THE DEVELOPMENT OF THE HINTIKKA PROGRAM

Ilkka Niiniluoto

One of the highlights of the Second International Congress for Logic, Methodology, and Philosophy of Science, held in Jerusalem in 1964, was Jaakko Hintikka's lecture "Towards a Theory of Inductive Generalization" (see [Hintikka, 1965a]). Two years later Hintikka published a two-dimensional continuum of inductive probability measures (see Hintikka, 1966), and ten years later he announced an axiomatic system with K $\geq$2 parameters (see [Hintikka and Niiniluoto, 1976]). These new original results showed once and for all the possibility of systems of inductive logic where genuine universal generalizations have non-zero probabilities in an infinite universe. Hintikka not only disproved Karl Popper's thesis that inductive logic is inconsistent (see [Popper, 1959; 1963]), but he also gave a decisive improvement of the attempts of Rudolf Carnap to develop inductive logic as the theory of partial logical implication (see [Carnap, 1945; 1950; 1952]). Hintikka's measures have later found rich applications in semantic information theory, theories of confirmation and acceptance, cognitive decision theory, analogical inference, theory of truthlikeness, and machine learning. The extensions and applications have reconfirmed — *pace* the early evaluation of Imre Lakatos [1974] — the progressive nature of this research program in formal methodology and philosophy of science.

1 INDUCTIVE LOGIC AS A METHODOLOGICAL RESEARCH PROGRAM

Imre Lakatos [1968a] proposed that Carnap's inductive logic should be viewed as a methodological "research programme". Such programs, both in science and methodology, are characterized by a "hard core" of basic assumptions and a "positive heuristics" for constructing a refutable "protective belt" around the irrefutable core. Their progress depends on the problems that they originally set out to solve and later "problem shifts" in their dynamic development. In a paper written in 1969, Lakatos claimed that Popper has achieved "a complete victory" in his attack against "the programme of an *a priori* probabilistic inductive logic or confirmation theory" - although, he added, "inductive logic, displaying all the characteristics of a degenerating research programme, is still a booming industry" [Lakatos, 1974, p. 259].[1] A similar position is still advocated by David Miller [1994], one of the leading Popperian critical rationalists.

[1] If the emphasis is on the term "a priori", Hintikka agrees with Lakatos. However, we shall see that Hintikka's "victory" over Carnap is completely different from Popper's.

Handbook of the History of Logic. Volume 10: Inductive Logic. Volume Editor: Dov M. Gabbay, Stephan Hartmann and John Woods.
General editors: Dov M. Gabbay, Paul Thagard and John Woods.

In this paper, I argue for a different view about inductive logic (see also [Niiniluoto, 1973; 1983]). Hintikka's account of inductive generalization was a meeting point of several research traditions, and as a progressive turn it opened new important paths in Bayesian methodology, epistemology, and philosophy of science. Its potential in artificial intelligence is still largely unexplored.

The roots of inductive logic go back to the birth of probability calculus in the middle of the seventeenth century. Mathematical probabilities were interpreted as relative frequencies of repeatable events, as objective degrees of possibility, and as degrees of certainty [Hacking, 1975]. For the classical Bayesians, like the determinist P. S. Laplace in the late eighteenth century, probability was relative to our ignorance of the true causes of events. The idea of probabilities as rational *degrees of belief* was defended in the 19^{th} century by Stanley Jevons. This Bayesian interpretation was reborn in the early 20^{th} century in two different forms. The Cambridge school, represented by W. E. Johnson, J. M. Keynes, C. D. Broad, and Harold Jeffreys, treated inductive probability $P(h/e)$ as a logical relation between two sentences, a hypothesis h and evidence e. Related logical views, anticipated already by Bernard Bolzano in the 1830s, were expressed by Ludwig Wittgenstein and Friedrich Waismann. The school of subjective or personal probability interpreted degrees of belief as coherent betting ratios (Frank Ramsey, Bruno de Finetti) (see [Skyrms, 1986]).

In Finland, pioneering work on "probability logic" in the spirit of logical empiricism was published by Eino Kaila in the 1920s (see [Kaila, 1926]). Kaila's student Georg Henrik von Wright wrote his doctoral dissertation on "the logical problem of induction" in 1941 (see [von Wright, 1951; 1957]). Von Wright, with influences from Keynes and Broad, tried to solve the problem of inductive generalization by defining the probability of a universal statement in terms of relative frequencies of properties (see [Hilpinen, 1989; Festa, 2003; Niiniluoto, 2005c]). Von Wright was the most important teacher of Jaakko Hintikka who, in turn with his students, continued the Finnish school of induction.

Rudolf Carnap came relatively late to the debates about probability and induction. Karl Popper had rejected induction in *Logik der Forschung* in 1934 (see [Popper, 1959]), and he was also sharply critical of the frequentist probability logic of Hans Reichenbach. Kaila had sympathies with Reichenbach's empiricist approach. In a letter to Kaila on January 28, 1929, Carnap explained that he would rather seek a positivist solution where "probability inferences are equally analytic (tautologous) as other (syllogistic) inferences" (see [Niiniluoto, 1985/1986]). Carnap — influenced by the ideas of Keynes, Jeffreys, and Waismann on objective inductive probabilities, and to the disappointment of Popper — started to develop his views about probability in 1942-44 (see [Carnap, 1945]).

Carnap's *Logical Foundations of Probability* (*LFP*, 1950) gave a detailed and precise account of the inductive probability measure c^*, which is a generalization of Laplace's famous "rule of succession". In 1952 Carnap published *A Continuum of Inductive Methods*. Its class of measures, defined relative to one real-valued parameter λ, contained c^* as a special case only. Another special case c^+, proposed by Bolzano and Wittgenstein, was rejected by Carnap since it does not make learning from experience possible. The same point had been expressed in the 19^{th} century by George Boole and Charles Peirce in their criticism of Bayesian probabilities. With John Kemeny, Carnap further showed how the λ-continuum can be justified on an axiomatic basis (see [Kemeny, 1963]). Part of Carnap's counterattack to Popper's praise of improbability was based on the new exact theory of semantic information that he developed with Yehoshua Bar-Hillel (see [Carnap and Bar-Hillel, 1952]).

Ian Hacking [1971] has argued that the typical assumptions of Carnap's inductive logic can be found already in the works of G. W. F. Leibniz in the late 17^{th} century:

(L1) There is such a thing as non-deductive evidence.

(L2) 'Being a good reason for' is a relation between propositions.

(L3) There is an objective and formal measure of the degree to which one proposition is evidence for another.

These assumptions can be found also in Keynes' *A Treatise on Probability* (1921). Carnap formulated them in the Preface to *LFP* as follows:

(C1) All inductive reasoning is reasoning in terms of probability.

(C2) Inductive logic is the same as probability logic.

(C3) The concept of inductive probability or degree of confirmation is a logical relation between two statements or propositions, a hypothesis and evidence.

(C4) The frequency concept of probability is used in statistical investigations, but it is not suitable for inductive logic.

(C5) All principles and theorems of inductive logic are analytic.

(C6) The validity of induction is not dependent upon any synthetic presuppositions.

The treatment of probabilities of the form $P(h/e)$, where h is a hypothesis and e is evidence, connects Carnap to the Bayesian tradition. Against the subjectivist school, Carnap's intention was to eliminate all "psychologism" from inductive logic — just as Gottlob Frege had done in the case of deductive logic. Carnap's C4 accepts a probabilistic dualism with both physical and epistemic probabilities. By C5 and C6, probability as partial entailment is independent on all factual assumptions. In practical applications of inductive logic, degrees of confirmation $P(h/e)$ have to be calculated relative to the total evidence e available to scientists.

Carnap's commitment to probabilistic induction (C1 and C2) leaves open the question whether the basic notion of induction is *support* (e confirms h) or *acceptance* (h is rationally acceptable on e). According to Carnap, David Hume was right in denying the validity of inductive inferences, so that the proper task of inductive logic is to evaluate probabilities of the form $P(h/e)$. Such probabilities can then be used in practical decision making by applying the rule of Maximizing Expected Utility (cf. [Stegmüller, 1973]). In normative decision theory, there are "quasi-psychological" counterparts to "purely logical" inductive probabilities (see [Carnap, 1971]). Carnap was followed by Richard Jeffrey in the denial of inductive acceptance rules (cf. the debate in [Lakatos, 1968b]).

In the second 1962 edition of *LFP,* Carnap defended himself against the bitter attacks of Popper by distinguishing two senses of "degree of confirmation": posterior probability $P(h/e)$ and increase of probability $P(h/e) - P(h)$. This was a clarification of the core assumption C3.

In *LFP*, Carnap demanded that inductive logic should give an account of the following types of cases:

(a) *Direct inference*: from a population to a sample

(b) *Predictive inference*: from a sample to another sample

(c) *Inference by analogy*: from one individual to another by their known similarity

(d) *Inverse inference*: from a sample to a population

(e) *Universal inference*: from a sample to a universal hypothesis.

He showed how the measure c^* helps to solve these problems. But Carnap did not wish to claim that c^* is "perfectly adequate" or the "only adequate" explicatum of inductive probability (*LFP*, p. 563). So his method can be characterized by the following heuristic principle:

(C7) Use logic to distinguish alternative states of affairs that can be expressed in a given formal language L. Then define inductive probabilities for sentences of L by taking advantage of symmetry assumptions concerning such states of affairs.

The systematic applications of C7 distinguish the Carnapian program of inductive logic from the more general Bayesian school which admits all kinds of prior probability measures.

As we shall see in the next section, Hintikka's work on inductive logic relies on the heuristic principle C7 in a novel way, so that the problem (e) of universal inference gets a new solution. Hintikka's system satisfies the core assumptions C1, C2, and C4. But, in his reply to Mondadori [1987], Hintikka himself urges that his studies did not just amount to "tinkering with Carnap's inductive logic" or removing some "anomalies" from it, but rather "it means to all practical purposes a refutation of Carnap's philosophical program in developing his inductive logic" [Hintikka, 1987b]. What Hintikka has in mind is the "logicism" involved in the Carnapian core assumptions C3, C5, and C6.

Hintikka's own move is to replace C3 and C6 with more liberal formulations:

(C3′) Inductive probability $P(h/e)$ depends on the logical form of hypothesis h and evidence e.

(C6′) Inductive probabilities, and hence inductive probabilistic inferences, may depend on extra-logical factors.

Here C6′ allows that inductive inferences may have contextual or "local" presuppositions (cf. [Bogdan, 1976]). Inductive probability is thus not a purely syntactical or semantical notion, but its explication involves pragmatic factors. However, in the spirit of what Hintikka calls "logical pragmatics", C3′ and C6′ should be combined with C7 so that the dependence and interplay of logical and extra-logical factors is expressed in an explicit and precise way. Then it turns out that C5 is ambiguous: some principles of induction may depend on pragmatic boundary conditions (like the extra-logical parameters), while some mathematical theorems of inductive logic turn out to be analytically true.

2 FROM CARNAP TO HINTIKKA'S TWO-DIMENSIONAL CONTINUUM

In inductive logic, probabilities are at least partly determined by symmetry assumptions concerning the underlying language [Carnap, 1962; Hintikka and Suppes, 1966; Niiniluoto and Tuomela, 1973]. In Carnap's λ-continuum the probabilities depend also on a free parameter λ which indicates the weight given to logical or language-dependent factors over and above purely empirical factors (observed frequencies) (see [Carnap, 1952]). Carnap's λ serves thus as an index of caution in singular inductive inference. In Hintikka's 1966 system one further parameter α is added to regulate the speed in which positive instances increase the probability of a generalization.

More precisely, let $Q_1, ..., Q_K$ be a K-fold classification system with mutually exclusive predicates, so that every individual in the universe U has to satisfy one and only one Q-predicate. A typical way of creating such a classification system is to assume that a finite monadic first-order language L contains k basic predicates $M_1, ..., M_k$, and each *Q-predicate* is defined by a k-fold conjunction of positive or negative occurrences of the M-predicates: $(\pm)M_1x\&...\&(\pm)M_kx$. Then $K = 2^k$. Each predicate expressible in language L is definable as a finite disjunction of Q-predicates. Carnap generalized this approach to the case where the dichotomies $\{M_j, \sim M_j\}$ are replaced by families of mutually exclusive predicates $\mathbf{M_j} = \{\mathbf{M_{j1}}, ..., \mathbf{M_{jm}}\}$, and a Q-predicate is defined by choosing one element from each family $\mathbf{M}_j$ (see [Jeffrey, 1980]). For example, one family could be defined by colour predicates, another by a quantity taking discrete values (e.g., age).

Assume that language L contains N individual names $a_1, ..., a_N$. Let L be interpreted in universe U with size N, so that each object in U has a unique name in L. A *state description* relative to individuals $a_1, ..., a_N$ tells for each a_i which Q-predicate it satisfies in universe U. A *structure description* tells how many individuals in U satisfy each Q-predicate. Every sentence within this first-order monadic framework L can be expressed as a disjunction of state descriptions; in particular, a structure description is a disjunction of state descriptions that can be obtained from each other just by permuting individual constants. The state descriptions in L that entail sentence h constitute the *range* $R(h)$ of h. *Regular* probability measures m for L define a non-zero probability $m(s)$ for each state description s of L. For each sentence h in $L, m(h)$ is the sum of all measures $m(s)$, $s \in R(h)$. A regular confirmation function c is then defined as conditional probability:

(1) $$c(h/e) = \frac{m(h\&e)}{m(e)}.$$

Let now e_n describe a sample of n individuals in terms of the Q-predicates, and let $n_i \geq 0$ be the observed number of individuals in cell Q_i (so that $n_1 + ... + n_K = n$). Carnap's λ-continuum takes the posterior probability $c(Q_i(a_{n+1})/e_n)$ that the next individual a_{n+1} will be of kind Q_i to be

(2) $$\frac{n_i + \lambda/K}{n + \lambda}.$$

This value is known as the *representative function* of an inductive probability measure. The probability (2) is a weighted average of n_i/n (observed relative frequency of individuals in Q_i) and $1/K$ (the relative width of predicate Q_i). The choice $\lambda = \infty$ gives Reichenbach's Straight Rule, which allows only the empirical factor n_i/n to determine posterior probability. The choice $\lambda = 4$ would give the range measure proposed in Wittgenstein's *Tractatus*, which divides probability evenly to state descriptions, but it makes the inductive probability (2) equal to $1/K$ which is *a priori* independent of the evidence e and, hence, does not allow for the learning from experience. When $\lambda < \infty$, predictive probability is asymptotically determined by the empirical factor:

(3) $$[c(Q_i(a_{n+1})/e_n) - n_i/n] \to 0, \text{ when } n \to \infty.$$

Principle (3) is known as *Reichenbach's Axiom* [Carnap and Jeffrey, 1971; Kuipers, 1978b]. It is known that (3) implies the principle of *Positive Instantial Relevance*:

(4) $$c(Q_i(a_{n+2})/e_n\&Q_i(a_{n+1})) > c(Q_i(a_{n+1})/e_n).$$

The choice $\lambda = K$ in (2) gives Carnap's measure c^*, which allocates probability evenly to all structure descriptions. The formula

(5) $$c^*(Q_i(a_{n+1}/e_n) = \frac{n_i + 1}{n + K}$$

includes as a special case ($n_i = n, K = 2$) Laplace's Rule of Succession

(6) $$\frac{n+1}{n+2}.$$

Laplace derived this probability of the next favourable instance after n positive ones by assuming that all structural compositions of an urn with white and black balls are equally probable.

If the Q-predicates are defined so that they have different relative widths q_i, such that $q_1 + ... + q_K = 1$, then (2) is replaced by

$$(2') \qquad \frac{n_i + q_i\lambda}{n+\lambda}.$$

[Carnap and Stegmüller, 1959; Carnap, 1980]. (2) is obtained from (2′) by choosing $q_i = 1/K$ for all $i = 1, ..., K$.[2]

If the universe U is potentially infinite, so that its size N may grow without limit, the probability $c(h/e)$ is defined as the limit of the values (1) in a universe of size N (when $N \to \infty$). Then it turns out that all measures of Carnap's λ-continuum assign the probability zero to universal generalizations h on singular evidence e. Carnap admitted that such a result "may seem astonishing at first sight", since in science it has been traditional to speak of "well-confirmed laws" (see [Carnap, 1945]). But he immediately concluded that "the role of universal sentences in the inductive procedures of science has generally been overestimated", and proposed to measure the *instance confirmation* of a law h by the probability that a new individual not mentioned in evidence e fulfills the law h.

