Pr(A_1) + \cdots + \Pr(E|A_n)\Pr(A_n).$$

Applied to (5.10), this yields:

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E|A_1)\Pr(A_1) + \cdots + \Pr(E|A_n)\Pr(A_n)}. \tag{5.11}$$

In particular, we can use $H, \neg H$ as elements of the partition, in which case we have:

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E|H)\Pr(H) + \Pr(E|\neg H)\Pr(\neg H)}.$$

[12] As almost always in philosophy, the devil is in the details, and various worries arise when one really wants to measure degrees of belief in terms of bets, but those issues lie beyond the scope of our survey.

The most interesting feature of Bayes' Theorem is that it determines the conditional probability of a hypothesis given a body of evidence in terms of other probabilities (which is quite helpful if those other probabilities are easier to ascertain). In general, determining $\Pr(E|H_i)$ is often much easier than determining $\Pr(H_i|E)$ (and there may be good reasons for assigning equal probabilities to all H_i).

An important role in Bayesianism is played by a procedure called *conditionalization*. It consists in changing our belief in a hypothesis H once new evidence E is obtained in the following way. Take your initial probabilities involved in the right-hand side of (5.11) at time t. If you already have the right-hand side probabilities, Bayes' Theorem allows you to calculate the probability of H conditional on E at time t: $\Pr_t(H \mid E)$. Now, if new evidence E is provided at some later time t', your $\Pr_{t'}(H)$ should be identical to $\Pr_t(H \mid E)$. That is, if at a certain time you believe that the probability of a certain hypothesis given E is k, this is the probability you should assign to that hypothesis once you find out that E (and you don't find out anything else that might have impact on the relevant probabilities).[13]

The Bayesian framework allows for a number of ways of making sense of the confirmation that a piece of evidence gives to a hypothesis. A piece of evidence E (incrementally) confirms hypothesis H if $\Pr(H \mid E) > \Pr(H)$ and the confirmation level of H by E is often identified either with the *difference measure* $\Pr(H \mid E) - \Pr(H)$ or the *ratio measure* $\Pr(H \mid E)/\Pr(H)$.

5.4.3 Arguments for Bayesianism

Why would a rational agent's degrees of belief satisfy the axioms of probability? The claim is supported by considerations meaning to show that the acceptance of the axioms of probability theory is required to avoid being susceptible to sure loss. A Dutch Book against an agent is a bet (or a series thereof) which, collectively taken, the agent has to lose. Agents are called *coherent* if they are not susceptible to a Dutch Book. De Finetti [16] proved that if one's degrees of belief do not comply to the axioms of probability theory, one is not coherent. Kemeny [41], Shimony [69] and Lehman [49] proved that the implication in the opposite direction also holds.

One might be worried that grounding an epistemic standard in pragmatic considerations is inappropriate. For people with such concerns, another class of arguments developed from the perspective of *epistemic utility theory*, is available [5]. Think of truth as 1 and falsehood as 0. Pick a measure of distance between a given degree of belief and the given sentence's truth-value (for instance, one can use squared difference). The lower the score, the greater the accuracy of your belief. Define some sensible way of aggregating inaccuracies of one's beliefs into

[13] Jeffrey [35] provides a more general formulation which applies also to cases where one only finds out that E is probable to a certain degree. A Dutch Book argument (see Sect. 5.4.3) for this general formulation has been given by Armendt [1] (see also [70]).

one global measure of inaccuracy. Now, *Accuracy theorem* is available to the effect that if a set of degrees of beliefs violates the axioms of probability, there is a set of probabilistic degrees of belief which are more accurate, no matter what truth-values the beliefs have, and the *Converse accuracy theorem* says that no probabilistic set of beliefs is so dominated by a non-probabilistic one. From this perspective, not being Bayesian is irrational, because it entails being further from the truth, no matter what the truth is.

5.4.4 Challenges to Bayesianism

Let's briefly list the main concerns that the Bayesians have to deal with (some of them apply also to Carnap's approach):

- ▷ Bayesianism does not say anything about the choice of initial probabilities of E, of H and of $E \mid H$, so the same evidence might legitimately motivate two researchers to assign quite different probabilities to a hypothesis, if their initial probabilities are sufficiently different.

  The Bayesian response to this difficulty is that one can prove that as the amount of evidence increases, probabilities assigned to relevant hypotheses will converge (almost) independently of what the initial subjective probabilities are [68]. The problem is that (i) this works only if the initial subjective probabilities are not 0 or 1, and (ii) extreme initial probabilities (close to 1 or close to 0) prevent rapid convergence and make the further search for evidence practically useless.
- ▷ In actual reasoning, rational agents rarely can assign (or even decently approximate) subjective probabilities to the relevant factors, and it is unclear whether betting preferences are a sufficient and correct way of discovering the priors [35].
- ▷ If $\Pr(E) = 1$, then $\Pr(H \mid E) = \Pr(H)$, so old evidence cannot confirm any hypothesis even if one realizes now that the evidence is relevant for the hypothesis, for example because it is implied by it (this is called the *problem of old evidence* [18, 23–25]).[14] Some (like [21]) try to avoid this by weakening the assumption that agents are logically omniscient,[15] but it is not clear what modifications of the Bayesian formal apparatus this move entails. Some try to apply pre-formal philosophical discussion to massage the phenomenon into the Bayesian framework [14, 48].
- ▷ Bayesianism in a sense disregards the structure of explanation and does not take into account such factors as its simplicity or the unity of the underlying theory. That is, all that is considered when we evaluate a given theory is our prior probabilities and available evidence in its favor: factors like simplicity or unity,

[14] For a discussion, see [1].