Carnap's attempted reduction of universal inference to predictive singular inference did not convince all his colleagues. In Lakatosian terms, this move was a regressive problem-shift in the Carnapian program. One of those who criticized Carnap's proposal was G. H. von Wright [1951a]. Von Wright knew, on the basis of Keynes, that universal generalizations h can be confirmed by positive singular evidence $e_n = i_1\&...\&i_n$ entailed by h if two conditions are satisfied: (i) the prior probability $P(h)$ is not minimal, and (ii) new confirmations of h are not maximally probable relative to previous confirmations [von Wright, 1951b]. The *Principal Theorem of Confirmation* thus states the following:

(7) If $P(h) > 0$ and $P(i_{n+1}/e_n) < 1$, then $P(h/e_n\&i_{n+1}) > P(h/e_n)$.

As Carnap's system does not satisfy this theorem, it is "no Confirmation-Theory at all" (see [von Wright, 1957, pp. 119, 215]). It also fails to solve the dispute of Keynes and Nicod about the conditions for the convergence of posterior probability to its maximum value one with increasing positive evidence:

(8) $P(h/e_n) \to 1$, when $n \to \infty$.

It is remarkable that Popper, the chief opponent of inductive logic, also argued for the zero logical probability of universal laws, i.e., the same result that shattered Carnap's system (see [Popper, 1959, appendices vii and viii]). Lakatos [1968a]

[2]Carnap's probabilities (2) and (2′) are known to statisticians as symmetric and non-symmetric Dirichlet distributions (see [Festa, 1993]). Skyrms [1993a] has observed that statisticians have extended such distributions to a "value continuum" (i.e., the discrete set of Q-predicates is replaced by a subclass of a continuous space).

called the assumption that $P(h) > 0$ for genuinely universal statements h "the Jeffreys-Keynes postulate", and Carnap's thesis about the dispensability of laws in inductive logic "the weak atheoretical thesis". If indeed $P(h) = 0$ for laws h, then $P(h/e) = 0$ for any evidence e; and equally well all other measures of confirmation (like Carnap's difference measure) or corroboration (cf. [Popper, 1959]) are trivialized (see [Niiniluoto and Tuomela, 1973, pp. 212–216, 242–243]).

Carnap's notion of instance confirmation restricts the applications of inductive logic to singular sentences. A related proposal is to accept the Carnapian framework for universal generalization in *finite* universes. This move has been defended by Mary Hesse [1974]. However, the applications of inductive logic would then depend on synthetic assumptions of the size of the universe — against the principle C6. Moreover, the Carnapian probabilities of finite generalizations behave qualitatively in a wrong way: the strongest confirmation is given to those universal statements that allow many Q-predicates, even when evidence seems to concentrate only on a few Q-predicates (see [Hintikka, 1965a; 1975]).

In his "Replies" in the Schilpp volume (see [Schilpp, 1963, p. 977]), Carnap told that he has constructed confirmation functions which do not give zero probabilities to universal generalizations, but "they are considerably more complicated than those of the λ-system". He did not published any of these results. In this problem situation, Hintikka's presentation of his "Jerusalem system" in the 1964 Congress was a striking novelty.

Hintikka solves the problem of universal generalization by dividing probability to constituents. He had learned this logical tool during the lectures of von Wright in Helsinki in 1947–1948. Von Wright characterized logical truth by means of "distributive normal forms": a tautology of monadic predicate logic allows all constituents, which are mutually exclusive descriptions of the constitution of the universe. Hintikka's early insight in 1948, at the age of 21, was the way of extending such distributive normal forms to the entire first-order logic with relations (see [Bogdan, 1987; Hintikka, 2006, p. 9]). This idea resulted in 1953 in a doctoral dissertation on distributive normal forms. Hintikka was thus well equipped to improve Carnap's system of induction.

Let L again be a monadic language with Q-predicates $Q_1, ..., Q_K$. A *constituent* C^w tells which Q-predicates are non-empty and which empty in universe U. The logical form of a constituent is thus

$$(9) \quad (\pm)(\exists x)Q_1(x)\&...\&(\pm)(\exists x)Q_K(x).$$

If $Q_i, i \in CT$, are precisely those Q-predicates claimed to be non-empty by (9), then (9) can be rewritten in the form

$$(10) \quad \bigwedge_{i \in CT} (\exists x)Q_i(x)\&(x)[\bigvee_{i \in CT} Q_i(x)].$$

The cardinality of CT is called the *width* of constituent (10). Often a constituent with width w is denoted by C^w. Then C^K is the maximally wide constituent which claims that all Q-predicates (i.e., all kinds of individuals which can be described by the resources of language L) are exemplified in the universe. Note that if C^K is true in universe U, then there are no true universal generalizations in L. Such a universe U is *atomistic* with respect to L, and C^K is often referred to as the atomistic constituent of L.

The number of different constituents of L is 2^K. Among them we have the empty constituent of width zero; it corresponds to a contradiction. Other constituents are maximally consistent and complete theories in L: each of them specifies a "possible word" by means of primitive monadic predicates, sentential connectives and quantifiers. Thus, constituents are mutually exclusive, and the disjunction of all constituents is a tautology. Note that in a language with finitely many individual constants each constituent can be expressed by a disjunction of state descriptions or by a disjunction of structure descriptions.

Each consistent generalization h in L (i.e., a quantificational statement without individual constants) can be expressed as a finite disjunction of constituents:

$$(11)\quad \vdash h = \bigvee_{i \in I_h} C_i$$

(11) is the *distributive normal form* of h. Constituents are *strong* generalization in L, and other generalizations in L are *weak*. By (11), the probability of generalizations reduces to the probabilities of constituents.

As above, let evidence e be a descriptions of a finite sample of n individuals, and let c be the number of different kinds of individuals observed in e. Sometimes we denote this evidence by e_n^c. Then a constituent C^w of width w is compatible with e_n^c only if $c \leq w \leq K$. By Bayes's formula,

$$(12)\quad P(C^w/e) = \frac{P(C^w)P(e/C^w)}{\sum_{i=0}^{K-c} \binom{K-c}{i} P(C^{c+i})P(e/C^{c+i})}.$$

Hence, to determine the posterior probability $P(C^w/e)$, we have to specify the prior probabilities $P(C^w)$ and the likelihoods $P(e/C^w)$.

In his first papers, Hintikka followed the heuristic principle C7. His *Jerusalem system* is obtained by first dividing probability evenly to all constituents and then dividing the probability-mass of each constituent evenly to all state descriptions belonging to it [Hintikka, 1965a]. His *combined system* is obtained by first dividing probability evenly to all constituents, then evenly to all structure descriptions satisfying a constituent, and finally evenly to state descriptions belonging to a structure description [Hintikka, 1965b]. In both cases, the prior probabilities $P(C^w)$ of all constituents are equal to $1/2^K$. It turns out that there is one and only one constituent C^c which has asymptotically the probability one when the size n of the sample e grows without limit. This is the "minimal" constituent C^c which states that the universe U instantiates precisely those c Q-predicates which are exemplified in the sample e:

$$(13)\quad \begin{aligned} &P(C^c/e_n^c) \to 1, \text{ if } n \to \infty \text{ and } c \text{ is fixed} \\ &P(C^w/e_n^c) \to 0, \text{ if } n \to \infty, c \text{ is fixed, and } w > c. \end{aligned}$$

It follows from (13) that a constituent which claims some uninstantiated Q-predicates to be exemplified in U will asymptotically receive the probability zero. A weak generalization h in L will receive asymptotically the probability one if and only if its normal form (11) includes the constituent C^c:

(14) Assuming that $n \to \infty$ and c is fixed,
$P(h/e_n^c) \to 1$ iff $C^c \vdash h$.

The Keynes-Nicod debate thus receives an answer by Hintikka's probability assignment.

In his two-dimensional continuum of inductive methods, Hintikka [1966] was able to formulate a system which contains as special cases his earlier measures as well as the whole of Carnap's λ-continuum. Hintikka proposes that likelihoods relative to C^w are calculated in the same way as in Carnap's λ-continuum (cf. (2)), but by restricting the universe to the w Q-predicates that are instantiated by C^w. Thus, if e is compatible with C^w, we have

(15) $$P(Q_i(a_{n+1})/e\&C^w) = \frac{n_i + \lambda/w}{n + \lambda}.$$

By (15), we can calculate that

(16) $$P(e/C^w) = \frac{\Gamma(\lambda)}{\Gamma(n+\lambda)} \prod_{j=1}^{c} \frac{\Gamma(n_j + \lambda/w)}{\Gamma(\lambda/w)},$$

where Γ is the Gamma-function. Note that $\Gamma(n + 1) = n!$.

For prior probabilities Hintikka proposes that $P(C^w)$ should be chosen as proportional to the Carnapian probability that a set of α individuals is compatible with C^w. This leads to the assignment

(17) $$P(C^w) = \frac{\frac{\Gamma(\alpha + w\lambda/K)}{\Gamma(w\lambda/K)}}{\sum_{i=0}^{K} \binom{K}{i} \frac{\Gamma(\alpha + i\lambda/K)}{\Gamma(i\lambda/K)}}.$$

The posterior probability $P(C^w/e)$ can then be calculated by (12), (16), and (17).

If $\alpha = 0$, then (17) gives equal priors to all constituents:

(18) $P(C^w) = 1/2^K$ for all C^w.

The Jerusalem system is then obtained by letting $\lambda \to \infty$. Small value of α is thus an indication of the strength of *a priori* considerations in inductive generalization — just as small λ indicates strong weight to *a priori* considerations in singular inference. But if $0 < \alpha < \infty$, then we have

(19) $P(C^w) < P(C^{w'})$ iff $w < w'$.

Given evidence e_n^c which has realized c Q-predicates, the minimal constituent C^c compatible with e_n^c claims that the universe is similar to the sample e_n^c. This constituent is the *simplest* of non-refuted constituents in the sense of ontological parsimony (cf. [Niiniluoto, 1994]). By (19) it is initially the least probable, but by (13) it is the only constituent that receives asymptotically the probability one with increasing but similar evidence.

If λ is chosen to be a function of w, so that $\lambda(w) = w$, then Hintikka's *generalized combined system* is obtained; the original combined system of [Hintikka, 1965b] is a special case with $\alpha = 0$. The formulas of the two-dimensional system are reduced to simpler equations:

$$(15') \quad P(Q_i(a_{n+1})/e\&C^w) = \frac{n_i + 1}{n + w}.$$

$$(16') \quad P(e/C^w) = \frac{(w-1)!}{(n+w-1)!} \prod_{j=1}^{c} (n_j!)$$

$$(17') \quad P(C^w) = \frac{\dfrac{(\alpha+w-1)!}{(w-1)!}}{\sum_{i=0}^{K} \binom{K}{i} \frac{(\alpha+i-1)!}{(i-1)!}}.$$

Hence, by (12),

$$(20) \quad P(C^w/e_n^c) = \frac{\frac{(\alpha+w-1)!}{(n+w-1)!}}{\sum_{i=0}^{K-c} \binom{K-c}{i} \frac{(\alpha+c+i-1)!}{(n+c+i-1)!}}.$$

In particular, when $n = \alpha$, (20) reduces to

$$\frac{1}{\sum_{i=0}^{K-c} \binom{K-c}{i}} = \frac{1}{2^{K-c}}.$$

If n and α are sufficiently large in relation to K, then using the approximation $(m+n)! \simeq m!m^n$, where m is sufficiently large in relation to n^2 (see [Carnap, 1950, p. 150]), we get from (20) an approximate form of $P(C^w/e)$:

$$(21) \quad P(C^w/e_n^c) \simeq \frac{(\alpha/n)^{w-c}}{(1+\alpha/n)^{K-c}}.$$

(See [Niiniluoto, 1987, p. 88].) Formula (21) shows clearly the asymptotic behaviour (13) of the posterior probabilities when n increases without limit.

The representative function of the generalized combined system is

$$(22) \quad P(Q_i(a_{n+1}/e_n^c) = (n_i+1) \frac{\sum_{i=0}^{K-c} \binom{K-c}{i} \frac{(\alpha+c+i-1)!}{(n+c+i)!}}{\sum_{i=0}^{K-c} \binom{K-c}{i} \frac{(\alpha+c+i-1)!}{(n+c+i-1)!}}.$$

If h is a universal generalization in L which claims that certain b Q-predicates are empty, and if h is compatible with e, then

$$(23)\quad P(h/e_n^c) = \frac{\sum_{i=0}^{K-b-c} \binom{K-b-c}{i} \frac{(\alpha+c+i-1)!}{(n+c+i-1)!}}{\sum_{i=0}^{K-c} \binom{K-c}{i} \frac{(\alpha+c+i-1)!}{(n+c+i-1)!}}.$$

Approximately, for sufficiently large α and n, (23) gives

$$(24)\quad P(h/e) \simeq \frac{1}{(1+\alpha/n)^b}.$$

In agreement with (14), the value of (24) approaches one when n increases without limit.

On the other hand, if $\alpha \to \infty$, we can see by (20) that $P(C^w/e) \to 1$ if and only if $w = K$. In fact, the same result holds for the prior probabilities of constituents:

$$(25)\quad \text{If } \alpha \to \infty, \text{ then } P(C^K) \to 1 \text{ and } P(C^w) \to 0 \text{ for } w < K.$$

More generally, we have the result that the probabilities of Hintikka's $\lambda - \alpha$-continuum approach the probabilities of Carnap's λ-continuum, when $\alpha \to \infty$. The result (25) explains why the probabilities of all universal generalizations are zero for all of Carnap's measures: his probabilities of universal generalizations are fixed purely *a priori* in the sceptical fashion that the prior probability of the atomistic constituent C^K is one. Carnap's λ-continuum is thus the only special case ($\alpha = \infty$) of Hintikka's two-dimensional continuum where the asymptotic behaviour (13) of posterior probabilities does not hold.

3 AXIOMATIC INDUCTIVE LOGIC

The aim of axiomatic inductive logic is to find general rationality principles which narrow down the class of acceptable probability measures. The first axiomatic treatment of this kind was presented by W. E. Johnson [1932] (cf. [Pietarinen, 1972]). His main results were independently, and without reference to him, rediscovered by Kemeny and Carnap in 1952-54 (see [Schilpp, 1963]).

Let P be a real-valued function P defined for pairs of sentences (h, e), where e is consistent, of a finite monadic language L. Assume that P satisfies the following:

(A1) *Probability axioms*

(A2) *Finite regularity*: For singular sentences h and $e, P(h/e) = 1$ only if $\vdash e \supset h$.

(A3) *Symmetry with respect to individuals*: The value of $P(h/e)$ is invariant with respect to any permutation of individual constants.

(A4) *Symmetry with respect to predicates*: The value of $P(h/e)$ is invariant with respect to any permutation of the Q-predicates.

(A5) *λ-principle*: There is a function f such that $P(Q_i(a_{n+1}/e) = f(n_i, n)$.

For the advocates of personal probability, A1 is the only general constraint of rational degrees of belief. It guarantees that probabilities serve as coherent betting ratios. A2 excludes that some contingent singular sentence has the prior probability one. A3 is equivalent to De Finetti's condition of *exchangeability* (cf. [Carnap and Jeffrey, 1971; Hintikka, 1971]). It entails that the probability $P(Q_i(a_{n+1}/e)$ depends upon evidence e only through the numbers $n_1, \ldots, n_K$, so that it is independent on the order of observing the individuals in e. A4 states that the Q-predicates are symmetrical: $P(Q_i)(a_j) = 1/K$ for all $i = 1, \ldots, K$. A5 is Johnson's [1932] "sufficientness postulate", or Carnap's "axiom of predictive irrelevance". It states that the representative function $P(Q_i(a_{n+1}/e)$ is independent of the numbers $n_j, j \neq i$, of observed individuals in other cells than Q_i (as long as the sum $n_1 + ... + n_K = n$).

The Kemeny–Carnap theorem states that axioms A1-A5 characterize precisely Carnap's λ-continuum with $\lambda > 0$: if A1–A5 hold for P, then

$$f(n_i, n) = \frac{n_i + \lambda/K}{n + \lambda},$$

where

$$\lambda = \frac{Kf(0,1)}{1 - Kf(0,1)}.$$

If $K = 2$, the proof requires the additional assumption that f is a linear function of n_i. The case $\lambda = 0$ is excluded by A2. By dropping A4, the function $f(n_i, n)$ will have the form ($2'$). Hence, we see that a regular and exchangeable inductive probability measure is Carnapian if and only if it satisfies the sufficiency postulate A5. In particular, the traditional Bayesian approach of Laplace with probability c^* satisfies A5.

Axiom A5 is very strong, since it excludes that predictive singular probabilities $P(Q_i(a_{n+1}/e_n^c)$ about the next instance depend upon the *variety of evidence* e_n^c, i.e., upon the number c of cells Q_i such that $n_i > 0$. As the number of universal generalizations in L which evidence e falsifies is also a simple function of c, axiom A5 makes induction purely *enumerative* and excludes the *eliminative* aspects of induction (see [Hintikka, 1968b]). We have already seen that the representative function (22) of Hintikka's generalized combined system depends on c. The inability of Carnap's λ-continuum to deal with inductive generalization is thus an unhappy consequence of the background assumption A5.

The Carnap-Kemeny axiomatization of Carnap's λ-continuum was generalized by Hintikka and Niiniluoto in 1974, who allowed that the inductive probability (2) of the next case being of type Q_i depends on the observed relative frequency n_i of kind Q_i and on the number c of different kinds of individuals in the sample e (see [Hintikka and Niiniluoto, 1976]):

A6 *c-principle*: There is a function f such that $P(Q_i(a_{n+1}/e^c_n) = f(n_i, n, c)$.

The number c expresses the variety of evidence e, and it also indicates how many universal generalizations e has already falsified. Hintikka and Niiniluoto proved that measures satisfying axioms A1–A4 and A6 constitute a *K-dimensional system* determined by K-parameters

$$\lambda = \frac{Kf(1,K+1,K)}{1-Kf(1,K+1,K)} - 1$$

$$\gamma_c = f(0,c,c), \text{ for } c = 1, ..., K-1.$$

Here $\lambda > -K$ and

(26) $0 < \gamma_c \quad \leq \dfrac{\lambda/K}{c+\lambda}$.

This K-dimensional system is called the NH-system by Kuipers [1978b]. (See also [Niiniluoto, 1977].) The upper bound of (26) is equal to the value of probability $f(0, c, c)$ in Carnap's λ-continuum; let us denote it by δ_c. It turns out that, for infinite universes, the probability of the atomistic constituent C^K is

$$P(C^K) = \frac{\gamma_1 ... \gamma)K-1}{\delta_1 ... \delta_{K-1}}.$$

Hence,

$$P(C^K) = 1 \text{ iff } \gamma_i = \delta_i \text{ for all } i = 1, ..., K-1.$$

In other words, Carnap's λ-continuum is the only special case of the K-dimensional system which does not attribute non-zero probabilities to some universal generalizations. Again, Carnap's systems turns out to be biased in the sense that it assigns *a priori* the probability one to the atomistic constituent C^K that claims all Q-predicates to be instantiated in universe U.

The reduction of all inductive probabilities to K parameters, which concern probabilities of very simple singular predictions, gives a counter-argument to Wolfgang Stegmüller's [1973] claim that it does not "make sense" to bet on universal generalizations (cf. [Hintikka, 1971]). In the K-dimensional system, a bet on a universal law is equivalent to a system of K bets on singular sentences on finite evidence.

The parameter $\gamma_c = f(0, c, c)$ expresses the predictive probability of finding a new kind of individual after c different successes. For such evidence e, the posterior probability of C^c approaches one when γ_c approaches zero. Further, $P(C^c)$ decreases when γ_c increases. Parameter γ_w thereby serves as an index of caution for constituents of width w. While Hintikka's two-dimensional system has one index α of overall pessimism about the truth of constituents $C^w, w < K$, in the K-dimensional system there is a separate index of pessimism for each width $w < K$.

The K-dimensional system allows more flexible distributions of prior probabilities of constituents than Hintikka's $\alpha - \lambda$-continuum. For example, principle

(19) may be violated. One may divide prior probability equally first to sentences $S^w(w = 0, ..., K)$ which state that there are w kinds individuals in the universe. Such "constituent-structures" S^w are disjunctions of the $\binom{K}{w}$ constituents C^w of width w. This proposal was made by Carnap in his comment on Hintikka's system (see [Carnap, 1968]; cf. [Kuipers, 1978a]).

Assuming that the parameters γ_c do not have their Carnapian values, one can show

(27) $P(Q_i(a_{n+1})/e\&C^w) = \dfrac{n_i + \lambda/K}{n + w\lambda/K}.$

Comparison with formula (15) shows that, in addition to the λ-continuum, the intersection of the K-dimensional system and Hintikka's $\alpha - \lambda$-system contains those members of the latter which satisfy the condition that λ as a function of w equals αw for some constant $a > 0$.[3] The case with $a = 1$ is Hintikka's generalized combined system (cf. (15′)). This new way of motivating this system shows its naturalness. The relations of different inductive systems are studied in detail by Theo Kuipers [1978b].[4]

It follows from (27) that the K-dimensional system satisfies Reichenbach's Axiom (3) and Instantial Positive Relevance (4).

The fundamental adequacy condition (13) of inductive generalization is satisfied whenever the parameters γ_i are chosen more optimistically than their Carnapian values:

(28) If $\gamma_i < \delta_i$, for $i = c, ..., K-1$, then $P(C^c/e) \to 1$ when $n \to \infty$ and c is fixed.

This result shows again that the much discussed result of Carnap's λ-continuum, viz. the zero confirmation of universal laws, is really an accidental feature of a system of inductive logic. We get rid of this feature by weakening the λ-principle A5 to the c-principle A6.

4 EXTENSIONS OF HINTIKKA'S SYSTEM

Hintikka's two-dimensional continuum was published in *Aspects of Inductive Logic* [1966], edited by Hintikka and Patrick Suppes. This volume is based on an International Symposium on Confirmation and Induction, held in Helsinki in late September 1965 as a continuation of an earlier seminar at Stanford University in the spring of the same year. Hintikka had at that time a joint appointment at Helsinki and Stanford. Besides essays on the paradoxes of confirmation (Max Black, Suppes, von Wright), the volume includes several essays on induction by

[3]This class of inductive methods is mentioned by Hilpinen [1968, p. 65]. Kuipers [1978b] calls them SH-systems.

[4]Zabell [1997] has developed Johnson's axiomatization so that constituents of width one receive non-zero probabilities. This result is a special case of the K-dimensional system with $\gamma_1 < \delta_1$.

Hintikka's Finnish students: Risto Hilpinen, Raimo Tuomela, and Juhani Pietarinen.

Like Carnap's continuum, Hintikka's two-dimensional system is formulated for a monadic first-order language with finitely many predicates. As a minor technical improvement, the Q-predicates may have different widths (cf. (2′)). The Q-predicates may be defined by families of predicates in Carnap's style, so that they allow discrete quantitative descriptions. Moreover, the number of predicates may be allowed to be countably infinite (see [Kuipers, 1978b]).

However, more challenging questions concern extensions of Hintikka's framework to languages which are essentially more powerful than monadic predicate logic. Hilpinen [1966] considers *monadic languages with identity.* In such languages it is possible to record that we have picked out different individuals in our evidence ("sampling without replacement"). Numerical quantifiers "there are at least d individuals such that" and "there are precisely $d-1$ individuals" can be expressed by sentences involving a layer of d interrelated quantifiers. The maximum number of nested quantifiers in a formula is called its quantificational *depth.* Hence, by replacing existential quantifiers in formula (9) by numerical quantifiers, constituents of depth d can specify that each Q-predicate is satisfied by either $0, 1, ..., d-1$, or at least d individuals (see [Niiniluoto, 1987, p. 59]). A constituent of depth d splits into a disjunction of "subordinate" constituents at the depth $d+1$: the claim that there are at least d individuals in Q_i means that there are precisely d or at least $d+1$ individuals in Q_i. For finite universes monadic constituents with identity are equivalent to Carnap's structure descriptions, but expressed without individual constants.

Hilpinen extends Hintikka's Jerusalem system to the monadic language with identity by dividing the probability mass evenly to all constituents at the depth d and then evenly to state descriptions entailing a constituent. Hilpinen shows that, on the basis this probability assignment, it is not reasonable to project that cells not instantiated in our evidence are occupied by more than d individuals. Also it is not rational to project that observed singularities (i.e., cells with only one observed individual) are real singularities in the whole universe. However, all constituents according to which in our universe there are unobserved singularities have an equal degree of posterior probability on any evidence, and these constituents are equally probable as the constituent which denies the existence of unobserved singularities. The last result is not intuitive. Hilpinen shows that it can be changed by an alternative probability assignment: distribute probability first evenly to constituents of depth 1, then evenly to all subordinate constituents of dept 2, etc. (cf. [Hintikka, 1965a]). Then the highest probability is given to the constituent which denies the existence of unobserved singularities.

Tuomela [1966] shows that Hintikka's main result (13) about inductive generalization can be achieved in an ordered universe. The decision problem for a first-order language containing the relation $Rxy =$"y is an immediate successor of x" is effectively solvable. The Q-predicates for such a language specify triples: predecessor, object, successor. The constituents state which kinds of triples there are in the universe. If all constituents are given equal prior probabilities, the simplest constituent compatible with evidence will have the greatest posterior probability.

Inductive logic for *full first-order logic* is investigated by Hilpinen [1971].[5] In principle, Hintikka's approach for monadic languages can be generalized to this situation, since Hintikka himself showed in 1953 how distributive normal forms can be defined for first-order languages L with a finite class of polyadic relations (cf. [Niiniluoto, 1987, pp. 61-80]). For each quantificational depth $d > 0$, i.e., the number of layers of quantifiers, a formula of L can be expressed as a disjunction of constituents of depth d. This normal form can be expanded to greater depths. A new feature of this method results from the undecidability of full first-order logic: some constituents are non-trivially inconsistent and there is no effective method of locating them. The logical form of a constituent of depth d is still (13), but now the Q-predicates or "attributive constituents" are trees with branches of the length d. A constituent of depth 1 tells what kinds of individuals there are in the universe, now described by their properties and their relations to themselves. A constituent $C^{(2)}$ of depth 2 is a systematic description of all different kinds of pairs of individuals that one can find in the universe. A constituent $C^{(d)}$ of depth d is a finite set of finite trees with maximal branches of the length d. Each such branch corresponds to a sequence of individuals that can be drawn with replacement from the universe. Such constituents $C^{(d)}$ are thus the strongest generalizations of depth $\leq d$ expressible in language L. Each complete theory in L can be axiomatized by a monotone sequence of subordinate constituents $\langle C^{(d)} | d < \infty \rangle$, where ... $C^{(d+1)} \vdash C^{(d)} \vdash ... \vdash C^{(1)}$.'indexDe Morgan, A.

Given the general theory of distributive normal forms, the axiomatic approach can in principle be applied to the case of constituents of depth d. Whenever assumptions corresponding to A1, A2, A3, and A6 can be made, there will be one constituent $C^{(d)}$ which receives asymptotically the probability one on the basis of evidence consisting of ramified sequences of d interrelated individuals. The general case has not yet been studied. Hilpinen's 1971 paper is still the most detailed analysis of inductive logic with relations.[6]

---

[5] Assignment of mathematical probabilities to formulas of a first-order language L is studied in *probability logic* (see [Scott and Krauss, 1966; Fenstad, 1968]). The main representation theorem, due to Jerzy Los in 1963, tells that the probability of a quantificational sentence can be expressed as a weighted average involving two kinds of probability measures: one defined over the class of L-structures, and the other measures defined over sets of individuals within each L-structure. Again exchangeability (cf. A3) guarantees some instantial relevance principles (cf. [Nix and Paris, 2007]), but otherwise probability logic has yet not lead to new fruitful applications in the theory of induction.

[6] Nix and Paris [2007] have recently investigated binary inductive logic, but they fail to refer to Hintikka's program in general and to Hilpinen [1971] in particular. The basic proposal of Nix and Paris is to reduce binary relations to unary ones: for example, 'x pollinates y' is treated as equivalent to a long sentences involving only unary predicates of x and y. As a general move this proposal is entirely implausible. The reduction of relations to monadic properties was a dogma of classical logic, until De Morgan and Peirce in the mid-nineteenth century started the serious study of the logic of relations (see [Kneale and Kneale, 1962]). The crucial importance of the distinction between monadic and polyadic first-order logic was highlighted by the metalogical results in the 1930s: the former is decidable, the latter is undecidable. This difference has

Hilpinen studies constituents of depth 2. Evidence e includes n observed individuals and a complete description of the relations of each pair of individuals in e. Now constituents C^w of depth 2 describe what kinds of individuals there are in the universe U. A statement D^v which specifies, for each individual a_i in e, which attributive constituent a_i satisfies, gives an answer to the question of how observed individuals are related to unobserved individuals. Hilpinen distributes inductive probability $P(C^w)$ to constituents C^w evenly. Probabilities of the form $P(D^v/C^w)$ are defined by the Carnap-Hintikka style representative function of the form (15). Likelihoods $P(e/D^v\&C^w)$ are also determined by the same type of representative function, but now applied to pairs of individuals. Again, corresponding to Hintikka's basic asymptotic result (13), the highest posterior probability on large evidence is given to the simplest conjunction $D^e\&C^e$ where D^e states that the individuals in e are related to unobserved individuals in the same ways as to observed individuals and C^e states that there are in the universe only those kinds of individuals that are, according to D^e, already exemplified in E. Hilpinen observes that there is another kind of "simplicity": the number of kinds of individuals in D^e may be reduced by assuming that all individuals in e bear the same relations to observed and some yet unobserved individuals. If this statement is denoted by D^1, and the corresponding constituent by C^1, then $P(D^v/C^v)$ is maximized by $D^1\&C^1$. Hence, inductive methods dealing with polyadic languages need at least two separate parameters which regulate the weights given to the two kinds of simplicity.'indexlawlike generalizations

An extension of Hintikka's system to *modal logic* has been proposed by Soshichi Uchii [1972; 1973; 1977] (cf. [Niiniluoto, 1987, pp. 91-102]). Such an account is interesting, if the formulation of *lawlike* generalizations requires intensional notions like necessity and possibility (see [Pietarinen, 1972]). Hintikka himself is one of the founders of the possible worlds semantics for modal logic (see [Bogdan, 1987; Hintikka, 2006]). Uchii is interested in a monadic language $L(\Box)$ with the operators of *nomic* or *causal necessity* $\Box$ and *nomic possibility* $\Diamond$. Here $\Diamond =\sim \Box \sim$. It is assumed that necessity satisfies the conditions of the Lewis system S5. The *nomic constituents* of $L(\Box)$ can now be defined in analogy with (10):

$$(29) \quad \bigwedge_{i \in CT} \Diamond(\exists x)Q_i(x)\&\Box(x)\left[\bigvee_{i \in CT} Q_i(x)\right].$$

Uchii calls (29) "a non-paradoxical causal law". (29) specifies which kinds of individuals are physically possible and which kinds are physically impossible. Even stronger modal statements can be defined by

$$(30) \quad \bigwedge_{i \in H} \Diamond C_i\&\Box\left[\bigvee_{i \in H} C_i\right].$$

where C_i are the ordinary constituents of the language L without $\Box$. The laws expressible in $L(\Box)$ are typically what John Stuart Mill called "laws of coexistence".

dramatic consequences to Hintikka's theory of distributive normal forms as well.

To express Mill's "laws of succession", some temporal notions have to be added to $L(\Box)$ (see [Uchii, 1977]).

Let us denote by B_i the nomic constituent (29) which has the same positive Q-predicates as the ordinary constituent C_i. As actuality entails possibility, there are $K - w$ nomic constituents compatible with an ordinary constituent C^w of width w. Uchii's treatment assumes that $P(C_i) = P(B_i)$ for all i. Further, the probability of evidence e, given knowledge about the actual constitution C_i of the universe, is not changed if the corresponding nomic constituent B_i is added to the evidence: $P(e/C_i) = P(e/C_i \& B_i)$. It follows from (13) that

(31) $P(B^c/e_n^c) \to 1$, when $n \to \infty$ and c is fixed, iff $P(B^c/C^c) = 1$.

Thus, if we have asymptotically become certain that C^c is the true description of the actual constitution of the universe, the same certainty holds for the nomic constituent B^c if and only if $P(C^c/B^c) = P(B^c/C^c) = 1$. Uchii makes this very strong assumption, which simply eliminates all the $K - c$ nomic constituents compatible with C^c and undermined by the asymptotic evidence. In fact, he postulates that $P((\exists x\phi(x))/\Diamond(\exists x)\phi(x)) = 1$ for all formulas ϕ. This questionable metaphysical doctrine, which says that all genuine possibilities are realized in the actual history, is known as the *Principle of Plenitude*.

An alternative interpretation is proposed by Niiniluoto [1987, pp. 101-102]. Perhaps the actual constitution of the universe is not so interesting, since evidence e obtained by active *experimentation* will realize new possibilities. As laws of nature have counterfactual force, experimentation can be claimed to be the key to their confirmation (see [von Wright, 1971]). So instead of the fluctuating true actual constituent C^c, we should be more interested in the permanent features of the universe expressed by the true nomic constituent. This suggests that the inductive approach of Sections 2 and 3 is directly formulated with nomic constituents, so that the axiomatic assumptions imply a convergence result for the constituent B^c on the basis of experimental evidence e_n^c.