[15] That is, the assumption that they know all logical consequences of what they know.

intuitively important for the evaluation of a theory, are not explicitly considered in the evaluation procedure. Sure, one can take these factors to be incorporated among prior probabilities, but if that is the case, bayesianism does not really explain how these factors are to be assessed and sweeps them under the carpet of unexplained prior belief degrees.

- ▷ Bayesianism tells a story about rationality and its relation to betting behavior. Yet, it does not say much about why being rational in this sense should put one in an epistemologically privileged position. Why does the fact that I obey rules which would help me to avoid a Dutch Book result in Bayesian updating being the best way to go about scientific reasoning? It is not immediately clear why scientific success and winning bets should be related [10].
- ▷ As already mentioned, Bayes' Theorem establishes a connection between certain probabilities. The connection is useful if the probabilities on the right-hand side are easier to ascertain than the one we attempt to assess. Perhaps, $\Pr(E \mid H)$ often can be easily assessed, but the other probabilities on the right-hand side of (5.11) may be more problematic. For example, $\Pr(E \mid \neg H)$ seems at least as mysterious as $\Pr(H \mid E)$ if H is a general hypothesis. For instance, it is not really clear how to establish the probability of observing a black raven if not all ravens are black or the probability of Eddington's observation if relativity theory is false.
- ▷ Conditionalization does not follow from Bayes' Theorem and is not justified as an a priori rule of rationality. It does not follow from Bayes' Theorem, because one can obey Bayes' Theorem at each moment while completely changing one's degrees of beliefs between moments. Nor does it seem a priori, because it is diachronic, which means that it incorporates a prediction about what will happen at a later time based on what has happened so far (and such moves are usually not considered a priori since Hume). Some diachronic Dutch Book arguments have been given by David Lewis (as reported by [73]), but they rely on stronger assumptions which themselves do not seem a priori.
- ▷ The claim that rational agents should obey the laws of probability implies their logical omniscience (insofar as deductive logic is involved). This difficulty Bayesianism shares with many formal approaches to epistemology.[16]

Given a variety of troubles that a fully subjectivist approach to priors encounters, various unorthodox versions of Bayesianism are being put forward [34, 66, 76] which try to put some additional constraints on priors without running into the problems that fully objectivist and syntactical accounts run into [for a survey of early papers of Bayesianism see [46], and for a survey and further refences see 3].[17]

[16] A twist to this problem is that once classical logic becomes the underlying logic, Bayesianism is unable to account for the possibility of the underlying logic being revised and to explain how evidence might motivate a change of underlying logic [72].

[17] It is also worth mentioning that one of the strength of Bayesianism lies in various applications of the framework to classical philosophical problems. For instance, the framework is used to describe and assess more precisely various arguments in the philosophy of religion (see e.g., [29]).

5.5 Popper

As is well-known, Popper rejected the early Vienna Circle's verificationism from his (1935) on. For him, the central mechanism of scientific methodology is falsification. Good scientists try to falsify theories, and our best theories are the outcome of such attempts. Popper also rejects the idea of confirmation in the sense in which it was used before in this paper. No finite set of observational data can justify one to raise one's degree of belief in a theory[18]; a single falsification to the contrary justifies one's rejection of the theory. The very idea of inductive logic is rejected. For Popper, all logic is deductive (and coincides with classical logic).

Popper's disagreement with the Vienna Circle, not to mention personalists, lies in his different conception of science. Scientific theories are not mere generalizations of observations, but express lawlike connections. They are not justified by (passive) observations, but by actions: attempts to falsify the theories. This requires that one looks for specific observations or, even more typically, performs specific experiments. Finding 'confirming instances' is too easy.[19] But so is the duplication of experiments that are likely to succeed. This is why Popper requires severe tests, tests that are most likely to lead to falsification. The stress on theories is Popper's. Separate generalizations cannot be tested because their falsification can always be reasoned away by modifying another generalization.[20] Popper pushed the idea of falsification to its extreme consequences—we shall see only part of that here.

Popper invoked formal methods to make all this precise. These methods invoke classical logic. They also invoke logical probabilities. This, however, did not cause any embarrassment of riches for Popper. As he explained in appendix $^*vii$ of [61], he considered Carnap's $m^\dagger$ as the only methodologically acceptable measure function for logical probability. All other measure functions can only be justified by non-logical considerations. So any occurrence of Pr in this section should be interpreted in terms of $m^\dagger$.

Testing a theory means trying to bring about an observable fact that falsifies the theory. So the first question for Popper's methodology is which theories one should test first.[21] A theory is *falsifiable* if a possible observable fact contradicts it—non-

[18] Compare this to the fact that if the number of constants is infinite, then every measure function m from Carnap's λ-continuum gives $m(h) = 0$ whenever h is a universally quantified formula, and gives $c(h, e) = 0$ whenever h is a universally quantified formula and e is the conjunction of finitely many singular formulas.

[19] In the appendix of (1979) Popper moreover rejects the common sense 'bucket theory' of knowledge.

[20] Compare this to Quine's arguments in "Two dogmas of empiricism" [63], which led Quine to a holistic position.