5 SEMANTIC INFORMATION

Hintikka was quick to note that his inductive probability measures make sense of some of Popper's ideas. Hintikka [1968b] observed that his treatment of induction is not purely enumerative, since the inductive probability of a generalization depends also on the ability of evidence to refute universal statements. This *eliminative* aspect of induction is related to the Popperian method of falsification. Popper [1959] argued that preferable theories should have a low absolute logical probability: good theories should be falsifiable, bold, informative, and hence improbable. In Hintikka's two-dimensional system with the prior probability assignment (17), the initially least probable of the constituents compatible with evidence e, i.e., constituent C^c (see (19)) eventually will have the highest posterior probability. The smaller finite value α has, the faster we switch our degrees of confirmation from initially more probable constituents to initially less probable constituents.

Hence, the choice of a small value of parameter α is "an indication of one aspect of that intellectual boldness Sir Karl has persuasively advocated" [Hintikka, 1966, p. 131].

A systematic argument in defending essentially the same conclusion comes from the theory of semantic information [Hintikka and Pietarinen, 1966]. It can be shown that the degree of information content of a hypothesis is inversely proportional to its prior probability. A strong generalization is the more informative the fewer kinds of individuals it admits of. Therefore, C^c is the most informative of the constituents compatible with evidence. For constituents, high degree of information and low prior probability — Popper's basic requirements — but also high degree of posterior probability go together.

The relevant notion of semantic information was made precise by Carnap and Bar-Hillel [1952] (cf. [Niiniluoto, 1987, pp. 147-155]). They defined the *information content* of a sentence h in monadic language L by the class of the content elements entailed by h, where content elements are negations of state descriptions. Equally, information content could be defined as the range $R(\sim h)$ of the negation $\sim h$ of h, i.e., the class of state descriptions which entail $\sim h$. If Popper's "basic sentences" correspond to state descriptions, this is equivalent to Popper's 1934 definition of empirical content. As the *surprise value* of h Carnap and Bar-Hillel used the logarithmic measure

(32) $\mathrm{inf}(h) = -\log P(h)$,

which is formally similar to Shannon's measure in statistical information theory. For the degree of *substantial information* of h Carnap and Bar-Hillel proposed

(33) $\mathrm{cont}(h) = P(\sim h) = 1 - P(h)$.

Substantial information is thus inversely related to probability, just as Popper [1959] also required.

As both Carnap and Popper thought that $P(h) = 0$ for all universal generalizations, they could not really use the cont-measure to serve any comparison between rival laws or theories. Hintikka's account of inductive generalization opened a way for interesting applications of the theory of semantic information. He developed these ideas further in "The Varieties of Information and Scientific Explanation" [Hintikka, 1968a] and in the volume *Information and Inference* (1970), edited together by Hintikka and Suppes.

Hintikka [1968a] defined measures of *incremental information* which tell how much information h adds to the information already contained in e:

(34) $\mathrm{inf}_{\mathrm{add}}(h/e) = \mathrm{inf}(h\&e) - \mathrm{inf}(e)$

$\mathrm{cont}_{\mathrm{add}}(h/e) = \mathrm{cont}(h\&e) - \mathrm{cont}(e)$.

Measures of *conditional information* tell how informative h is in a situation where e is already known:

(35) $\inf_{\text{cond}}(h/e) = -\log P(h/e)$
$\text{cont}_{\text{cond}}(h/e) = 1 - P(h/e)$.

Hence, $\inf_{\text{add}}$ turns out to be same as $\inf_{\text{cond}}$. Measures of *transmitted information* tell how much the uncertainty of h is reduced when e is learned, or how much substantial information e carries about the subject matter of h:

(36) $\text{transinf}(h/e) = \inf(h) - \inf(h/e) = \log P(h/e) - \log P(h)$
$\text{transcont}_{\text{add}}(h/e) = \text{cont}(h) - \text{cont}_{\text{add}}(h/e) = 1 - P(h \vee e)$
$\text{transcont}_{\text{cond}}(h/e) = \text{cont}(h) - \text{cont}_{\text{cond}}(h/e) = P(h/e) - P(h)$.

Thus, e transmits some positive information about h, in the sense of transinf and $\text{transcont}_{\text{cond}}$, just in case $P(h/e) > P(h)$, i.e., e is *positively relevant* to h. In the case of $\text{transcont}_{\text{add}}$, the corresponding condition is that $h \vee e$ is not a tautology, i.e., h and e have some *common* information content.

Hilpinen [1970] used these measures to given an account of the information provided by observations. His results provide an information-theoretic justification of the principle of total evidence.

Measures of transmitted information have also an interesting application to measures of *explanatory power* or *systematic power* (see [Hintikka, 1968a; Pietarinen, 1970; Niiniluoto and Tuomela, 1973]). In explanation, the *explanans* h is required to give information about the *explanandum* e. With suitable normalizations, we have three interesting alternatives for the explanatory power of h with respect to e:

(37) $\text{expl}_1(h, e) = \text{transinf}(e/h)/\inf(e) = \dfrac{\log P(e) - \log P(e/h)}{\log P(e)}$

$$\text{expl}_2(h, e) = \text{transcont}_{\text{add}}(e/h)/\text{cont}(e) = \frac{1 - P(h \vee e)}{1 - P(e)} = P(\sim h/\sim e)$$

$$\text{expl}_3(h, e) = \text{transcont}_{\text{cond}}(e/h)/\text{cont}(e) = \frac{P(e/h) - P(e)}{1 - P(e)}.$$

Here $\text{expl}_2(h, e)$ is the measure of systematic power proposed by Hempel and Oppenheim in 1948 (see [Hempel, 1965]). Note that all of these measures receive their maximum value one if h entails e, so that they cannot distinguish between alternative deductive explanations of e. One the other hand, if *inductive explanation* is explicated by the positive relevance condition (cf. [Niiniluoto and Tuomela, 1973; Festa, 1999]), then they can used for comparing rival inductive explanations h of data e.[7]

[7]For measures of systematic power relative to sets of competing hypotheses, see Niiniluoto and Tuomela [1973].

6 CONFIRMATION AND ACCEPTANCE

Hintikka's basic result shows that universal hypotheses h can be confirmed by finite observational evidence. This notion of confirmation can be understood in two different senses (cf. [Carnap, 1962; Niiniluoto, 1972]): h may have a *high posterior probability* $P(h/e)$ on e, or e may *increase the probability* of h.

(HP) *High Probability*: e confirms h if $P(h/e) > q \geq \frac{1}{2}$.

(PR) *Positive Relevance*: e confirms h iff $P(h/e) > P(h)$.

PR is equivalent to conditions $P(h\&e) > P(h)P(e), P(h/e) > P(h/\sim e)$, and $P(e/h) > P(e)$. The basic difference between these definitions is that HP satisfies the principle of Special Consequence:

(SC) If e confirms h and $h \vdash g$, then e confirms g,

while PR satisfies the principle of Converse Entailment:

(CE) If a consistent h entails a non-tautological e, then e confirms h.

In Peirce's terminology, hypothetical inference to an explanation is called *abduction*, so that by "abductive confirmation" one may refer to the support that a theory receives from its explanatory successes. By Bayes's Theorem, PR satisfies the abductive criterion, when $P(h) > 0$ and $P(e) < 1$:

(38) If h deductively or inductively explains or predicts e, then e confirms h.

(see [Niiniluoto, 1999]). It is known that no reasonable notion of confirmation can satisfy SC and CE at the same time.

In the spirit of PR, Carnap [1962] proposed that quantitative *degrees of confirmation* are defined by the difference measure:

(39) $\text{conf}(h, e) = P(h/e) - P(h)$.

We have seen in (36) that (39) measures the transmitted information that e carries on h. As Hintikka [1968b] points out, many other measures of confirmation, evidential support, and factual support are variants of (39) (see also [Kyburg, 1970). This is the case also with Popper's proposals for the degree of *corroboration* of h by e (see [Popper, 1959, p. 400]). Popper was right in arguing that degrees of corroboration should not be identified with prior or posterior probability. But Hintikka's system has the interesting result that, in terms of measure (39) and its variants, with sufficiently large evidence e the minimal constituent C^c at the same time maximizes posterior probability $P(C^w/e)$ and the information content $\text{cont}(C^w)$. Hence, it also maximizes the difference (39), which can be written in the form $P(h/e) + \text{cont}(h) - 1$ (see [Hintikka and Pietarinen, 1966]).

Hintikka [1968a] proposed a new measure of corroboration which gives an interesting treatment of weak generalizations (see also [Hintikka and Pietarinen, 1966; Niiniluoto and Tuomela, 1973]). Assume that h is equivalent to the disjunction of constituents $C_1, ..., C_m$, and define $\text{corr}(h/e)$ as the minimum of the posterior probabilities $P(C_i/e)$:

(40) $\text{corr}(h, e) = \min \{P(C_i/e)|i = 1, ..., m\}$.

Measure (40) guarantees that, unlike probability, corroboration covaries with logical strength:

(41) If $e \vdash h_1 \supset h_2$, then $P(h_1/e) \leq P(h_2/e)$

(42) If $e \vdash h_1 \supset h_2$, then $\text{corr}(h_1, e) \geq \text{corr}(h_2, e)$.

Further, (40) favours C^c among all (weak and strong) generalizations in language L:

(43) With sufficiently large evidence e_n^c with fixed $c, \text{corr}(h, e_n^c)$ has its maximum value when h is the constituent C^c.

Hintikka inductive probability measures can be applied to the famous and much debated *paradoxes of confirmation*. In Hempel's paradox of ravens, the universal generalization "All ravens are black" is confirmed by three kinds of instances: black ravens, black non-ravens, and non-black non-ravens. The standard Bayesian solution, due to Janina Hosiasson-Lindenbaum in 1940 (see [Hintikka and Suppes, 1966; Niiniluoto, 1998]), is that these three instances give different incremental confirmation to the hypothesis, since in a finite universe these cells are occupied by different numbers of objects. Instead of such an empirical assumption, one could also make a conceptual stipulation to the effect that the predicates "black" and "non-black", and "raven", and "non-raven", have different widths, and then apply the formula (2′). Hintikka [1969a] proposes that the "inductive asymmetry" of the relevant Q-predicates could be motivated by assuming an ordering of the primitive predicates (cf. [Pietarinen, 1972]).

Another famous puzzle is Nelson Goodman's paradox of grue. Here Hintikka's solution appeals to the idea that parameter α regulates the *confirmability* of universal laws. More lawlike generalizations can be more easily confirmed than less lawlike ones. If we associate a smaller α to the conceptual scheme involving the predicate "green" than to the scheme involving the odd predicate "grue", then differences in degrees of confirmation of the generalizations "All emeralds are green" and "All emerald are grue" can be explained (see [Hintikka, 1969b; Pietarinen, 1972; Niiniluoto and Tuomela, 1973]).

Carnap's system of induction does not include *rules of acceptance*. Rather, the task of inductive logic is to evaluate the epistemic probabilities of various hypotheses. These probabilities can be used in decision making (see [Carnap, 1980; Stegmüller, 1973]). Carnap agrees here with many statisticians — both frequentists (Jerzy Neyman, E. S. Pearson) and Bayesians (L. J. Savage) — who recommend that inductive inferences are replaced by *inductive behaviour* or probability-based actions. In this view, the main role of the scientists is to serve as advisors of practical decision makers rather than as seekers of new truths. On the other hand, according to the *cognitivist* model of inquiry, the tentative results of scientific research constitute a body of accepted hypotheses, the so-called "scientific knowledge" at a given time. In the spirit of Peirce's *fallibilism*, they may be at any time questioned and revised by new evidence or novel theoretical insights. But, on some conditions, it is rational to tentatively accept a hypothesis on the basis of evidence. One of the tasks of inductive logic is define such rules of acceptance for corrigible factual statements. Hintikka, together with Isaac Levi [1967], belongs to the camp of the cognitivists.

The set of accepted hypotheses is assumed to consistent and closed under logical consequence (cf. [Hempel, 1965]). Henry E. Kyburg's lottery paradox shows then that high posterior probability alone is not sufficient to make a generalization h acceptable. But in Hintikka's system one may calculate for the size n of the sample e a threshold value n_0 which guarantees that the informative constituent C^c has a probability exceeding a fixed value $1 - \varepsilon$:

(44) Let n_0 be the value such that $P(C^c/e) \geq 1 - \varepsilon$ if and only if $n \geq n_0$.
Then, given evidence e, accept C^c on e iff $n \geq n_{0.}$

(See [Hintikka and Hilpinen, 1966; Hilpinen, 1968].) Assuming logical closure, all generalizations entailed by C^c are then likewise acceptable on e.[8] In Hintikka's two-dimensional continuum, n_0 can be defined as the largest integer n for which

$$\varepsilon' \leq \max \sum_{i=1}^{K-c} \binom{K-c}{i} (\frac{c}{c+i})^{n-\alpha},$$

where the maximum is taken over values of $c, 0 \leq c \leq K-1$, and $\varepsilon' = \varepsilon/(1-\varepsilon)$.

Hintikka and Hilpinen [1966] argue further that a singular hypothesis is inductively acceptable if and only if it is a substitution instance of an acceptable generalization:

(45) A singular hypothesis of the form $\phi(a_i)$ is acceptable on e iff the generalization $(x)\phi(x)$ is acceptable on e.

This principle reduces singular inductive inferences (Mill's "eduction") to universal inferences.

7 COGNITIVE DECISION THEORY

According to *Bayesian decision theory*, it is rational for a person X to accept the action which maximizes X's subjective expected utility. Here the relevant utility

[8]Note that (44) is a factual detachment rule in the sense that it concludes a factual statement from factual and probabilistic premises. This kind of inductive rule should be distinguished from probabilistic detachment rules which can be formulated as deductive arguments within the probability calculus (see [Suppes, 1966]). An example of the latter kind of rules is the following:

$$\begin{array}{l} P(h/e) = r \\ \underline{P(e) = 1} \\ P(h) = r. \end{array}$$

function express quantitatively X's subjective preferences concerning the outcomes of alternative actions, usually in terms of some practical goals. The probabilities needed to calculate expected utility are X's personal probabilities, degrees of belief concerning the state of nature. *Cognitive decision theory* adopts the same Bayesian decision principle with a new interpretation: the relevant actions concern the acceptance of rival hypotheses, and the utilities express some cognitively important values of inquiry. Such *epistemic utilities* may include truth, information, explanatory and predictive power, and simplicity. With anticipation by Bolzano, the basic ideas of cognitive decision theory were suggested in the early 1960s independently by Hempel and Levi (cf. [Hempel, 1965; Levi, 1967]). Inductive logic is relevant to this project, since it may provide the relevant epistemic probabilities [Hilpinen, 1968; Niiniluoto and Tuomela, 1983: Niiniluoto, 1987].

Let us denote by $B = \{h_1, ..., h_n\}$ a set of mutually exclusive and jointly exhaustive hypotheses. Here the hypotheses in B may be the most informative descriptions of alternative states of affairs or possible worlds within a conceptual framework L. For example, they may be state descriptions, structure descriptions or constituents of a monadic language, or complete theories expressible in a finite first-order language.[9] If L is interpreted on a domain U, so that each sentence of L has a truth value (true or false), it follows that there is one and only true hypothesis (say h^*) in B. Our *cognitive problem* is to identify the target h^* in B. The elements h_i of B are the potential *complete answers* to the cognitive problem. The set $D(B)$ of *partial answers* consists of all non-empty disjunctions of complete answers. The *trivial* partial answer in $D(B)$, corresponding to 'I don't know', is represented by a tautology, i.e., the disjunction of all complete answers.

For any $g \in D(B)$ and $h_j \in B$, we let $u(g, h_j)$ be the epistemic utility of accepting g if h_j is true. We also assume that a rational probability measure P is associated with language L, so that each h_j can be assigned with its epistemic probability $P(h_j/e)$ given the available evidence e. Then the best hypothesis in $D(B)$ is the one g which maximizes the *expected epistemic utility:*

$$(46)\quad U(g/e) = \sum_{j=1}^{n} P(h_j/e)u(g, h_j).$$

Expected utility gives us a new possibility of defining inductive *acceptance rules*:

(EU) Accept on evidence e the answer $g \in D(B)$ which maximizes the value $U(g/e)$.

Another application is to use expected utility as a criterion of epistemic preferences and *cognitive progress*:

(CP) Step from answer $g \in D(B)$ to another answer $g' \in D(B)$ is cognitively progressive on evidence e iff $U(g/e) < U(g'/e)$.

[9] The framework also includes situations where B is a subset of some quantitative space like the real numbers $\mathbb{R}$, but then sums are replaced by integrals.

(See [Niiniluoto, 1995].)