[21] Many of Popper's ideas stem from (what since Kuhn is called) revolutionary science and this requires conceptual change. Yet Popper's formal criteria (like all approaches discussed in the previous sections) presuppose a given language.

falsifiable theories are deemed unscientific.[22] A theory is more falsifiable (has a higher degree of testability) to the extent that more logically possible facts contradict it. This brings Popper to two criteria: generality and specificity. A hypothesis is more general to the extent that it concerns a logically larger set of objects; it is more precise to the extent that it specifies more about those objects. To get the flavor: where P, Q, and R are logically independent predicates, $\forall x(Px \supset Qx)$ is more general than $\forall x((Px \wedge Rx) \supset Qx)$: the former is contradicted by every sentence of the form $P\alpha \wedge \neg Q\alpha$ whereas the latter is only contradicted by sentences of the form $P\alpha \wedge R\alpha \wedge \neg Q\alpha$; $\forall x(Px \supset (Qx \wedge Rx))$ is more specific than $\forall x(Px \supset Qx)$: the former is contradicted by sentences of the form $P\alpha \wedge \neg Q\alpha$ as well as by sentences of the form $P\alpha \wedge \neg R\alpha$, whereas the latter is only falsified by the former sentences.[23] Popper identifies the *content* of a sentence A with the falsifiability of A and measures it, for example, by $1 - \Pr(A)$, which is $\Pr(\neg A)$. (By the way, despite using probability to define the content of a sentence, Popper did not use probability to explicate the notion of confirmation.) Note that, where A and B are logically independent,[24] $A \wedge B$ has an intuitively higher content than A and indeed $\Pr(\neg(A \wedge B)) > \Pr(\neg A)$.

Needless to say, $m^{\dagger}$ is unable to capture the differences between the general sentences from the previous paragraph if the domain is infinite. All those sentences have probability zero. These probabilities are defined by a limit for the number of elements of the domain going to infinity. In appendix $^{*}vii$ of (1935), Popper introduces a "fine-structure of probability". Even if $\Pr(A) = \Pr(B) = 0$, it is possible that $\Pr(A \mid B) > \Pr(B \mid A)$, and this indicates that B has a higher content than A. A ready example is obtained by letting A be $\forall x(Px \supset Qx)$ and letting B be $\forall x(Px \supset (Qx \wedge Rx))$. In this case $1 = \Pr(A \mid B) > \Pr(B \mid A)$. A different way to look at the criterion is by noting that the limits of the probabilities of both A and B converge to zero as the domain increases, but that the ratio of these probabilities is always larger than 1 (and goes to infinity).

So the objective is clear: formulate and test *bold hypotheses*. If the hypothesis survives the tests, one obtains a corroborated informative hypothesis—see below. If it fails, one may still move to a non-falsified hypothesis that has the next highest content (degree of falsifiability). Of course, no single (non-falsified) hypothesis has the highest content. In the propagandistic style that was usual for those days, Popper does not stress this. Here (as elsewhere), he is a free-market pal: pick yours and go for it. The market (sorry, the facts) will decide.

To compare theories, Popper [59, 60] introduced (and in [62] elaborated on) the notion of *verismilitude* or *truthlikeness*. Its qualitative version (as opposed to the quantitative formulation, mentioned below) is as follows: take an interpreted theory

[22] But Popper hastens to relativize this 'demarcation criterion'. 'Metaphysical' ideas play a central role in generating scientific theories.

[23] Compare also "all heavenly bodies move in circles" to "all planets of the sun move in ellipses", remembering that all circles are ellipses.

[24] Note this entails they're contingent.

T and let T^1 (T^0) be the set of its true (false) sentences. T is more truthlike than a theory S iff both $S^1 \subseteq T^1$ and $T^0 \subseteq S^0$, and either $S^1 \neq T^1$ or $T^0 \neq S^0$ (That is, a theory to be more truthlike has to surpass the other in its truth content without surpassing it in its falsity content, or to have smaller falsity content without being ahead in its truth content.) Miller [52] and Tichý [74] provide a compelling criticism of Popper's definitions.[25]

Suppose then some theories survived the imposed tests. How good are they? Here too, Popper formulates a measure, which he calls the degree of corroboration of a hypothesis. Here is a definition from appendix $*ix$ of [61]:

$$C(H, E) = \frac{\Pr(E \mid H) - \Pr(E)}{\Pr(E \mid H) - \Pr(E \wedge H) + \Pr(E)}$$

So, where E is the conjunction of the available empirical evidence, the degree of corroboration of the hypothesis H is proportional to the difference between the probability of the evidence given the hypothesis and the absolute probability of the evidence. The denominator is a normalizing factor, which keeps the values between -1 and $+1$. If E contradicts H, $\Pr(E \mid H) = \Pr(E \wedge H) = 0$. So the degree of corroboration of H is -1. This indicates that H is falsified. The maximal value to which H may be corroborated is obtained if E is identical to H—this will apply if H is a singular statement or if, being God, you see that H obtains. In this case $\Pr(E/H) = 1$ and $\Pr(E \wedge H) = \Pr(H)$. The degree of corroboration of H then reduces to $1 - \Pr(H)/1$, in other words $\Pr(\neg H)$. So the maximal degree to which a hypothesis H may be corroborated is the content of H (the falsifiability degree of H). The higher the content of a hypothesis, the higher its potential degree of corroboration.

It is amusing to see that falsifiability turns up again here. Yet, putting the formal machinery in perspective, Popper stresses that the corroboration of H is only significant if H was subjected to the severest possible tests. We have seen before that these are the tests that are most likely to falsify the hypothesis. That Popper never offered a formal criterion for this, is presumably related to a weak spot in his formalisms. Intuitively, repeating an experiment that did not lead to falsification is not a severe test. But why is that? Apparently because, in view of previous instances of the test, the next instance is likely not to lead to falsification. But why is that so? Apparently this conclusion can only be drawn if we presume that the outcome of the next instance of the test is likely similar to the outcome of previous instances.