Assume now that g is a partial answer in $D(B)$ with

(47) $\vdash g \equiv \bigvee_{i \in I_g} h_i,$

If truth is the only relevant epistemic utility, then we may take $u(g, h_j)$ simply to be the truth value of g relative to h_j:

$$\begin{aligned} u(g, h_j) &= 1 \text{ if } h_j \text{ is in } g \\ &= 0 \text{ otherwise.} \end{aligned}$$

Hence, $u(g, h^*)$ is the real (and normally unknown) truth value $tv(g)$ of g relative to the domain U. It follows from (46) that the expected utility $U(g/e)$ equals the posterior probability $P(g/e)$ of g on e:

(48) $U(g/e) = \sum_{i \in I_g} P(h_j/e) = P(g/e)$

In this sense, we may say that posterior probability equals expected truth value. The rule of maximizing expected utility leads now to an extremely conservative policy: the best hypotheses g on e are those that satisfy $P(g/e) = 1$, i.e., are completely certain on e. For example, e itself and a tautology are such statements. If we are not certain of the truth, then it is always progressive to change an uncertain answer to a logically weaker one. The argument against using probability as a criterion of theory choice was made already by Popper in 1934 (see [Popper 1959]). He proposed that good theories should be bold, improbable, and informative (cf. Section 6).

However, it is likewise evident that information cannot be the only relevant epistemic utility. Assume that the information content of g is measured by $\text{cont}(g) = 1 - P(g)$ (see (33)). If we now choose $u(g, h_j) = \text{cont}(g)$, then the expected utility $U(g/e)$ equals $1 - P(g)$, which is maximized by a contradiction with probability zero. Further, any false theory could be cognitively improved by adding new falsities to it. Similar remarks can be made about explanatory and systematic power, if they were chosen as the only relevant utility.

Levi [1967] measures the information content $I(g)$ of a partial answer g in $D(B)$ by the number of complete answers it excludes. With a suitable normalization, $I(g) = 1$ if and only if g is one of the complete answers h_j in B, and $I(g) = 0$ for a tautology. Levi's proposal for epistemic utility is the weighted combination of the truth value $tv(g)$ of g and the information content $I(g)$ of g:

(49) $tv(g) + qI(g),$

where $0 < q \leq 1$ is an "index of boldness", indicating how much the scientist is willing to risk error, or to "gamble with truth", in her attempt to relieve from agnosticism. The expected epistemic utility of g is then

(50) $P(g/e) + qI(g)$.

By using the weight q, formula (50) expresses a balance between two mutually conflicting goals of inquiry. It has the virtue that all partial answers g in $D(B)$ are comparable with each other.

If epistemic utility is defined by information content $\mathrm{cont}(g)$ in a truth-dependent way, so that

$$(51)\quad \begin{aligned} u(g,e) &= \mathrm{cont}(g) \text{ if } g \text{ is true} \\ &= -\mathrm{cont}(\sim g) \text{ if } g \text{ is false,} \end{aligned}$$

(i.e., in accepting hypothesis g, we gain the content of g if g is true, but we lose the content of the true hypothesis $\sim g$ if g is false), then the expected utility $U(g/e)$ is equal to

(52) $P(g/e) - P(g)$.

This proposal, originally made by Levi in 1963 but rejected in Levi [1967], was defended by Hintikka and Pietarinen [1966]. This measure combines the criteria of boldness (small prior probability P(g)) and high posterior probability $P(g/e)$. Similar results can be obtained when $\mathrm{cont}(g)$ is replaced by Hempel's [1965] measure of systematic power (37): if

$$(53)\quad \begin{aligned} u(g,e) &= P(\sim g/\sim e) \text{ if } g \text{ is true} \\ &= -P(g/\sim e) \text{ if } g \text{ is false,} \end{aligned}$$

then the expected utility of g is $[P(g/e) - P(g)]/P(\sim e)$, which again is a variant of the difference measure (52) (see [Pietarinen, 1970]).

Hilpinen [1968] proposed a modification of (49) and (51):

$$(54)\quad \begin{aligned} u(g,e) &= 1 - P(g) \text{ if } g \text{ is true} \\ &= -qP(g) \text{ if } g \text{ is false.} \end{aligned}$$

Then the expected utility is

(55) $U(g/e) = P(g/e) - qP(g)$,

where again q serves as an index of boldness.

In Hintikka's system, with sufficiently large evidence e, the answer which maximizes the values (50), (52), and (55) is the minimal constituent C^c. As C^c is also the simplest of the constituent compatible with e, Hintikka posterior probabilities favour simplicity — so that simplicity need not be added to the framework as an extra condition (cf. [Niiniluoto, 1994]).

8 INDUCTIVE LOGIC AND THEORIES

Inductive logic has the reputation that it is a formal tool of narrowly *empiricist* methodology. Although induction was discussed already by Aristotle, his account was intimately connected to concept formation (see [Hintikka, 1980; Niiniluoto, 1994/95]). The role of induction in science was emphasized especially by the British philosophers from Francis Bacon to William Whewell, John Stuart Mill, and Stanley Jevons in the 19^{th} century and the Cambridge school in the 20^{th} century. Many empiricists were also *inductivists* in the sense that they restricted science to empirical observations and generalizations that could be discovered and justified by enumerative induction.

The *hypothetico-deductive* (HD) *method* of science allows scientists to freely invent hypothetical theories to explain observed data, but requires that such hypotheses are indirectly tested by their empirical consequences. Bayes's theorem provides a method of evaluating the performance of hypotheses in observational tests. In the basic result about indirect confirmation (38), the hypothesis h may be a theory with makes postulates about unobservable entities and processes. In this sense, the theory of inductive probabilities is not committed to narrow empiricism and inductivism [Hempel, 1965; Niiniluoto and Tuomela, 1973].

For Carnap, inductive logic was part of his program of logical empiricism. The typical assumption of Carnap's system is that evidence is given by singular observational sentences. His negative results about the zero confirmation of laws seemed further to strengthen the "atheoretical thesis" [Lakatos, 1968a] which makes laws and theories dispensable in the theory of induction.

In his replies to Hilary Putnam, Carnap acknowledged that it would be desirable to construct inductive logic to "the total language of science, consisting of the observational language and the theoretical language" [Schilpp, 1963, p. 988]. This would allow that inductive predictions could take into account "also the class of the actually proposed laws". In particular, Carnap admitted that the meaning postulates of theories should be given the m-value one.

Carnap was one of architects of the view which claims that scientific theories include theoretical terms that are not explicitly definable by observational terms, but still theories formulated in the total language L of science should be testable by the consequences that they have in the observational sublanguage L_o of L. Hempel agreed, but raised in 1958 in "The Theoretician's Dilemma" the following puzzle for scientific realists: if a theory T in L achieves *deductive systematization* between observational sentences e and e' in L_o (i.e., $(T\&e) \vdash e'$, but not $e \vdash e'$), then the elimination methods of Ramsey and Craig show that the same deductive systematization is achieved by an observational subtheory of T in L_o. Therefore, theoretical terms are not after all logically indispensable for observational deductive systematization (see [Hempel, 1965]).

Hempel suggested that theoretical terms may nevertheless be logically indispensable for inductive systematization. Consider a simple theory T:

$$T = (x)(Mx \supset O_1x)\&(x)(Mx \supset O_2x),$$

where O_1 and O_2 are observational terms and M is a theoretical predicate. Then, by the first law, from O_1a one may inductively infer Ma, and, by the second law, from Ma one may infer O_2a. However, this attempt to establish an inductive link between O_1a and O_2a via T is not justified, since it relies at the same time on the incompatible principles CE and SC (see [Niiniluoto, 1972]).

The theoretician's dilemma was the starting point Raimo Tuomela in his studies on the deductive gains of the introduction of theoretical terms (see [Hintikka and Suppes, 1970]). The issue of inductive gains of theoretical terms was the topic of the doctoral dissertation of Ilkka Niiniluoto in 1973.

Let eIh state that h is "inducible" from e. Then theory T in L achieves *inductive systematization* between e and e' in L_o, if $(T\&e)Ie'$ but not eIe'. By using PR as the explication of I, these conditions can be written::

(IS) $P(e'/e\&T) > P(e')$
$P(e'/e) = P(e')$.

The first condition has an alternative interpretation:

(IS′) $P(e'/e\&T) > P(e'/T)$.

Both conditions presuppose that inductive probabilities can take genuine theories as their conditions. The relevant calculations are given in [Niiniluoto and Tuomela, 1973] by using Hintikka's generalized combined system. For example, if h a universal generalization in L_o, and T is a theory in L using theoretical terms, then the probability $P(h/e\&T)$ depends on the number b' of Q-predicates of L which are empty by h but not by T. For large values of n and α, we have approximately

(56) $$P(h/e\&T) \simeq \frac{1}{(1+\alpha/n)^{b'}}.$$

Comparison with (24) shows that $P(h/e) < P(h/e\&T)$ iff $b > b'$ (*op. cit.*, p. 38). In particular, if T is an explicit definition of M in terms of L_o, then $b = b'$. Hence,

(57) If T is an explicit definition of the theoretical terms in L by L_o, then $P(h/e\&T) = P(h/e)$.

The main result proved in [Niiniluoto and Tuomela, 1973, Ch. 9] is that theoretical terms can be logically indispensable for observational inductive systematization. This result shows that Hintikka's inductive logic helps to solve the theoretician's dilemma and thereby to give support to scientific realism.

Probabilities of the form $P(h/e\&T)$ can be used for *hypothetico-inductive* inferences, when T is a tentative theory. When $T\&e$ is accepted as evidence, e gives *observational support* and T gives *theoretical support* for h. Many formulas discussed in earlier sections can be reformulated by taking into account the background theory T. It is also possible to calculate directly probabilities of the form $P(C^w/e)$, where C^w is a constituent in the full language L with theoretical terms and e is observational evidence in L_o (see [Niiniluoto, 1976]). These applications prove that inductive logic can developed as a *non-inductivist* methodological program [Niiniluoto and Tuomela, 1973 Ch. 12].

Comparison of probabilities of the form $P(h/e)$ and $P(h/e\&T)$ illuminate also the inductive effects of conceptual change. Changes in the values of inductive parameters λ and α cannot be modelled in terms of Bayesian conditionalization (cf. Hintikka, 1966; 1987b; 1997]. The strategy of Niiniluoto and Tuomela [1973] is to keep parameters fixed, and to study the changes due conditionalization on conceptual or theoretical information. Result (57) shows that inductive logic satisfies a reasonable condition of linguistic invariance: if the new predicate M in L is explicitly definable by predicates of L_o, and T is the meaning postulate expressing this definition, then the probabilities in L *conditional on* T are equal to probabilities in L_o. Similarly, if $L1$ and $L2$ are intertranslatable, their inductive logics are equal [Niiniluoto and Tuomela, 1973, p. 175].

9 ANALOGY AND OBSERVATIONAL ERRORS

Inference by analogy is a traditional form of non-demonstrative reasoning. It can be regarded as a generalization of the deductive rule for *identity*:

$$\text{(RI)} \quad \frac{\begin{array}{c} F(a) \\ b = a \end{array}}{F(b).}$$

If identity = is replaced by the weaker condition of *similarity*, RI is replaced by

$$\text{(RS)} \quad \frac{\begin{array}{c} F(a) \\ b \text{ is similar to } a \end{array}}{F(b).}$$

The classical idea of explicating similarity is by *partial identity*: objects a and b share some of their properties. Using the terms of Keynes, the attributes that a and b agree on belong to their *positive analogy*, and the attributes that a and b disagree belong to their *negative analogy*. If objects a and b are known to agree on k attributes and disagree on m attributes, then J. S. Mill in his *A System of Logic* proposed to measure the strength of the analogical inference by its probability $k/(k+m)$. This suggestion means that analogical inference is simple enumerative induction with respect to properties.

Carnap required that a system of inductive logic should be able to handle inference by analogy. It turns out that Carnap's and Hintikka's systems give a satisfactory treatment of simple positive analogy, but fail if there is some negative analogy (cf. [Hesse, 1964]). Assume that, in a monadic language with k primitive predicates,

$$\begin{array}{l} F_1(x) = M_1(x)\&...\&M_m(x) \\ F_2(x) = M_1(x)\&...\&M_m(x)\&M_{m+1}(x)\&...\&M_n(x), \end{array}$$

where $1 \leq m < n$. Then the width of F_1 is $w_1 = 2^{k-m}$ and the width of F_2 is $w_2 = 2^{k-n}$. Hence, in the K-dimensional system of inductive logic,

$$P(F_2(b)/F_1(b)) = \frac{w_2}{w_1}$$

$$(58)\quad \begin{array}{rcl} P(F_2(b)/F_1(b)\&F_2(a)) & = & \frac{1-(K-w_2)f(0,1,1)}{1-(K-w_1)f(0,1,1)} \\ & > & \frac{w_2}{w_1} = P(F_2(b)/F_1(b)). \end{array}$$

(See [Niiniluoto, 1981, p. 7].) By choosing $f(0,1,1)$ as equal to its value in Carnap's λ-continuum $(\lambda/K)/(1+\lambda)$, the probability (58) is

$$(59)\quad \frac{1+w_2\lambda/K}{1+w_1\lambda/K}.$$

By putting $\lambda = K$ in (59), we obtain Carnap's analogy formula for his measure c^*:

$$(60)\quad \frac{1+w_2}{1+w_1}$$

(See [Carnap, 1950, pp. 569-570].) However, as soon as the evidence includes some known difference between objects a and b, the analogy influence disappears in (58).

Carnap's first attempt to handle the difficulty was a new "analogy parameter" η (see [Carnap and Stegmüller, 1959]). This device was applied in Hintikka's system by Pietarinen [1972] (see also [Festa, 2003]). In his posthumously published "Basic System", Carnap developed the idea that analogy can be accounted for by *distances between predicates* (see [Carnap, 1980]). Natural measures of such distances can be defined by means of Q-predicates. For two Q-predicates Q_u and Q_v of the form $(\pm)M_1x\&...\&(\pm)M_kx$, the distance $d(Q_u, Q_v) = d_{uv}$ is m/k if they disagree on m of the k primitive predicates $M_1, ..., M_k$. For example, in a monadic language L with two primitive predicates, we have $d(M_1x\&M_2x, M_1x\& \sim M_2x) = \frac{1}{2}$, and $d(M_1x\&M_2x, \sim M_1x\& \sim M_2x) = 1$. For languages with families of predicates, each primitive dichotomy $\{Mj, \sim M_j\}$ is replaced by a family $\mathbf{M}_j$ of predicates, $j = 1, ..., k$, and each family is assumed to have its internal distance d_j (e.g., distance between colours, distance between discrete values of age). Then the earlier definition of d_{uv} can be generalized by defining the distance between two Q-predicates by the Euclidean metric relative to the k dimensions. With this distance function, the set of Q-predicates becomes a metric space of concepts (see [Niiniluoto, 1987]).

If d_{uv} is normalized so that $0 \leq d_{uv} \leq 1$, the *resemblance* r_{uv} between two Q-predicates Q_u and Q_v can be defined by $r_{uv} = 1 - d_{uv}$. Alternatively, we can use $r_{uv} = 1/(1+d_{uv})$. Two individuals a and b, which satisfy the Q-predicates Q_u and Q_v, respectively, are then completely similar (relative to the expressive power of language L) if and only if $r_{uv} = 1$.

The principle of Positive Instantial Relevance (4) states that the observation of individuals of kind Q_i increases our expectation to find completely similar individuals in the universe. A natural modification of this principle is to allow *similarity influence*: the observation of individuals of kind Q_i increases the expectation to find individuals in cells Q_j that are close to Q_i. In terms of singular probabilities, similarity influence should allow that

$$P(Q_2(b)/Q_1(a)) > P(Q_3(b)/Q_1(a)) \text{ iff } d_{12} < d_{13} \text{ iff } r_{12} > r_{13}.$$

(See [Carnap, 1980, p. 46].) Note that this principle already helps to solve the problem of negative analogy. Similarly, in the context of Hintikka's system or the K-dimensional system, the probability $P(C^w/e)$ of the constituent C^w should reflect the distances of the cells CT^w claimed to be non-empty by C^w and the cells CT_e already exemplified in evidence e.