[25] Popper introduced also two quantitative notions of verisimilitude, which employed the notion of probability. Tichý [74] argues that both attempts have highly counterintuitive consequences. This is not to say that the project of defining truthlikeness is doomed. There are various interesting attempts to define the concept after Popper's initial failure (see e.g., [54, 58]). Even though no particular account is currently agreed on by everyone, certain progress has been made, and the issue is a lively topic (for a survey, see [55, 56]).

To presume so, however, is to presume a measure function different from Carnap's $m^{\dagger}$, viz. one that assigns a non-zero weight to the empirical factor. And this Popper does not want.

Indeed, Popper always stressed that a (non-falsifying) degree of corroboration of H should not affect our degree of rational belief in H. He nevertheless advised one to use the best tested theory as basis for action, and some take this to mean that the degree of rational belief of those theories is raised. In Section 9 of [62], Popper tried to remove this confusion. He distinguished preferring the best tested theory as basis for action from relying on that theory. Preferring such theory is justified, because of the merits the theory proved to have in the past. But this says nothing about the future. So we cannot rely on the theory; no theory was shown true or can be shown true. Our present most corroborated theories embody the best knowledge available today. Only fools take alternative theories as better. But even our best theories may be falsified tomorrow.

5.6 Inductive Generalization in Terms of a Logic

The approaches discussed before have clearly sensible application contexts. More problematic is their explicit or implicit claim on universality. Why, for example, should the decision to act on a certain scientific theory be arrived at by the same method as the decision to participate in a certain lottery?

Once a plurality of methods is accepted, there can hardly be any objection against phrasing some of these as logics: functions that assign a consequence set to every premise set. In the present section, we shall present such an approach, the one we are most familiar with: adaptive logics. The discussion will be restricted to logics of inductive generalization. These are logics that enable one to infer hypotheses of the form "all A are B" from sets of data, and which (most importantly) provide such consequence operation with a proof theory.[26]

An advantage of this approach to the 'acceptance' of scientific hypotheses is that it is more realistic than approaches in terms of degrees of (rational) belief. While data, hypotheses, and theories may be rejected in view of new observations or in view of a new systematization, they are provisionally considered as 'given'. Next, it is easily possible to consider scientific research as problem solving in the presence of provisionally accepted background knowledge. Moreover, there is no need to assign specific degrees of certainty to data, background knowledge, and inferred generalizations. Still, as we shall see, it is possible to express that some background theories are more 'reliable' than others.

[26]To complete the picture, we would need adaptive logics that enable one to derive hypotheses of the form $\Pr(A \mid B) = r$, in which Pr is an objective probability, and we would need adaptive logics that enable one to derive predictions that do not follow from derived general hypotheses.

5.6.1 An Example

A logic that enables one to derive general hypotheses from sets of data is obviously ampliative with respect to **CL** (Classical Logic); the derived general hypotheses are not derivable from the premises by **CL**. Moreover, such derivations are risky in several senses. For one thing, new evidence may become available and it may falsify some of the formerly derived hypotheses. Moreover, in the presence of background knowledge (formerly 'accepted' theories), it may be impossible to show, within a finite period of time, that a certain generalization is incompatible with the data and background knowledge.[27]

This situation has several consequences for the proofs of logics of inductive generalization. All ampliative conclusions drawn at some point in a proof, may later have to be revoked for one of the two reasons mentioned in the previous paragraph. This means that such proofs are dynamic and hence that one needs a device to control this dynamics. In adaptive logics, the control is exerted by, on the one hand, introducing ampliative conclusions on a (non-empty) condition and, on the other hand, providing a marking definition. At every stage of a proof, the marking definition settles which lines are marked and which unmarked. Marked lines are considered as OUT: the formula of a line that is marked at a stage of the proof is considered as not derived on that line at the stage. A stage of a proof is a sequence of lines that are correct according to the rules of the logic. A stage s' extends another stage s if all lines that occur in s occur in the same order in s'.

Marks may come and go with every new stage of the proof; a line may unmarked at a certain stage, marked at a later stage, unmarked again at a still later stage, and so on. So, apart from *derivability at a stage*, we need a stable notion of derivability, which is called *final derivability*. The premises may not enable one to show by means of a finite proof that a generalization is finally derived. If this is the case and we applied the right heuristic means, the proof provides us with a reason to *prefer* the derived generalizations as basis for action, but it does not allow us to *rely* on them—we borrow this distinction from Popper, as the reader will remember from Sect. 5.5.

Introducing adaptive logics in general would take too much space; we refer for example to [4, 7] for that. Here we shall start with a toy example proof and next introduce the machinery in as far as we need it. The idea behind the proof will be falsification. One may introduce any generalization in the proof on the condition that the negation of the generalization can be considered as false in view of the

[27] There is a mechanical procedure which for any particular inconsistent premise set will show that it is inconsistent. But there is no mechanical procedure which for any particular premise set will decide whether it is inconsistent. So, technically speaking, the set of inconsistent sets of formulas is semi-recursive (or semi-decidable) but not recursive (not decidable).

premises.[28] Let the premises be those of lines 1–4 of the proof, in which we also introduce some consequences.

1	$Pa \wedge Qa \wedge Sa$	premise	$\emptyset$
2	$Pb \wedge \neg Qb \wedge \neg Rb \wedge Sb$	premise	$\emptyset$
3	$\neg Sc$	premise	$\emptyset$
4	$\neg Pd$	premise	$\emptyset$
5	$\forall x(Px \supset Sx)$	RC	$\{\neg\forall x(Px \supset Sx)\}$
6	$\forall x(Px \supset Qx)$	RC	$\{\neg\forall x(Px \supset Qx)\}$
7	$\forall x(Px \supset (Qx \wedge Sx))$	5, 6; RU	$\{\neg\forall x(Px \supset Sx), \neg\forall x(Px \supset Qx)\}$
8	$\forall x Px$	RC	$\{\neg\forall x Px\}$
9	$\forall x \neg Rx$	RC	$\{\neg\forall x \neg Rx\}$
10	$\forall x(Qx \supset Rx)$	RC	$\{\neg\forall x(Qx \supset Rx)\}$
11	$\neg Pc$	3, 5; RU	$\{\neg\forall x(Px \supset Sx)\}$

Apart from the premise rule, two generic rules are used in the proof. The conditional rule RC may be read provisionally as: introduce any generalization A on the condition $\{\neg A\}$. The unconditional rule RU may be read as: if B is **CL**-derivable from $A_1, \ldots, A_n$ and $A_1, \ldots, A_n$ occur in the proof (on some conditions), one may derive B on the condition that is the union of the conditions of $A_1, \ldots, A_n$. Formulas introduced by the premise rule receive the empty set as their condition: premises need never be revoked. Note that 11 is a prediction derived from a derived generalization in view of the premises (the data).

The reader will have noted that some of the generalizations are falsified by the premises. In the subsequent extension of the proof, we show how this is handled. We do not repeat the premises.

5	$\forall x(Px \supset Sx)$	RC	$\{\neg\forall x(Px \supset Sx)\}$	
6	$\forall x(Px \supset Qx)$	RC	$\{\neg\forall x(Px \supset Qx)\}$	✓12
7	$\forall x(Px \supset (Qx \wedge Sx))$	5, 6; RU	$\{\neg\forall x(Px \supset Sx), \neg\forall x(Px \supset Qx)\}$	✓12
8	$\forall x Px$	RC	$\{\neg\forall x Px\}$	✓13
9	$\forall x \neg Rx$	RC	$\{\neg\forall x \neg Rx\}$	
10	$\forall x(Qx \supset Rx)$	RC	$\{\neg\forall x(Qx \supset Rx)\}$	
11	$\neg Pc$	3, 5; RU	$\{\neg\forall x(Px \supset Sx)\}$	
12	$\neg\forall x(Px \supset Qx)$	2; RU		
13	$\neg\forall x Px$	4; RU	$\emptyset$	

By deriving 12 from 2, we obtain a member of the conditions of lines 6 and 7. So the conditions of these lines cannot be considered as false, because a member of them has to be true if the premises are true. Similarly, line 8 is marked in view of line 13. These are plain cases of falsification.

There are also some unexpected features, which we illustrate in the following extension of the proof, in which we leave out the marked lines for reasons of pagination.

[28] We do not say "on the condition that the generalization is not falsified by the premises". Soon, the reason will become clear.

5	$\forall x(Px \supset Sx)$	RC	$\{\neg\forall x(Px \supset Sx)\}$	
...				
9	$\forall x\neg Rx$	RC	$\{\neg\forall x\neg Rx\}$	✓16
10	$\forall x(Qx \supset Rx)$	RC	$\{\neg\forall x(Qx \supset Rx)\}$	✓16
11	$\neg Pc$	3, 5; RU	$\{\neg\forall x(Px \supset Sx)\}$	
12	$\neg\forall x(Px \supset Qx)$	2; RU		
13	$\neg\forall x Px$	4; RU	$\emptyset$	
14	$Ra \vee \neg Ra$	RU	$\emptyset$	
15	$Ra \vee (Qa \wedge \neg Ra)$	1, 14; RU	$\emptyset$	
16	$\neg\forall x\neg Rx \vee \neg\forall x(Qx \supset Rx)$	15; RU	$\emptyset$	
17	$Pc \vee \neg Pc$	RU	$\emptyset$	
18	$(Pc \wedge \neg Sc) \vee \neg Pc$	3, 17; RU	$\emptyset$	
19	$\neg\forall x(Px \supset Sx) \vee \neg\forall x Px$	18; RU	$\emptyset$	

Note that 14 and 17 are theorems of **CL** and that RU allows one to introduce these anywhere in a proof. The first interesting case is line 16, at which a disjunction of negations of generalizations is derived. What this tells us is that either $\neg\forall x\neg Rx$ or $\neg\forall x(Qx \supset Rx)$ is true, but is does not tell us which of them is true. So should we mark lines 9 and 10 or not? We cannot have both unmarked because they jointly contradict the premises (as line 16 shows). And no logical consideration allows us to mark one rather than the other. So we had better mark both lines.

Why does line 19 not lead to marking line 5 (together with line 8)? Line 19 tells us that either $\neg\forall x(Px \supset Sx)$ or $\neg\forall x Px$ is true and it also does not tell us which of them is true. But line 13 tells us that: $\neg\forall x Px$ is true (on these premises). So $\neg\forall x(Px \supset Sx)$ is off the hook: if we know that A as well as $A \vee B$ are true and start to consider as many members of $\{A, B\}$ as false as is possible, we can safely consider B as false.

For the present logic, negations of generalizations are the *abnormalities*; the formulas of which as many as possible are considered as false. Let the set $U_s(\Gamma)$ comprise all disjuncts of minimal disjunctions of abnormalities that are derived on the condition $\emptyset$ at stage s of the proof. The members of $U_s(\Gamma)$ are the abnormalities that are *unreliable* at stage s. The marking definition goes as follows: a line is marked at a stage s if and only if its condition contains a member of $U_s(\Gamma)$.[29]

In general, an adaptive logic in standard format is defined as a triple: a lower limit logic (required to have certain properties), a set of abnormalities characterized by a logical form, and an adaptive strategy. The logic informally introduced before is called $\mathbf{LI}^r$—see [5] and [8]. Its lower limit logic is **CL**, its set of abnormalities is the set comprising all negations of generalizations, and its strategy is Reliability.