Proposals to this effect were given in [Niiniluoto, 1980; 1981]: the probability $P(Q_i(a_{n+1})/e)$ can be modified by multiplying it with a factor which expresses the minimum distance of Q_i from the cells CT_e or the weighted influence of the observed individuals in the cells CT_e. It turned out that these specific proposals did not always satisfy the principle of Positive Instantial Relevance (4) and Reichenbach's Axiom (3). An alternative, proposed by Kuipers [1984; 1988], is to allow that the analogy influence gradually vanishes when the evidence increases. Kuipers replaces Carnap's representative function (2) by

(61) $$\frac{n_i + \alpha_i(e_n) + \lambda/K}{n + \alpha(n) + \lambda}$$

where $\alpha_i(e_n)$ is the analogy profit of cell Q_i from sample e_n and

$$\alpha(n) = \sum\nolimits_{i=1}^{K} \alpha_i(e_n)$$

is the analogy in the first n trials. To guarantee the validity of Reichenbach's axiom, the marginal analogy of the nth trial $\alpha(n) - \alpha(n-1)$ is assumed to decrease from its positive initial value to zero when n grows without limit. To obtain a theory of inductive generalization with analogy influence, probabilities of the form $P(Q_i(a_{n+1})/e_n\&C^w)$ in the K-dimensional system can be modified in a similar way by allowing initial and evidence-based analogy profits (see [Niiniluoto, 1988]).[10]

Distances between Q-predicates have been applied also in a proposal concerning the treatment of *observational errors* in inductive logic (see [Niiniluoto, 1997]). In Hintikka's system, the likelihoods $P(Q_i(a_{n+1})/e_n\&C^w)$ are defined by restricting the Q-predicates Q_i to those allowed by the constituent C^w (see (15)). Thereby the possibility of observational errors is excluded. The so called Jeffrey conditionalization handles uncertain evidence by directly modifying probabilities, without conditionalization relative to a possibly mistaken evidence statement. Instead, standard applications of Bayesian statistical inference include an error distribution, which allows the observed data to deviate from the true value of a parameter. To build this idea to inductive logic, let $Q_j(a)$ state that object a is Q_j, and S_j(a) state that object a seems to be Q_j. Then it is natural to assume that errors of observation depend on distances between predicates:

(62) The probability $p_{ij} = P(S_i(a)/Q_j(a))$ is a decreasing function of the distance $d_{ij} = d(Q_i, Q_j)$.

[10]Other attempts to handle inductive analogy include Skyrms [1993b], di Maio [1996], Festa [1996], Maher [2000], and Romeyn [2005].

The probability of correct observation, or observational reliability, can be assumed to be a constant $p_{ii} = P(S_i(a)/Q_i(a)) = \beta$ for all $i = 1, ..., K$. In standard inductive logic without observational errors, $p_{ij} = 1$ if $i = j$ and 0 otherwise. By Bayes's Theorem, probabilities of the form $P(Q_i(a)/S_j(a))$ can be calculated. Further, probabilities of the form $P(S_i(a_{n+1})/e_n(S)\&C^w)$ have to be defined, when $e_n(S)$ is like an ordinary sample description but Q_js replaced by S_js.

10 TRUTHLIKENESS

In his campaign against induction and inductive logic, Karl Popper defined in 1960 comparative and quantitative notions of *truthlikeness* or *verisimilitude* (see [Popper, 1963]). Instead of assessing the inductive probability that a hypothesis is true, truthlikeness is an objective concept expressing how "close" to the truth a scientific theory is. Maximum degree of truthlikeness is achieved by a theory which is completely and comprehensively true, so that verisimilitude combines the ideas of truth and information content while probability combines truth with lack of content. According to Popper, in some situations we may have strong arguments for claiming that we have made progress toward the truth, and he proposed his own measure of corroboration as an epistemic indicator of verisimilitude.

Popper's attempted definition failed, since it did not allow the comparison of two false theories. Since 1974, a new program for defining degrees of truthlikeness was initiated by Risto Hilpinen, Pavel Tichy, Ilkka Niiniluoto, Raimo Tuomela, and Graham Oddie. All of them employed the notion of similarity or likeness between possible worlds or between Hintikkian constituents describing such possible worlds. Niiniluoto [1977] further proposed that inductive probabilities can be applied in the estimation of degrees of verisimilitude.

In the likeness approach (see [Niiniluoto, 1987]), truthlikeness is defined relative to a cognitive problem $B = \{h_i | i \in I\}$, where the elements of B are mutually exclusive and jointly exhaustive (see Section 7). The unknown true element h* of B is the *target* of the problem B. The basic step is the introduction of a real-valued function $\Delta : B \times B \rightarrow \mathbf{R}$ which expresses the *distance* $\Delta(h_i, h_j) = \Delta_{ij}$ between the elements of B (i.e., complete answers). Here $0 \leq \Delta_{ij} \leq 1$, and $\Delta_{ij} = 0$ iff $i = j$. This distance function Δ has to be specified for each cognitive problem B separately, but there are canonical ways of doing this for special types of problems. First, if B is the set of state descriptions, the set of structure descriptions, or the set of constituents of a first-order language L, the distance Δ can be defined by counting the differences in the standard syntactical form of the elements of B. For example, a monadic constituent tells that certain kinds of individuals (given by Q-predicates) exist and others do not exist; the simplest distance between monadic constituents is the relative number of their diverging claims about the Q-predicates. If a monadic constituent C_i is characterized by the class CT_i of Q-predicates that are non-empty by C_i, then the *Clifford distance* between C_i and C_j is the size of the symmetric difference between CT_i and CT_j:

(63) $|CT_i \Delta CT_j|/K$.

Secondly, Δ may be directly definable by a natural metric underlying the structure of B. For example, if the elements of B are point estimates of an unknown real-valued parameter, their distance can be given simply by the geometrical or Euclidean metric on $\mathbf{R}$ (or $\mathbf{R}^K$). If the elements of B are quantitative laws, then their distance is given by the Minkowski metrics between functions.

The next step is the extension of Δ to a function $B \times D(B) \to \mathbf{R}$, so that $\Delta(h_i, g)$ expresses the distance of a partial answer $g \in D(B)$ from $h_i \in B$. Let $g \in D(B)$ be a potential answer with

$$\vdash g = \bigvee_{i \in I_g} h_i,$$

where $I_g \subseteq I$. Define the *minimum distance* of g from h_i by

(64) $\Delta_{\min}(h_i, g) = \min_{j \in I_g} \Delta_{ij}.$

Then g is *approximately true* if $\Delta_{\min}(h^*, g)$ is sufficiently small. Degrees of approximate truth can be defined by

(65) $AT(g, h^*) = 1 - \Delta_{min}(h^*, g)$.

Hence, $AT(g, h^*) = 1$ if and only if g is true. Instead, the notion of truthlikeness Tr should have its maximum when g is identical with the complete truth h^*:

(66) $Tr(g, h^*) = 1$ iff $\vdash g \equiv h^*$.

To introduce a concept which satisfies the condition (66), truthlikeness should include a factor which tells how effectively a statement is able to exclude falsities. This can be expressed by the relativized sum-measure which includes a penalty for each mistake that g allows, and weights this mistake by its distance from the target:

(67) $\Delta_{\text{sum}}(h_i, g) = \sum_{j \in I_g} \Delta_{ij} / \sum_{j \in I} \Delta_{ij}$.

At the same time, a truthlike statement should preserve truth as closely as possible. As a sufficient condition, one might suggest that a partial answer g' is more truthlike than another partial answer g if g' is closer to the target h^* than g with respect to both the minimum distance and the sum distance, but only few answers in $D(B)$ would be comparable by this criterion. Full comparability is achieved by the *min-sum* measure Δ_{ms}, where the weights γ and γ' indicate our cognitive desire of finding truth and avoiding error, respectively:

(68) $\Delta_{ms}^{\gamma\gamma'}(h_i, g) = \gamma \Delta_{\min}(h_i, g) + \gamma' \Delta_{\text{sum}}(h_i, g)$ $(\gamma > 0, \gamma' > 0)$.

Then a partial answer g is *truthlike* if its min-sum distance from the target h^* is sufficiently small. One partial answer g' is *more truthlike* than another partial answer g if $\Delta_{ms}^{\gamma\gamma'}(h^*, g') < \Delta_{ms}^{\gamma\gamma'}(h^*, g)$.

The *degree of truthlikeness* $Tr(g, h^*)$ of $g \in D(B)$ (relative to the target hj in B) is defined by

(69) $Tr(g, h^*) = 1 - \Delta_{ms}^{\gamma\gamma'}(h^*, g)$.

This measure Tr has many nice features. For a tautology t, we have $Tr(t, h^*) = 1 - \gamma'$. For a complete answer h_i, we get the expected result that $Tr(h_i, h^*)$ decreases with the distance $\Delta(h_i, h^*)$. This gives a practical rule for the choice of the parameters: γ and γ' should be chosen so that the complete answers closest to h^* have a degree of truthlike larger than $1 - \gamma'$, while the most misleading ones should be worse than ignorance. (For example, we may choose $\gamma'/\gamma \approx 1/2$.)

If the distance function Δ on B is trivial, i.e., $\Delta_{ij} = 1$ for all $i \neq j$, then $Tr(g, h^*)$ reduces to a special case of Levi's [1967] definition of epistemic utility.

As the target h^* is unknown, the value of $Tr(g, h^*)$ cannot be directly calculated by our formulas (68) and (69). However, there is a method of making rational comparative judgments about verisimilitude, if we have - instead of certain knowledge about the truth — rational degrees of belief about the location of truth. Thus, to *estimate* the degree $Tr(g, h^*)$, where h^* is unknown, assume that there is an epistemic probability measure P defined on B, so that $P(h_i/e)$ is the rational degree of belief in the truth of h_i given evidence e. The *expected degree of verisimilitude* of $g \in D(B)$ given evidence e is then defined by

(70) $\text{ver}(g/e) = \sum_{i \in I} P(h_i/e) Tr(g, h_i)$.

If B includes constituents of a monadic language and P is chosen to be Hintikka's inductive probability measure, explicit formulas for calculating the value of ver(g/e) can be provided (see [Niiniluoto, 1987, Ch.9.5; Niiniluoto, 2005a]).

If the probability distribution P on B given e is even, so that we are completely ignorant of the true answer to the cognitive problem, the values of ver(h_j/e) are also equal to each other. On the other, the difference between expected verisimilitude and posterior probability is highlighted by the fact that ver(g/e) may be high even when e refutes g and thus $P(g/e) = 0$. This is also a crucial difference between ver and most probabilistic measures of confirmation and corroboration (e.g., (39) and (40)).

Equation (70) gives us a comparative notion of estimated verisimilitude: g' *seems more truthlike* than g on evidence e, if and only if ver$(g/e) <$ ver(g'/e). Function ver defines also an acceptance rule: given evidence e, accept as the most truthlike theory g which maximizes ver(g/e). This rule is comparable to the Bayesian decision theory, when the loss function is proportional to distances from the truth. Besides comparisons to point estimation, a theory of Bayesian interval estimation can be based upon the rule of accepting the most truthlike hypothesis (see [Niiniluoto, 1986; 1987; Festa, 1986]).

The relation between the functions Tr and ver is analogous to the relation between truth value tv (1 for true, 0 for false) and probability P, i.e.,

(71) $Tr : \text{ver} = tv : P$.

This can be seen from the fact that the posterior probability $P(g/e)$ equals the *expected truth value* of g on e:

$$\sum_{i \in I} P(h_i/e) tv(g, h_i) = \sum_{i \imath I_g} P(h_i/e) = P(g/e).$$

(Cf. (48).) By (71), expected verisimilitude ver(g/e) is an estimate of real truthlikeness $Tr(g, h^*)$ in the same sense in which posterior epistemic probability is an estimate of truth value. The standard form of Bayesianism which evaluates hypotheses on the basis of their posterior probability can thus be understood as an attempt to maximize expected truth values, while ver replaces this goal by the maximization of expected truthlikeness.

Definition (70) guarantees that ver$(g/e) = Tr(g, h_j)$ if evidence e entails one of the strong answers h_j. In Hintikka' s system of inductive logic, the result (13) guarantees that asymptotically it is precisely the boldest constituent compatible with the evidence that will have the largest degree of estimated verisimilitude:

(72) ver$(g/e) \to Tr(g, C^c)$, when $n \to \infty$ and c is fixed.

(73) ver$(g/e) \to 1$ iff $\vdash g \equiv C^c$, when $n \to \infty$ and c is fixed.

11 MACHINE LEARNING

Besides expected verisimiltude ver, other ways of combining closeness to the truth and epistemic probability include the concepts of *probable verisimilitude* (i.e., the probability given e that g is truthlike at least to a given degree) and *probable approximate truth* (i.e., the probability given e that g is approximately true within a given degree). (See [Niiniluoto, 1987; 1989].)

Let g in $D(B)$ be a partial answer, and $\varepsilon \in 0$ a small real number. Define

(74) $V_\varepsilon(g) = \{h_i \text{ in } B | \Delta_{\min}(h_i, g) \leq \varepsilon\}$.

Denote by g^ε the "blurred" version of g which contains as disjuncts all the members of the neighborhood $V_\varepsilon(g)$. Then $g \vdash g^\varepsilon$, and g is approximately true (within degree ε) if and only if g^ε is true. The probability that the minimum distance of g from the truth h^* is not larger than ε, given evidence e, defines at the same time the posterior probability that the degree of approximate truth $AT(g, h^*)$ of g is at least $1 - \varepsilon$:

(75) $PAT_{1-\varepsilon}(g/e) = P(h^* \in V_\varepsilon(g)/e) = \sum_{h_i \in V_\varepsilon(g)} P(h_i/e)$.

PAT defined by (75) is thus a measure of *probable approximate truth*. Clearly we have always $P(g/e) \leq PAT_{1-\varepsilon}(g/e)$. When ε decreases toward zero, in the limit we have $PAT_1(g/e) = P(g/e)$. Further, $PAT_{1-\varepsilon}(g/e) > 0$ if and only if $P(g^{\varepsilon}) > 0$. Unlike ver, PAT shares with P the property that logically weaker answers will have higher PAT-values than stronger ones. An important feature of both probable verisimilitude and probable approximate truth is that their values can be non-zero even for hypotheses with a zero probability on evidence: it is possible that $PAT_{1-\varepsilon}(g/e) > 0$ even though $P(g/e) = 0$.

The notion of probable approximate truth is essentially the same as the definition of PAT in the theory of *machine learning* (see [Niiniluoto, 2005b]). What the AI community calls "concepts" are directly comparable to monadic constituents, and thereby "concept learning" can be modelled by Hintikka-style theory of inductive generalization. While AI accounts of concept learning are purely eliminative, as they recognize positive and negative instances, the treatment by Hintikka's system allows one to include enumerative and analogical considerations as well.

The convergence properties of Hintikka's inductive probabilities are also comparable to the work on *formal learning theory*. Recall that in Hintikka's system the posterior probability $P(C^c/e)$ approaches one when c is fixed and n grows without limit. But this result (13) states only that our degrees of belief about C^c *converge to certainty* on the basis of inductive evidence. It does not yet guarantee that C^c is identical with the *true* constituent C^*. Similarly, by (73) we know that the expected verisimilitude $ver(C^c/e)$ converges to one when c is fixed and n grows without limit. Again this does not guarantee that the "real" truthlikeness $Tr(C^c, h^*)$ of C^c is maximal, i.e., that C^c is true. For these stronger results an additional *evidential success condition* is needed:

(ESC) Evidence e is true and fully informative about the variety of the world w.

ESC means that e is exhaustive in the sense that it exhibits (relative to the expressive power of the given language L) all the kinds of individuals that exist in the world (see [Niiniluoto, 1987, p. 276]). With ESC we can reformulate our results so that they concern *convergence to the truth*:

(13′) If ESC holds, $P(C^*/e) \to 1$, when c is fixed and $n \to \infty$.

(72′) If ESC holds, $ver(g/e) \to Tr(g, C^*)$, when $n \to \infty$ and c is fixed.

(73′) If ESC holds, $ver(g/e) to 1$ iff $\vdash g \equiv C^*$, when $n \to \infty$ and c is fixed.

Similar modifications can be made in the convergence results about probable approximate truth.

Now we can see that the convergence results of formal learning theory are not stronger than those of probabilistic approaches, even though they demand success with respect to all data streams, since they *presuppose* something like the success condition ESC: decidability in the limit assumes that the data streams are "compelete in that they exhaust the relevant evidence" [Earman, 1992, p. 210] or "perfect" in that "all true data are presented and no false datum is presented" and all objects are eventually described [Kelly, 1996, p. 270]. Without ESC even "global underdetermination" cannot be avoided (*ibid.*, p. 17), since we cannot be certain that even an infinite sample of swans refutes the false generalization 'All swans are white': it is logically possible that an infinite stream of white swans is picked out from a world containing white and black swans.

A more formal way of expressing these conclusions is to note that the learner of Hintikka's system uses *epistemic* probabilities in her prior distribution $P(C^w)$ and likelihoods $P(e/C^w)$. The result (13) as such needs no assumption that behind these likelihoods there are some objective conditions concerning the sampling method.[11] The same observation can be made about the famous results of de Finetti and L.J. Savage about the convergence of opinions in the long run, when the learning agents start from different non-dogmatic priors ([Howson and Urbach, 1989; Earman, 1992]; cf. [Niiniluoto, 1984, p. 102]). But it is possible to combine a system of inductive logic with the assumption that the evidence arises from a *fair* sampling procedure which gives each kind of individual an objective non-zero chance of appearing the evidence e [Kuipers, 1977b], where such chance is defined by a physical probability or *propensity*. As such propensities do not satisfy the notorious Principle of Plenitude, claiming that all possibilities will sometimes be realized, they do not exclude infinite sequences which violate ESC (see [Niiniluoto, 1988b]). But such sequences are extremely improbable by the convergence theorems of probability calculus (cf. [Festa, 1993, p. 76]). Suppose that we draw with replacement a fair sample of objects from an urn w. Let r be the proportion of objects of kind A in w. Then the objective probability of picking out an A is also r. The Strong Law of Large Numbers states now that the observed relative frequency k/n converges *with probability one* to the unknown value of r. Such "almost sure" convergence is weaker than convergence in the ordinary sense. The reason for using this notion of convergence is that there are no logical reasons for excluding such non-typical sequences of observations that violate ESC — even though their measure among all possible sequences is zero.

Formal learning theory and probabilistic theories of induction, as plausible attempts to describe scientific inquiry, are in the same boat with respect to the crucial success conditions: ESC is precisely the reason why inductive inference is always non-demonstrative or fallible even in the ideal limit, since there are no logical reasons for excluding the possibility that ESC might be incorrect. The conclusion to be drawn from these considerations can be stated as follows: the best results for a fallibilist "convergent realist" do not claim decidability in the limit or even gradual decidability, but rather convergence to the truth with probability one.

[11] So we need not follow Kuipers [1977b] in the claim that the appropriate applications of the K-dimensional system are restricted to "multinomial contexts" which have underlying objective probability distributions associated with repeatable experiments (see [Niiniluoto, 1983]). But of course this condition may be a contextual presupposition of inductive logic.

12 EVALUATION OF THE HINTIKKA PROGRAM

Inductive probability measures, or their systems, can be viewed as a sequence of successive "theories" about rational inferential practices. There is a remarkable continuity among the proposals of inductive logicians: the K-dimensional system is a generalization of Hintikka's generalized combined system, which is a generalization of and modification of Carnap's c^*, which is a generalization of Laplace's rule of succession. Moreover, Hacking suggest that "a Leibnizian ought to like Carnap's preferred c^*" [Hacking, 1975, p. 141]. The successor "theories" in this sequence contain their predecessors as special or limiting cases. The sequence is also progressive by the Lakatosian standards: each member solves more problems than its predecessor. In particular, Hintikka's system is able to handle inductive generalization, while the K-dimensional system shows why the systems from Laplace to Carnap failed in this task. Hintikka's system can be modified so that it accounts also for analogical inference and observational errors, and it can be applied to more complex methodological situations than those listed by Carnap — in particular, to cases involving theoretical premises and conclusions.

Hintikka's philosophical conclusion from his two-dimensional continuum was that inductive probabilities are always relative to extra-logical parameters. When α is small, the posterior probability of universal generalizations grows rapidly. In this sense, the choice of a small α is an index of boldness of the investigator, or the choice of a large α is an index of caution. Alternatively, the choice of α can be regarded as a regularity assumption about the lawlikeness of the relevant universe U.[12] According to Hintikka, this abolishes the hopes for purely logical probabilities in the sense of Carnap's core assumptions C3, C5, and C6 (see [Hintikka, 1987b; 1997, p. 317]).

After the introduction of the λ-continuum, Carnap's position was in fact quite similar. He argued that, for certain kinds of processes with objective probabilities, there is an optimal value of λ [Carnap, 1952]; methods for estimating such an optimal value in different contexts are elaborated by Festa [1995]. In his later work, Carnap [1980] related the choice of λ to "objectivist" matters of attribute distance, but also admitted that the choice of λ may be a symptom of the investigator's personality. He concluded that λ should be included in the interval between $\frac{1}{2}$ and K, and finally gave the recommendation to take $\lambda = 1$.

The contextual or extra-logical character of inductive probabilities brings inductive logic close to the mainstream of Bayesianism. From this perspective, inductive logic is a form of Bayesianism which investigates epistemic probabilities in relation to several kinds of structural assumptions — such as the expressive power of languages, the choice of inductive parameters, order and disorder in the universe, and similarity influence. Using terms from Bayesian decision theory, such "structural axioms" do no belong to the "pure theory of rationality" [Suppes, 1969, p.

[12] This regularity is relative to the expressive power of universal generalizations. Similarly, Walk [1966] argues that Carnap's λ is directly proportional to the statistical entropy of the universe.

95], but rather limit those situations where the theory is applicable [Niiniluoto, 1977]. In any case, inductive logic finds more structure (depending on linguistic frameworks and contextual assumptions) to the probabilities than the personalist Bayesians. Instead of trying to find one and only one system of inductive reasoning, or to fix a unique prior probability distribution for each problem as some "objective Bayesians" suggest, the study of probability measures in relation to different kinds of structural assumptions leads to a rich framework with potential applications to many types of methodological situations (cf. [Niiniluoto, 1977; 1983, 1988a]).

In comparison with standard Bayesian statistics, the Hintikka program has a special character (cf. [Niiniluoto, 1983]). A statistician might suggest that constituents could be replaced by statistical hypotheses: for example, the claim that all individuals belong to cell Q_1 could be expressed by the claim that the proportion of individuals in Q_1 belongs to a short interval $(q, 1]$. The latter hypothesis can be studied by the techniques of Bayesian statistics, i.e., by introducing a prior probability distribution over the parameter space and applying Bayes's Theorem relative to observed data, so that no separate inductive logic is needed. However, there is an important difference between the two cases: while the observation of one counterexample falsifies the constituent, it does not refute the statistical counterpart of a constituent. Further, the Hintikkian prior distributions which assign non-zero probabilities to constituents give positive weight to subsets of the parameter space with the geometrical measure zero. To illustrate, in a classification system with two Q-predicates, the parameter space consists of pairs $\langle p_1, p_2 \rangle$, where p_i is the proportion of individuals in Q_i, and $p_{1+}p_2 = 1$. There are three constituents in this case, and two of them correspond to singular sets of measure zero:

$$
\begin{aligned}
C_1 &= \{\langle 1, 0 \rangle\} \\
C_2 &= \{\langle 0, 1 \rangle\} \\
C_3 &= \{\langle p, 1-p \rangle | 0 < p < 1\}.
\end{aligned}
$$

Laplace's prior, i.e., a uniform distribution over the whole parameter space, gives probability one to the atomistic constituent C_3. This argument illuminates how Hintikka's improvement of Carnap's approach constitutes a break away from the standard Bayesian statistics.

Hintikka [1997] further points out that inductive logic with parameters is not "purely Bayesian", since changes of probabilities due to new choices of the values of λ and α cannot be described by conditionalization. In this sense, inductive logic can be a tool for studying language change (cf. [Niiniluoto and Tuomela, 1973]). Hintikka [1997] also suggests the move of introducing all extra-logical assumptions (such as the choice of parameters) as explicit premises, so that inductive logic would become a part of deductive logic. The same point is repeated in Hintikka's "Replies" in [Auxier and Hahn, 2006].

It is interesting that, in his later work, Hintikka has turned out to be a staunch critic of traditional treatments of induction. He has developed an "interrogative model of inquiry", based on a dialogue between two players, "the inquirer" and "Nature" (see [Hintikka, 1981; 1987a; 1988; 1992]). Within this model, induction plays only a very modest role in the micro-level of experimental inquiry, and Hume's problem seems to disappear from the "serious theory of the scientific method". Hintikka argues that in typical controlled experiments Nature provides us answers that are already general or have at least AE-complexity. Thus, we should reject the "super-dogma" of the "Atomistic Postulate" which restricts Nature's answers to negated or unnegated atomistic propositions.

In my view, Hintikka's study of inquiry on different levels of complexity is highly fruitful. However, one may doubt whether the step of inductive generalization can really dispensed in all important contexts by explicitly formulated strong regularity assumptions.[13] Moreover, Hintikka's own system of inductive logic is not tied to the Atomistic Postulate in the same way as Carnap's, since one can study probabilities of the form $P(g/e\&T)$ and $P(g/T)$ where T is a background theory (see [Niiniluoto, 1997]).

Carnap's original motivation for developing a system of inductive logic came from his logical empiricism. Many philosophers of science feel that this motivation is largely outdated, as it is too much oriented to naive and old-fashioned empiricism. However, the picture looks different when we take notice of Hintikka's ability to deal with genuine inductive generalizations and scientific theories. To use the slogan of Niiniluoto and Tuomela [1973], inductive logic can be "non-inductivist" — free of simple assumptions of the role of induction in scientific inference. With its applications to theoretical systematization, semantical information, explanation, abduction, and truthlikess, inductive logic can be a tool for a critical scientific realist.

Lakatos made fun of inductive logicians who, he argued, desperately attempt to calculate exact numerical degrees of confirmation for specific scientific theories. But the introduction of quantitative degrees of probability (or systematic power, truthlikeness, etc.) with absolute values can be viewed primarily as a tool of making comparative judgements between rival theories. Moreover, inductive logic need not be understood to provide such manuals of calculation. Rather, it is a study of the general principles of probabilistic and uncertain reasoning (cf. [Niiniluoto, 1980, p. 226]). In this respect, even though Hintikka's system has not yet been fully extended to first-order languages, the generalization of distributive normal forms and constituents indicates that the most important general features of induction in monadic cases extend to the richer languages as well.

On the other hand, at least for the manageable cases of monadic languages, it is certainly possible to implement the formulas of Hintikka's system as computer programs. The development of computer science, robotics, and artificial intelligence has opened the further perspective that inductive logic could be useful as a

[13]In his "Replies" [Auxier and Hahn, 2006, p. 778], Hintikka states: "If these premises are spelled out explicitly, we no longer need any independent rules of inductive inference. Ordinary deductive logic with probability theory does the whole job." Here Hintikka seems to ignore the distinction between factual and probabilistic rules of detachment (see note 8 above).

framework of machine learning.[14] In their attempt to model uncertain inference by computer programs, the AI community is reinventing many logical approaches proposed earlier by philosophers. In addition to its applications in the philosophy of science, it may this turn out that the future paradigm of systems of induction can be found in the field of machine learning.

BIBLIOGRAPHY

[Auxier and Hahn, 2006] R. E. Auxier and L. E. Hahn, eds. *The Philosophy of Jaakko Hintikka*. Chicago and La Salle: Open Court, 2006.

[Bogdan, 1976] R. J. Bogdan, ed. *Local Induction*. Dordrecht: Reidel, 1976.

[Bogdan, 1987] R. J. Bogdan, ed. *Jaakko Hintikka*. Dordrecht: Reidel, 1987.

[Carnap, 1945] R. Carnap. On Inductive Logic. *Philosophy of Science* **12**, 72-97, 1945.

[Carnap, 1950] R. Carnap. *The Logical Foundations of Probability*. Chicago: University of Chicago Press, 1950. (2^{nd} ed. 1962.)

[Carnap, 1952] R. Carnap. *The Continuum of Inductive Methods*. Chicago: The University of Chicago Press, 1952.

[Carnap, 1968] R. Carnap. The Concept of Constituent-Structure. In Lakatos (1968b), pp. 218-220, 1968.

[Carnap, 1971] R. Carnap. Inductive Logic and Rational Decisions. In [Carnap and Jeffrey, 1971, pp. 5-32].

[Carnap, 1980] R. Carnap. A Basic System of Inductive Logic Part II. In [Jeffrey, 1980, pp. 7-155].

[Carnap and Bar-Hillel, 1952] R. Carnap and Y. Bar-Hillel. An Outline of the Theory of Semantic Information. In Y. Bar-Hillel, ed., *Language and Information*, pp. 221–274. Reading, Mass.: Addison-Wesley, 1952.

[Carnap and Jeffery, 1971] R. Carnap and R. C. Jeffrey, eds. *Studies in Inductive Logic and Probability vol. I*. Berkeley and Los Angeles: University of California Press, 1971.

[Carnap and Stegmüller, 1959] R. Carnap and W. Stegmüller. *Induktive Logik und Wahrscheinlichkeit*. Springer: Wien, 1959.

[Cohen and Hesse, 1980] L. J. Cohen and M. Hesse, eds. *Applications of Inductive Logic*. Oxford: Oxford University Press, 1980.

[Earman, 1992] J. Earman. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, Mass.: The MIT Press 1992.

[Fenstad, 1968] J. E. Fenstad. The Structure of Logical Probabilities. *Synthese* **18**, 1-27, 1968.

[Festa, 1986] R. Festa. A Measure for the Distance Between an Interval Hypothesis and the Truth. *Synthese* **67**, 273–320. 1986.

[Festa, 1993] R. Festa. *Optimum Inductive Methods. A Study in Inductive Probabilities, Bayesian Statistics, and Verisimilitude*. Dordrecht: Kluwer, 1993.

[Festa, 1995] R. Festa. Verisimilitude, Disorder, and Optimum Prior Probabilities. In T. A. F. Kuipers and A. R. Mackor, eds., *Cognitive Patterns in Science and Common Sense*. Amsterdam: Rodopi, pp. 299–320, 1995.

[Festa, 1996] R. Festa. Analogy and Exchangeability in Predictive Inferences. *Erkenntnis* **45**, 229–252, 1996.

[Festa, 1999] R. Festa. Bayesian Confirmation. In M. C. Galavotti and A. Pagnini, eds., *Experience, Reality, and Scientific Explanation*. Dordrecht: Kluwer, pp. 55–87, 1999.

[Festa, 2003] R. Festa. Induction, Probability, and Bayesian Epistemology. In L. Haaparanta and I. Niiniluoto, eds., *Analytic Philosophy in Finland*. Amsterdam: Rodopi, pp. 251-284, 2003.

[Gavroglu *et al.*, 1989] K. Gavroglu, Y. Goudaroulis, and P. Nicolacopulos, eds. *Imre Lakatos and Theories of Scientific Change*. Dordrecht: Kluwer, 1989.

[Hacking, 1971] I. Hacking. The Leibniz-Carnap Program for Inductive Logic. *The Journal of Philosophy* **68**, 597-610, 1971.

[14] It is remarkable that already Carnap [1971] spoke about "inductive robots".

[Hacking, 1975] I. Hacking. *The Emergence of Probability*. Cambridge: Cambridge University Press, 1975.

[Helman, 1988] D. H. Helman, ed. *Analogical Reasoning*. Dordrecht: Kluwer, 1988.

[Hempel, 1965] C. G. Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press, 1965.

[Hesse, 1964] M. Hesse. Analogy and Confirmation Theory. *Philosophy of Science* **31**: 319-324, 1964.

[Hesse, 1974] M. Hesse. *The Structure of Scientific Inference*. London: Macmillan, 1974.

[Hilpinen, 1966] R. Hilpinen. On Inductive Generali1zation in Monadic First-Order Logic with Identity. In Hintikka and Suppes (1966), pp. 133–154, 1966.

[Hilpinen, 1968] R. Hilpinen. *Rules of Acceptance and Inductive Logic,* Acta Philosophica Fennica **21**. Amsterdam: North-Holland 1968.

[Hilpinen, 1970] R. Hilpinen. On the Information Provided by Observations. In [Hintikka and Suppes, 1970, pp. 97–122].

[Hilpinen, 1971] R. Hilpinen. Relational Hypotheses and Inductive Inference. *Synthese* **23**, 266–286, 1971.

[Hilpinen, 1972] R. Hilpinen. Decision-Theoretic Approaches to Rules of Acceptance. In R. E. Olson and A. M. Paul, eds., *Contemporary Philosophy in Scandinavia*. Baltimore: The John Hopkins Press, pp. 147–168, 1972.

[Hilpinen, 1973] R. Hilpinen. Carnap's New System of Inductive Logic. *Synthese* **25**, 307–333, 1973.

[Hilpinen, 1989] R. Hilpinen. Von Wright on Confirmation Theory. In [Schilpp and Hahn, 1989, pp. 127–146].

[Hintikka, 1965a] J. Hintikka. Towards a Theory of Inductive Generalization. In Y. Bar-Hillel, ed., *Proceedings of the 1964 International Congress for Logic, Methodology, and Philosophy of Science*. Amsterdam: North-Holland, pp. 274–288, 1965.

[Hintikka, 1965b] J. Hintikka. On a Combined System of Inductive Logic. *Studia Logico-Mathematica et Philosophica in Honorem Rolf Nevanlinna,* Acta Philosophica Fennica **18**. Helsinki: Societas Philosophica Fennica, pp. 21–30, 1965.

[Hintikka, 1966] J. Hintikka. A Two-dimensional Continuum of Inductive Methods. In [Hintikka and Suppes, 1966, pp. 113–132].

[Hintikka, 1968a] J. Hintikka. The Varieties of Information and Scientific Explanation. In B. van Rootselaar and J. F. Staal (Eds.), *Logic, Methodology and Philosophy of Science III, Proceedings of the 1967 International Congress*: Amsterdam: North-Holland, pp. 151–171, 1968.

[Hintikka, 1968b] J. Hintikka. Induction by Enumeration and Induction by Elimination [with Reply]. In [Lakatos, 1968b, pp. 191–216 and 223–231. 1968.

[Hintikka, 1968c] J. Hintikka. The possibility of rules of acceptance (Comments on Kyburg). In [Lakatos, 1968b, pp. 144–146].