We promised to introduce final derivability. Let A be the formula of line l of a stage s of a proof from Γ. A is finally derived from Γ at line l if and only if l is

[29] This is the so-called Reliability strategy. The Minimal Abnormality strategy is slightly different from Reliability and offers a few more consequences than Reliability (and never less consequences). We shall not introduce it here.

unmarked at stage s and every stage s' that extends s and in which l is marked, can be further extended in such a way that l is unmarked. It is handy to see this in game-theoretic terms: the proponent writes s, the opponent extends it to s', and then the proponent is allowed to further extend s'. For Reliability both extensions are finite.

Adaptive logics also have a semantics. Consider the minimal disjunctions of abnormalities that are **CL**-derivable from Γ and let $U(\Gamma)$ be the set of their disjuncts. A reliable model of Γ is a **CL**-model of Γ that verifies no other abnormalities than members of $U(\Gamma)$. For logics that have Reliability as their strategy, semantic consequences of Γ are the formulas verified by all Reliable models of Γ. An interesting feature of the (occasionally mentioned) standard format is that it provides every adaptive logic in standard format with its proof theory, its semantics, soundness and completeness proofs, and proofs for lots of other metatheoretic properties. The standard format provides also certain criteria for final derivability: if one follows a certain (proof or tableau) procedure, final derivability will, for some premise sets and consequences, be established after finitely many steps.

Back to inductive generalization. Basic moves underlying $\mathbf{LI}^r$ can arguably provide guidance in actual research. The logic suggests one to obtain certain observations, possibly by experimental means. These observations might falsify certain generalizations. If that happens, shorter disjunctions of abnormalities become derivable, and by the same token, other, often more specific generalizations may become derivable.

5.6.2 Some Alternatives

Logics should not make methodological decisions, but should offer means to express methods in a precise way. So there should be many adaptive logics of inductive generalization, from which a scientist may choose on methodological grounds. This is indeed the case.

A first series of them is obtained by varying the elements of the adaptive logic. The Minimal Abnormality strategy was mentioned in Footnote 29; no other strategies seem sensible in the present context. Variants of the set of abnormalities have been studied. For some logics, this set comprises the formulas of the form $\exists x A(x) \wedge \exists x \neg A(x)$ in which $A(x)$ is a disjunction of literals. The effect of this change is a logic richer than $\mathbf{LI}^r$ in which $\forall x(Px \supset Qx)$ can only be introduced in a proof if a formula of the form $\neg P\alpha \vee Q\alpha$ (an instance of the generalization in the logician's sense) is derivable from the premises. A still richer logic is obtained if the set of abnormalities comprises the formulas of the form $\exists x(A(x) \wedge \pi x) \wedge \exists x(A(x) \wedge \neg \pi x)$, In which πx is a literal and $A(x)$ is a conjunction of literals. In this case $\forall x(Px \supset Qx)$ can only be introduced in a proof if a 'positive instance' of it is available, a formula of the form $P\alpha \wedge Q\alpha$. One might of course also vary the lower limit logic.

A different series of variants is obtained by combining adaptive logics (not necessarily those mentioned so far). There are several ways to do so. One of them goes as follows: first apply an adaptive logic that has one of the above sets of abnormalities restricted to the case where only one predicate occurs in the abnormality; to the resulting consequence set, apply an adaptive logic that has abnormalities in which at most two predicates occur; and so on. This leads to a serious enrichment and agrees with Poppers requirement that one should first test hypotheses that have the highest content.

An interesting extension is where the person applying the logic is allowed to introduce certain preferences. A scientist may have several reasons to do so, going from relying on 'established' science to personal preferences and mere guesses. These preferences may be expressed by defeasibly denying certain abnormalities in a prioritized way: this abnormality is almost certainly false, that one is probably false, etc. The priorities are expressed in the language by operators, which basically have a comparative effect. Several adaptive logics to handle such prioritized rejections are available in the literature [4, 6, 10] and each of them can be combined with an adaptive logic for inductive inference.

Most realistic applications require that handling background knowledge is combined with inductive inference. Background knowledge will drastically extend the data, but it may be falsified by them. Sometimes this is a reason to reject the whole theory, sometimes this is a reason to reject only falsified consequences of the theory and provisionally go with the others, hoping for a new systematization in the future. Again, adaptive logics to handle both cases (even jointly) are available. For a more complete overview of materials in this section and for further references we refer to Batens [6].

References

1. Armendt, B. (1980). Is there a Dutch Book argument for probability kinematics? *Philosophy of Science, 47*, 583–589.
2. Ayer, A. (1972). *Probability and evidence*. London: Macmillan.
3. Batens, D. (1975). *Studies in the logic of induction and in the logic of explanation containing a new theory of meaning relations*. Brugge: De Tempel.
4. Batens, D. (2005). On a logic of induction. In R. Festa, A. Aliseda, & J. Peijnenburg (Eds.), *Confirmation, empirical progress, and truth approximation* (Essays in debate with Theo Kuipers, Vol. 1, pp. 221–242). Amsterdam/New York: Rodopi.
5. Batens, D. (2006). On a logic of induction. *L&PS – Logic & Philosophy of Science, IV*(1), 3–32.
6. Batens, D. (2011). Logics for qualitative inductive generalization. *Studia Logica, 97*(1), 61–80.
7. Batens, D. (2018). *Adaptive logics and dynamic proofs. Mastering the dynamics of reasoning*. [Manuscript in progress, available online at http://logica.ugent.be/adlog/book.html].
8. Batens, D., & Haesaert, L. (2001). On classical adaptive logics of induction. *Logique et Analyse, 173–175*, 255–290. Appeared 2003.

9. Batens, D., & Haesaert, L. (2003). On classical adaptive logics of induction. *Logique et Analyse, 46*, 225–290.
10. Bird, A. (1998). *Philosophy of science*. New York: Routledge.
11. Carnap, R. (1950). *Logical foundations of probability*. London: Routledge.
12. Carnap, R. (1952). *The continuum of inductive methods*. Chicago: University of Chicago Press.
13. Carnap, R. (1959). *Induktive Logik und Wahrscheinlichkeit*. Wolfgang Stegmüller.
14. Christensen, D. (2004). *Putting logic in its place: Formal constraints on rational belief*. Oxford: Clarendon Press.
15. Constantini, D. (1983). Analogy by similarity. *Erkenntnis, 20*, 103–114.
16. De Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré, 7*, 1–68. (Translated as "Foresight: Its Logical Laws, Its Subjective Sources" in Kyburg & Smokler (1964)).
17. Eberhardt, F., & Glymour, C. (2009). Hans Reichenbach's probability logic. In J. W. Stephan Hartmann & D. Gabbay (Eds.), *Handbook of the history of logic. Volume 10: Inductive logic* (pp. 357–389). Amsterdam/Boston: Elsevier.
18. Eells, E. (2005). Confirmation theory. In J. Pfeifer & S. Sarkar (Eds.), *Philosophy of science: An encyclopedia* (pp. 144–150). New York: Routledge.
19. Fitelson, B. (2001). *Studies in Bayesian confirmation theory*. Ph.D. thesis, University of Wisconsin-Madison.
20. Fitelson, B. (2005). Inductive logic. In J. Pfeifer & S. Sarkar (Eds.), *Philosophy of science: An encyclopedia* (pp. 384–394). New York: Routledge.
21. Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. In J. Earman (Ed.), *Testing scientific theories. Midwest studies in the philosophy of science* (Vol. X, pp. 99–131). Minneapolis: University of Minnesota Press.
22. Glaister, S. (2002). Inductive logic. In D. Jacquette (Ed.), *A companion to philosophical logic* (pp. 565–581). Malden: Blackwell.
23. Glymour, C. (1980). *Theory and evidence*. Princeton: Princeton University Press.
24. Good, I. (1968). Corroboration, explanation, evolving probability, simplicity, and a sharpened razor. *British Journal of Philosophy of Science, 19*, 123–143.
25. Good, I. (1985). A historical comment concerning novel confirmation. *British Journal of Philosophy of Science, 36*, 184–186.
26. Goodman, N. (1946). A query on confirmation. *Journal of Philosophy, 43*, 383–385.
27. Goodman, N. (1978). *Fact, fiction and forecast*. New Yor: The Bobbs-Merrill Company, Inc.
28. Hájek, A. (2010). Interpretations of probability. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Spring 2010 edition. Stanford: The Metaphysics Research Lab, Stanford University.
29. Harrison, V., & Chandler, J. (Eds.). (2012). *Probability in the philosophy of religion*. Oxford: Oxford University Press.
30. Hintikka, J. (1966). A two-dimensional continuum of inductive methods. In J. Hick (Ed.), *Aspects of inductive logic* (pp. 113–132). Amsterdam: North Holland Publishing Company.
31. Hintikka, J., & Niiniluoto, I. (1980). An axiomatic foundation for the logic of inductive generalization. In R. Jeffrey (Ed.), *Studies in inductive logic and probability* (pp. 157–181). Berkeley: University of California Press.
32. Huber, F. (2007). Confirmation and induction. In J. Fieser & B. Dowden (Eds.), *Internet encyclopedia of philosophy*.
33. Hume, D. (1739). *A treatise of human nature* (2nd ed.). Clarendon Press; 1985 edition, edited by L.A. Selby-Bigge.
34. Jaynes, E. (2003). *Probability theory: The logic of science*. Cambridge: Cambridge University Press.
35. Jeffrey, R. (1965). *The logic of decision*. Chicago: University of Chicago.
36. Johnson, W. E. (1924). *Logic, part III. The logical foundations of science*. Cambridge: Cambridge University Press.
37. Johnson, W. E. (1932). Probability: The deductive and inductive problems. *Mind, 41*, 409–423.