[Hintikka, 1968d] J. Hintikka. Conditionalization and Information (Comments on Carnap). In [Lakatos, 1968b, pp. 303–306].

[Hintikka, 1969a] J. Hintikka. Inductive Independence and the Paradoxes of Confirmation. In N. Rescher (Ed.), *Essays in Honor of Carl G. Hempel*. Dordrecht: Reidel, pp. 24–46, 1969.

[Hintikka, 1969b] J. Hintikka. Statistics, Induction, and Lawlikeness: Comments on Dr. Vetter's Paper. *Synthese* **20**, 72–85, 1969.

[Hintikka, 1970] J. Hintikka. On Semantic Information. In [Hintikka and Suppes, 1970, pp. 3–27].

[Hintikka, 1971] J. Hintikka. Unknown Probability, Bayesianism, and de Finetti's Representation Theorem. In R. C. Buck and R. S. Cohen (Eds.), *In Memory of Rudolf Carnap, Boston Studies in the Philosophy of Science VIII*. Dordrecht: Reidel, pp. 325–341, 1971.

[Hintikka, 1975] J. Hintikka. Carnap and Essler versus Inductive Generalization. *Erkenntnis* **9**, 235–244, 1975.

[Hintikka, 1981] J. Hintikka. On the Logic of an Interrogative Model of Scientific Inquiry. *Synthese* **47**, 60–84, 1981.

[Hintikka, 1987a] J. Hintikka. The Interrogative Approach to Inquiry and Probabilistic Inference. *Erkenntnis* **26**, 429–442, 1987a.

[Hintikka, 1987b] J. Hintikka. Replies and Comments. Mondadori and my Work on Induction and Probability. In Bogdan (1987), pp. 302–307, 1987.

[Hintikka, 1988] J. Hintikka. What is the Logic of Experimental Inquiry? *Synthese* **74**, 173–190, 1988.

[Hintikka, 1992] J. Hintikka. The Concept of Induction in the Light of the Interrogative Approach to Inquiry. In J. Earman (Ed.), *Inference, Explanation, and Other Frustrations*. Berkeley: University of California Press, pp. 23–43, 1992.

[Hintikka, 1997] J. Hintikka. Replies. In [Sintonen, 1997, pp. 309–338].

[Hintikka, 1999] J. Hintikka. *Inquiry as Inquiry: A Logic of Scientific Discovery*. Dordrecht: Kluwer, 1999.

[Hintikka, 2006] J. Hintikka. Intellectual Autobiography. In [Auxier and Hahn, 2006, pp. 1-84].

[Hintikka and Hilpinen, 1966] J. Hintikka and R. Hilpinen. Knowledge, Acceptance and Inductive Logic. In [Hintikka and Suppes, 1966, pp. 1–20].

[Hintikka and Hilpinen, 1971] J. Hintikka and R. Hilpinen. Rules of Acceptance, Indices of Lawlikeness, and Singular Inductive Inferences: Reply to a Critical Discussion. *Philosophy of Science* **38**, 303–307, 1971.

[Hintikka and Niiniluoto, 1976] J. Hintikka and I. Niiniluoto. An Axiomatic Foundation of Inductive Generalization. In Przełecki, Szaniawski, and Wójcicki (1976), pp. 57–81, 1976. Reprinted with Postscript in [Jeffrey, 1980, pp. 157-181].

[Hintikka and Pietarinen, 1966] J. Hintikka and J. Pietarinen. Semantic Information and Inductive Logic. In [Hintikka and Suppes, 1966, pp. 96–112].

[Hintikka and Suppes, 1966] J. Hintikka and P. Suppes, eds. *Aspects of Inductive Logic*. Dordrecht: Reidel, 1966.

[Hintikka and Suppes, 1970] J. Hintikka and P. Suppes, eds. *Information and Inference*. Dordrecht: Reidel, 1970.

[Howson and Urbach, 1989] C. Howson and P. Urbach. *Scientific Reasoning: The Bayesian Approach*. La Salle, Ill.: Open Court, 1989.

[Jeffrey, 1980] R. C. Jeffrey, ed. *Studies in Inductive Logic and Probability vol. II*. Berkeley and Los Angeles: University of California Press, 1980.

[Johnson, 1932] W. E. Johnson. Probability: The Deductive and Inductive Problems. *Mind* 41, 403-423, 1932.

[Kaila, 1926] E. Kaila. *Die Prinzipien der Wahrscheinlichkeitslogik*. Turku: Annales Universitatis Fennicae Aboensis, 1926.

[Kelly, 1996] K. Kelly. *The Logic of Reliable Inquiry*. New York: Oxford University Press, 1996.

[Kemeny, 1963] J. Kemeny. Carnap's Theory of Probability and Induction. In [Schilpp, 1963, pp. 711–738].

[Keynes, 1921] J. M. Keynes. *A Treatise on Probability*. London: Macmillan, 1921.

[Kneale and Kneale, 1962] W. Kneale and M. Kneale. *The Development of Logic*. Oxford: Oxford University Press, 1962.

[Kuipers, 1978a] T. Kuipers. On the Generalization of the Continuum of Inductive Methods to Universal Hypotheses. *Synthese* **37**, 255-284, 1978.

[Kuipers, 1978b] T. Kuipers. *Studies in Inductive Probability and Rational Expectation*. Dordrecht: Reidel, 1978.

[Kuipers, 1984] T. Kuipers. Two Types of Inductive Analogy by Similarity. *Erkenntnis* **21**, 63-87, 1984.

[Kuipers, 1988] T. Kuipers. Inductive Analogy by Similarity and Proximity. In [Helman, 1988, pp. 299–313].

[Kuipers, 1997] T. Kuipers. The Carnap-Hintikka Programme in Inductive Logic. In [Sintonen, 1997, pp. 87-99].

[Kuipers, 2006] T. Kuipers. Inductive Aspects of Confirmation, Information, and Content. In [Auxier and Hahn, 2006, pp. 855-883].

[Kyburg, 1970] H. E. Kyburg, Jr. *Probability and Inductive Logic*. London: The Macmillan Company, 1970.

[Lakatos, 1968a] I. Lakatos. Changes in the Problem of Inductive Logic. In [Lakatos, 1968b, pp. 315-417].

[Lakatos, 1968b] I. Lakatos, ed. *The Problem of Inductive Logic*. Amsterdam: North-Holland, 1968.

[Lakatos, 1974] I. Lakatos. Popper on Demarcation and Induction. In P. A. Schilpp, ed., *The Philosophy of Karl Popper*, La Salle: Open Court, pp. 241-273, 1974.

[Levi, 1967] I. Levi. *Gambling with Truth*. New York: Alfred A. Knopf, 1967.

[Maher, 2000] P. Maher. Probabilities for Two Properties. *Erkenntnis* **52**, 63-91, 2000.

[di Maio, 1996] M. C. di Maio. Predictive Probability and Analogy by Similarity in Inductive Logic. *Erkenntnis* **43**, 369-394, 1996.

[Miller, 1994] D. Miller. *Critical Rationalism: A Restatement and Defence*. Chicago: Open Court, 1994.

[Mondadori, 1987] M. Mondadori. Hintikka's Inductive Logic. In [Bogdan, 1987, pp. 157-180].

[Niiniluoto, 1972] I. Niiniluoto. Inductive Systematization: Definition and a Critical Survey. *Synthese* **25**, 25-81, 1972.

[Niiniluoto, 1973] I. Niiniluoto. Review of A. C. Michalos, *The Popper–Carnap Controversy*. *Synthese* **25**, 417–436, 1973.

[Niiniluoto, 1976] I. Niiniluoto. Inductive Logic and Theoretical Concepts. In [Przełecki *et al.*, 1976, pp. 93–112].

[Niiniluoto, 1977] I. Niiniluoto. On a K-Dimensional System of Inductive Logic. In F. Suppe and P. D. Asquith, eds., *PSA 1976*, Vol. 2. East Lansing: Philosophy of Science Association, 425–477, 1977.

[Niiniluoto, 1980] I. Niiniluoto. Analogy, Transitivity, and the Confirmation of Theories. In [Cohen and Hesse, 1980, pp. 218–234].

[Niiniluoto, 1981] I. Niiniluoto. Analogy and Inductive Logic. *Erkenntnis* **16**, 1–34, 1981.

[Niiniluoto, 1983] I. Niiniluoto. Inductive Logic as a Methodological Research Programme. *Scientia: Logic in the 20th Century*. Milano, 77–100, 1983.

[Niiniluoto, 1984] I. Niiniluoto. *Is Science Progressive?* Dordrecht: D. Reidel, 1984.

[Niiniluoto, 1985/1986] I. Niiniluoto. Eino Kaila und der Wiener Kreis. In G. Gimpl, ed., *Weder-Noch. Tangenten zu den finnish-österreichischen Kulturbeziehungen*. Jahrbuch für deutsch-finnische Literaturbeziehungen 19/20, pp. 223-241, 1985/1986.

[Niiniluoto, 1987] I. Niiniluoto. *Truthlikeness*. Dordrecht: Reidel, 1987.

[Niiniluoto, 1988a] I. Niiniluoto. Analogy by Similarity in Scientific Reasoning. In [Helman, 1988, pp. 271–298].

[Niiniluoto, 1988b] I. Niiniluoto. Probability, Possibility, and Plenitude. In J. H. Fetzer, ed., *Probability and Causality*. Dordrecht: Reidel, pp. 91–108, 1988.

[Niiniluoto, 1994] I. Niiniluoto. Descriptive and Inductive Simplicity. In W. Salmon and G. Wolters, eds., *Logic, Language, and the Structure of Scientific Theories: Proceedings of the Carnap–Reichenbach Centennial*. Pittsburgh: University of Pittsburgh Press, pp. 147–170, 1994.

[Niiniluoto, 1994/1995] I. Niiniluoto. Hintikka and Whewell on Aristotelian Induction. *Grazer Philosophische Studien* **49**, 49–61, 1994/1995.

[Niiniluoto, 1995] I. Niiniluoto. Is There Progress in Science?. In H. Stachowiak, ed., *Pragmatik: Pragmatische Tendenzen in der Wissenschaftstheorie*. Hamburg: Felix Meiner, pp. 30-58, 1995.

[Niiniluoto, 1997] I. Niiniluoto. Inductive Logic, Atomism, and Observational Error. In [Sintonen, 1997, pp. 117–131].

[Niinuloto, 1998] I. Niiniluoto. Induction and Probability in the Lvow-Warsaw School. In K. Kijania-Placek and J. Woleński, eds., *The Lvov-Warsaw School and Contemporary Philosophy*. Dordrecht: Kluwer, pp. 323–335, 1998.

[Niiniluoto, 1999] I. Niiniluoto. Defending Abduction. *Philosophy of Science (Proceedings)* **66:** S436–S451, 1999.

[Niiniluoto, 2005a] I. Niiniluoto. Abduction and Truthlikeness. In R. Festa, A. Aliseda, and J. Peijnenburg (Eds.), *Confirmation, Empirical Progress, and Truth Approximation*. Amsterdam: Rodopi, pp. 255-275, 2005.

[Niiniluoto, 2005b] I. Niiniluoto. Inductive Logic, Verisimilitude, and Machine Learning. In P. Hájek, L. Valdés-Villanueva, and D. Westerstal, eds., *Logic, Methodology, and Philosophy of Science: Proceedings of the Twelfth International Congress*. London: King's College Publications, pp. 295-314, 2005.

[Niiniluoto, 2005c] I. Niiniluoto. G. H. von Wright on Probability and Induction. In I. Niiniluoto and R. Vilkko (Eds.), *Philosophical Essays in Memoriam Georg Henrik von Wright*. Acta Philosophica Fennica **77**, Helsinki: The Philosophical Society of Finland, pp. 11-32, 2005.

[Niiniluoto and Tuomela, 1973] I. Niiniluoto and R. Tuomela. *Theoretical Concepts and Hypothetico-Inductive Inference*. Dordrecht: Reidel, 1973.

[Nix and Paris, 2007] C. J. Nix and J. B. Paris. A Note on Binary Inductive Logic. *Journal of Philosophical Logic* **36**, 735-771, 2007.

[Pietarinen, 1970] J. Pietarinen. Quantitative Tools for Evaluating Scientific Systematizations. In [Hintikka and Suppes, 1970, pp. 123–147].

[Pietarinen, 1972] J. Pietarinen. *Lawlikeness, Analogy and Inductive Logic.* Acta Philosophica Fennica **26**. Amsterdam: North-Holland, 1972.

[Pietarinen, 1974] J. Pietarinen. Inductive Immodesty and Lawlikeness. *Philosophy of Science* **41**, 196–198, 1974.

[Popper, 1959] K. R. Popper. *The Logic of Scientific Discovery.* London: Hutchinson, 1959.

[Popper, 1963] K. R. Popper. *Conjectures and Refutations*. London: Hutchinson, 1963.

[Przełecki *et al.*, 1976] M. Przełecki, K. Szaniawski, and R. Wójcicki, eds. *Formal Methods in the Methodology of Empirical Sciences*. Dordrecht: Reidel, 1976.

[Romeyn, 2005] J.-W. Romeyn. *Bayesian Inductive Logic*. Groningen: University of Groningen, 2005.

[Schilpp, 1963] P. A. Schilpp, ed. *The Philosophy of Rudolf Carnap*. Lassalle, Ill.: Open Court, 1963.

[Schilpp and Hahn, 1989] P. A. Schilpp and L. E. Hahn, eds. *The Philosophy of Georg Henrik von Wright*. La Salle, Illinois: Open Court, 1989.

[Scott and Krauss, 1966] D. Scott and P. Krauss. Assigning Probabilities to Logical Formulas. In [Hintikka and Suppes, 1966, pp. 219-264].

[Sintonen, 1997] M. Sintonen, ed. *Knowledge and Inquiry. Essays on Jaakko Hintikka's Epistemology and Philosophy of Science*. Amsterdam: Rodopi, 1997.

[Skyrms, 1986] B. Skyrms. *Choice and Chance. An Introduction to Inductive Logic. Third Edition.* Belmont, California: Wadsworth Publishing Company, 1986.

[Skyrms, 1993a] B. Skyrms. Carnapian Inductive Logic for a Value Continuum. In H. Wetterstein, ed., *The Philosophy of Science*, Midwest Studies in Philosophy **18**. South Bend: University of Notre Dame Press, pp. 78-89, 1993.

[Skyrms, 1993b] B. Skyrms. Analogy by Similarity in HyperCarnapian Inductive Logic. In J. Earman, ed., *Philosophical Problems of the Internal and External Worlds. Essays in the Philosophy of Adolf Grünbaum.* Pittsburgh: University of Pittsburgh Press, pp. 273–282, 1993.

[Stegmüller, 1973] W. Stegmüller. Carnap's Normative Theory of Inductive Probability. In P. Suppes *et al.*, eds., *Logic, Methodology, and Philosophy of Science IV*. Amsterdam: North-Holland, pp. 501-513, 1973.

[Suppes, 1966] P. Suppes. Probabilistic Inference and the Concept of Total Evidence. In [Hintikka and Suppes, 1966, pp. 49-65].

[Suppes, 1969] P. Suppes. *Studies in the Methodology and Foundations of Science: Selected Papers from 1952 to 1969.* Dordrecht: D. Reidel, 1969.

[Tuomela, 1966] R. Tuomela. Inductive Generalization in an Ordered Universe. In [Hintikka and Suppes, 1966, pp. 155–174].

[Uchii, 1972] S. Uchii. Inductive Logic with Causal Modalities: A Probabilistic Approach. *Philosophy of Science* **39**, 162-178, 1972.

[Uchii, 1973] S. Uchii. Inductive Logic with Causal Modalities: A Deterministic Approach. *Synthese* **26**, 264-303, 1973.

[Uchii, 1977] S. Uchii. Induction and Causality in Cellular Space. In F. Suppe and P. D. Asquith, eds., *PSA 1976*, vol. 2, East Lansing: Philosophy of Science Association, pp. 448-461, 1977.

[Walk, 1966] K. Walk. Simplicity, Entropy and Inductive Logic. In [Hintikka and Suppes, 1966, pp. 66-80].

[von Wright, 1951] G. H. von Wright. Carnap's Theory of Probability. *Philosophical Review* **60**, 362–374, 1951.

[von Wright, 1951b] G. H. von Wright. *A Treatise on Induction and Probability.* London: Routledge and Kegan Paul, 1951.

[von Wright, 1957] G. H. von Wright. *The Logical Problem of Induction and Probability*, 2^{nd} revised edition. Oxford: Basil Blackwell, 1957. [First Edition: 1941. In *Acta Philosophica Fennica* **3**, Helsinki.]

[von Wright, 1971] G. H. von Wright. *Explanation and Understanding.* Ithaca: Cornell University Press, 1971.

[Zabell, 1997] S. Zabell. Confirming Universal Generalizations. *Erkenntnis* **45**: 267-283, 1997.