38. Juhl, C. (1994). The speed-optimality of Reichenbach's straight rule of induction. *British Journal of Philosophy of Science, 45*, 857–863.
39. Kemeny, J. (1953). A logical measure function. *Journal of Symbolic Logic, 18*, 289–308.
40. Kemeny, J. G. (1953). Review of Carnap (1952). *Journal of Symbolic Logic, 18*, 168–169.
41. Kemeny, J. (1955). Fair bets and inductive probabilities. *Journal of Symbolic Logic, 20*, 263–273.
42. Kemeny, J., & Oppenheim, P. (1952). Degrees of factual support. *Philosophy of Science, 19*, 307–324.
43. Keynes, J. (1921). *Treatise on probability*. London: Macmillan. (Reprinted in 1962).
44. Kolmogorov, A. N. (1956). *Foundations of the theory of probability*. New York: Chelsea Publishing Company. (Translated by Nathan Morrison; first edition published in 1950; the original published in 1933 in the Ergebnisse Der Mathematik).
45. Kuipers, T. (1984). Two types of analogy by similarity. *Erkenntnis, 21*, 63–87.
46. Kyburg, H. E. (1970). *Probability and inductive logic*. London: Macmillan.
47. Kyburg, H., & Smokler, H. (Eds.). (1964). *Studies in subjective probability*. New York: Wiley.
48. Lange, M. (1996). Calibration and the epistemological role of Bayesian conditionalization. *Journal of Philosophy, 96*, 294–324.
49. Lehman, R. S. (1955). On confirmation and rational betting. *Journal of Symbolic Logic, 20*(3), 251–262.
50. Maher, P. (2000). Probabilities for two properties. *Erkenntnis, 52*, 63–91.
51. Maher, P. (2001). Probabilities for multiple properties: The models of Hesse and Carnap and Kemeny. *Erkenntnis, 55*, 183–216.
52. Miller, D. (1974). Popper's qualitative theory of verisimilitude. *British Journal for the Philosophy of Science, 25*(2), 166–177.
53. Niiniluoto, I. (1981). Analogy and inductive logic. *Erkenntnis, 16*, 1–34.
54. Niiniluoto, I. (1987). *Truthlikeness*. Dordrecht: Reidel.
55. Niiniluoto, I. (1998). Verisimilitude: The third period. *British Journal for the Philosophy of Science, 49*(1), 1–29.
56. Niiniluoto, I. (2005). Verisimilitude. In J. Pfeifer & S. Sarkar (Eds.), *Philosophy of science: An encyclopedia* (pp. 854–857). New York: Routledge.
57. Niiniluoto, I. (2009). The development of the Hintikka program. In Dov M. Gabbay, Stephan Hartmann, & John Woods (Eds.), *Handbook of the history and philosophy of logic* (Inductive logic, Vol. 10 pp. 265–309). Elsevier: Amsterdam.
58. Oddie, G. (1986). *Likeness to truth*. Dordrecht: Reidel.
59. Popper, K. (1962). Some comments on truth and the growth of knowledge. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress* (pp. 285–292). Stanford: Stanford University Press.
60. Popper, K. (1963). *Conjectures and refutations*. London: Routledge.
61. Popper, K. R. (1935). *Logik der Forschung*. Wien: Verlag von Julius Springer.
62. Popper, K. R. (1979). *Objective knowledge: An evolutionary approach*. Oxford: Oxford University Press.
63. Quine, W. V. O. (1953). *From a logical point of view*. Cambridge, MA: Harvard University Press.
64. Ramsey, F. (1978). Truth and probability. In D. H. Mellor (Ed.), *Foundations: Essays in philosophy, logic, mathematics and economics* (pp. 58–100). London: Routledge. [Originally published in 1926].
65. Reichenbach, H. (1938). *Experience and prediction*. Chicago: University of Chicago Press.
66. Rosenkrantz, R. (1981). *Foundations and applications of inductive probability*. Atascadero: Ridgeview Publishing.
67. Salmon, W. (1966). *The foundations of scientific inference*. Pittsburgh: University of Pittsburgh Press.
68. Savage, L. (1954). *The foundations of statistics*. New York: Wiley.
69. Shimony, A. (1955). Coherence and the axioms of confirmation. *Journal of Symbolic Logic, 20*(1), 1–28.

70. Skyrms, B. (1984). *Pragmatics and empiricism*. New Haven: Yale University Press.
71. Skyrms, B. (1993). Analogy by similarity in hyperCarnapian inductive logic. In J. Earman, A. I. Janis, G. J. Massey, & N. Rescher (Eds.), *Philosophical problems of the internal and external worlds* (pp. 273–282). Pittsburgh: University of Pittsburgh Press.
72. Talbott, W. (2008). Bayesian epistemology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Fall 2008 edition. Stanford: The Metaphysics Research Lab, Stanford University.
73. Teller, P. (1976). Conditionalization, observation, and change of preference. In *Foundations of probability theory, statistical inference, and statistical theories of science*. Dordrecht: Reidel.
74. Tichý, P. (1974). On Popper's definitions of verisimilitude. *British Journal for the Philosophy of Science, 25*(2), 155–160.
75. Vickers, J. (2010). The problem of induction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Fall 2010 edition. Stanford: The Metaphysics Research Lab, Stanford University.
76. Williamson, J. (2010). *In defence of objective Bayesianism*. Oxford: Oxford University Press.
77. Zabell, S. (1996). Confirming universal generalizations. *Erkenntnis, 45*, 267–283.
78. Zabell, S. (1997). The continuum of inductive methods revisited. In J. Earman & J. D. Norton (Eds.), *The cosmos of science* (pp. 351–385). Pittsburgh: University of Pittsburgh Press.
79. Zabell, S. L. (2009). Carnap and the logic of inductive inference. In J. W. Stephan Hartmann & D. Gabbay (Eds.), *Handbook of the history and philosophy of logic* (Inductive logic, Vol. 10, pp. 265–309). Elsevier: Amsterdam.

Recommended Readings

1. Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.
2. Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge: Cambridge University Press.
3. Hájek, A., & Hall, N. (2002). Induction and probability. In P. K. Machamer & M. Silberstein (Eds.), *The Blackwell guide to the philosophy of science* (pp. 149–172). Malden: Blackwell.
4. Kuipers, T. (1978). *Studies in inductive probability and rational experimentation*. Dordrecht: Reidel.
5. Pettigrew, R. (2016). Epistemic utility arguments for probabilism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Spring 2016 edition. Stanford: Metaphysics Research Lab, Stanford University.
6. Reichenbach, H. (1949). *The theory of probability*. Berkeley: University of California Press.